8-2018

Mechanism of Incorporation and Repair of Uracil at Highly Transcribed Genes In Saccharomyces Cerevisiae

Norah Auma Owiti

Recommended Citation

MECHANISM OF INCORPORATION AND REPAIR OF URACIL AT HIGHLY TRANSCRIBED GENES IN SACCHAROMYCES CEREVISIAE

By

Norah Auma Owiti, B.A

APPROVED:

_____

Nayun Kim, Ph.D.
Advisory Professor

_____

Theresa M. Koehler, Ph.D.

_____

Grzegorz Ira, Ph.D.

_____

Lei Li, Ph.D.

_____

Jiqiang Ling, Ph.D.

APPROVED:

_____
Dean, The University of Texas
MD Anderson Cancer Center UTHealth Graduate School of Biomedical Sciences

MECHANISM OF INCORPORATION AND REPAIR OF URACIL AT HIGHLY TRANSCRIBED GENES IN SACCHAROMYCES CEREVISIAE

A

DISSERTATION

Presented to the Faculty of

The University of Texas

MD Anderson Cancer Center UTHealth

Graduate School of Biomedical Sciences

in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

By

Norah Auma Owiti, B.A

Houston, TX

August 2018

To my Parents,

Lister Boaz Owiti and Perez Aoko Okatch

Who instilled in me the value of education.

I hope I made you proud.

You are always and forever in my heart.

Many are the plans in a person's heart, but it is the LORD's purpose that prevails.
~ Proverbs 19:21 (NIV)

I have been fortunate to be surrounded by a lot of really great friends. To Yi Liu, going through this journey with you has been liberating. Joan, your positive outlook on life has been contagious, thanks for always supporting and cheering me on. Anna Statz, you are a gift and my sister from another mother. I was lucky to have met you when I first moved to the United States. Thanks for your 10 years of friendship! Faith and Sheillah, you are some of the most incredible women I have ever met. Faith thanks for showing me around Houston and letting me share your friends! Sheillah thanks for all your free homemade dinners! Gerald, Kam and Yves knowing you all added so much life to my life. Thanks Gerald for always organizing activities that provided a much needed break and recharge during my dissertation journey. Harriet Jenkins thank you for welcoming me into your family and giving me a home away from home in Iowa during all the breaks. You all, and many others, not highlighted, were my source of motivation and support in my journey and I am so grateful.

Last, but definitely not least, I would like to thank my family. To my uncle, Dr. Joseph Owiti, thank you for your constant encouragements and support. You have been my cheerleader and I couldn't thank you enough for all that you have done to ensure I get to where I am today. To my siblings Edwin, June, Selma and Lydiah, thanks for your unconditional love. Thanks to all my cousins, nieces, nephews, aunts, and grandparents. I would not be where I am today without all your love and support. Nawapenda wote!

MECHANISM OF INCORPORATION AND REPAIR OF URACIL AT HIGHLY TRANSCRIBED GENES IN SACCHAROMYCES CEREVISIAE

Norah Auma Owiti, B.A.

Advisory Professor: Nayun Kim, Ph.D.

Recombination and mutagenesis are elevated by high levels of transcription. The correlation between transcription and genome instability is largely explained by the topological and structural changes in DNA and the associated physical obstacles generated by the transcription machinery. However, such explanation does not directly account for the unique types of mutations originating from the non-canonical residues such as uracil, which are also elevated at highly transcribed regions. Apurinic/Apyrimic or Abasic (AP) sites derived from uracil excision, accumulate at a higher rate in actively transcribed regions of the genome in *S. cerevisiae* and are primarily repaired by base excision repair (BER) pathway. I have demonstrated that transcription-coupled nucleotide excision repair (NER) pathway can functionally replace BER to repair those AP sites located on the transcribed strand much like the strand specific repair of UV-induced pyrimidine dimers.

This thesis reveals that the DNA composition can be modified to include higher uracil-content through the non-replicative, repair-associated DNA synthesis. I show here a positive correlation between the level of transcription and the density of uracil residues in the yeast genome indirectly through the mutations generated by the glycosylase that excise undamaged cytosine as well as uracil. The higher uracil-density at actively transcribed regions is confirmed by the long-amplicon qPCR analysis. I also show that the uracil-associated mutations at highly transcribed regions are elevated by the induced DNA damage and reduced by the overexpression of a dUTP-catalyzing enzyme, Dut1, in G1- or G2-phases of the cell cycle.

Additional roles of transcription elongation factor Dst1 and RNAPII degradation factor Def1 in AP induced transcription arrest is also revealed. I report that Def1 directs NER to AP lesions on the transcribed strand of an actively transcribed gene but that its function is dependent on metabolic state of the yeast cells. I additionally show that Dst1, a homolog of mammalian transcription elongation factor TFIIS, interferes with NER-dependent repair of AP lesions while suppressing homologous recombination pathway.

In summary, this thesis elucidates a novel mechanism of introducing uracil into DNA during damage-induced repair synthesis and provides further insights onto how AP sites on the transcribed DNA strand are repaired.

Table of Contents:

List of Figures:

List of Tables:

List of Important Abbreviations:

Abbreviations	Full Name
AP site	apurinic/apyrimidic site
DSB	double strand break
NTS	non-transcribed strand
TAR	transcription-associated recombination
TAM	transcription-associated mutagenesis
TS	thymidylate synthase
5-FU	5-fluorouracil
Ung1/UDG	uracil DNA glycosylase
BER	base-excision repair
NER	nucleotide-excision repair
TCR	transcription-coupled (nucleotide-excision) repair
GGR	global genome (nucleotide-excision) repair
TLS	translesion synthesis
UDS	unscheduled DNA synthesis
4NQO	4-nitroquinoline
CPT	camptothecin
RNAPII	RNA polymerase II
Dox	Doxycycline

Chapter 1

Introduction

1.1 Sources of DNA Modifications

DNA serves as the template for the essential cellular metabolic processes of DNA replication and transcription; therefore, maintaining its integrity is important for viability. The genome is exposed to different kinds of changes that can alter its structure and/or the information encoded in the sequence of nucleotides. These modifications have the potential to promote mutagenesis, the process of changing the normal DNA sequence, and to facilitate genomic instability, which can be both beneficial and detrimental to organisms depending on the context. One benefit of genomic instability is that the changes ranging from simple base substitutions to the rearrangements or loss of chromosomes can act as the raw materials of evolution. Additionally, genetic variation is important in processes such as immunoglobulin diversification (Maizels, 2005). However, these changes can also be harmful to the cell because they can disrupt essential cellular processes. In higher eukaryotes, genomic instability is associated with premature aging, predisposition to various types of cancers, and inherited diseases.

The genome is subject to both endogenous and exogenous sources of damage. It has been known for a long time, even before the discovery of the double helical structure of DNA in 1953, that exogenous (environmental) damaging agents have mutagenic effects and can lead to cancer predisposition ("An early suggestion of DNA Repair. Effect os sublethal doses of monochromatic ultraviolet radiation on bacteria in liquid suspensions. By Alexander Hollaender and John T. Curtis. Proc Soc Exp Biol Med, 33,61-62(1935)," 1975; Friedberg, 2008; Kelner, 1949). It took yet another decade after the DNA structure was revealed to recognize that DNA is also subject to endogenous (spontaneous) damage during normal cellular metabolism (Lindahl, 1993; Lindahl & Nyberg, 1972). Both exogenous and endogenous sources of DNA damage can induce damage in a variety of forms such as: altered bases, base adducts, abasic sites (apurinic/ apyrimidinic, or AP sites), or single- or double-stranded breaks.

2

DNA modifications can arise from various environmental factors and lead to various types genome instability events. Exogenous sources of DNA damage include ultraviolet (UV) radiation originating from sunlight, ionizing radiation derived from both natural or clinical sources, food, chemicals and inhaled cigarette smoke. UV-induced damage can lead to formation of cyclobutane pyrimidine dimers (CPDs) and 6-4 photoproducts (6-4PPs) formed between adjacent pyrimidine dimers (Pfeifer, 1997). These photoproducts can cause the collapse of the replication fork, leading to the generation of DNA double stranded breaks (DSBs) (G. L. Chan et al., 1985; Cordeiro-Stone et al., 1997). Smoking has been associated with mutations in the lung and development of lung cancer (Ozlu & Bulbul, 2005).

A variety of processes inside the cells can also lead to DNA damage. Reactive oxygen species (ROS) that are produced as intermediate products of cellular metabolic processes can cause oxidative damage to DNA leading to the formation of single- or double-stranded breaks (Sallmyr et al., 2008). Changes to genomic DNA can also occur during normal cellular processes such as transcription or DNA replication. For example, DNA polymerases are prone to adding incorrect bases, which can lead to mutagenesis if not properly removed. In addition, because transcription and replication utilize the same template, collisions between the transcription machinery and the replication machinery are inevitable and can result in genomic instability. Finally, spontaneous hydrolysis of the DNA bases in the form of deamination or the removal of bases to create AP sites can also lead to genomic instability.

There are approximately 70,000 DNA lesions formed in each human cell per day (Lindahl & Barnes, 2000). Given this large amount of daily damage to the cell's genome, preserving the integrity of the genome is a monumental but critical task. Cells have developed several mechanisms to sense and repair DNA damage and preclude genomic instability. Years after the discovery of the double helical structure of DNA, Francis Crick wrote, "We totally missed the possible role of enzymes in repair although… I later came to

realize that DNA is so precious that probably many distinct mechanisms would exist. Nowadays one could hardly discuss mutations without considering repair at the same time" (Crick, 1974).

1.2 Transcription as a source of genomic instability

Transcription is a fundamental cellular function where the genetic information for both protein expression and non-coding function is converted from DNA to an RNA molecule. There are many aspects of transcription that allow it to be a source of genomic instability. Unlike DNA replication, in which one copy of each DNA strand is made throughout the genome once per cell division, transcription is a very nonuniform process that occurs at specific segments of the genome. Also unlike replication, the rate of transcription is highly variable and only one strand of the DNA is serves as a template.

Transcription involves more than just copying the DNA template. It also has the ability to alter the genomic landscape by enhancing the rates of recombination and as well as mutagenesis, processes referred to as transcription-associated recombination (TAR) and transcription-associated mutagenesis (TAM) respectively. Several studies performed in the 1970s using reversion assays in *Escherichia coli* were the first to demonstrate that highly transcribed genomic loci are more prone to mutagenesis (Herman & Dworkin, 1971; Savic & Kanazir, 1972). These studies showed that mutations were more efficiently induced by exogenous mutagens such as ICR-191 and ultraviolet radiation when the reporter gene was highly transcribed. In the last two decades, several studies have demonstrated that transcription has the ability to stimulate spontaneous mutagenesis and recombination in eukaryotes (Datta & Jinks-Robertson, 1995; Keil & Roeder, 1984; Voelkel-Meiman et al., 1987). The earlier experiments in yeast utilized a reversion *LYS2* allele that detects one-base pair frameshift reversions to demonstrate that reversion frequencies increased with the increasing rates of transcription (Datta & Jinks-Robertson, 1995). In addition, transcription

4

increased the hyper-recombinogenic effects of 4-Nitroquinoline (4-NQO) and methyl methane-sulphonate (MMS) in a synergistic manner suggesting that transcription can render DNA more susceptible to genotoxic agents (Garcia-Rubio et al., 2003). Currently, TAR is thought to be associated with collisions between transcription and replication machinery, and the formation of non-canonical DNA structures such as R-loops and G4-DNA structures while TAM is thought to be associated with increased access to mutagens as well change in the nucleotide composition of DNA.

The question of how transcription can sensitize DNA to damage remains to be addressed. One possible mechanism could involve the damage-prone single-stranded DNA regions that occur in highly transcribed regions of the genome. Transcription leads to changes in DNA topology by introducing positive and negative supercoils to the DNA. During transcription, RNA polymerase II (RNAPII) generates positive supercoiling ahead of the transcription bubble and negative supercoiling behind the transcription bubble, which may cause torsional stress (Aguilera, 2002; Kim & Jinks-Robertson, 2012). This superhelical stress generated during transcription opens up the double helix and facilitates the formation of non-canonical DNA structures that can encourage genome instability. In addition, DNA strands are separated during transcription, temporarily leaving the DNA in a single-stranded state, which is more susceptible to damages than double-stranded DNA. In summary, transcription-related stress can lead to the accumulation of different kinds of damage in the DNA and facilitate genomic instability.

Co-transcriptionally formed RNA/DNA hybrids have been implicated as a major source of TAR in yeast. Original studies that provided evidence for TAR in yeast showed a strong increase in TAR in *hpr1* mutant of the THO complex (Chavez & Aguilera, 1997; Prado et al., 1997). Additionally, it was shown that the hyper-recombination observed in the *hpr1* mutants was associated with the accumulation of co-transcriptionally formed RNA-DNA hybrids (R-loops) (Huertas & Aguilera, 2003). R-loops are stable hybrids of the transcribed

DNA strand (TS) and the nascent RNA that form during transcription leaving the non-transcribed DNA strand (NTS) single-stranded. Genome-wide analyses in yeast confirmed that R-loops, which have been associated with mutagenesis, recombination and rearrangements, are formed at highly transcribed genomic regions (Y. A. Chan et al., 2014; El Hage et al., 2014; Wahba et al., 2016). One potential mechanism for R-loop induced genomic instability is that R-loops can block/stall replication fork progression and induce recombination. Other studies have shown that R-loops mediate genomic instability in yeast and human cells by inducing transcription-replication conflicts (Castellano-Pozo et al., 2012; Herrera-Moyano et al., 2014). R-loops are degraded by specialized enzymes called RNaseH that specifically degrade the RNA that is hybridized to DNA (Pallan & Egli, 2008). If not removed, these R-loops can cause the stalling and eventual collapse of the replication fork, which must be restarted/ repaired via homologous recombination. In addition, R-loops expose the non-transcribed DNA strand (which is single-stranded) further subjecting it to the endogenous or exogenous DNA damaging agents that can increase the frequency of recombination and genomic rearrangements by interfering with replication (Aguilera & Garcia-Muse, 2012).

A high level of transcription facilitates the formation of non-B DNA structures that promote TAR. Single-stranded DNA formed during transcription can facilitate the formation of non-canonical DNA structures depending on the sequence of the DNA (Hall et al., 2017; Zhao & Usdin, 2015). An example of a non-canonical DNA structure is G-quadruplex structure (G4-DNA) which consist of stacks of guanine bases bound to each other. These structures are formed when several runs of guanine are present within a single-stranded segment of DNA *via* the Hoogstein bonds. G4-DNA structures are fairly stable and have been shown to stabilize R-loops and also block transcription *in vitro* and could potentially block replication as well (Belotserkovskii et al., 2010). The rate of recombination was significantly increased, in a transcription-dependent manner when a G-rich murine switch-

6

region fragment was introduced into yeast. The increase was especially significant when the G-rich sequence was located on the NTS (Kim & Jinks-Robertson, 2011).

Interestingly, TAR was shown to depend on replication in yeast and mammalian cells (Gottipati et al., 2008; Prado & Aguilera, 2005). This observation raised the question of how replication forks contend with transcription, especially because both replication and transcription utilize the same template. DNA replication complexes move faster in comparison to the transcription machinery; therefore, collisions between these two complexes are inevitable. When the lagging strand is being transcribed, the replication forks and the transcription machinery converge, and when the leading strand is being transcribed, both the replication machinery and the transcription machinery are moving in the same direction. The resulting conflicts are referred to as head on and co-directional collisions respectively, and they can each trigger TAR. Head-on collisions have been shown *in vitro* to be a stronger hindrance to the progression of the replication machinery compared to the co-directional collisions (Liu & Alberts, 1995). Additionally, the direct conflicts observed between the two machineries generate positive supercoils ahead of these machineries. These supercoils lead to the accumulation of superhelical stresses that can trigger replication fork reversal and eventually accumulation of double strand breaks (Postow, Crisona, et al., 2001; Rudolph et al., 2007). Therefore, transcription-replication conflicts can be a major source of TAR and therefore genomic instability.

Even though transcription, replication, and DNA repair are distinct and separate metabolic processes, they are all interconnected with genomic DNA being the common template. In addition to stimulating recombination, transcription has also been shown to stimulate mutagenesis (i.e. transcription-associated mutagenesis or TAM). The enhanced single-stranded nature of the non-transcribed strand during transcription is an important source of TAM.   Most importantly, increased spontaneous cytosine deamination was observed on the non-transcribed DNA strand compared to the transcribed strand in bacterial

7

cells (Beletskii & Bhagwat, 1996). In eukaryotes, it was observed that cytosine deaminases prefer the cytosines that are located on the non transcribed single stranded DNA of the R-loops structures. Correspondingly, the enzymatic deamination of cytosine residues to uracil has primarily been observed on the NTS of DNA in yeast THO mutant (Gomez-Gonzalez & Aguilera, 2007).

Transcription also appears to alter the nucleotide composition of DNA. Using a frameshift assay, TAM was shown to be a result of damage to DNA. A reduction in TAM was observed in yeast strains lacking the error-prone translesion synthesis polymerase (TLS) and increased in the absence of the error-free repair pathway (Datta & Jinks-Robertson, 1995; Morey et al., 2000). Further studies using a mutation reporter system modified with a tetracycline-regulatable *pTET* promoter demonstrated a dramatically elevated mutagenesis upon activation of transcription in repair-deficient strains (Kim & Jinks-Robertson, 2009, 2010). This TAM was abrogated when uracil DNA glycosylase, Ung1, was disabled indicating that the observed mutagenesis was the result of excision of uracil from DNA. This uracil-dependent mutagenesis was further reduced when the free dUTP level was reduced by the overexpression of dUTPase, Dut1. These results suggested that the uracil derived from dUMP incorporated into DNA by DNA polymerases is the major source of uracil associated with highly transcribed genomic regions. In addition to enhanced uracil-derived mutagenesis when transcription is activated, increased incorporation of ribonucleotides into DNA has also been shown to be enhanced when transcription was activated (Nick McElhinny et al., 2010). Transcription, therefore, apparently alters the underlying DNA nucleotide composition. Overall, studies have indicated transcription as a major source of endogenous damage to DNA.

1.3 Uracil in DNA

DNA encodes, stores and transmits genetic instructions that are necessary for cellular functions. The building blocks of DNA are deoxyribonucleotides with four different bases: the pyrimidine bases are adenine and guanine and purine bases are thymine and cytosine. DNA forms an anti-parallel double helical structure, where two strands of DNA are held together by hydrogen bonds that form between pyrimidines and purines. Adenines (A) base-pairs with the thymine (T); guanine (G) with cytosine (C). An exceptional case exists in some rare bacteriophages where all thymine residues are completely replaced by uracil (U) residues (Kiljunen et al., 2005; Skurnik et al., 2012; Warner & Duncan, 1978). In contrast to DNA, RNA is usually found as a single-stranded molecule containing the 5-carbon sugar ribose in the sugar-phosphate backbone instead of a deoxyribose. RNA also utilizes the purine, uracil in place of thymine bases. Uracil and Thymine are very similar in structure except for the existence of a methyl group in the 5' C atom position in a thymine base. The lack of the 5-methyl group in uracil does not interfere with the interaction and base pairing of the uracil residue with adenine (Figure 1).

Uracil can also be found in DNA. Spontaneous or enzymatic deamination of cytosine to uracil, can introduce uracil residues into genomic DNA, creating a pre-mutagenic U:G mispair. Uracil in DNA can also arise through direct incorporation into DNA in place of thymine during DNA replication creating a U:A mispair. The uracil residue is then excised by uracil DNA glycosylase, Ung1, to create an abasic site. Repeated cycles of Ung1-initiated uracil repair might result in accumulation of abasic sites leading to DNA strand breaks and ultimately cell death (el-Hajj et al., 1988; Gadsden et al., 1993).

Depending on the context, uracil in DNA can cause unfavorable or beneficial events. On one hand, uracil incorporation at specific regions of the immunoglobulin gene is an essential intermediate in antibody affinity maturation during adaptive diversification processes. On the other hand, uracil can be present in the DNA as a lesion that might lead

to deleterious mutations if not properly removed. Presence of uracil in DNA can interferes with the binding of some transcription factors thereby altering the expression of those genes (Luhnsdorf et al., 2014).
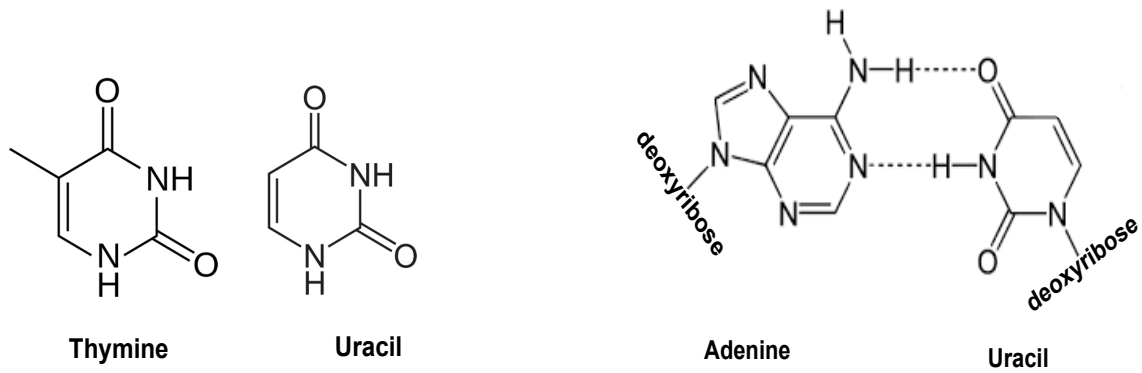
Figure 1: Structure and base pairing of uracil with adenine

Uracil and Thymine are very similar in structure except for the 5' methyl group in Thymine.

Due to these structure similarities, uracil can successfully base pair with Adenine as the 5'

methyl is not necessary for base pairing.

Uracil derived from cytosine deamination is necessary for adaptive immunity but can also lead to deleterious genomic instability events. 100-500 cytosine deamination events have been predicted to occur per human cell per day (Lindahl, 1993). Studies have shown that enzymatic deamination of cytosine at the immunoglobulin loci is necessary for antibody diversification (Muramatsu et al., 1999). In antibody-expressing B lymphocytes, Activation Induced Deaminase (AID) enzyme triggers the deamination of cytosine to uracil and subsequent uracil removal is the first phase of hypermutation driving immunoglobulin gene diversification (Longerich et al., 2006; Petersen-Mahrt et al., 2002), (Klemm et al., 2009; Maul & Gearhart, 2010). In this case, the presence of uracil in the genome and the resulting mutation increases diversity in the immunoglobulin genes and serves an important role in adaptive immunity. In some instances, the unregulated expression of AID has been associated with increased uracil-dependent mutations and chromosomal translocations in genes that are important for cancer development in several cancer types (Alexandrov et al., 2013; Tsai et al., 2008; Zhang et al., 2010).

Uracil in DNA can also result from its incorporation in place of thymine during DNA replication. dUTP is a precursor required for the *de novo* synthesis of thymidine (Hochhauser & Weiss, 1978; Tye et al., 1977). Because of the structure similarity between uracil and thymine bases, most DNA polymerases cannot distinguish between these two nucleotides and will readily incorporate dUMP instead of dTMP (Bessman et al., 1958; Warner et al., 1981). Incorporation of dUMP depends on the [dUTP]/[dTTP] ratio which was estimated to be less than 0.1 under normal physiological conditions (Traut, 1994).

Even though most replicative polymerases cannot distinguish between uracil and thymine, a subset of B family polymerases of archaea can recognize the presence of uracil in DNA (Lasken et al., 1996). Furthermore, the DNA polymerases of thermophilic archaebacteria bind DNA template containing uracil with higher affinity causing stalled replication forks (Greagg et al., 1999; Wardle et al., 2008) (Firbank et al., 2008; Fogg et al.,

2002). These studies using primer extension assays demonstrated that replicative polymerases from bacteria and eukaryotes completed a full extension of a DNA template containing uracil residue indicating that they could not recognize the uracil residue; however, the thermophilic archaea DNA polymerases stalled 4 to 6 nucleotides upstream of the uracil residue. These studies also suggest that thermophilic archaeabacteria might have developed mechanisms to tolerate high uracil DNA content. For example, it has been shown that the rate of cytosine deamination is enhanced at higher temperatures. This suggest that cytosine deamination occurs at a higher frequency in thermophilic archaea; therefore, a greater number of uracil residues accumulate at these conditions (Greagg et al., 1999; Lindahl & Nyberg, 1974). The ability of these polymerases to recognize uracil might be a mechanism that evolved to prevent mutagenesis that might arise from increased uracil density from enhanced cytosine deamination.

In summary, it is necessary for the cell to regulate the levels of uracil in DNA to avoid unfavorable outcomes resulting from its presence in the genome.

1.4 Regulation of dUTP and dTTP nucleotide biosynthesis

Fine-tuned regulation of cellular nucleotide metabolism is vital for maintaining the deoxynucleotide triphosphate (dNTP) pool and sustaining DNA replication. In addition, excluding non-canonical bases from DNA is necessary in maintaining genome stability. As a result, multiple enzymes with temporal regulation throughout the cell cycle are in place to allow for efficient production and balance of nucleotides (McIntosh et al., 1986; Storms et al., 1984) (Figure 2).

In the dTTP biosynthesis pathway, actions of ribonucleotide reductase (RNR) result in the biosynthesis of dUTP which can be directly incorporated into DNA to create A:U base pairs. In addition, the dUTP produced is a precursor for dTTP synthesis (Warner et al.,

1981). Deoxyuridine triphosphatase (dUTPase), an essential enzyme, selectively catalyzes the conversion of free dUTP to dUMP and pyrophosphate, PPi, thereby eliminating it from the dNTP pool (Gadsden et al., 1993; Shlomai & Kornberg, 1978). dUTPase therefore serves two functions; (i) lowering the dUTP:dTTP ratio thereby preventing incorporation of uracil into DNA during replication or repair and (ii) providing a dUMP substrate for thymidine synthesis (Lindahl, 1993).

dUTPase deficiency leads to the accumulation of uracil into DNA. Subsequent rounds of repair initiation by Ung1 cause DNA strand breaks and consequently, cell death (Gadsden et al., 1993; Kavli et al., 2007). For *S. cerevisiae* cells defective in dUTPase, viability was restored by exogenous supplementation with dTMP (Guillet et al., 2006). However, the cell death associated with *dut1* mutant is not due to the lack of dTMP: disruption of Ung1, a uracil DNA glycosylase that excises uracil and initiates excision repair, was able to restore survival in *dut1* mutants in yeast, bacteria and flies (Guillet et al., 2006; Lari et al., 2006; Muha et al., 2012). In *dut1 ung1* double mutants, uracil content in the DNA is significantly elevated but the cells can survive because uracil does not alter the coding characteristics of the DNA. Similalrly, in wild-type *Drosophila melanogaster,* the dUTPase expression is observed only in imaginal (undifferentiated) tissues and in the central nervous system but not in most of the larval tissues (Horvath et al., 2015; Muha et al., 2012). But, due to the absence of Ung1, the fly can tolerate elevated genomic uracil content (Adams et al., 2000).
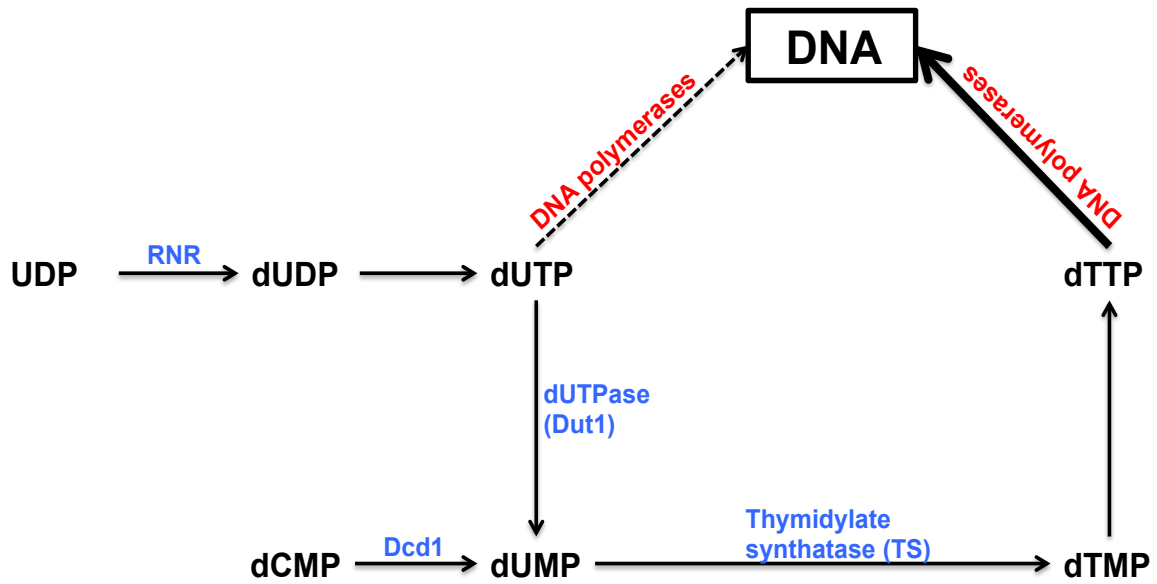
Figure 2: Key Enzymes in the Thymidine biosynthesis pathway

dUTP is a precursor of dTTP synthesis and can be directly incorporated into DNA as well as dTTP. This figure illustrates the key enzymes in the thymidine biosynthesis pathway discussed in this thesis including: dUTPase, Dcd1 and Thymidylate synthatase (TS).

The role of dUTPase in preventing accumulation of uracil in DNA is conserved in several organisms including: *Escherichia coli, Saccharomyces cerevisiae, Arabidopsis thaliana, Trypanosoma brucei, Caenorhabditis elegans Drosophila melanogaster* and humans, indicating that dUTPase is ubiquitous and essential in DNA replication across organisms (Castillo-Acosta et al., 2008; Dengg et al., 2006; Dubois et al., 2011; el-Hajj et al., 1988; Guillet et al., 2006; Muha et al., 2012; Vertessy & Toth, 2009). dUTPase also plays an important role in virus replication. A form of dUTPase is encoded by poxviruses, herpesviruses and some retroviruses to minimize the incorporation of uracil into the viral DNA by ensuring a low [dUTP]/[dTTP] ratio (Chen et al., 2002). In most organisms, dUTPase is highly expressed in S-phase both at the protein and mRNA level ensuring that dUTP levels are kept low during replication (Ladner & Caradonna, 1997; Pardo & Gutierrez, 1990).

In addition to the hydrolysis of dUTP by dUTPases, dUMP, required for dTTP synthesis, can also be synthesized by another highly conserved enzyme deoxycytidine monophosphate deaminase, Dcd1, which converts dCMP to dUMP (L. Wang & Weiss, 1992). The dUMP produced by Dcd1 is sufficient in generating adequate dTTP to sustain replication (McIntosh & Haynes, 1984). Unlike dUTPase, which is cell-cycle regulated, Dcd1 is constitutively expressed throughout the cell cycle (McIntosh et al., 1986). Studies in *S. cerevisiae* indicated that deletion of *DCD1* led to a significant increase in dCTP and reduction in dTTP pool without affecting the viability of the cells (Kohalmi et al., 1991; Sanchez et al., 2012). In addition, an increase in mutagenesis was observed following *DCD1* deletion, possibly due to misinsertion of nucleotides into DNA and most likely uracil (Kohalmi et al., 1991). Dcd1 therefore, plays an important role in dTTP synthesis without utilizing the dUTP pool as the intermediate.

dUMP produced by dUTPase or Dcd1 is subsequently converted to dTMP, which is the substrate for dTTP synthesis. This reaction, catalyzed by thymidylate synthase (TS)

16

enzyme, is the sole *de novo* source of dTMP essential for DNA replication and repair and therefore is the most critical step in thymidine biosynthesis. TS uses 5,10 methylenetetrahydrofolate (5,10-CH$_2$THF) as a co-factor to add a methyl group to C5 of uracil base (Costi et al., 2005). During this reaction, 5,10-CH$_2$THF is oxidized to dihydrofolate (DHF) which is thenrecycled by dihydrofolate reductase (DHFR) and serine hydroxymethyltransferase (SHMT) to produce more 5,10-CH$_2$THF. TS, similar to dUTPase, is an essential enzyme which is tightly regulated with highest activity in S-phase of the cell cycle (Storms et al., 1984). Overall the cell has developed several mechanisms to ensure maintenance of the nucleotide pool during DNA synthesis and also to avoid incorporation of non-canonical nucleotides, such as uracil, into DNA. An illustration of the thymidine biosynthesis pathway is shown in Figure 2.

1.5 Role of uracil metabolism in Cancer Chemotherapy

Inhibitors of thymidine biosynthesis pathway are one of the most effective chemotherapy agents in the treatment of cancer since the 1940s (Farber & Diamond, 1948). Since TS enzyme has an important role in thymidylate metabolism, it is a frequent chemotherapy target. Inhibition of TS leads to depletion of dTMP and the imbalance of the nucleotide pool leads to impaired DNA synthesis and repair resulting in damage (Chu et al., 2003). Inhibitors of this pathway include fluoropyrimidines such as 5-fluorouracil (5-FU), fluorodeoxyuridine (FUdR) and capecitabine and antifolates such as raltitrexed, pemetrexed and methotrexate classes of anti-cancer agents (P. M. Wilson et al., 2014). These agents block dTMP production by either directly inhibiting TS or indirectly inhibiting TS by targeting DHFR, thereby limiting the availability of 5,10-CH$_2$THF required for TS reaction. 5-FU is one of the oldest and still widely used fluoropyrimidine that was first synthesized in 1957 (Heidelberger et al., 1957). It has broad activity towards many solid tumors including colorectal, head and neck, gastric, ovarian, pancreatic and breast tumors. Methotrexate, an

antifolate, was discovered in the 1940s for treatments against Leukemia (Farber & Diamond, 1948; Freireich, 1967; Goldin et al., 1955).

In addition to damage induced by dNTP pool imbalance, TS inhibition can induce cytotoxicity *via* various mechanisms. TS inhibition causes the accumulation of dUMP which leads to the production of dUTP through phosphorylation and subsequent incorporation of uracil into DNA (Caradonna & Cheng, 1980). Ung1-mediated cleavage of the uracil residue creates abasic sites (Lindahl, 1979). Persistence of high [dUTP]/[dTTP] leads to continuous rounds of uracil incorporation into DNA and attempted repair by Ung1, accumulation of DNA breaks and ultimately cell death (Brynolf et al., 1978). In addition to uracil-induced cytotoxicity, several other mechanisms of anti-tumor activity have been observed. For 5-FU, the metabolite fluorodeoxyuridine monophosphate (FdUMP) also binds to and inhibit TS (Danenberg & Lockshin, 1981). FdUTP is also incorporated into DNA albeit to a low extent and repair is initiated by Ung1 excision of the FdUTP nucleoside analog (Caradonna & Cheng, 1980; Danenberg et al., 1981; Seiple et al., 2006). Finally, fluorouridine triphosphate (FUTP) is also misincorporated into RNA leading to cytotoxicity *via* the disruption of RNA synthesis (Glazer & Lloyd, 1982; Herrick & Kufe, 1984). The precise mechanism of cytotoxicity of fluoropyrimidines and antifolates are often debated. However, it is possible that a combination of dUTP and FdUTP in DNA, thymine deficiency, and FUTP in RNA contribute to the cytotoxicity in the clinical setting.

In TS targeted therapies, it has been demonstrated that dUTPase can protect tumor cells from TS-induced cytotoxicity (Tinkelenberg et al., 2002; P. M. Wilson et al., 2012). In fact, dUTPase overexpression has been demonstrated in most cancer cells treated with TS inhibitors, thereby protecting the cells from uracil-derived cytotoxicity as a result of TS inhibition (Kawahara et al., 2009; Longley et al., 2003; P. M. Wilson et al., 2014). Recent studies have demonstrated that targeting dUTPase augments the efficacy of the TS inhibitors, presumably by enhancing the misincorporation of uracil into DNA and

subsequently cell death (Hagenkort et al., 2017; Koehler & Ladner, 2004; Miyahara et al., 2012). The first dUTPase inhibitor, TAS-114, is currently in clinical trials and has demonstrated safety and synergistic effect in combination with current TS inhibitors (Saito et al., 2014). These studies altogether highlight the important role of uracil in cancer therapy.

1.6 Mechanism of Uracil DNA Repair (BER and TCR)

Since the DNA is subject to damage on a regular basis, DNA repair mechanisms are critical for maintaining genome stability. Consequently, the cell has developed several mechanisms to deal with the different types of damages. The importance of DNA repair was highlighted in 2015 Nobel prize in Chemistry to Thomas Lindahl for pioneering studies on the mechanisms of Base Excision Repair, Aziz Sancar for his contributions to understanding Nucleotide Excision Repair and Paul Modrich for Mismatch Repair (Lindahl et al., 2016). Because DNA repair is critical for survival, most of these repair mechanisms are highly conserved from bacteria to humans. These repair mechanisms are, for the most part, efficient but can also be error-prone and inadvertently encourage genome instability.

Abasic sites are the most common, endogenous DNA lesions that if left unrepaired can block transcription or replicative polymerases and induce mutagenesis (Boiteux & Guillet, 2004; De Bont & van Larebeke, 2004; Swenberg et al., 2011). It is estimated that approximately 10,000 abasic sites are formed per human cell per day (Lindahl, 1993; Lindahl & Nyberg, 1972). To prevent mutagenesis, AP sites must be processed and repaired prior to the next round of DNA replication. There are three major pathways the cell utilizes to deal with abasic sites and resume DNA replication: (i) Abasic sites can be processed in an error-free manner by Base- Excision Repair (BER), (ii) If present on the transcribed DNA strand and in the absence of BER, abasic sites are repaired by Transcription-Coupled Nucleotide Excision Repair (TCR) and finally, (iii) If the repair

pathways are overwhelmed, specialized translesion synthesis (TLS) polymerases can be recruited to bypass AP sites in a mutagenic manner.

The BER pathway is highly conserved from bacteria to humans indicating its necessity in maintaining the integrity of the genome (Krokan & Bjoras, 2013). This repair pathway is initiated by the recognition of damaged or incorrect base by DNA glycosylases, that cleave and release the incorrect base creating an apurinic/apyrimidic or Abasic (AP) site. Uracil DNA glycosylase, Ung1, in yeast recognizes and excises uracil residues present in either single or double-stranded DNA leaving an AP site (Percival et al., 1989). The resulting AP site is recognized by AP endonucleases, which bind to and cleave AP sites on duplex DNA leaving a sugar attached to the 5' side of the lesion. In yeast there are two AP endonucleases, Apn1 and Apn2 with Apn1 accounting for >95% of endonuclease activity (Popoff et al., 1990). AP site repair can also be initiated by glycosylase-associated AP lyase in an Apn1-independent manner. In yeast, the AP lyases Ntg1 and Ntg2 can nick the sugar-phosphate backbone on the 3' side of the AP site (Boiteux & Guillet, 2004). Once the DNA backbone is cleaved by either Apn1, Ntg1 or Ntg2, the remaining deoxyribose phosphate residue is removed by a 3' or 5' phosphodiesterase. DNA polymerases then fills the gap created and the remaining nick is sealed by DNA ligase (Krokan & Bjoras, 2013).

When BER is overwhelmed or disabled, Nucleotide Excision Repair (NER) can act as a backup in the repair of AP sites (Swanson et al., 1999; Torres-Ramos et al., 2000). NER is characterized with removing the helix-distorting lesions with the ability to stall replication and transcription such as UV-induced cyclobutane pyrimidine dimers (CPDs), 6-4 photoproducts as well as other bulky adducts (Cadet et al., 2005). However, it was later recognized that NER can also repair small lesions such as AP sites and oxidized bases (Gellon et al., 2001; Scott et al., 1999; Swanson et al., 1999; Torres-Ramos et al., 2000). Loss of NER is associated with the disease Xeroderma Pigmentosum (XP) characterized by increased sensitivity to sunlight and cancer predisposition. NER pathway is subdivided into

two pathways, global genome nucleotide-excision repair (GGR) and transcription-coupled nucleotide excision repair (TCR) (Hanawalt, 2002). In GGR, when there is a damage to DNA, repair proteins directly recognize the distortion to DNA and initiate repair. The lesions are repaired regardless of the transcriptional status of the gene. Rad7 and Rad16 in yeast specifically participate in GGR (Verhage et al., 1996).

TCR is specific in repairing lesions which are present on the transcribed strand of DNA and can stall the elongating RNA polymerase (Hanawalt, 2002). This stalling of the RNA polymerases signals for the recruitment of repair proteins to remove the lesions and allow for continued transcription elongation (Hanawalt, 2002). TCR therefore, promotes the rapid removal of transcription blocking lesions to prevent detrimental effects. There are several mechanisms the cell uses to resolve these lesions that block transcription elongation; backtracking of the stalled RNA polymerase, bypassing the damage or degradation of the RNA polymerase. These mechanisms are discussed in more details in Chapter 4 of this thesis. TCR was first discovered in mammalian cells. The repair of CPDs was shown to be faster in the highly expressed *DHFR* gene than in the less transcribed genes downstream indicating that the lesions on the transcribed DNA sequences were preferentially removed (Bohr et al., 1985; Mellon et al., 1986). In yeast, it was shown that CPDs were removed faster in the actively expressed MATα locus relative to the inactive HMLα locus (Terleth et al., 1990). Philip Hanawalt was first to clearly demonstrate the preferential repair of CPDs on the transcribed strand relative to the non-transcribed strand in the *RPB2* gene (Sweder & Hanawalt, 1992). Rad26 in yeast functions specifically for the initiation of TCR, however, its effect is only partial unlike its human homolog CSB that is required for TCR (Kim & Jinks-Robertson, 2010; Verhage et al., 1996). A second protein Rpb9 has been implicated in TCR of UV-induced damages in a Rad26 independent manner (S. Li & Smerdon, 2002).

Following recognition of the lesion for both GGR and TCR, the downstream repair mechanisms are similar for both subpathways. Dual incisions around the lesion are made, and a single-stranded DNA of ~25-30 nucleotides is released. The remaining gap is filled by DNA polymerase and the nick sealed by a DNA ligase. Rad14 is required for the lesion binding step of both GGR and TCR subpathways of NER and, thus, the loss of Rad14 completely abolishes NER (Guzder et al., 1995).

When DNA polymerases encounters the unrepaired AP sites during replication, specialized translesion synthesis (TLS) DNA polymerases can be recruited to bypass the damages and allow for the continued replication. TLS polymerases have very low fidelity hence the translesion synthesis pathway is highly mutagenic. However, for the cell, introducing mutations in DNA is a preferable risk in comparison to the drastic ramifications that result from the complete replication block. In yeast, Rev1 and Polζ are important in the bypass of AP sites (Gibbs et al., 2005; Kow et al., 2005). Biochemical characterization revealed that Rev1 utilizes a dCMP transferase activity during AP site bypass (Nelson et al., 1996). These results were further confirmed *in vitro* and *in vivo* with experiments that showed that the insertion of C nucleotide across the AP site accounts for ~85% of bypass events at abasic sites (Gibbs et al., 2005; Kim, Mudrak, et al., 2011; Kow et al., 2005; Pryor & Washington, 2011).

1.7 Summary and Significance

DNA is susceptible to both endogenous and exogenous damage which can lead to genomic instability in the form of mutations or rearrangements. Distribution of genome instability events is not random, there are certain hotspots such as highly transcribed genomic loci, that show a larger load of DNA damage (Aguilera, 2002; Kim & Jinks-Robertson, 2012). Until recently, the elevated damages associated with highly transcribed genomic loci was attributed to the single-stranded nature of highly transcribed genomic loci

and the consequently elevated susceptibility to DNA damage. Additionally, the enhanced formation of non-canonical structures such as R-loops or G4-DNA during highly transcribed genomic loci contributes to genomic instability. Recently, a novel mechanism of transcription-associated DNA lesion, not involving chemical modification to DNA, was identified through genetic studies in yeast. This novel category of DNA lesions is composed of non-canonical DNA nucleotide, uracil (Kim & Jinks-Robertson, 2009).

Thesis work described here first aims to elucidate the mechanism by which the mutagenic imbalance in nucleotide composition is achieved at highly transcribed genomic loci leading to specific elevation in the mutagenesis associated with uracil-misincorporation into DNA. The thesis further attempts to expand the current understanding of TCR repair of AP sites, specifically the roles played by some transcription elongation factors.

Changes in the genome can suppress essential cellular processes and lead to defects such as cancer and aging. Understanding the causes of genome instability and how different repair pathways are regulated is therefore, fundamental to understanding both normal, and abnormal developmental processes. The results of this study provide additional insights into how uracil is incorporated into DNA as well as the factors involved in the uracil repair process. In addition to clarifying the molecular mechanisms of transcription-associated genomic instability, mechanistic insights into the regulation of uracil into DNA will enhance our understanding of the mechanisms of action of TS- and dUTPase-inhibitors as well as the mechanisms of resistance to these therapeutics leading to designing of more targeted and effective therapeutic strategies.

Chapter 2

Materials and Methods

2.1 Yeast Strains and Plasmids

Yeast strains were derived from YPH45 (*MATa, ura3-52 ade2-101 trp1Δ1).* The construction of strains containing the *his4Δ::pTET-lys2-TAA* allele was previously described (Kim & Jinks-Robertson, 2010). The *his4Δ::pTET-lys2-TAG* allele was introduced using the pop-in/pop-out two-step allele replacement method to replace the *his4Δ::pTET-LYS2* allele on Chr III using Bgl*II*-digested pSR982. Further gene deletions of the yeast strains containing the *pTET-lys2*-TAA or -TAG allele were carried out using the standard one-step gene disruption method. The subsequent Cre/*lox*P-mediated deletion of the marker gene was carried out as appropriate (Gueldener et al., 2002).

pCDG and pTDG are 2-micron plasmids each with a mutant human UDG encoding sequence under the *pGAL* control with the *TRP1* marker (P. A. Auerbach & Demple, 2010) and were gifts from Dr. Bruce Demple (Stony Brooke School of Medicine, Stony Brooke, NY). The cell-cycle specific Dut1-overexpression plasmids were constructed by digesting p426-GAL1-DUT1 (Kim & Jinks-Robertson, 2009) with BamH*I* and Sac*I* to remove and replace the pGAL promoter with the promoters of yeast genes *CLN2*, *HHO1* or *CLB2*. Sequences of primers used to amplify the promoters of *CLN2*, *HHO1* or *CLB2* genes from the yeast genome were described previously (Prado & Aguilera, 2005; Wellinger et al., 2006). The 537 nt sequence encoding the *CLN2* PEST domain was synthesized through the Invitrogen GeneArt Gene Synthesis service and was inserted into the EcoR*I*/BamH*I* digested pCLN2-DUT1, pHHO1-DUT1, and pCLB2-DUT1 plasmids. Each of these plasmids were further modified by the addition of 3XHA-encoding sequence to the C-terminal ends.

2.2 Mutation and Recombination Rates

Mutation and recombination rates were determined by fluctuation analysis and the method of the median; the 95% confidence intervals were calculated as described

25

previously (Spell & Jinks-Robertson, 2004). Each rate was based on data obtained from 12-24 independent cultures and two independently derived isolates. For the pCDG and pTDG expression experiments, the indicated strains were transformed with either the empty vector pYES2 or pCDG or pTDG and plated selectively on the synthetic complete medium with 2% dextrose media lacking tryptophan (SCD-Trp). Individual colonies were inoculated into a 1-mL SC-Trp culture supplemented with 2% galactose and 1% raffinose. After 4 days of growth at 30°C, appropriate dilutions were plated on SCD-Trp to determine total cell numbers and on SCD-Trp-Lys to determine the number of Lys+ revertants in each culture. For the Dut1-overexpression experiments, the indicated strains were transformed with pRS426 or pGAL-DUT1, pCLN2-DUT1-PEST-HA, pHHO1-DUT1-PEST-HA, or pCLB2-DUT1-PEST-HA. After culuring in SC-Ura media with 2% galactose/1% raffinose or SC-Ura with 2% glycerol/2% ethanol for 4 days at 30°C, the Lys+ revertants were selected on SC-Ura-Lys plates. Where "low transcription" is indicated, doxycyline (2 µg/mL) was added to the media to lower the transcription of *LYS2* gene.

For Dst1, Srs2 and Sub1 experiments, 1 mL of yeast extract-peptone (YEP) medium supplemented with 2% glycerol and 2% ethanol (YEPGE) was inoculated with 250,000 cells from an overnight culture grown in the same medium. For *DEF1*-deleted strains, 500,000 cells from an overnight culture were used to inoculate. Following growth at 30°C in YEP with 2% dextrose (YEPD) for 3 days or YEPGE for 4 days, cells were washed with water and appropriate dilutions were plated either on synthetic, lysine-deficient medium containing 2% dextrose (SCD-Lys) to select Lys+ revertants or on SCD-Leu medium to determine the total number of cells in each culture. *CAN1* forward mutation rates were determined by plating cells on SCD-Arg medium supplemented with 60µg L-canavanine sulfate per mL (SCD-Arg+Can; Sigma). Since fluctuation tests are non-random and unbiased, all data with

26

mutation rates and recombination rates are presented using median and 95% confidence

intervals of the median.

2.3 Mutation Frequency

To determine the mutation frequency following drug treatments, cells were first

grown in YEPD (1% yeast extract, 2% peptone, 2% glucose) medium overnight. The

overnight cultures were diluted to an OD of 0.2 and grown for 4 hrs at 30°C. 5-FU, 4NQO,

or CPT was added to the yeast culture to final concentrations of 10 μM, 0.2 μg/mL and 100

μM, respectively, and incubated at 30°C for 20 hrs with shaking. The cell cultures were spun

and washed twice with sterile water to remove any residual drug and plated on SCD-Lys

plates to select for revertants and on YEPD plates to determine the total cell number.

Colonies were counted after 48 hrs and the mutation frequency was calculated by method of

the median as described above.

2.4 Mutation Spectra

To determine the mutation spectra, individual colonies were used to inoculate 0.3 mL

YEPGE cultures (or SC-Trp with 2% galactose and 1% raffinose for CDG and TDG

expression experiments). After 2 or 3 days of growth at 30°C, an appropriate fraction of

each culture was plated on SCD-Lys (or SCD-Trp-Lys in CDG and TDG expression

experiments). A single Lys$^+$ revertant from each culture was purified on YEPD plates, and

genomic DNA was prepared using a 96-well format in microtiter plates. The l*ys2-TAA or -*

TAG reversion window was amplified using LYSWNF and LYSWNR primers and the PCR

product was sent to Eurofins Genomics for sequencing using LYSEQ primer (primer

sequences listed in Table 2). The rates of A>C and T>G were calculated by multiplying the

proportion of the events by the total Lys$^+$ mutation rate. The 95% confidence intervals for the

rate of A>C or T>G mutation type were calculated first by obtaining the 95% confidence

intervals for the proportion of each mutation type (Vassarstats; (Newcombe, 1998)) and applying the root-sum-squared (RSS) method.

2.5 Quantitative reverse transcriptase PCR (qRT-PCR)

Total RNA was extracted using the standard hot acid phenol method and treated with DNase 1 (New England Biolabs). For qRT-PCR performed in chapter III, cDNA synthesized using the AmfiRivert cDNA Synthesis Platinum Master Mix from GenDepot was used in the subsequent qPCR performed using amfiSure qGreen Master Mix from GenDepot and Biorad CFX Connect instrument. The qRT-PCR conditions were as follows: 45 °C for 10 mins and 95 °C for 2 mins followed by 40 cycles of 95 °C for 5 s, 60 °C for 10 s and 72 °C for 5 s. The primers used for amplification are listed in Table 1. Relative RNA levels were determined by *ΔΔCq* analysis using *ALG9* as the reference gene. For the qRT-PCR performed in chapter IV, SensiFAST SYBR No-ROX one-step kit from Bioline was used and *UBC6* or *ACT1* used as reference gene as indicated.

2.6 Quantitative long-amplicon PCR

The uracil density in DNA was quantified using the long amplicon quantitative real-time PCR approach as described previously with the following modifications (Hunter et al., 2010). DNA samples (5 µg) were digested with 1 unit of UDG (New England Biolabs) for 30 min to remove uracil residues followed by an incubation with EndoVIII (New England Biolabs) for 1 hr at 37°C. After DNA was precipitated and dissolved in water, 100 ng of DNA from each sample was used to carry out qPCR in triplicates. Primers used for the amplification of the yeast *LYS2, CAN1,* and *TDH3* are listed in the Table S1. The amplification was performed using Bioline SensiFAST SYBR No-ROX kit and Biorad CFX Connect Real-Time PCR machine. Cycling parameters were as follows: For the short amplicons: 95°C for 3 min followed by 40 cycles of 95°C for 5 s, 60°C for 10 s and 72°C for

10 s. For the long amplicons: 95°C for 3 min followed by 40 cycles of 95°C for 5 s, 60°C for 10 s and 72°C for 1 min.

2.7 Determination of the induced lesion frequency

The frequency of uracil in DNA was calculated by assuming that the UDG/EndoVIII treatment leads to the strand breaks specifically at the location of uracil, which results in the quantitative loss of the template DNA and consequently the reduced qPCR amplification efficiency. The uracil density at *pTET-LYS2*, *CAN1*, and *TDH3* was inferred from the reduction in the amplification of the UDG/EndoVIII-treated samples relative to the untreated samples when amplifying a large 3- to 4-kb region of each gene. For each gene, qPCR amplification of ~ 100-bp target area was used to normalize for the template DNA loading. Primers used for the amplification of the 100-bp or the 3- or 4-kb regions of the yeast *LYS2, CAN1,* and *TDH3* are listed in the Table 2. Assuming the Poisson distribution of uracil in the large amplicons, the density of uracil was calculated using the following equation where the amplification percent of the large amplicons in the UDG/EndoVIII-treated and in the untreated controls, relative to the amplification of the small ~100-bp amplicons, are represented by *At* and *Au*, respectively.

$$\text{Uracils per 10-kb DNA} \ = \ \frac{-\ln\left(\frac{At}{Au}\right) x\ 10000(bp)}{Size\ of\ long\ amplicon\ (bp)}$$

2.8 Labelling and quantification of uracil

Fluorescent labelling of the uracil-derived AP sites was performed as described previously with minor modifications (Wei et al., 2015). Briefly, genomic DNA was isolated from the *ung1Δ* yeast cells treated with the indicated concentrations of 5-FU or 4NQO and treated with 10 mM methoxyamine to reduce the endogenous AP sites. Then, 5 µg of each

DNA sample was treated with 1 unit of UDG (New England Biolabs) for 30 mins at 37°C to generate the uracil-specific AP sites and labeled by incubation with 5 mM AA3 for an additional hour. Following the addition of Cy5 azide (Lumiprobe) to the final concentration of 0.5 mM and the freshly prepared CuBr/TBTA (1:4 in DMSO/t-BuOH 3:1, 0.5 mM, Sigma), the mixture was shaken at 37°C for 2 hrs. The Cy5/AA3-labeled DNA was purified using ethanol precipitation, heated at 95°C, and transferred to a positively charged nylon membrane using the Bio-Dot microfiltration apparatus (Biorad). The membrane was scanned using the ChemiDoc MP imaging system (Biorad) with a Cy5 filter and quantified using the Image Lab software.

2.9 Cell Synchronization

Synchronization of yeast cells at G1 was carried out by arresting *bar1Δ* cells with α-factor as described previously (Rosebrock, 2017). Briefly, yeast cells were grown in either YEPD or SC-Ura + 2% glucose overnight. The overnight cultures were diluted to an OD_{600} of 0.2 (for YPD) or 0.4 (for SC-Ura + glucose) and grown at 30°C until reaching an OD_{600} of 0.8. The cells were washed twice followed by the addition of α-factor (Sigma) to a final concentration of 50ng/mL and grown for 2 hrs or until ~100% of the cells were unbudded with a typical pear/ schmoo shape characteristic of α-factor arrest. To remove the α factor, cells were washed twice with water and resuspended in the medium supplemented with 50 µg/mL pronase. Cells were collected after every 15 minutes (YEPD) or 20 minutes (SC-Ura + 2% glucose) following the release and used for RNA analysis.

2.10 5-FU and 4-NQO survival assays

The overnight cultures were diluted to an OD_{600} of 0.2 and grown at 30°C for 4 hrs before adding 5-FU, 4NQO or DMSO to a final concentration of 10 µM, 0.2 µg/mL, and 0.1%, respectively. Following incubation at 30°C for 20 hrs with shaking, cells were spun,

washed twice, and plated on YEPD plates. Colonies were counted and the percent survival

of 5-FU- or 4NQO-treated cultures was calculated relative to the DMSO-added cultures.

Table 1: Primers used in this study

Primer name	Sequence 5'-3'	Used for:
LYS2 3KBF	CTTTCAGTGTTACCACATGA	Long amplicon qPCR
LYS2 3KBR	CAAATTTTTCGTTCCAAGTACC	Long amplicon qPCR
LYS2 4KBF	CTTTCAGTGTTACCACATGA	Long amplicon qPCR
LYS2 4KBR	GTTCTATACTTGGCAGTGGAAG	Long amplicon qPCR
CAN1 3KBF	CAGTCCTATTCGGAGATACAG	Long amplicon qPCR
CAN1 3KBR	CTAACTCAGACATTATCGGAAC	Long amplicon qPCR
TDH3 3KBF	GTTCTCACAC GGAACACCAC	Long amplicon qPCR
TDH3 3KBR	GTGGCAGCAAGTGATAAGCAAGC	Long amplicon qPCR
5'BGLF	GAGTAACCGGTGACGATGATATT	Long amplicon qPCR and qRT-PCR
5'BGLR	CATTAAATGACCACGTTGGTTGA	Long amplicon qPCR and qRT-PCR
CAN1F	GAGTTCTGGGTCGCTTCCAT	Long amplicon qPCR and qRT-PCR
CAN1R	GGCACCTGGGTTTCTCCAAT	Long amplicon qPCR and qRT-PCR
TDH3F	CATGGGGTTCTTCCAACGTTG	Long amplicon qPCR and qRT-PCR
TDH3R	GGAAGATGGAGCAGTGATAAC	Long amplicon qPCR and qRT-PCR
DUT1F	GGTTCTGCCACTGCCGCGGG	qRT-PCR
DUT1R	GGCGCAATACGACCGTAGGT	qRT-PCR
ALG9F	CACGGATAGTGGCTTTGGTGAACAATTAC	qRT-PCR
ALG9R	TATGATTATCTGGCAGCAGGAAAGAACTTGGG	qRT-PCR
HTA2F	CCCAGTTGGTAGAGTGCACAG	qRT-PCR
HTA2R	CTCTAGCAGCATTACCAGCC	qRT-PCR
UBC6F	GATACTTGGAATCCTGGCTGGTCTGTCTC	qRT-PCR
UBC6R	AAAGGGTCTTCTGTTTCATCACCTGTATTTGC	qRT-PCR

ACT1F	ATTCTGAGGTTGCTGCTTTGG	qRT-PCR
ACT1R	TGTCTTGGTCTACCGACGATAG	qRT-PCR
LYSWNF	AGCTCGATGTGCCTCATGATAG	Amplification before sequencing
LYSWNR	CATCACACATACCATCAAATCC	Amplification before sequencing
LYSEQ	TAGAGTAACCGGTGACGATG	Sequencing for mutation spectra

Chapter 3

Unscheduled DNA synthesis leads to elevated uracil residues

at highly transcribed genomic loci

3.1 Introduction

Transcription, a fundamental cellular process, can inappropriately pose a serious threat to genome stability. Highly transcribed genomic regions have been reported to be hotspots for mutagenesis and recombination, phenomena referred to as transcription-associated mutagenesis (TAM) and transcription-associated recombination (TAR), respectively (Aguilera, 2002; Datta & Jinks-Robertson, 1995; Herman & Dworkin, 1971; Kim & Jinks-Robertson, 2012).

Several different ways by which transcription promotes genomic instability have been described. First, the single strand DNA generated by DNA strand-separation during transcription is much more chemically labile than double-stranded DNA, leading to the mutations resulting from the spontaneous base modifications such as deamination (Fix & Glickman, 1987; Klapacz & Bhagwat, 2002, 2005; Skandalis et al., 1994). Second, transcription necessitates a change in DNA topology and the ensuing accumulation of both negative and positive supercoils promotes the formation of non-canonical secondary structures such as R-loops, the stable hybrids of transcribed DNA and nascent RNA or G-quadruplex DNAs (G4 DNA), the four-stranded DNA configuration held together by Hoogsteen bonds among guanine bases (Duquette et al., 2004; Wu et al., 1988). These structures leave the non-transcribed DNA in the susceptible single-stranded state and frequently lead to stalling and eventual collapse of the replication fork, which must be restarted/ repaired *via* homologous recombination (Aguilera & Garcia-Muse, 2012; Cox et al., 2000; Huertas & Aguilera, 2003). Finally, because replication and transcription occur on the same template, collisions between the replication fork and transcription machinery are possible (Prado & Aguilera, 2005; Takeuchi et al., 2003). These collisions induce helical stress and can trigger replication fork reversal giving rise to non-canonical DNA structures that can be processed into double stranded breaks (Postow, Ullsperger, et al., 2001; Rudolph et al., 2007).

35

Recently, a novel mechanism of transcription-associated mutagenesis involving the non-canonical DNA nucleotide, uracil, was identified through genetic studies in yeast (Kim & Jinks-Robertson, 2009, 2010). Due to its close structural resemblance to thymine, uracil can be directly incorporated into DNA by DNA polymerases that cannot distinguish between the two bases, leading to U:A base mispairs (Guillet & Boiteux, 2003). Subsequent uracil removal by an uracil-DNA glycosylase (Ung1 in yeast) creates potentially toxic apurinic/apyrimidic (AP) sites (Percival et al., 1989). AP sites are the most prevalent endogenous DNA lesions produced from either spontaneous or DNA N-glycosylases-catalyzed hydrolysis of the base-glycosidic linkage and can act as a potent block to the transcription machinery and the replicative DNA polymerases. (Boiteux & Guillet, 2004). Blocked DNA synthesis can be rescued by the recruitment of translesion synthesis (TLS) DNA polymerases Rev1 and Polζ that together bypass the lesion by typically incorporating a C nucleotide across from the AP site in yeast and metazoans (P. A. Auerbach & Demple, 2010; Lindahl & Nyberg, 1972; Otsuka et al., 2002; Waters et al., 2009). In addition to their misincorporation into DNA by DNA polymerases during replication or repair, uracil in DNA can result from either spontaneous or enzymatic deamination of cytosines to create U:G mispairs, which cause G:C to A:T transitions. Although uracil in DNA leads to deleterious mutations if not properly repaired, it is an essential intermediate in antibody affinity maturation during the adaptive diversification processes at the immunoglobulin genes (Z. Li et al., 2004).

The detrimental mutagenic outcome of uracil in DNA is prevented by Base Excision Repair (BER) pathway, which initiates repair *via* the AP endonuclease-catalyzed cleavage of the sugar-phosphate backbone at the 5' side of the AP lesion. In yeast, the major AP endonucleases are Apn1, which carries out 95% of the repair, and Apn2 (Demple & Harrison, 1994; D. M. Wilson, 3rd & Barsky, 2001). N-glycosylases such as Ntg1 and Ntg2 with the associated AP lyase activity can also create breaks at the DNA backbone adjacent

to the AP site in certain instances such as when the AP endonucleases activity is diminished or overwhelmed (Kim & Jinks-Robertson, 2010; Senturker et al., 1998). Subsequent steps in BER involve the removal of the blocked DNA ends, gap filling by a DNA polymerase and ligation of the remaining nick by a DNA ligase. Although Nucleotide Excision Repair (NER) usually removes bulky, helix-distorting lesions such as UV-induced DNA damage, the transcription-coupled repair (TCR) sub-pathway of NER has been implicated in the repair of AP sites when BER is overburdened or disrupted (Kim & Jinks-Robertson, 2010). TCR specifically repairs the RNA polymerase-stalling lesions in the transcribed strand of active genes encouraging a rapid removal of damage and preventing the accumulation of mutations on the transcribed strand. In yeast, Rad14 is absolutely required for TCR repair of AP sites, while Rad26 and Def1 contribute partially (Kim & Jinks-Robertson, 2010; Owiti et al., 2017).

Deoxyuridine triphosphatase (dUTPase), a ubiquitous enzyme that is essential for viability in both prokaryotic and eukaryotic organisms, catalyzes the conversion of dUTP to dUMP and pyrophosphate (Ppi), thereby reducing the pool of free dUTP and preventing the incorporation of uracil into DNA (Gadsden et al., 1993). Following the conversion of dUTP to dUMP by the dUTPase, Thymidylate Synthase (TS), using tetrahydrofolate as a methyl donor, converts dUMP to dTMP, an intermediate that is required for dTTP synthesis. dUTPase serves two essential functions; maintaining a low intracellular [dUTP]/[dTTP] ratio, thus minimizing incorporation of uracil into DNA, and providing an important intermediate, dUMP, for the *de novo* synthesis of thymidylate. In addition to pyrophosphorlysis of dUTP by dUTPase, the pools of dUMP required for TS reaction is partly supplied by the deamination of dCMP by deoxycytidylate deaminase, Dcd1 in yeast (L. Wang & Weiss, 1992).

Recent studies in *Saccharomyces cerevisiae* demonstrated an increase in uracil-derived mutations following the activation of transcription at a defined reporter gene (Kim & Jinks-Robertson, 2009). These mutations were highly elevated by the disruption of BER and

eliminated by the deletion of *UNG1* gene or the overexpression of yeast dUTPase, Dut1, suggesting that the uracil-dependent mutations result from the AP sites generated by excision of uracil incorporated into DNA. Repressing the transcription of the reporter gene lowered the uracil-associated mutations, suggesting a link between the extent of uracil incorporation into DNA and the process of transcription. However, a clear demonstration of higher uracil content at highly transcribed genomic loci is still lacking. In addition, how the nucleotide composition is affected by transcription remains to be deciphered. In the current study, I used the mutagenesis reporter with a regulatable promoter to further investigate the link between active transcription and uracil DNA content. The results show that there are significantly more uracil residues present at the highly transcribed genomic locus and that the DNA glycosylase activity is slightly enhanced when transcription is elevated. Furthermore, I show that the DNA repair synthesis, induced by DNA damaging agents, leads to an increase in uracil residues present in DNA and that the overexpression of Dut1 in G1- and G2-phases of the cell cycle leads to a significant reduction in the uracil-dependent mutations at the highly transcribed site. Overall, the data in this chapter strongly support a model in which various transcription-associated damages induce unscheduled DNA synthesis, particularly in G1 and G2, subsequently leading to the elevated uracil residues at highly transcribed genomic loci.

3.2 Results

Previously, a mutation reporter system with a tetracycline-regulatable promoter *(pTET)* was used to examine the effect of transcription on the mutagenesis in BER/NER-deficient yeast strains (Kim & Jinks-Robertson, 2010). Specifically, in the *pTET-lys2-TAA* system, the in-frame insertion of "TAA" stop codon disrupting the *LYS2* ORF renders the yeast cells auxotrophic for lysine (Lys⁻). Those mutations abrogating the TAA stop codon and allowing the translation read-through are then selected by the reversion to Lys+

38

phenotype. In the repair-deficient, *apn1, apn1 rad14,* and *apn1 ntg1 ntg2* backgrounds, I observed a dramatic elevation in mutagenesis, particularly of A>C and T>G transversions, when the *pTET* promoter was activated. These mutations were mostly eliminated when Ung1 was disabled, or when transcription was repressed by the addition of the tetracycline analog doxycycline (+DOX). The rate of A>C and T>G mutations was also significantly reduced when the dUTPase-encoding *DUT1* gene was highly overexpressed to reduce the level of free dUTP available for incorporation into the genome. When uracil is misincorporated in place of thymine in DNA and subsequently excised by Ung1 to generate AP sites, the net result of Rev1/Polζ-dependent translesion bypass synthesis, inserting predominantly C nucleotides opposite AP sites, is expected to be A>C or T>G mutations. Therefore, the highly elevated rate of A>C and T>G mutations observed under high transcription conditions at the *pTET-lys2-TAA* reporter is originating from the uracil residues in DNA that are excised by Ung1 to generate the mutagenic AP sites.

There are two possible explanations for the enhanced uracil-dependent mutations when transcription is activated: (i) highly active transcription leads to elevated dUTP incorporation into the genome or, (ii) rather than affecting the number of uracil residues incorporated into the DNA, active transcription leads to the enhanced uracil glycosylase activity. These two hypotheses, each leading to the accumulation of mutagenic AP sites specifically at highly transcribed genes, are summarized in Figure 3.
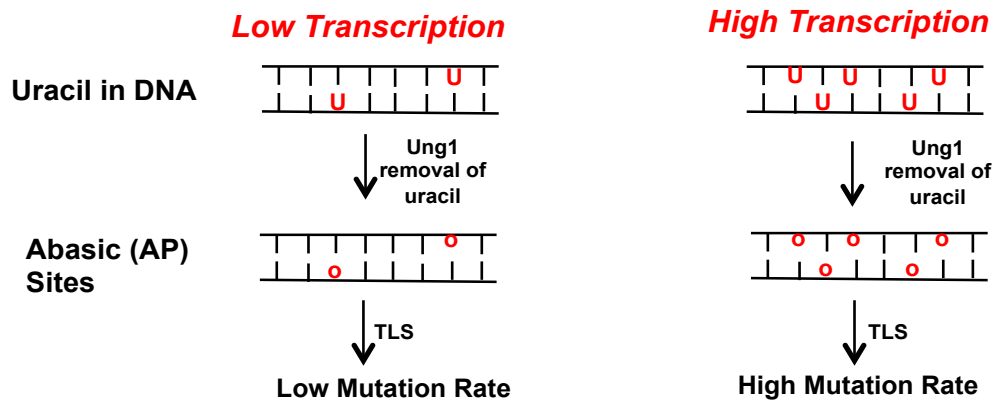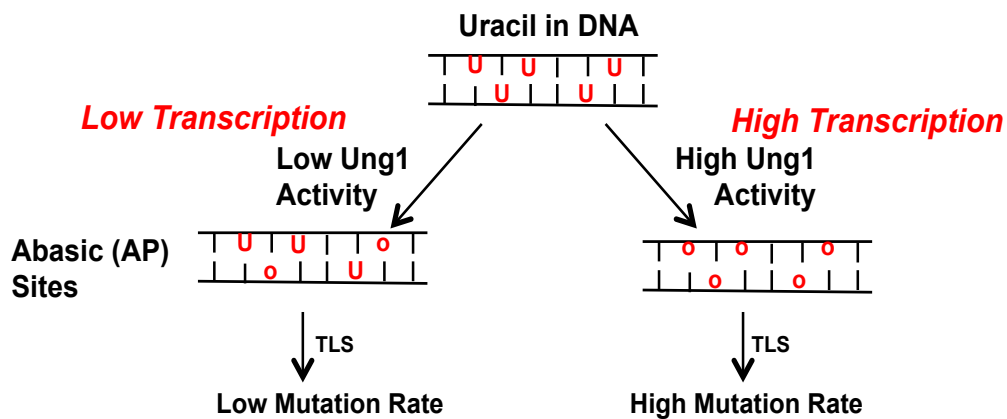
Figure 3: Alternative hypotheses for the transcription-associated elevation in uracil-dependent mutations.

A) Higher number of uracil residues are present at highly transcribed loci resulting in the higher rate of uracil-dependent mutations. **B)** Uniform number of uracil residues are present regardless of the transcription level; the uracil DNA glycosylase activity is enhanced at highly transcribed genomic loci. In both instances, the uracil residue (***U***) is recognized and excised by Ung1 to create AP sites (***o***). The resulting abasic site is bypassed by translesion synthesis (***TLS***) polymerases in a mutagenic manner.

3.2.1 CDG glycosylase causes uracil- and cytosine- derived mutations.

In order to investigate whether the main cause underlying the transcription-associated elevation in uracil-dependent mutations is the higher uracil density or the enhanced uracil excision, I used a modified human uracil DNA glycosylase (UDG). The mutant enzyme, hereon referred to as CDG, was generated by introducing Asn204 to Asp mutation in the substrate binding pocket of the UDG (P. Auerbach et al., 2005; Kavli et al., 1996). CDG is able to excise unmodified cytosine residues from oligonucleotide substrates *in vitro.* Expression of CDG in the TLS-proficient yeast cells was previously shown to induce the accumulation of A:T > C:G and G:C > C:G transversions, resulting from the excision of uracils and cytosines, respectively. I expressed CDG to generate AP sites through excision of cytosines and uracils in yeast cells containing the mutation reporter *pTET-lys2-TAG*. This reporter contains the in-frame TAG stop codon inserted into the *LYS2* ORF and mutations at this stop codon is required for the reversion to Lys$^+$ phenotype. The *pTET-lys2-TAG* reporter can be transcribed at a high or low level by the absence or presence of doxycycline in the media, respectively. As illustrated in Figure 4A, a A>C or T>G mutation is expected when an AP site is generated from the excision of uracil that is in place of thymine. And a G>C mutation is expected when an AP site is generated from the excision of cytosine (Fig. 4B). I expressed CDG in the *apn1∆ ung1* or *apn1 rad14 ung1* strain backgrounds, where the endogenous *UNG1* is deleted so that all uracil- or cytosine-associated mutations are resulting from the activity of ectopically expressed CDG. The rates of uracil-dependent, A>C and T>G mutations as well as cytosine-dependent, G>C mutations were determined under the high (no DOX) and low (+DOX) transcription conditions to determine the effect of transcription on these mutations. Since the number of cytosine residues in DNA should not be affected by the level of transcription, the rate of mutations caused by the AP sites generated by the cytosine excision (G>C mutations) should not change whether under the high or low transcription conditions, unless the enzymatic efficiency of cytosine excision by

41

CDG is affected by activated transcription. When CDG was expressed in the *apn1 ung1* strain, the overall mutation rate was ~7-fold higher under the high transcription conditions than under the low transcription conditions (Fig. 4C and Table 2). The mutation spectra showed a ~10-fold increase in the rate of uracil-dependent, A>C and T>G mutations but only a ~1.7 -fold increase in the rate of the cytosine-dependent, G>C mutations (Fig. 4D and 4E). In a BER/NER- deficient, *apn1 rad14 ung1* strain*,* the CDG expression under high transcription conditions led to ~10- and ~ 2.5-fold increases in the rates of uracil- and cytosine-dependent mutations, respectively, compared to the CDG expression under low transcription conditions. Overall, the active transcription resulted in the significantly greater increase in the uracil-associated mutations compared to the cytosine-associated mutations. Altogether, these data suggest that the elevated glycosylase efficiency is only a minor factor contributing to the transcription-dependent elevation in the uracil-associated mutations.

A. Uracil-dependent mutations

```
———————— TAG ————————→
←———————— ATC ————————
```
↓ DNA polymerase incorporation of uracil
```
———————— UAG ————————→
←———————— AUC ————————
```
↓ CDG removal of uracil
```
———————— oAG ————————→
←———————— AoC ————————
```
↓ TLS bypass of AP site
```
———————— oCG ————————→
←———————— CoC ————————
```
↓ Replication
```
———————— GCG ————————→
←———————— CGC ————————
```
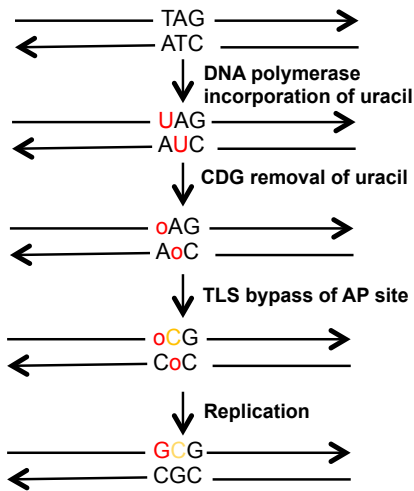
Resulting mutations
T>G (uracil in top strand)
A>C (Uracil in bottom strand)

B. Cytosine-dependent mutations

```
———————— TAG ————————→
←———————— ATC ————————
```
↓ CDG removal of cytosine
```
———————— TAG ————————→
←———————— ATo ————————
```
↓ TLS bypass of AP site
```
———————— TAC ————————→
←———————— ATo ————————
```
↓ Replication
```
———————— TAC ————————→
←———————— ATG ————————
```

Resulting mutations
G>C (cytosine in bottom strand)

C. Overall mutation rates

Legend: High trx, Low trx
Y-axis: Mutation Rate X 10^{-8} (0–50)
X-axis: apn1 ung1, apn1 rad14 ung1

D. Uracil-mutation rates

Legend: High trx, Low trx
Y-axis: Mutation Rate X 10^{-8} (0–30)
X-axis: apn1 ung1, apn1 rad14 ung1
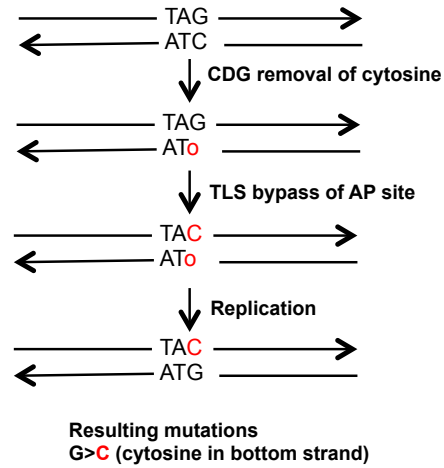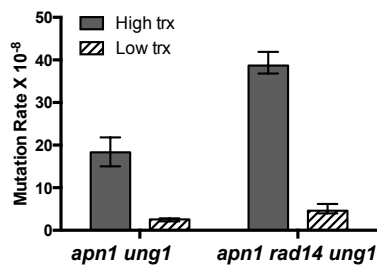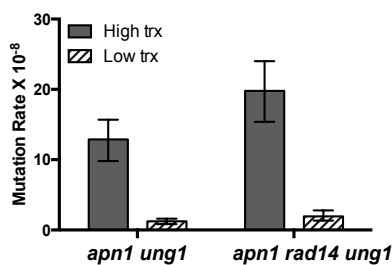
E. Cytosine-mutation rates

Legend: High trx, Low trx
Y-axis: Mutation Rate x 10^{-8} (0–10)
X-axis: apn1 ung1, apn1 rad14 ung1
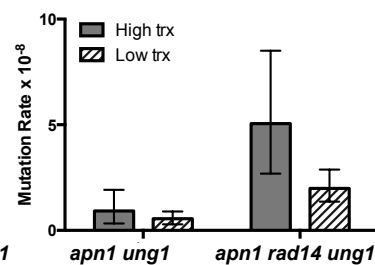
Figure 4: The effect of CDG expression on uracil- and cytosine-derived mutations

Schematic representations of the mutations originating from **A)** the uracil and **B)** the cytosine excised by the CDG glycosylase. Uracil or cytosine residue is removed by CDG creating an AP site. The AP site (**o**) is bypassed by TLS polymerases inserting predominantly C across the AP site. **C)** Overall mutation rates in yeast strains with CDG glycosylase-expressing plasmid under high-transcription (no doxycycline) or low-transcription (doxycycline added) conditions. Error bars represents 95% confidence intervals. **D)** and **E)** Rates of the uracil-dependent A>C and T>G mutations and the cytosine-dependent G>C mutations, respectively.

3.2.2 CDG glycosylase activity in the absence of Topoisomerase 1 or RNase Hs.

To further probe the effect of transcription on the activity of glycosylases, I repeated the CDG expression in *apn1 top1 ung1* and *apn1 rad14 top1 ung1* strains and determined the mutation rates under the high and low transcription conditions. Topoisomerase 1 (Top1) functions to relieve topological stress, including transcription-associated supercoils, by creating transient strand breaks and then rejoining DNA strands (J. C. Wang, 2002). Deletion of *TOP1* leads to an accumulation of negative helical stress, particularly in the highly transcribed areas. Upon CDG expression in the cells of *apn1 top1 ung1* background, the rate of overall mutation at the *pTET-lys2-TAG* reporter was 7.9-fold higher under high transcription conditions than under low transcription conditions (Fig. 5A). This is comparable to the 7.1-fold difference between the rates of CDG-induced mutations under high and low transcription conditions in *apn1 ung1* background.

Additionally, I sequenced the *pTET-lys2-TAG* allele in the Lys+ revertants to identify the specific nucleotide substitutions. In *apn1 top1 ung1* background, CDG-expression under the high transcription conditions resulted in 16.3- and 27.6-fold increases in A>C and T>G mutations, respectively, compared to the CDG-induced mutations occurring under low transcription conditions (Figure 5B, 5C and Table 2). These are substantially higher than the high-transcription associated elevation in A>C and T>G mutations observed in *apn1 ung1* background, which were 4.2- and 11.2-fold increases, respectively (Table 2). Additionally, when comparing the CDG-induced mutations, the transcription-dependent fold elevation (high/low) of A>C, T>G, and C>G mutations were significantly higher in *apn1 rad14 top1 ung1* than in *apn1 rad14 ung1*. As illustrated in Fig. 5A, the uracil-derived A>C and T>G mutations result from the excision of uracil present on the transcribed (bottom) strand and the non-transcribed (top) strand, respectively. Since *TOP1* deletion resulted in the higher transcription-dependent elevation of both A>C and T>G mutations, the overall DNA topology changes in the Top1-deficient cells appear to increase the access of

glycosylase to the uracil residues regardless of whether they are located on the transcribed or non-transcribed strand.

To determine whether uracil residues in the single-stranded DNA are more easily accessed by DNA glycosylases, I determined the rates of uracil- and cytosine-dependent mutations in the *apn1 rnh1 rnh201 ung1* strain following CDG expression. *RNH1* and *RNH201* encode the RNaseH enzymes that degrade RNA hybridized to DNA (Cerritelli & Crouch, 2009). During transcription, RNA-DNA hybrids form when the nascent RNA anneals to the template, transcribed DNA strand leaving the non-transcribed DNA strand single-stranded. In the absence of RNaseH enzymes, the transcription-associated RNA-DNA hybrids persist and accumulate to form "R-loops" (Aguilera & Garcia-Muse, 2012). In *apn1 rnh1 rnh201 ung1* strain backgrounds, the rate of overall mutations induced by the CDG-expression at the *pTET-lys2-TAG* reporter under the high transcription conditions was not significantly greater than the rate in *apn1 ung1* (Fig. 5D). And the transcription-dependent increase in the rates of A>C or C>G mutations was not significantly different in these two backgrounds (Fig. 5E, 5F and Table 2). However, the transcription-dependent fold-increase (high/low) in the rate of T>G mutation was 39.1 in *apn1 rnh1 rnh201 ung1* compared to 11.2 in *apn1 ung1* (Table 2). The greater effect of the disruption of RNase H enzymes on the T>G mutations relative to the nominal effect it had on the A>C or C>G mutations indicate that CDG has a greater access to the target (U or C) on the single-stranded non-transcribed strand of DNA.
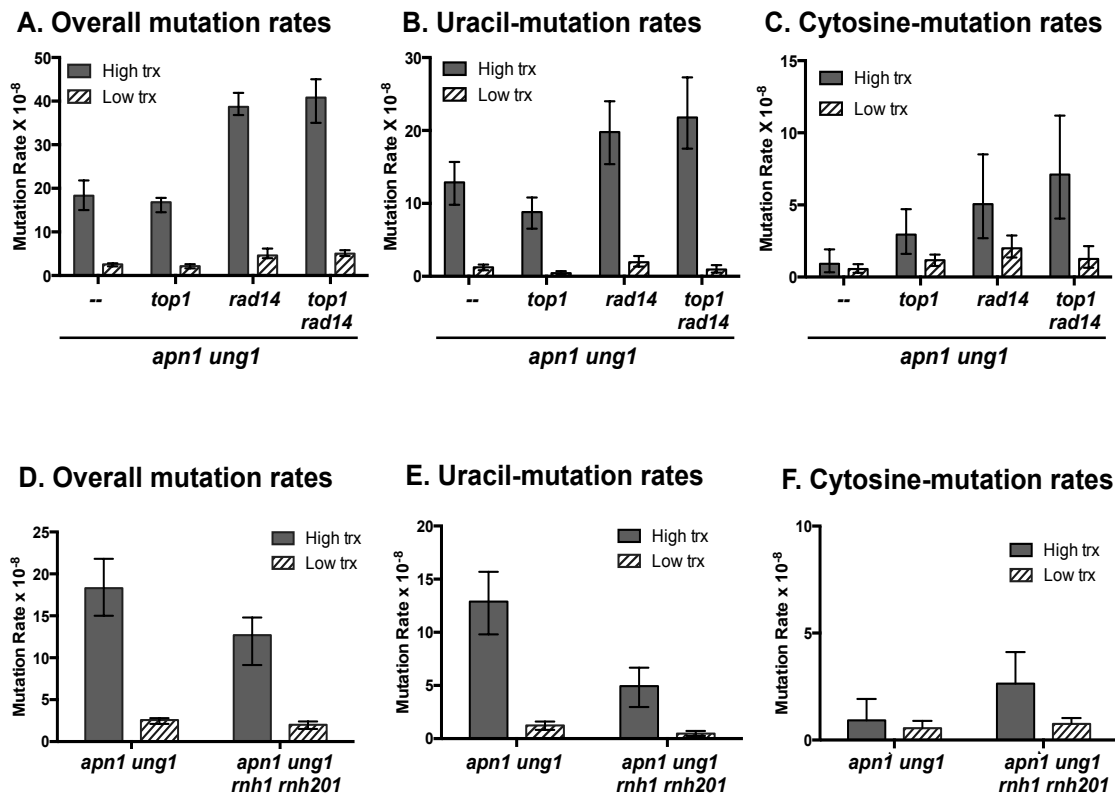
Figure 5: The effect of CDG expression on uracil- and cytosine-derived mutations in the absence of Topoisomerase 1 or RNase Hs.

A) and **D)** Overall mutation rates in yeast strains with CDG glycosylase-expressing plasmid in under high-transcription (no doxycycline) or low-transcription (doxycyline added) conditions. Error bars represents 95% confidence intervals. **B), C), E) and F)** Rates of the uracil-dependent A>C and T>G mutations and the cytosine-dependent G>C mutations.

Table 2: Mutation Rates, Mutation Spectra and Fold Change following CDG expression

Relevant Genotype	Mutation Rate (x10 $^{-8}$)	High rate* (95%CI)	N*** (High)	Low rate** (95% CI)	N*** (Low)	Rate ratio (High/Low)
apn1 ung1	Overall	18.3 (15-21.8)	91	2.57 (2.11-2.8)	56	7.12
	A>C	3.3	25	0.78	17	4.2
	T>G	5.14	39	0.46	10	11.2
	A>C + T>G	12.9	64	1.24	27	10.4
	G>C	0.92	7	0.55	12	1.67
	Other	4.02	20	0.78	17	5.15
apn1 ung1 rad14	Overall	38.7 (36.8-41.9)	92	4.61 (3.99-6.17)	67	8.4
	A>C	18.1	43	1.58	23	11.5
	T>G	1.68	4	0.34	5	4.94
	A>C + T>G	19.8	47	1.93	28	10.24
	G>C	5.05	12	1.99	29	2.53
	Other	13.9	33	0.68	10	20.4
apn1 ung1 top1	Overall	16.8 (14.5-17.8)	80	2.14 (1.63-2.62)	55	7.85
	A>C	4.41	21	0.27	7	16.3
	T>G	4.41	21	0.16	4	27.56
	A>C + T>G	8.82	42	0.43	11	20.8
	G>C	2.94	14	1.17	30	2.52
	Other	5.04	24	0.97	25	5.19
apn1 ung1 rad14 top1	Overall	40.8 (35-45)	86	5.03 (4.49-5.81)	65	7.95
	A>C	18.03	38	0.31	4	58.2
	T>G	3.8	8	0.62	8	6.12
	A>C + T>G	21.8	46	0.93	12	23.44
	G>C	7.11	15	0.93	12	7.64
	Other	11.9	25	3.17	41	3.74
apn1 ung1 rnh1 rnh201	Overall	12.7 (9.14-14.8)	77	2.01 (1.52-2.41)	75	6.32
	A>C	1.81	11	0.40	15	4.5
	T>G	3.13	19	0.08	3	39.13
	A>C + T>G	4.95	30	0.48	18	10.3
	G>C	2.64	16	0.75	28	3.52
	Other	5.11	31	0.78	29	6.55

*Rates of mutations under high transcription conditions (no doxycycline)
**Rates of mutations under low transcription conditions (2µg/mL doxycycline)
***Rates of specific mutation types were calculated by multiplying the overall rate by the proportion of the relevant mutation type in the corresponding spectrum.

3.2.3 The long-amplicon qPCR experimental strategy to quantify uracil residues in DNA.

To determine whether there is a correlation between the level of transcription and the uracil content in DNA, I sought to measure the density of uracil in DNA at specific genomic sites using the long-amplicon quantitative PCR approach. This method, which measures the reduction in the amplification efficiency resulting from the polymerase-blocking damage to the template DNA, was previously used to quantify the damage in the mitochondrial DNA of various vertebrate species (Santos et al., 2006) and the uracil residues at the mouse immunoglobulin loci in B lymphocytes (Maul et al., 2011). Genomic DNA isolated from yeast cells lacking the endogenous Ung1 was treated *in vitro* with recombinant UDG to create AP sites specifically at the uracil residues. The resulting AP site was then converted into a single-strand break by the Endo VIII-treatment. The presence of uracil is measured as the relative loss of qPCR signal or the relative reduction in the amplification efficiency when the UDG/Endo VIII-treated DNA is amplified compared to when an untreated DNA sample is amplified. First, to validate this strategy, I measured uracil content in *ung1* yeast cells treated with 0, 1, 5 and 10 µM 5-fluorouracil (5-FU), an inhibitor of thymidylate synthase (TS) that leads to an increase in the cellular dUTP pool and the accumulation of uracil residues in DNA. The DNA samples purified before or after UDG/Endo VIII-treatment were used as the template in qPCR reactions with *LYS2* primers targeting ~100-bp, 3-kb, and 4-kb regions (Fig. 6A). The Ct values from the "LYS2 100-bp" primers were used as normalizing controls, under the assumption that it is highly unlikely there is a significant number of the polymerase-blocking lesions within the approximately 100-bp region targeted by these primers. For the DNA sample from cells treated with no or 1µM 5-FU, the amplification of the UDG/EndoVIII-treated DNA were about 75% of the untreated DNA. For the samples treated with 5 and 10 µM 5-FU, there was a dose-dependent reduction in the amplification efficiency compared to that of untreated samples (Fig. 6A). As expected, the amplification

of the 4-kb amplicon ("LYS2 4kb" primer set) had a significantly reduced amplification in comparison to the 3-kb amplicon ("LYS2 3kb" primer set) since the larger amplicon size increases the likelihood of a lesion being encountered during PCR. A Poisson equation was used to further estimate the average uracil frequency at the *LYS2* locus from the long-amplicon amplification efficiency and observed that the uracil frequency is significantly elevated when treated with higher concentrations of 5-FU (Fig. 6B). The frequency of uracil in DNA was about 1 per 10 kb in cells treated with 0 or 1 µM 5-FU and was elevated to >2 per 10 kb with the treatment of 5 or 10 µM 5-FU. Overall, these results show that the long-amplicon qPCR strategy can be used to quantitatively measure the uracil frequency in yeast genomic DNA.
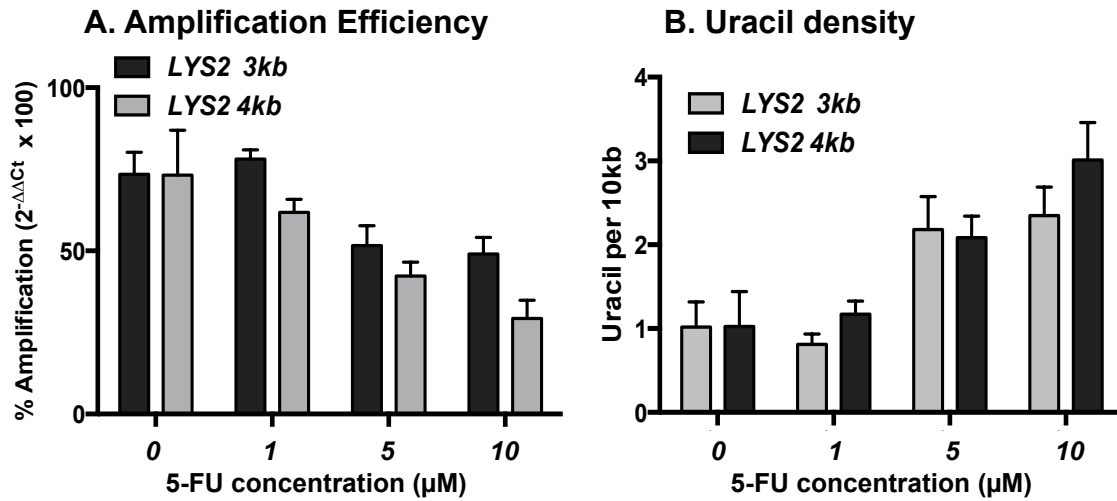
Figure 6: Validation of the long-amplicon qPCR approach

A) Relative percent amplification of the genomic DNA samples from the yeast cells treated with the indicated concentration of 5-FU for 24 hrs. **B)** The uracil-density of the genomic DNA samples calculated from the relative percent amplification shown in **A).**

3.2.4 Uracil residues enhanced at highly transcribed genomic loci.

The long amplicon qPCR method was extended to determine the endogenous uracil levels at several genomic loci. In order to first determine the level of transcription of the genes where the uracil frequency was measured, I isolated RNA from *ung1* and *ung1 dcd1* strains grown in the presence or absence of doxycycline. *DCD1* is a gene that encodes a deoxycytidylate deaminase (Dcd1), which converts dCMP to dUMP, a substrate for the dTTP production (Gadsden et al., 1993). The deletion of *DCD1* increases the [dUTP]/[dTTP] ratio, leading to the increased incorporation of uracil into DNA. As expected from the elevated level of uracil-incorporation into DNA, the deletion of *DCD1* in *apn1* strain led to a two-fold increase in the rate of mutation at the *pTET-lys2-TAA* reporter under high transcription conditions (Fig. 7A). Following mRNA extraction from *ung1* and *ung1 dcd1* strains, I performed RT-qPCR to determine the expression levels of *pTET-lys2-TAA, CAN1* and *TDH3* in the presence and absence of doxycycline. While the expression levels of *CAN1* and *TDH3* genes, which are not regulated by the *pTET* promoter, was not affected by the presence of doxycycline, the level of *pTET-lys2-TAA* transcripts from no DOX samples was ~170-fold higher than that from +DOX samples (Fig. 7B).

A. Overall mutation rate

B. mRNA expression levels +/- doxycyline

C. Primer sets for long- QPCR

Ch III — LYS2 (4Kb)

Ch V — CAN1 (1.77Kb)

Ch VII — HIP1 — PDX1 — TDH3 (1Kb)

3Kb
4Kb
~90-120 bp
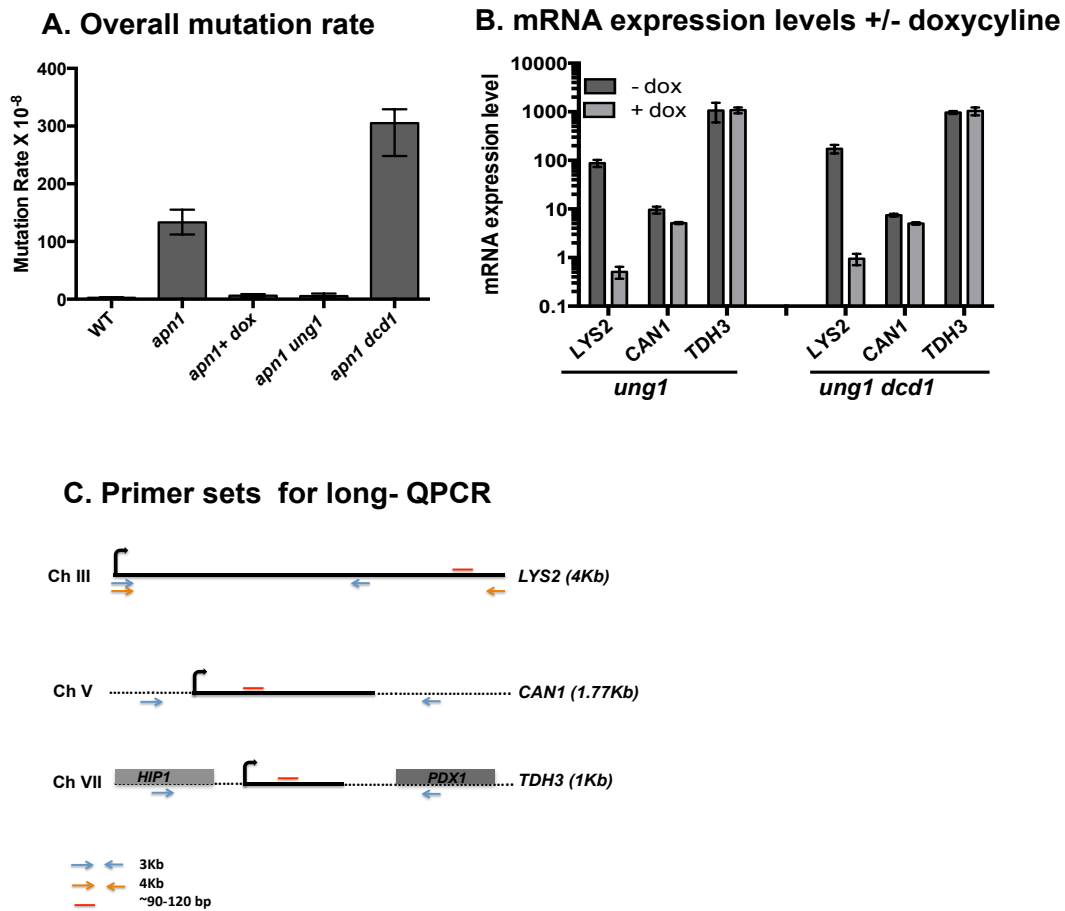
Figure 7: The rate of Lys+ mutations in *apn1 dcd1* strain and expression levels of target genes and primer locations used in the long-amplicon qPCR

A) Overall mutation rates of the indicated yeast strains under the high transcription conditions (no doxycycline). Error bars indicate 95% confidence intervals. **B)** The expression level of the indicated genes in the presence (+) or absence (-) of doxycycline in *ung1* or *ung1 dcd1* strains as determined by qRT-PCR with the *ALG9* gene as the control. Error bars indicate standard deviations and all measurements are from N=6. **C)** The locations of primer sets used in the long-amplicon qPCR are indicated. The sequences of the primers are listed in Table 1.

For the long-amplicon qPCR analysis, DNA samples were prepared from *WT*, *ung1,* and *ung1 dcd1* strains grown in the presence and absence of doxycycline. Under the conditions where the transcription of *pTET-lys2-TAA* is repressed (+ DOX), the density of uracil in DNA as inferred from the amplification carried out with the "LYS2 3kb" primers were 0.08, 0.93, and 1.1 per 10 kb in *WT*, *ung1,* and *ung1 dcd1,* respectively (Fig. 8A). Under high transcription conditions (no DOX), the uracil-density was significantly higher in all three strain backgrounds with 0.45, 1.5, and 2.1 per 10 kb in *WT*, *ung1,* and *ung1 dcd1* strains, respectively. In WT cells, the activity of endogenous Ung1 accounts for the relatively low level of uracil in DNA in WT as determined by the long-amplicon PCR approach. When "LYS2 4kb" primers were used for the analyses on the same DNA samples, the uracil-densities calculated were 0.53, 1.3, and 2.1 per 10 kb in *WT*, *ung1,* and *ung1 dcd1,* respectively, under high transcription conditions and statistically the same as those calculated using "LYS2 3kb" primers (Fig. 8B). For the *CAN1* or *TDH3* loci, a set of primers targeting a ~3 kb region encompassing each gene was used in the long amplicon PCR with another set of primers targeting a ~100 bp region within each gene as the control PCR reactions (Figure 7C). In a manner similar to the *pTET-lys2-TAA,* the density of the uracil in DNA were highest in the *ung1 dcd1* background and lowest in WT background (Fig. 8C and 8D). Unlike the *pTET-lys2-TAA,* the addition of doxycycline did not affect the uracil density at *CAN1* or *TDH3* loci. In the *ung1* background, the uracil density was about 3-fold lower at *CAN1* than *TDH3*. However, in the *ung1 dcd1* background, the uracil density derived by the long-amplicon PCR approach at *TDH3* and *CAN1* were not significantly different. In both *ung1* and *ung1 dcd1* backgrounds, *TDH3* is transcribed at ~100-fold higher rate than *CAN1* (Fig. 7B).
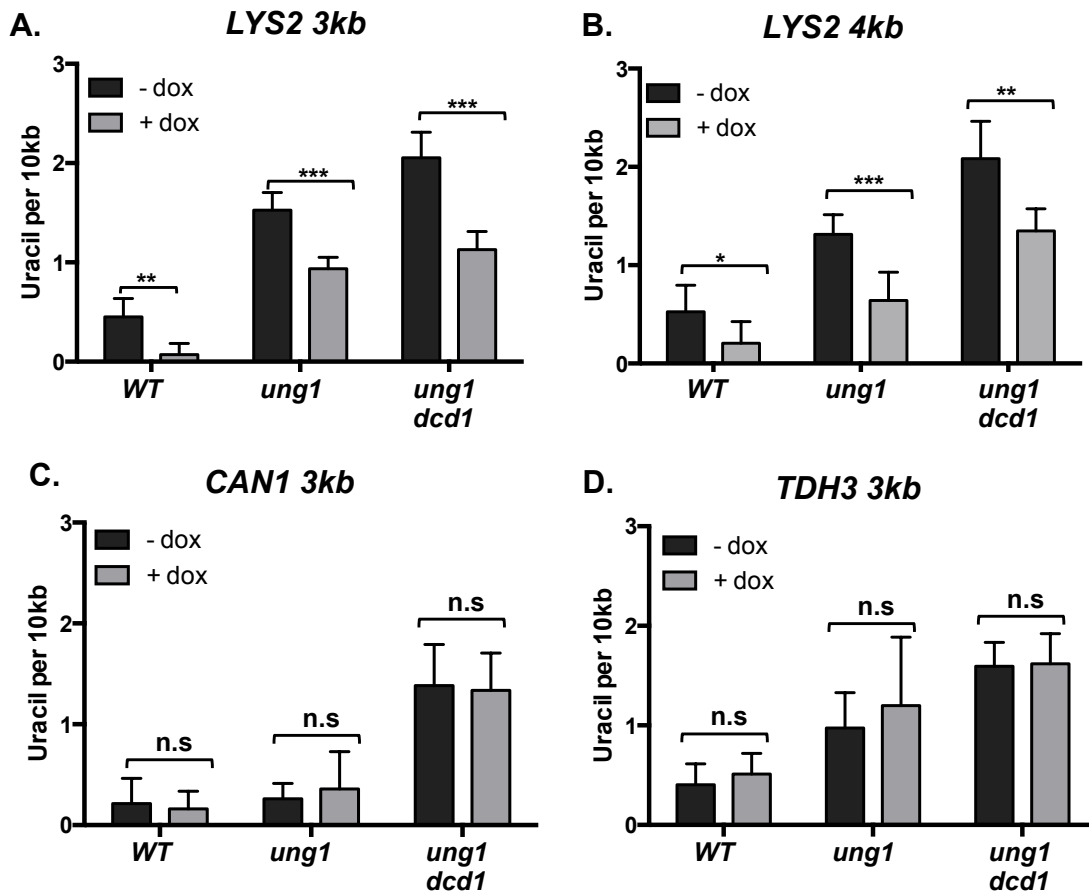
Figure 8: Uracil-density measured by the long-amplicon qPCR approach

A) – D), the genomic DNA samples isolated from the indicated yeast strains grown in the absence (-dox) or presence (+dox) of 2 µg/mL doxycycline in the media were used for qPCR and the calculation of uracil-density. **A)** Uracil density at *LYS2* calculated from qPCR using "LYS2 3 kb" primers. **B)** Uracil density at *LYS2* calculated from qPCR using "LYS2 4 kb" primers. **C)** Uracil density at *CAN1* calculated from qPCR using "CAN1 3 kb" primers. **D)** Uracil density at *TDH3* calculated from qPCR using "TDH3 3 kb" primers. For **A) to D),** error bars represent standard deviation (*$P < 0.05$, **$P < 0.005$, ***$P < 0.0005$, n.s - not significant; unpaired student t-test) and all measurements are from N=6.

3.2.5 Uracil-dependent mutations increase when repair synthesis is induced by 4NQO or CPT.

We postulate that spontaneous DNA damage associated with high transcription lead to cycles of unscheduled DNA synthesis (UDS) leading to the increase in uracil-associated mutations. In order to test this hypothesis, UDS was induced using three different types of DNA damaging agents, and the mutation frequency of uracil-dependent mutations calculated. First, I treated the cells with 5-fluorouracil (5-FU), which imbalances the [dUTP]/[dTTP] ratio and thus enhance the incorporation of uracil residues in DNA. The frequency of mutations at the reporter was determined in *WT*, *ung1*, *apn1,* and *apn1 ung1* background cells treated with 10μM 5-FU. Compared to the DMSO control, the 5-FU treatment led to a remarkable elevation in the rate of *pTET-lys2-TAA* mutations in the BER-deficient *apn1Δ* strain but not in the BER-proficient *WT* or *ung1Δ* strain. As would be expected of mutations arising from the uracil-derived AP sites, the rate of mutations was far reduced in the uracil-excision incapacitated *apn1 ung1* compared to that in *apn1.* When the Lys[+] mutants were sequenced, the majority (67/72) in the *apn1* strain were A>C or T>G, as expected of uracil-associated mutations (Fig. 9A and 9B).

In order to determine whether the repair induced by a broad spectrum of DNA damage can increase the uracil-incorporation into DNA and the consequent uracil-derived mutations, the frequency of the *pTET-lys2-TAA* mutations was measured in yeast cells treated with two DNA damaging agents without previously reported effects on the dUTP/dTTP metabolic pathway -camptothecin (CPT) and 4-nitroquinoline 1-oxide (4NQO). CPT is a Top1 inhibitor that traps the Top1-DNA cleavage complex and is known to elevate recombination and copy number variations in the eukaryotic genomes (Andersen et al., 2015; Balestrieri et al., 2001). Following CPT-treatment, the overall mutations were elevated by ~3 fold in *WT*, *ung1, apn1,* and *apn1 ung1* backgrounds (Fig. 9C). The mutation spectra and the frequency of the specific type of mutations were determined by sequencing the

pTET-lys2-TAA reporter in Lys+ revertants. In *apn1* cells, the increase in the rate of A>C and T>G mutations due to CPT-treatment was statistically significant with the 95% confidence intervals not overlapping (Fig. 9D). In *apn1 ung1* cells, the frequency of A>C and T>G mutations in the CPT-treated samples was slightly higher than the untreated control but the difference was not statistically significant. Also, A>C and T>G mutations in the CPT-treated *apn1 ung1* were significantly reduced compared to those in *apn1,* indicating that the mutations resulting from Ung1-mediated excision of uracil in DNA do occur at a significant level upon CPT-treatment.

4NQO is a mutagenic heterocyclic chemical that forms covalent bulky adducts to dG or dA, which are predominantly repaired by NER (Ikenaga et al., 1975). Following treatments with 4NQO, the overall mutation frequency at the *pTET-lys2-TAA* reporter was elevated ~15- to 50-fold in *WT*, *ung1, apn1,* and *apn1 ung1* backgrounds (Fig. 9E). The mutation spectra revealed that a majority of the mutations elevated by the 4NQO-treatment were A:T > T:A transversions (Fig. 10), the type of mutations that had previously been associated with 4NQO (Ryu et al., 1999). The uracil-dependent mutations (A>C and T>G) were significantly elevated only in the *apn1* cells treated with 4NQO (Fig. 9F). Similar to the observation in the CPT-treated cells, the frequency of 4NQO-induced A>C and T>G mutations in the *apn1 ung1* background was significantly lower than that in *apn1,* indicating that uracil-incorporation into DNA and the resulting mutations also occur at a significant level following the 4NQO-treatment.

56

Figure 9: Mutation frequencies following the 5-FU-, 4NQO-, and CPT-treatment.

A), **C),** and **E)** The frequencies of overall Lys+ mutations following treatments with 5-FU (10 µM), CPT (100 µM), or 4NQO (0.2 µg/mL), respectively, for 24 hrs. **B), D),** and **F)** The frequencies of the uracil-dependent A>C and T>G mutations following treatments with 5-FU, CPT and 4NQO, respectively. Error bars indicate 95% confidence intervals.

A. A>T mutation frequency
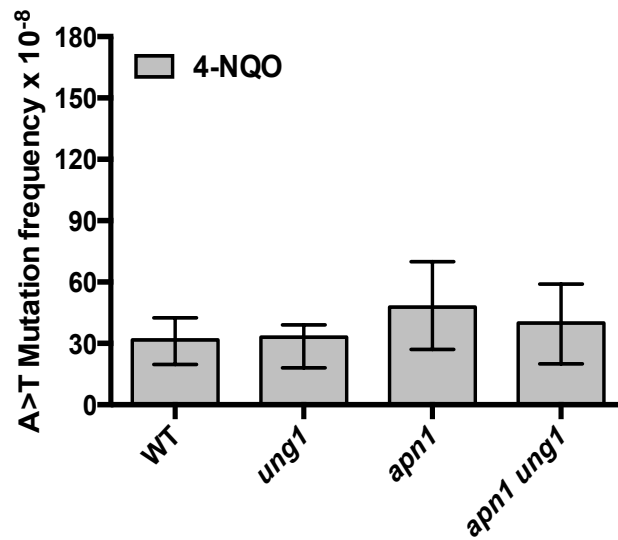
Figure 10: The frequencies A to T mutations following 4NQO treatment.

The frequencies of A>T Lys+ mutations following treatments with 0.2 µg/mL 4NQO. Error

bars indicate 95% confidence intervals

3.2.6 The level of uracil in DNA increases when repair synthesis is induced using 4NQO.

In order to quantify the uracil residues present in the genomic DNA, we used an AP-reactive alkoxyamine compound AA3 (Wei et al., 2015). This compound contains the alkyne group, through which a variety of compounds can be attached *via* click chemistry. The approach for detecting uracil-residues in DNA was first to label the uracil-derived AP sites with AA3 and then to attach the fluorescent dye cyanine 5 (Cy5) to AA3-AP conjugates (Fig. 11). To demonstrate its efficacy, the approach was used in the *ung1* cells treated with various concentration of 5-FU and a dose-dependent increase in the Cy5 signal when yeast cells were treated with 10, 50, and 100 μM 5-FU was observed (Fig. 12A). Genomic DNA isolated from Hela cells and the AID-expressing Daudi cells was used as negative and positive controls, respectively. Daudi is a B-cell lymphoma cell line with highly elevated level of uracil in DNA resulting from the overexpression of the APOBEC family of cytosine deaminases (Pettersen et al., 2015; Wei et al., 2015). While there was no significant difference between the level of uracil in yeast cells without 5-FU treatment and that in Hela cells, the level of uracil in DNA in yeast cells treated with 100 μM 5-FU increased to the level comparable to that in Daudi cells (Fig. 12A). Using the same AA3-Cy5 labeling approach, I measured the level of uracil in DNA in yeast cells treated with 1, 5, 10, and 20 μg/mL 4NQO. The levels of uracil detected in cells treated with 10 or 20 μg/mL 4NQO, but not with 1 or 5 μg/mL 4NQO, were significantly elevated compared to the untreated sample (Fig. 12B).

A. Schematic for AA3 labeling

DNA from *ung1Δ* strains

    + Methoxyamine

DNA with blocked AP sites

    + UDG

DNA with uracil-derived AP sites

    + AA3

AA3 labeled AP sites

    + Cy5 Azide (Click chemistry)

Cy5 labeled DNA
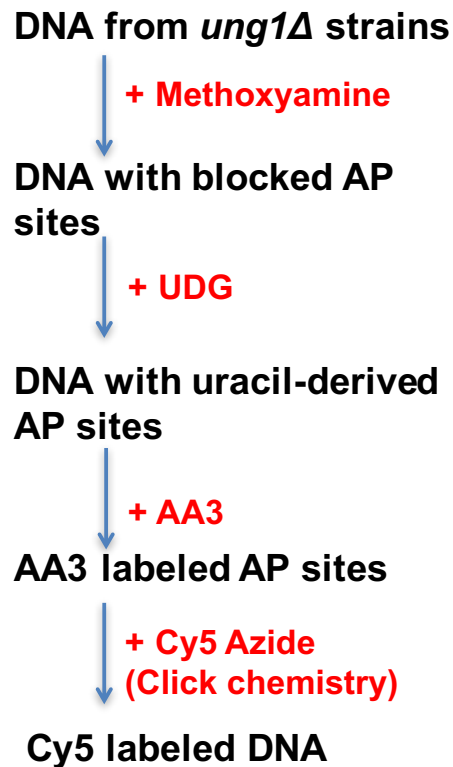
Figure 11: Schematic representation of AA3-labeling of AP sites

Uracil is excised by UDG to create AP sites. AP sites are labelled with AA3 followed by reaction with Cy5-azide and quantitation of the fluorescence on a nylon membrane.

* Uracil quantitation using AA3 (for 5-FU treated cells) were performed by Shanqiao Wei at Wayne State University.

**A. Uracil detection using AA3
(5-FU treated)**

B. Uracil detection using AA3 after 4NQO treatment

C. 5-FU treatment survival

D. 4-NQO treatment survival
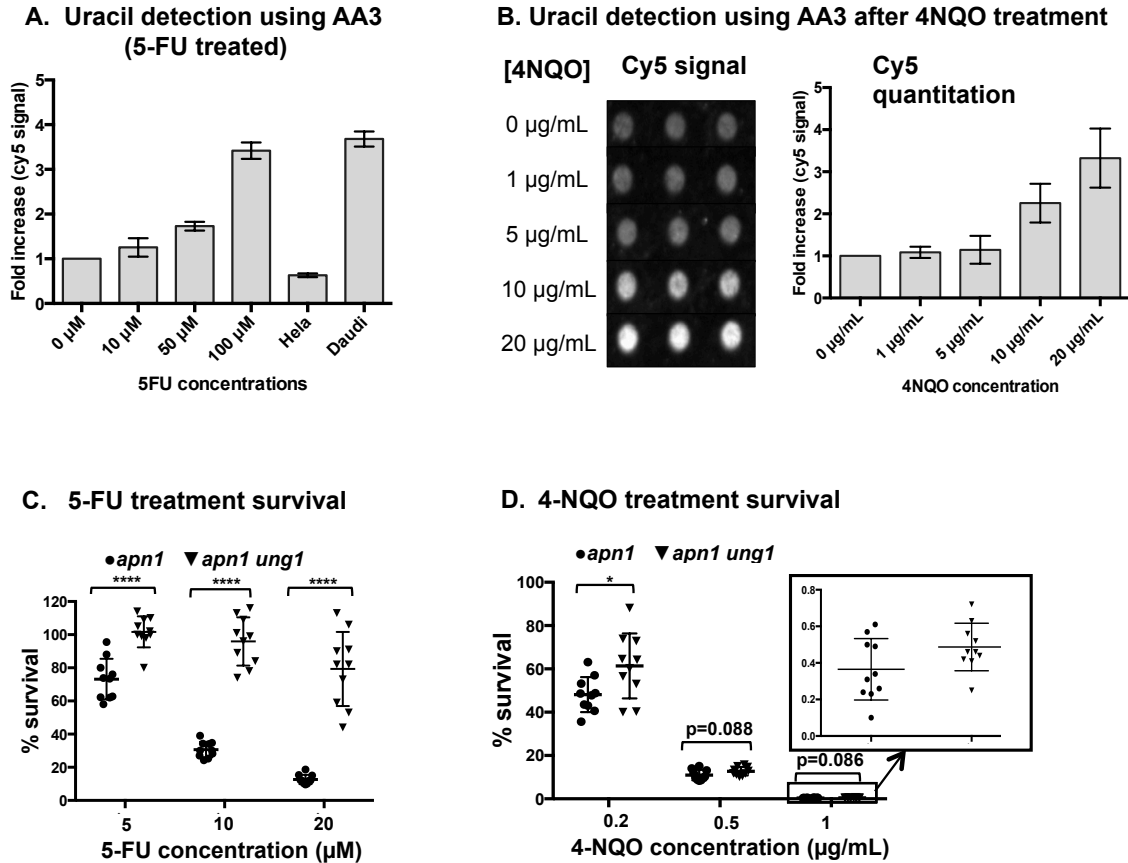
Figure 12: Uracil residues in the DNA following the 4NQO-treatment

A) Quantification of uracil residues in DNA from *ung1Δ* yeast cells treated with the indicated concentrations of 5-FU for 24 hrs. The genomic DNA samples from untreated Hela and Daudi are used as negative and positive controls, respectively. All measurements are from N=6. **B)** Quantification of uracil residues in DNA from yeast cells treated with the indicated concentrations of 4NQO. **Left**- Cy5 signal from AA3-labeled DNA dot-blotted on a nylon membrane. **Right**- Quantification of the cy5 signal shown on the left. Cy5 quantity is represented as relative to untreated (0μg/mL) sample. Error bars indicate standard deviations and all measurements are from N=6. Survival after growth in **C)** the 5-FU- or **D)** 4NQO-supplemented liquid culture, represented as the relative percentage compared to the untreated cells (*P<0.05, ****P < 0.0001; unpaired student t-test). N=10.

3.2.7 Sensitivity to 4NQO is partially reduced by UNG1 deletion.

To further show that the uracil-incorporation into DNA is a significant component of the repair synthesis associated with 4NQO treatment, I tested whether uracil in DNA is relevant to the cytotoxicity of 4NQO. The main cytotoxic lesions of 5-FU treatment are the AP sites derived from the uracil incorporated into the genomic DNA. When treated with 5-FU, yeast cells in the BER-deficient *apn1* backgrounds are highly sensitive (Fig. 12C). But the 5-FU sensitivity is very significantly reduced in *apn1 ung1* background, where uracil in DNA cannot be removed to create the toxic AP sites. If 4NQO treatment elevates the level of uracil in DNA, *apn1 ung1* strains would have a reduced level of sensitivity to 4NQO in comparison to *apn1Δ* strains. I calculated the number of surviving colony-forming cells after culturing them in the liquid medium with or without 0.2, 0.5, or 1 µg/mL concentration of 4NQO and observed a slight survival advantage in *apn1 ung1* compared to *apn1,* although only the 2 µg/mL 4NQO concentration elicited a statistically significant increase (Fig. 12D). Together, these results support the hypothesis that uracil incorporation into DNA occurs at a significant level when DNA repair synthesis is induced.

3.2.8 Dut1 is cell-cycle regulated in yeast.

We hypothesized that the transcription-associated increase in the density of uracil in DNA as determined by the long-amplicon qPCR or the CDG expression experiments above is due to the incorporation of uracil into DNA during unscheduled DNA synthesis (UDS) that can occur outside of the genome duplication in S phase. It has been shown in mammalian cells and plants that *DUT1*, a gene that encodes for dUTPase, is cell-cycle regulated with its highest expression in S-phase (Ladner & Caradonna, 1997; Pardo & Gutierrez, 1990). In

yeast, a large-scale high through-put analysis previously has shown that the dUTPase-encoding *DUT1* gene begins to be upregulated in late G1 phase ensuring that dUTP levels are kept low during replication (Cho et al., 1998). Conversely, DNA synthesis occurring outside of S phase (i.e. G1 and G2) will be subject to the dNTP pool with the relatively higher dUTP levels. To confirm the cell-cycle regulated *DUT1* expression in yeast, I arrested cells in G1 using the mating pheromone α-factor and collected cells every 15 min after release for the RNA isolation and qRT-PCR. The expression level of the histone H2-encoding *HTA2* gene, which was previously shown to be upregulated during S phase, was determined as a control (Heintz, 1991). The expression levels of both *DUT1 and HTA2* genes were highest at 45 min after the release from α-factor and declined to the lowest point at 75 mins after the release indicating that the *DUT1* expression is cell-cycle regulated in a manner similar to the *HTA2* gene (Fig. 13A).
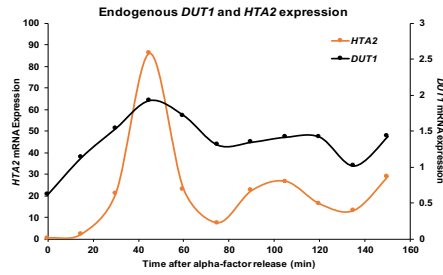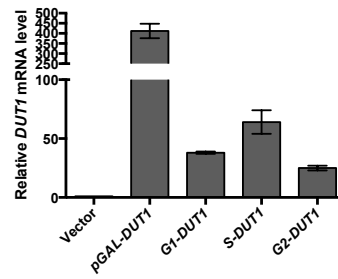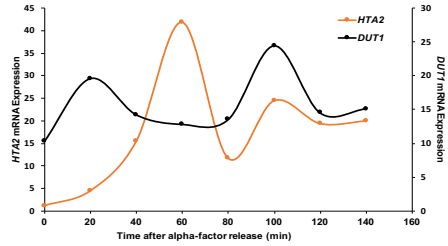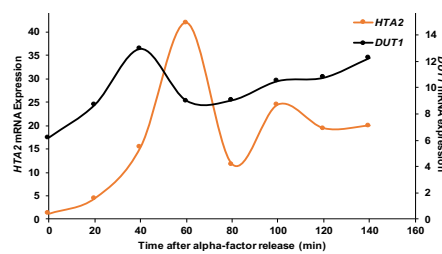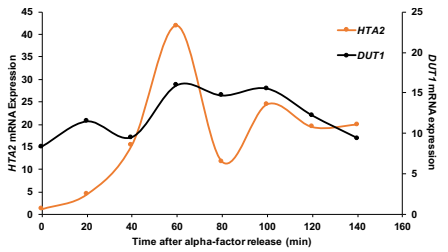
Figure 13: *DUT1* expression levels

A) Relative mRNA level of endogenous *DUT1* or endogenous *HTA2* gene expression in *bar1Δ* cells synchronized with α-factor. RNA was collected every 20 min after the release from α-factor. N=6 for all data points. **B)** Relative expression level of *DUT1* was measured by qRT-PCR from the asynchronous cells transformed with the plasmids from *pGAL* and the G1-, S-, G2-specific promoters. Error bars indicate standard deviations and all measurements are from N=3. **C)** Relative mRNA level of endogenous *HTA2* or *DUT1* overexpresed from *pCLN2* promoter, *pHHF01* promoter **(D)** or *pCLB2* promoter **(E).** Expression levels were assessed by qPCR and normalized to *ALG9.* N=6 for all data points.

3.2.9 Overexpression of Dut1 in G1 or G2 significantly reduces the transcription-associated mutations.

It was reported previously that the plasmid-mediated overexpression of *DUT1* from the galactose-inducible *pGAL* promoter can greatly reduce uracil-associated mutations at the *pTET-lys2-TAA* reporter (Kim & Jinks-Robertson, 2009), indicating that the cellular [dUTP] is a critical determinant in the transcription-associated uracil-dependent mutations. When induced by the addition of galactose to the media, the *pGAL*-regulated genes are highly expressed regardless of the cell cycle. To test whether uracil incorporation into DNA during G1 and G2 phases is a significant contributor to the transcription-associated uracil-dependent mutations, I modulated [dUTP] in G1, S and G2 phases by the cell-cycle specific overexpression of *DUT1* gene and measured the effect on the mutation rate at the *pTET-lys2-TAA* reporter. For the cell-cycle specific expression of *DUT1* in G1, S, or G2 phase, the *pGAL* promoter was replaced with the promoters of *CLN2*, *HHO1*, or *CLB2* genes, respectively (Prado & Aguilera, 2005; Wellinger et al., 2006). In order to reduce the protein half-life and thereby ensure cell-cycle specific presence of the overexpressed Dut1 protein, I added a protein destabilization domain (PEST) to the plasmid constructs (Berset et al., 2002). I then expressed *DUT1-PEST* from these plasmid constructs in yeast, isolated mRNA from asynchronous cells and performed qRT-PCR to determine the expression level. *DUT1* was expressed 38-, 64-, and 25-fold higher than the endogenous level from the G1, S, and G2 constructs, respectively. For all three constructs, the level of the overexpressed *DUT1* mRNA was substantially lower than the *DUT1* expression from *pGAL*-construct, which was 412-fold higher than the endogenous level (Fig, 13B). In order to confirm that *DUT1* is expressed in the cell cycle-specific manner from these promoters, we performed qRT-PCR with RNA samples isolated every 20 minutes after the release of cells arrested at G1 with α-factor. The *DUT1* mRNA expression was highest at 100 , 40, and ~60-100 mins after release from α-factor (Fig. 13C, 13D, and 13E). The S-phase time point under this specific

65

condition was determined to be 60 mins after the release from the α-factor by analyzing the mRNA level of *HTA2* gene.

I transformed the BER-deficient, *apn1 ntg1 ntg2* cells with the plasmids containing p*CLN2 (G1-)*, p*HHO1 (S-)*, or p*CLB2 (G2)*-constructs and calculated the mutation rates at the *pTET-lys2-TAA* reporter. I first carried out the fluctuation analysis for the determination of mutation rates in media supplemented with galactose and raffinose under high transcription conditions in order to compare the effect of Dut1 expression from p*CLN2*, p*HHO1*, or p*CLB2* to its expression from the previously studied *pGAL*-construct. While the *pGAL*-construct resulted in ~10-fold decrease in the rate of mutations compared to the vector-only control under these conditions, the G1-specific, p*CLN2*-construct led to a ~ 2-fold reduction in mutation rates in cells (Fig. 14A). Although the G2-specific, p*CLB2*-construct led to a <2-fold reduction, the mutation rate in the cells with this construct was significantly different from that in the cells with vector alone. For the cells with S-specific *pHHO1*-construct, there was no significant reduction in the mutation rate compared to the vector control. When the fluctuation analysis was repeated in the media supplemented with glycerol and ethanol, the G1-, S-, and G2-specific constructs all resulted in <2-fold, but statistically significant, reductions in the rates of mutation at the *pTET-lys2-TAA* reporter compared to the vector control (Fig. 14B). When the transcription of the *pTET-lys2-TAA* was repressed by adding doxycycline, the rates of mutation was unchanged with the Dut1-overexpression from the G1-, S-, or G2-specific promoters or from the *pGAL* promoter (Fig. 14C and 14D). These results suggest that the shift in the free dUTP pool affects the uracil-composition and the uracil-associated mutations more substantially at highly transcribed genes.
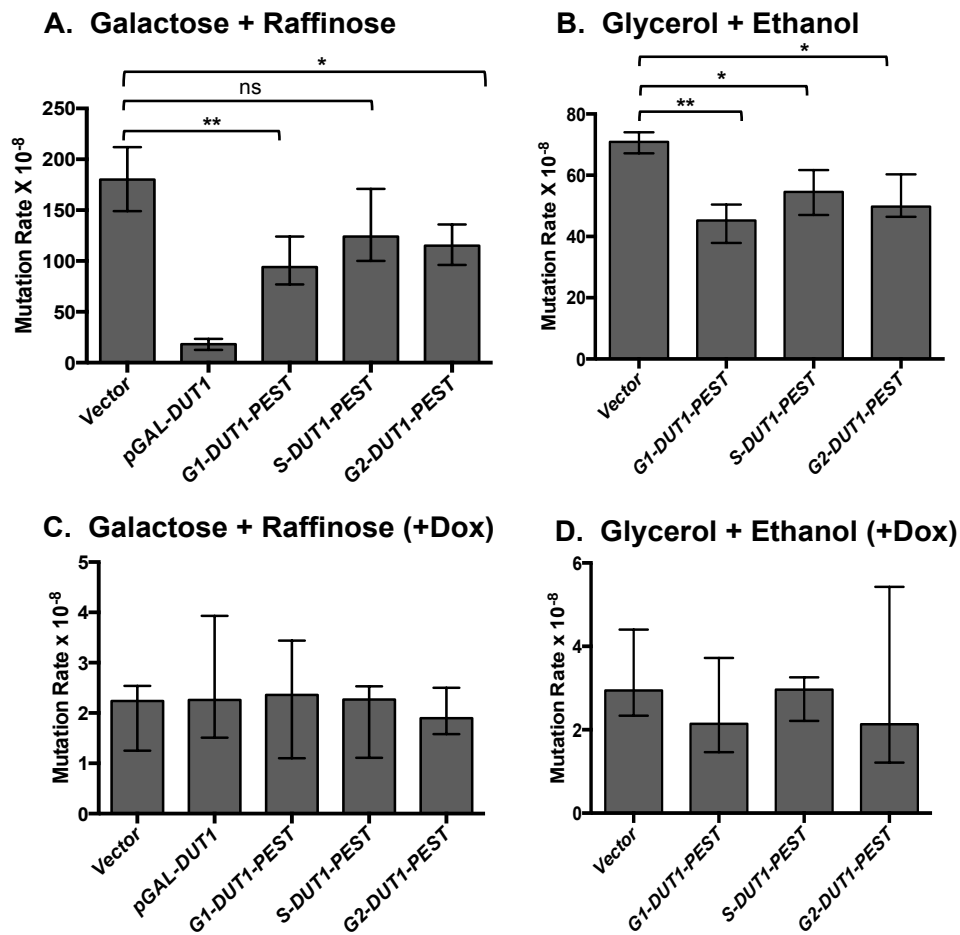
Figure 14: The rates of mutation in cells overexpressing Dut1 from the cell-cycle regulated promoters

The rates of Lys+ reversion mutations of the *apn1Δ ntg1Δ ntg2Δ* strain transformed with plasmids expressing *DUT1* from *pGAL*, p*CLN2 (G1)*, p*HHO1 (S)*, or p*CLB2* (G2). The growth conditions were in rich media supplemented with galactose and raffinose **(A)**, glycerol and ethanol **(B)**, galactose and raffinose plus doxycycline **(C)**, or glycerol and ethanol plus doxycycline **(D)**. The error bars indicate 95% confidence intervals. (*$P <$ 0.05, **$P <$ 0.005, ns – not significant; unpaired student t-test).

3.3 Discussion

High levels of transcription have been implicated previously as a major source of genomic instability in various organisms (reviewed in (Gaillard & Aguilera, 2016; Jinks-Robertson & Bhagwat, 2014; Kim & Jinks-Robertson, 2012). In yeast, when a reporter construct with the tetracycline-regulatable promoter (*pTET*) was used to determine the rate of mutations at several different levels of transcription, a linear and proportional relationship between the level of transcription and the rate of mutation was observed (Kim et al., 2007). Subsequent studies indicated that a majority of these mutations were derived from unrepaired AP sites (Kim & Jinks-Robertson, 2009). When BER is disabled, as in *apn1* or *apn1 ntg1 ntg2* strains, there was a unique elevation in specific types of mutations. Namely, when the mutations at the *pTET-lys2-TAA* allele was studied, TAA to GAA, TCA, or TAC mutations were elevated by ~200- and 500-fold in *apn1* or *apn1 ntg1 ntg2* strains respectively (Kim & Jinks-Robertson, 2010). The highly elevated rates of these T>G and A>C mutations were dramatically reduced by the disruption of the uracil DNA glycosylase Ung1 or by the overexpression of the dUTPase Dut1, indicating that these types of mutation are originating from uracil in DNA. There are two distinct ways by which the non-canonical uracil residues appear in DNA; by the deamination of cytosine residues present in DNA or by the incorporation into DNA by DNA polymerase utilizing dUTP in place of dTTP. The location of mutations at T:A or A:T pairs suggests the latter route of uracil appearance in DNA. Further genetic studies showed that AP sites generated by the excision of uracil by Ung1 is bypassed by the TLS polymerases Rev1 and Polζ to bring about the T>G and A>C mutations (Kim, Mudrak, et al., 2011). The most remarkable finding about these uracil-derived T>G and A>C mutations was that they are almost completely suppressed when the transcription of the *pTET-lys2-TAA* mutation reporter is repressed by the addition of doxycycline. This transcription-dependent elevation of mutations originating from uracil

residues in DNA led to the hypothesis that the chemical composition of the DNA can be changed to include a higher number of uracil residues when actively transcribed.

I tested the hypothesis of the transcription-dependent elevation of uracil residues in DNA by directly quantifying uracil residues at a defined genomic locus under high or low transcription conditions using the long-amplicon qPCR method. In *ung1* strains, where uracil residues, once incorporated into the DNA, cannot be excised out, there was a statistically significant 2-fold difference between the densities of uracil detected at the *pTET-lys2-TAA* under high and low transcription conditions (Fig. 8). The disruption of Dcd1, which has been shown to elevate the uracil incorporation into DNA in previous reports (Bryan & Hesselberth, 2015) and the uracil-associated mutations in the current study (Fig. 6A), resulted in significant elevation of the uracil density at the *pTET-lys2-TAA* under both high and low transcription conditions (Fig. 8). The densities of uracil at the *CAN1* and *TDH3* genes were also elevated by the disruption of Dcd1. These data demonstrate that the difference in the uracil density calculated using the long-amplicon qPCR approach adequately reflect the change in the DNA composition. At *CAN1* and *TDH3* genes, the levels of uracil as well as the rates of transcription did not change when doxycycline was added to repress transcription from the *pTET* promoter (Fig. 6 and 8). There was a ~170-fold difference in the level of transcription at the *pTET-lys2-TAA* between high and low transcription conditions while the level of uracil is elevated by ~2-fold. At *TDH3,* which is transcribed at ~100-fold higher rate than *CAN1* according to the RT-qPCR analysis, the uracil density was detected to be ~3-fold higher than at *CAN1* in *ung1* background, providing further corroboration for the transcription-dependent mechanism of uracil incorporation into DNA. However, when the uracil density at the *pTET-lys2-TAA* is compared to those at *CAN1* and *TDH3* genes, we observed that the transcription rate does not have a linear correlation with the level of uracil residues. Under the low transcription conditions, the *pTET-lys2-TAA* is transcribed at a considerably lower rate than the *CAN1* gene. However,

69

under the same conditions, there was no statistical difference in the uracil densities at these two genomic sites. On the other hand, the uracil level at the *pTET-lys2-TAA* under the high transcription conditions was slightly higher than that at the *TDH3* gene although the latter is transcribed at about ~10-higher rate than the *pTET-lys2-TAA*. These discrepancies suggest that there might be factors additional to transcription that modulate the level of uracil-incorporation into DNA such as position of the replication fork, replication timing, orientation of the transcription machinery.

In an *apn1* strain, the rate of uracil-derived mutations is elevated by ~20-fold when transcription from the *pTET* promoter is activated (Kim & Jinks-Robertson, 2009). The approximately 2-fold difference in the uracil density cannot wholly account for the dramatic increase in the mutation rate. An alternative, but not mutually exclusive, explanation for the transcription-dependent elevation in the mutations arising from uracil in DNA is that transcription affects the activity of the glycosylase converting the mutation-neutral uracil residues into the mutagenic AP sites. In order to test this hypothesis, we studied the mutations induced by the glycosylase CDG, which excises undamaged cytosines in addition to uracil residues, at *the pTET-lys2-TAG* mutation reporter under high and low transcription conditions (Fig. 4 and 5). If the base-excision by the glycosylase is not affected by the state of transcription, the rate of those mutations initiated by the excision of undamaged cytosine (G>C) should remain the same whether the transcription of *pTET-lys2-TAG* reporter is activated or repressed. In five different genetic backgrounds, the CDG expression led to significant transcription-dependent elevations of the uracil-dependent (A>C and T>G) mutations as expected from the greater uracil density under high transcription conditions. However, we also observed smaller but still significant transcription-dependent elevations in the rate of cytosine-dependent mutations. The fold-difference between the high and low transcription conditions was 1.7 and 2.5 in *apn1 ung1* and *apn1 rad14 ung1* backgrounds, respectively, and increased to 2.5 and 7.6 when *TOP1* gene was deleted from each of these

strains (Fig. 4E, 5C and Table 2). Compared to the transcription-dependent elevation in the CDG-induced A>C and T>G mutations, which ranges from 10- to 23-fold, the elevation of CDG-induced mutations at cytosine residues is relatively small but still significant, indicating that the efficiency of glycosylase activity is somehow affected by transcription (Fig. 4D, 5B and Table 2).

For the uracil-associated A>C and T>G mutations, the transcription-dependent elevation is further augmented by ~2-fold when the topoisomerase I is disrupted, implicating the transcription-associated change in the local DNA topology as one of the major factors affecting the glycosylase activity (Fig. 5B). When RNase Hs were disrupted as in *apn1 rnh1 rnh201 ung1* strain, the cytosine-dependent mutations at *the pTET-lys2-TAG* were elevated by 3.5-fold when transcription was highly activated (Fig. 5F). For the uracil-derived mutations, only the mutations originating from the excision of uracil located on the top, non-transcribed strand (i.e. T>G) were further elevated by the disruption of RNase Hs (Table 2). The R-loop accumulation in the absence of RNase Hs affects the two DNA strands within the transcribed regions asymmetrically; the bottom, transcribed strand forms stable hybrid with the nascent RNA and the top, non-transcribed strand is left unpaired. This asymmetry is directly reflected on the specific elevation of T>G over A>C mutations at *the pTET-lys2-TAG* and corroborates the biochemical analysis where CDG was shown to excise uracil or cytosine from single-stranded oligonucleotide substrates ~10-fold more efficiently than from double-stranded substrates (Kavli et al., 1996). The accumulations of helical stress and single-stranded DNA patches associated with active transcription appear to be major factors in enhancing the activity of uracil DNA glycosylase, contributing to the elevated uracil-dependent mutations at highly transcribed genes.

With the long-amplicon qPCR and the analysis of CDG-induced mutations, an increase in the density of uracil upon activation of transcription has been demonstrated by two independent approaches (Fig. 4, 5, and 8). The mechanism underlying such

phenomena, however, is still not clear. One plausible explanation can be found in the previously reported evidence of transcription-induced endogenous DNA damage (reviewed in (Gaillard & Aguilera, 2016)). The topological changes and DNA strand-separation necessitated by DNA is also responsible for the elevated susceptibility to genotoxic agents and the consequent accumulation of base damage. The highly transcribed areas of the genome are more prone to the replication fork stalls and collapse, which can be significantly aggravated at repetitive sequences where transcription facilitates the formation of non-B DNA structures. While the DNA polymerases utilizing dUTP in place of dTTP during replication can account for the stochastic presence of uracil throughout the genome, DNA synthesis associated with correcting the damage or resolving the stalled replication fork at highly transcribed regions provides additional opportunities to incorporate uracil into DNA and to affect the locus-specific elevation in uracil content. I demonstrated here that exogenously engendering damage to DNA with CPT and 4NQO, thereby inducing rounds of repair-dependent DNA synthesis, led to an accumulation of uracil specific mutations into DNA in the absence of BER at the *pTET-lys2-TAA* mutation reporter (Fig. 9). In case of 4NQO-treated cell, we also showed the uracil-accumulation in the genome by chemically probing for the uracil-derived AP sites (Fig. 12B). Another important implication of these experiments is that uracil in DNA could be a significant factor in not only the mutations but also the cytotoxicity induced by various DNA damaging chemicals not specifically targeting the pyrimidine biosynthesis pathway. For the TS-targeting drug 5-FU, the main cytotoxic mechanism involves the AP sites generated from the highly frequent uracil residues incorporated into DNA due to the imbalance in [dUTP]/[dTTP] ratio (Longley et al., 2003). The yeast cells of *apn1* background with the severely compromised BER pathway and the inability to efficiently repair AP lesions are acutely sensitive to 5-FU. However, the *apn1* cells become highly resistant to 5-FU when *UNG1* is deleted so that the uracil-to-AP

72

conversion cannot occur (Fig. 12C). We showed that *UNG1*-deletion can also reduce the cell sensitivity to 4-NQO treatment at a low drug concentration (Fig. 12D).

The repair-associated DNA synthesis occurring in G1- or G2-phase of cell-cycle would be a particularly potent way of uracil incorporation into DNA because the expression of dUTPase is upregulated in S-phase (Fig. S4A and (Cho et al., 1998; Ladner & Caradonna, 1997; Pardo & Gutierrez, 1990)). This cell-cycle dependent regulation of the available [dUTP] would ensure the minimal uracil-incorporation into DNA during replication (Fig. 15), but comparatively increases the [dUTP]/[dTTP] ratio and thus the possibility of dUTP being used by DNA polymerases during the repair synthesis occurring outside of S-phase. When we lowered the [dUTP]/[dTTP] ratio by overexpressing the dUTPase-encoding gene *DUT1* from the G1- or G2-specific promoters, the rates of mutations at the *pTET-lys2-TAA* reporter under the high transcription conditions in BER-deficient cells were significantly reduced (Fig. 14A and 14B), indicating that the uracil incorporated into DNA during G1 or G2 comprise a substantial source of transcription-associated mutations (Fig. 15). The *DUT1*-overexpresion from the ubiquitous *pGAL* promoter, however, resulted in much greater reduction in the mutation rate at the *pTET-lys2-TAA* reporter under high transcription conditions. This effect could largely be explained by the higher overall level of expression from *pGAL* promoter but also could indicate that a considerable level of uracil-incorporation does occur in all cell cycles including S. The effect of the *DUT1*-overexpression from S-specific promoter in reducing uracil-associated mutations when the *pTET-lys2-TAA* reporter under high-transcription conditions was relatively smaller compared to the G1- or G2-specific promoters. And the overexpression of *DUT1* from *pGAL* or S-specific promoter as well as G1- or G2-specific promoters had no effect on the mutation rates when the transcription of the *pTET-lys2-TAA* was repressed, indicating that the level of uracil incorporated into DNA during replication, which would be uniform throughout the

genome and only affected by the *DUT1*-expression from the *pGAL* or S-specific promoter, cannot induce uracil-associated mutations to a significant degree.

In summary, we found a novel mechanism of introducing uracil into DNA during the damage-induced repair synthesis during G1- or G2-phase of cell cycle. The repair-coupled uracil-incorporation would be a way to non-uniformly alter the nucleotide composition of genomic DNA. The degree of alteration and the extent of uracil-incorporation would depend on the extent of repair synthesis occurrence and would be expected to be greater at regions of frequent endogenous DNA damage i.e. highly transcribed genomic loci. Such a role played by transcription in changing the nucleotide composition locally would apply to other types of non-canonical residues. We speculate that a similar mechanism is involved in specifically elevating the ribonucleotides at highly transcribed regions. Previous studies have shown that ribonucleotide-dependent mutations are highly elevated by transcription (Kim, Huang, et al., 2011; Nick McElhinny et al., 2010). And similar to *DUT1*, *RNR1,* the gene encoding the essential, regulatory subunit of the ribonucleotide reductase is regulated in a cell-cycle dependent manner to ensure the optimal [dNTP]/[rNTP] ratio for the replication during S-phase (Elledge & Davis, 1990). There are several remaining questions to be answered through further studies such as whether the correlation between uracil and transcription apply linearly genome-wide and whether other sources of the endogenous DNA damage such as non-B DNA structures could elevate uracil-content through DNA repair synthesis. Further work is also needed to determine the specific repair pathway directing the uracil incorporation at highly transcribed regions.
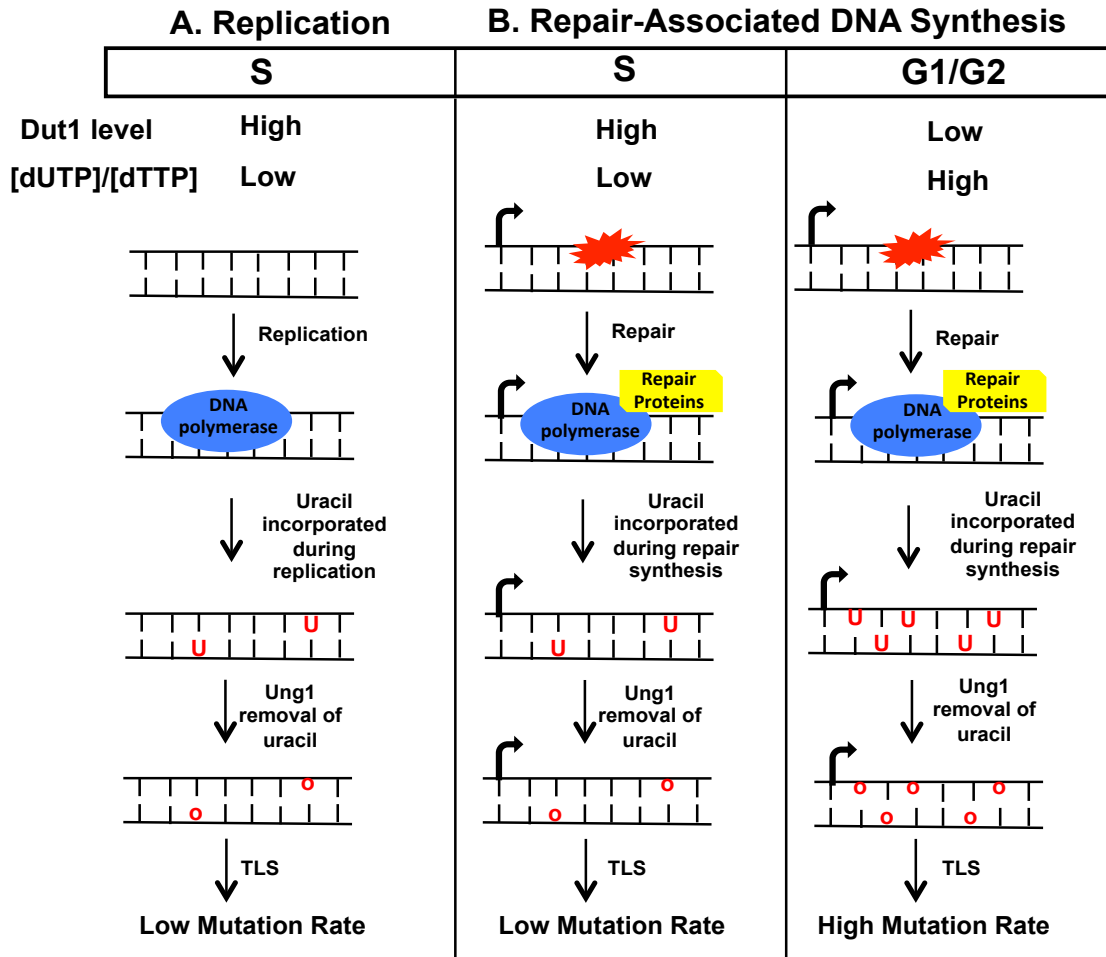
Figure 15: The model of uracil-incorporation into DNA during G1, S, or G2 phases of the cell cycle.

Dut1 level is higher and the ratio of [dUTP]/[dTTP] is lower in S phase compared to G1 or

G2. **A)** Uracil-incorporation during replication in S phase. The replicative synthesis occurs

in S phase; the extent of uracil-incorporation and the ensuing uracil-associated

mutagenesis is minor due to the low [dUTP]/[dTTP]. **B)** Uracil-incorporation during the

repair-associated DNA synthesis. In S-phase (middle), the repair synthesis induced by

transcription-associated endogenous DNA damage is subject to the low [dUTP]/[dTTP];

the extent of uracil-incorporation and the ensuing uracil-associated mutagenesis is minor

due to the low [dUTP]/[dTTP]. In G1 or G2 phase, the repair synthesis induced by

transcription-associated endogenous DNA damage is subject to the relatively high [dUTP]/[dTTP]; the extent of uracil-incorporation and the ensuing uracil-associated mutagenesis is significant.

Chapter 4

Def1 and Dst1 play distinct roles in repair of AP lesions in highly transcribed genomic regions

4.1 Introduction

Abasic (AP) sites comprise a major portion of endogenously occurring DNA damages and are produced when a variety of lesions including uracil, modified bases and spontaneous or enzymatic deamination products are excised by specific DNA N-glycosylases thereby releasing the damaged bases from DNA (Boiteux & Jinks-Robertson, 2013). In yeast, genetic studies showed that spontaneous AP sites prevalently originate from uracil incorporated during replication by DNA polymerases (Guillet & Boiteux, 2003). Following the excision of uracil base by DNA glycosylase, Ung1, AP sites ensue and become a potent block to transcription and replication machineries. Failure to repair AP lesions, therefore, can be both mutagenic and cytotoxic. As an alternative to repair, AP lesions can be bypassed by translesion synthesis (TLS) polymerases. In yeast and metazoans, Rev1 and Polζ ensure continued replication albeit at a high mutation cost by inserting mostly C nucleotides opposite the AP lesions (Gibbs et al., 2005; Kow et al., 2005). Base Excision Repair (BER), the major repair pathway for removing AP sites, is initiated when an AP endonuclease, Apn1 in yeast, cleaves the DNA phosphodiester bond at the 5' side of the lesion (Popoff et al., 1990). Ntg1 and Ntg2 AP lyases can alternatively initiate BER by nicking the sugar phosphate backbone on the 3' side of the AP site (Meadows et al., 2003). To complete repair, the deoxyribose phosphate residue is removed by a phosphodiesterase and a DNA polymerase fills the gap followed by a DNA ligase that seals the remaining nick.

When BER is disrupted or overwhelmed, Nucleotide Excision Repair (NER) can provide an alternative mechanism for the repair of AP sites (Gellon et al., 2001; Scott et al., 1999; Swanson et al., 1999; Torres-Ramos et al., 2000). NER pathway is primarily involved in removing helix-distorting lesions such as UV-induced thymidine dimers (Cadet et al., 2005). The loss of NER has been associated with increased sensitivity to sunlight and predisposition to skin cancer in humans. NER can be divided into two sub-pathways based

on the lesion recognition step, global genome repair (GGR) and transcription-coupled repair (TCR)(Hanawalt, 2002). In GGR, repair proteins directly recognize a distortion of the DNA helix and are recruited to the lesion site; Rad7 and Rad16 form a complex required for this step in yeast. GGR is involved in the repair of lesions throughout the genome regardless of the transcriptional status of the genes (Guzder et al., 1997; Sugasawa et al., 2001). In contrast, TCR is initiated by the stalled RNA polymerase complex at a lesion and thus repair only those lesions located on the transcribed strand of a gene blocking transcription. Rad26, a yeast homolog of human CSB and a DNA dependent ATPase, is uniquely required for the TCR sub-pathway. The abolishment of Rad26 results in significant but partial disruption of TCR (Verhage et al., 1996). Rad14, a homolog of human XPA, has been shown to be essential for the lesion verification step directly following the lesion recognition and is required for both GGR and TCR (Guzder et al., 1995). The subsequent steps are comparable for GGR and TCR pathways; structure-specific endonucleases together with NER-specific helicases create dual incisions around the lesion to release the lesion-containing single strand DNA of 25-30 nucleotides, the resulting gap is filled by DNA polymerases and remaining gap ligated by a DNA ligase.

AP sites are not recognized as helix distorting lesions by Rad7-Rad16 complex but can cause robust transcription block by T7 RNA polymerase or mammalian RNA polymerase II (Tornaletti et al., 2006). In yeast, our previous findings demonstrated that, while AP sites are predominantly repaired by Apn1-mediated BER pathway, TCR pathway contributes significantly to repair AP lesions and reduce the AP-associated mutations in the context of a highly transcribed gene. When Apn1 is disrupted, the defect in the repair of AP sites result in transcription-associated mutations (TAM), which is further elevated upon the disruption of back-up repair pathway involving Ntg1 and Ntg2, the N-glycosylases/AP lyases. The disruption of NER by the deletion of *RAD14* also led to a dramatic increase in TAM in *apn1* background suggesting the involvement of NER in AP site repair. The deletion

of *RAD26*, a TCR specific gene, also led to an increase in TAM although still significantly lower than the increase in mutation due to *RAD14* deletion. In contrast, the deletion of *RAD7*, a GGR specific gene, in *apn1* background did not affect mutagenesis suggesting that TCR, but not GGR, is involved in the repair of AP sites. The increase in TAM observed in *apn1 rad14* or *apn1 rad26* strain was specifically due to unrepaired lesions on the transcribed strand, further confirming TCR repair of AP sites which is likely triggered by stalling of the RNA polymerase II complex (RNAPII) at the AP sites (Kim & Jinks-Robertson, 2010). This overlap between TCR and BER was also demonstrated in human cells for the repair of the 8-oxoguanines (GO lesions), which is the most common type of oxidative DNA lesions (Guo et al., 2013).

The mechanism of what happens to RNAPII stalled at a UV-induced lesion and how the repair machinery accesses the damage obstructed by the stalled RNA polymerase have been extensively studied but still not completely understood (Gaillard & Aguilera, 2013). The events following RNAPII stalling at an AP lesion are even less understood. Multiple pathways exist that could help RNAPII contend with obstacles. First, the stalled RNAPII can be backtracked thereby allowing access for repair enzymes to the lesion. In *E. coli,* UvrD, a DNA helicase/translocase and NER protein, was shown *in vitro* to facilitate backtracking by binding to RNAPII and forcing it to slide backwards along DNA thereby exposing the DNA lesions (Epshtein et al., 2014). Second, alternative to promoting repair, the RNAPII backtracking can promote lesion bypass. I*n vitro* studies have shown that transcription elongation past the lesion on the template is promoted by a general transcription elongation factor, TFIIS. Mammalian TFIIS or yeast Dst1 functions to stimulate the intrinsic endonuclease activity of RNAPII, which cleaves the nascent RNA transcript allowing for the transcript re-alignment with the active site (Charlet-Berguerand et al., 2006). Finally, if RNAPII stalls for a prolonged period of time, the RNAPII degradation factor (Def1) promotes the ubiquitination and degradation of Rpb1, the largest subunit of RNAPII (M. D. Wilson et

al., 2013; Woudstra et al., 2002). The degradation of RNAPII could enhance repair by clearing the path for DNA repair enzymes especially since the stalled RNAPII can prevent access to the lesion by repair factors (Beaudenon et al., 1999; Lee et al., 2002; Woudstra et al., 2002).

While in mammalian cells, CSB, the homolog of Rad26, is sufficient in activating TCR, Rad26 only partially accounts for TCR of AP lesions in *S. cerevisiae* suggesting that additional factors are involved in recruiting NER proteins to the stalled RNAPII (Kim & Jinks-Robertson, 2010; Verhage et al., 1996). In the current report, we took advantage of a genetic assay designed to monitor the repair of AP sites within a highly transcribed gene in a strand-specific manner. This assay allowed us to examine the role of multiple factors that were previously implicated in the TCR of UV-induced DNA damage or in the rescue of the arrested RNAPII. Consistent with its function in the repair of UV lesions, the disruption of Def1 resulted in an increase in mutagenesis due to the accumulation of unrepaired AP sites specifically on the transcribed strand. However, I show here that, the disruption of Dst1 resulted in an unexpected decrease in the AP-associated mutagenesis suggesting a complex role for Dst1 when transcription elongation and damage repair are both at stake. These results provide new insights into factors that are involved in the repair or transcriptional bypass of the AP lesions located within transcribed genes.

4.2 Results

Reversion mutation assays were previously used to examine TAM in BER- and/or NER-deficient yeast strains and demonstrated that, under high transcription conditions, AP sites derived from uracil replacing thymine is a major source of TAM (Kim & Jinks-Robertson, 2009, 2010). In this report, the *pTET-lys2-TAA* allele, which has a TAA stop codon inserted in-frame into the *LYS2* ORF, was used in the quantitation of TAM. In this reversion mutation assay, a mutation in the TAA stop codon allows the production of

functional Lys2 protein required for lysine synthesis, which can be distinguished by selection on the lysine-lacking media. When the major AP endonuclease Apn1 is disrupted, the A>C and T>G transversion mutations are significantly elevated. These mutations are dependent on the expression of uracil DNA glycosylase Ung1 and significantly reduced when the cellular dUTP level is reduced upon overexpression of the dUTP pyrophosphatase Dut1. Overall, the previous investigation led to the model that A>C and T>G mutations are the net result of Rev1/ Polζ-dependent translesion bypass synthesis across from the uracil-derived AP sites during which predominantly C nucleotides are inserted opposite AP (Kim, Mudrak, et al., 2011). The *pTET-lys2-TAA* reversion assay allows us to infer from the mutation spectra whether the relevant AP sites were located on the transcribed or the non-transcribed strand. As diagramed in Fig. 1A, AP sites on the transcribed strand and non-transcribed strand each result in the A>C and T>G mutation signatures, respectively. Of the uracil-associated mutations, A>C but not T>G mutations are greatly elevated when NER factor Rad14 is disrupted (Kim & Jinks-Robertson, 2010). A>C mutations were also significantly elevated upon the disruption of TCR-factor Rad26, but not of GGR-factor Rad7 indicating that only TCR sub-pathway of NER is involved in the repair of AP lesions.

4.2.1 Non uracil-derived AP sites can be repaired by TC-NER.

To determine whether those AP lesions originating from sources other than uracil excision could be processed in a TCR-dependent manner, I generated AP sites in the yeast genome by the expression of modified human glycosylase TDG, which was shown to excise unmodified thymine residues from oligonucleotide substrates *in vitro* and to induce the accumulation of AT>CG transversions *in vivo* (P. Auerbach et al., 2005; Kavli et al., 1996). Expected mutations from TDG-induced AP lesions at the *pTET-lys2-TAA* reversion allele are shown in Fig. 16A. Compared to the vector control, TDG expression in yeast repair-deficient strains, *apn1 ung1* (BER[-]), *apn1 rad26 ung1* (BER[-] TCR[-]), *apn1 rad14 ung1* (NER[-]

(TCR⁻ GGR⁻)) and *apn1 ntg1 ntg2 ung1 (*BER⁻*),* resulted in elevation of the overall mutation rate by 5- to 11-fold (Fig. 16B). These elevated mutations are specifically due to TDG glycosylase activity and not from the uracil excision by yeast Ung1 as the experiments were carried out in *ung1* backgrounds. TDG-expression in *apn1 ntg1 ntg2 ung1* strains led to an accumulation of T>G mutations.   Compared to the vector-control in the same strain background, the rates of T>G and A>C mutations were elevated by 150- and 13-folds, respectively. These results indicate that, in the absence of Apn1 activity, the Ntg1/Ntg2-mediated sub-pathway of BER largely repairs those AP sites generated by excision of T's in the non-transcribed strand as previously observed with uracil-derived mutations (Fig. 16C and 16D) (Kim & Jinks-Robertson, 2010). In *apn1 rad14 ung1* strains, the rates of A>C and T>G mutations were elevated by 25.3-and 2.5-fold, respectively, indicating that a defect in NER led to accumulation of unrepaired AP sites mostly on the transcribed strand. In *apn1 rad26 ung1* strains, the rate of TDG-induced A>C mutations (42×10^{-8}) was higher compared to that in *apn1 ung1* (32×10^{-8}) albeit considerably lower compared to that in *apn1 rad14 ung1* (121×10^{-8}).   Overall, TCR sub-pathway of NER is a major back-up mechanism to repair TDG-induced AP lesions on transcribed strand within a highly transcribed gene and Rad26 is partially involved in mediating TCR.

A. Schematic for T>G and A>C mutations

B. Overall mutation rate C. Rate of A>C mutations D. Rate of T>G mutations
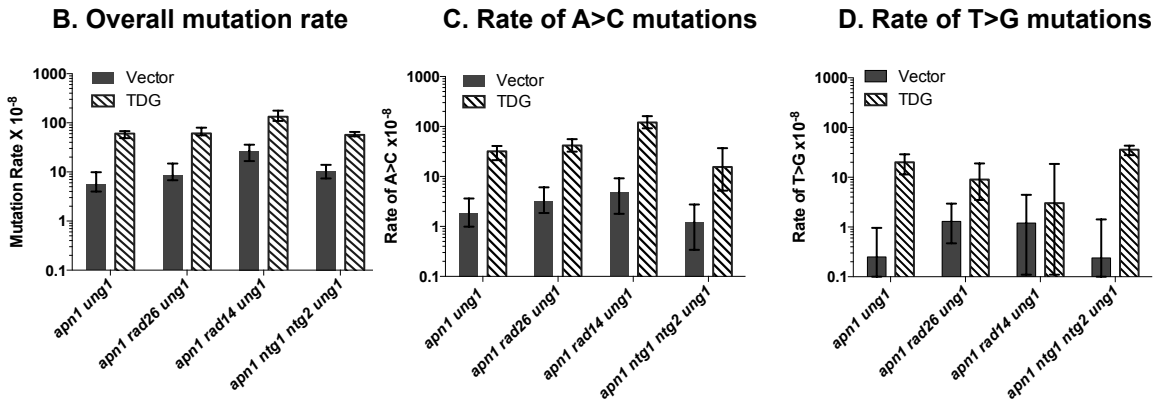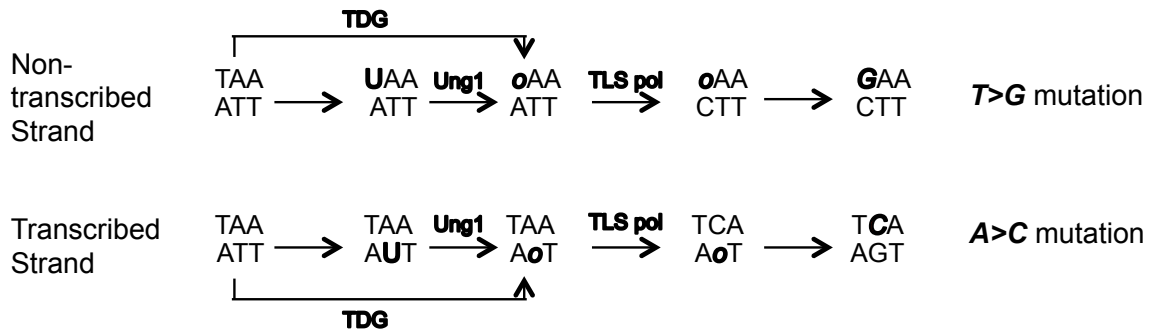
Figure 16: The effect of TDG expression on the rates of mutation at the *pTET-lys2-TAA* reporter

A) Schematic representation of the substitution mutations arising from base excision by Ung1 or TDG. *o* Indicates AP sites*.* B**)** Overall mutation rates in yeast strains with empty vector or TDG glycosylase-expressing plasmid. Error bars represents 95% confidence intervals. **C)** and **D)** Rates of A>C and T>G transversions, respectively. For the actual numbers see Table 3.

Table 3: Mutation rates with hTDG expression

pTET-*lys2-TAG* (Mutation rate x 10^{-8})				
Relevant genotype	**Total Mutation Rate (95%CI)**	**A>C rate (95%CI) (Fraction)**	**T>G rate (95%CI) (Fraction)**	**Other (Fraction)**
apn1 ung1 (vector)	5.5 (3.99 – 9.91)	1.88 (0.99-3.62) (15/44)	0.25 (0.03-0.96) (2/44)	3.38 (27/44)
apn1 rad26 ung1 (vector)	8.46 (6.77 – 14.8)	3.25 (1.87-6.06) (15/39)	1.30 (0.47-2.95) (6/39)	3.9 (18/39)
apn1 rad14 ung1 (vector)	26.9 (16.6 – 35.8)	4.78 (1.79-9.17) (8/45)	1.20 (0.11-4.44) (2/45)	20.9 (35/45)
apn1 ntg1 ntg2 ung1 (vector)	10.4 (7.36 – 14)	1.21 (0.34-2.76) (5/43)	0.24 (0.002-1.43) (1/43)	8.95 (37/43)
apn1 ung1 (hTDG)	60.7 (48.9 – 67.9)	32.0 (21.3-40.7) (26/45)	20.0 (11.4-29) (16/45)	4.04 (3/45)
apn1 rad26 ung1 (hTDG)	61.1 (55.4 – 78.9)	41.8 (31.6-55.7) (30/40)	9.00 (3.5-19) (6/40)	6.11 (4/40)
apn1 rad14 ung1 (hTDG)	133 (109 – 177)	120.9 (92.3-162) (40/44)	3.02 (0.12-18.6) (1/44)	9.07 (3/44)
apn1 ntg1 ntg2 ung1 (hTDG)	57.0 (53.2 – 65.4)	15.5 (5.22-37) (4/34)	36.0 (28-43.1) (28/34)	3.35 (2/34)

4.2.2 Srs2 or Sub1 is not involved in TCR-dependent repair of AP sites in yeast.

When RNAPII encounters DNA damage, it stalls at the damage site and the lesion must be repaired to ensure continued elongation. TCR of AP lesions is only partially affected by the disruption of Rad26 and additional factors must be involved in mediating efficient repair of AP sites by TCR (Kim & Jinks-Robertson, 2010). *E. coli* protein UvrD is a DNA helicase/translocase with known roles in NER and replication-associated processes (Kumura & Sekiguchi, 1984; Van Houten & McCullough, 1994). UvrD was recently shown *in vitro* to bind stalled RNA polymerase and to force it to slide backwards along DNA allowing greater access for the DNA repair factors to the lesion for more efficient repair (Epshtein et al., 2014). Srs2, a yeast paralog of UvrD, is a helicase and a DNA dependent ATPase that is involved in multiple facets of genome maintenance (i.e. homologous recombination, checkpoint recovery, and replication fork progress) (Krejci et al., 2003; Marini & Krejci, 2010; Van Komen et al., 2003). In order to determine whether Srs2, like *E. coli* UvrD, promotes TCR repair of AP lesions, I deleted *SRS2* gene in *WT*, *apn1* and *apn1 rad26* strains and determined the rates of spontaneous reversion mutations at the *pTET-lys2-TAA* allele. Overall mutation rate was not affected by Srs2 disruption in *WT* or *apn1* backgrounds and reduced by ~1.7-fold in *apn1 rad26* background (Fig. 17A). In *WT* or *apn1* backgrounds, mutation spectra also showed no significant change due to Srs2 disruption. However, in *apn1 rad26* background, there was a 14-fold decrease in the rate of T>G mutations (from AP sites on the non-transcribed strand) with no significant change in the rate of A>C mutations (from AP sites on the transcribed strand) (Figs. 17B and 17C). This strand specific effect of Srs2-disruption indicates that this factor is likely involved in promoting repair of AP lesions on the non-transcribed strand.

The transcriptional co-activator Sub1 was shown to be involved in the tolerance to the hydrogen peroxide-induced DNA damage in a TCR-dependent manner (J. Y. Wang et al., 2004). The mutation rates and spectra at the *pTET-lys2-TAA* allele were determined in

86

WT, *apn1* and *apn1 rad26* backgrounds with additional deletion of the *SUB1* gene. The rates of overall Lys+ mutations or A>C transversions were not affected by Sub1-disruption in all three strain backgrounds (Figs. 17D and 17E and Table 7). The rate of T>G mutation rates in *apn1 rad26 sub1* strains was ~4-fold reduced compared to *apn1 rad26* strains but this difference was not statistically significant (Fig. 17F).

Figure 17: The effect of Srs2 or Sub1 disruption on the rates of mutations

A) and **D)** Overall rates of mutations at *pTET-lys2-TAA.* Error bars represent 95% confidence intervals. **B)** and **E)** Rates of A>C mutations. **C)** and **F)** Rates of T>G mutations. *Zero mutation event of this class was detected and the mutation rate was calculated assuming 1 event. For the actual numbers see Table 7. Total, A>C, and T>G mutation rates in *apn1* and *apn1 rad26* backgrounds in **D) – F)** is identical to the rates shown in **A - C**.

4.2.3 *DST1* deletion reduces mutations at *pTET-lys2-TAA*.

During the DNA damage-induced transcription arrest, TFIIS can promote backtracking of the stalled RNAPII by stimulating the cleavage of nascent RNA but can also promote the resumption of transcription elongation and lesion bypass by RNAPII by realigning mRNA termini and DNA template (Gaillard & Aguilera, 2013; Reines & Mote, 1993). To determine whether TFIIS is involved in the TCR repair of AP sites, I deleted *DST1*, the gene encoding TFIIS in yeast and determined the mutation rates. I will from hereon refer to TFIIS as Dst1. Dst1-disruption resulted in a 3.1-, 3.4- and 6-fold reduction in the overall mutation rates in WT, *apn1,* and *apn1 rad26* backgrounds, respectively (Fig. 18A). The rates of both A>C and T>G mutations were reduced suggesting that this reduction is not specific to AP lesions on the transcribed strand (Figs. 18B and 18C). I also disrupted Dst1 in the BER and NER deficient backgrounds (*apn1 ntg1 ntg2* and *apn1 rad14,* respectively*)* and determined the mutation rates as well as the mutation spectra. I observed a ~ 6-fold reduction in the mutation rates in both *apn1 ntg1 ntg2 dst1* and *apn1 rad14 dst1* strains (Fig. 18A). There was a ~6-fold reduction in the rates of A>C in *apn1 rad14* background (Fig. 18B). For the T>G mutation rates, Dst1-disruption resulted in a 10-fold reduction in *apn1 ntg1 ntg2* background, further indicating that the function of Dst1 is not limited to the AP sites on the transcribed strand or to the TCR pathway (Fig. 18C).

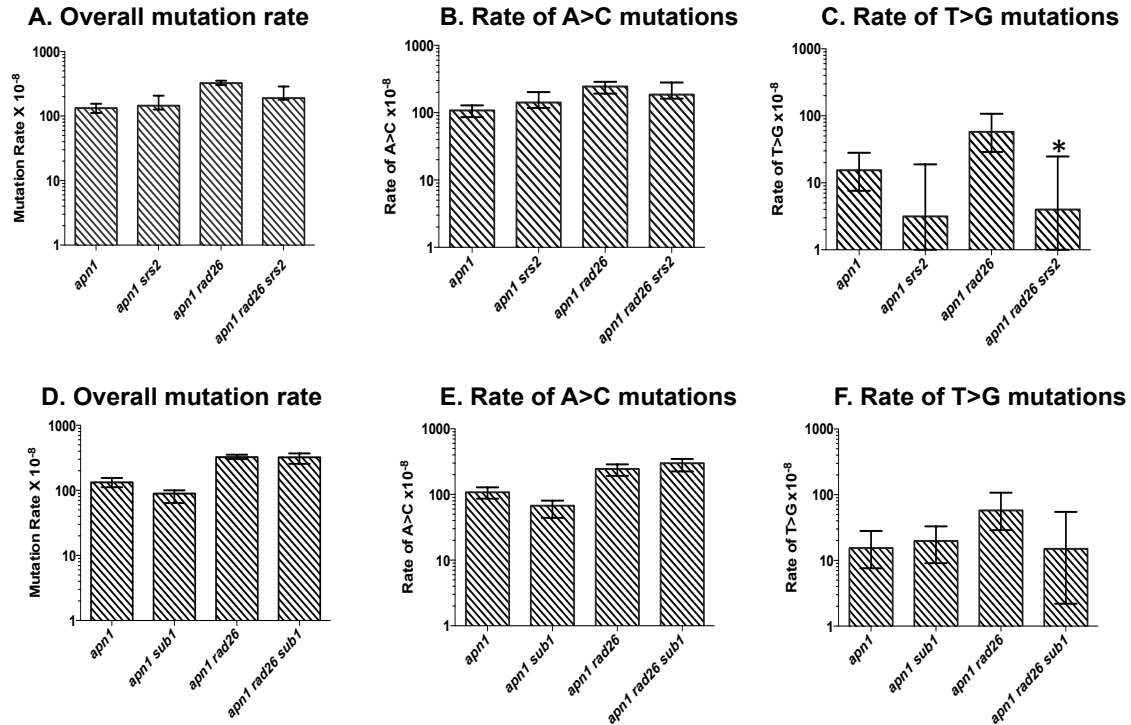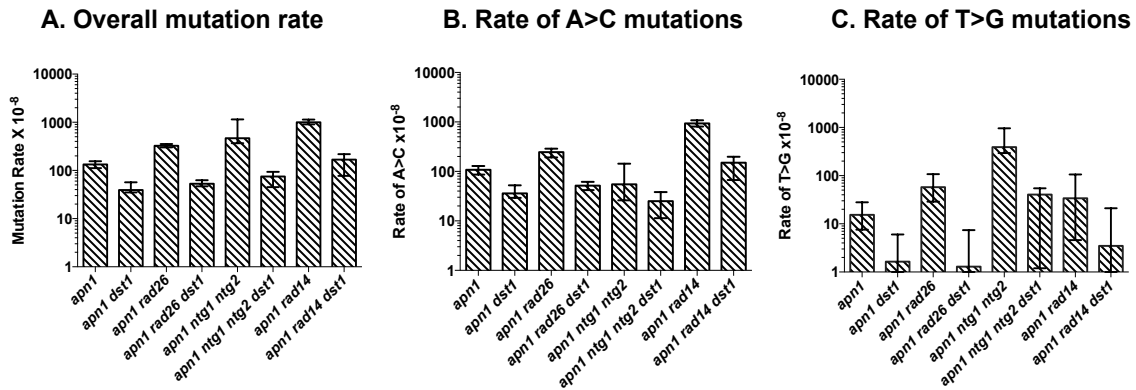**Figure 18: The effect of Dst1 disruption on the rates of mutations**

A) Overall rates of mutations at *pTET-lys2-TAA*. Error bars represents 95% confidence intervals. **B)** Rates of A>C mutations. **C)** Rates of T>G mutations. For the actual numbers see Table 7. Total, A>C, and T>G mutation rates in *apn1* and *apn1 rad26* backgrounds in this figure is identical to the rates shown in Figure 17.

4.2.4 The effect of Dst1 disruption is not due to change in transcription activity.

In order to determine whether the decrease in mutagenesis upon Dst1-disruption is dependent on the level of transcription, I looked at the mutation rates at *CAN1,* a moderately transcribed locus. *CAN1* encodes an arginine transporter and deleterious mutations at this locus confer resistance to the toxic arginine analog canavanine. Dst1-disruption led to a ~2-fold reduction in the canavanine-resistant (Can^R) mutation rates in *apn1 rad26* and a ~4.5 fold reductions in both *apn1 rad14* and *apn1 ntg1 ntg2* backgrounds (Fig. 19A). The Can^R mutation rate in *apn1* background was not affected.

It was previously demonstrated that mutation rate at the *pTET*-regulated *lys2* reporter linearly correlates with the rate of transcription; repressing transcription by addition of doxycycline resulted in the proportional decrease in the mutation rate (Kim et al., 2007). In order to determine whether the decrease in mutation rate upon *DST1* deletion was due to a decrease in the transcription rate, I determined the *LYS2* expression levels in *WT, dst1* and *apn1 dst1* strains using the qRT-PCR method. The *LYS2* expression level was not significantly affected upon *DST1*-deletion indicating that the effect of *DST1* deletion on mutation rate is not due to changes in the transcription levels of *LYS2 (*Fig. 19B). The expression level of *CAN1* was also not affected by *DST1*-deletion (Fig. 19C).

4.2.5 Dst1 disruption increases recombination.

A reduction in mutation rate upon *DST1*-deletion could be due to the repair of AP lesions by pathways other than BER and NER. Thus, Dst1 was disrupted in yeast strains containing three different homologous recombination reporter constructs that vary in their sequence context (G-rich, C-rich or G/C-rich). Detailed description of these reporter constructs as well as the recombination rates in *WT* strain backgrounds were previously reported (Yadav et al., 2014). *DST1*-deletion resulted in 2- to 3-fold increases in recombination rates for all three constructs (Fig. 19D).

91
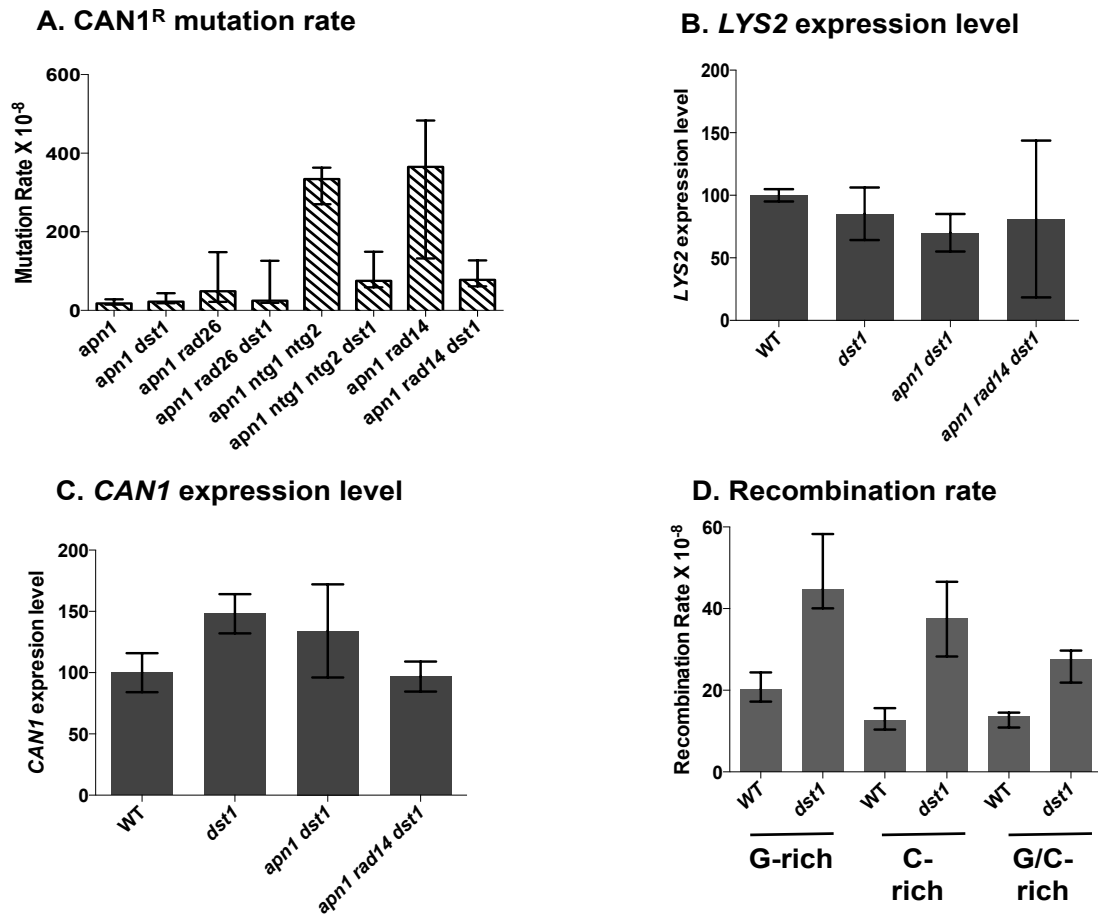
Figure 19: The effect of Dst1 disruption on transcription and recombination

A) Overall rate of *CAN1* forward mutations in the indicated genetic backgrounds. Error bars represent 95% confidence intervals. For the actual numbers see Table 4. **B)** *LYS2* and **C)** *CAN1* expression levels calculated by RT-QPCR using *UBC6* as the reference gene. RNA was extracted from yeast cells with the indicated genetic strains. Error bars indicate the standard deviation calculated from 3 or 4 independent experiments. For the actual numbers see Table 5. **D)** The rates of recombination at the *G-rich*, *C-rich* or *G/C-rich* cassettes were determined using the method of the median (Table 6.). Error bars represent 95% confidence intervals.

Table 4: Can1^R mutation rates and confidence intervals

Relevant genotype	CAN1^R (Mutation rate x 10^{-8})
apn1	17.9 (14.5 – 28)
apn1 rad26	48.5 (21.8 – 148)
apn1 ntg1 ntg2	334 (270 – 363)
apn1 rad14	365 (132 – 483)
apn1 dst1	22 (18.2 – 43.6)
apn1 rad26 dst1	24.5 (19 – 126)
apn1 ntg1 ntg2 dst1	75.3 (58.9 – 149)
apn1 rad14 dst1	77.7 (61.3 – 127)

Table 5: *LYS2* and *CAN1* expression levels

Relevant genotype	*LYS2* expression (%)	*CAN1* expression (%)
WT	100 ± 5	100 ± 16
dst1	85.2 ± 21	148 ± 16
apn1 dst1	70 ± 15	134 ± 38
apn1 rad14 dst1	81.1 ± 62.7	96.8 ± 12.3
apn1	100 ± 66	N/D
apn1 rad26	71.1 ± 23.3	N/D
apn1 rpb9	102.3 ± 16.1	N/D
apn1 rad26 rpb9	158.8 ± 7.3	N/D

We additionally determined the effect on AP-associated mutagenesis of disrupting Rpb9, which was previously reported to increase the UV sensitivity of *rad26* and *rad7* strains indicating its involvement in TCR (Gaillard et al., 2009; S. Li & Smerdon, 2002). Rpb9 is also a non-essential subunit of RNAPII that shares homology with Dst1 and was reported to be required for Dst1 to stimulate the transcript cleavage and the read-through by RNAPII (Awrey et al., 1997). Similar to *DST1*-deletion, the deletion of *RPB9* in WT, *apn1,* and *apn1 rad26* backgrounds significantly reduced the rates of A>C and T>G mutations but did not affect the transcription rate of *pTET-lys2-TAA* reporter (Fig. 20A - D). Rpb9-disruption, however, did not result in a significant elevation in recombination (Fig. 20E). This could be due to the technical difficulty associated with the severely affected growth rate in this strain.
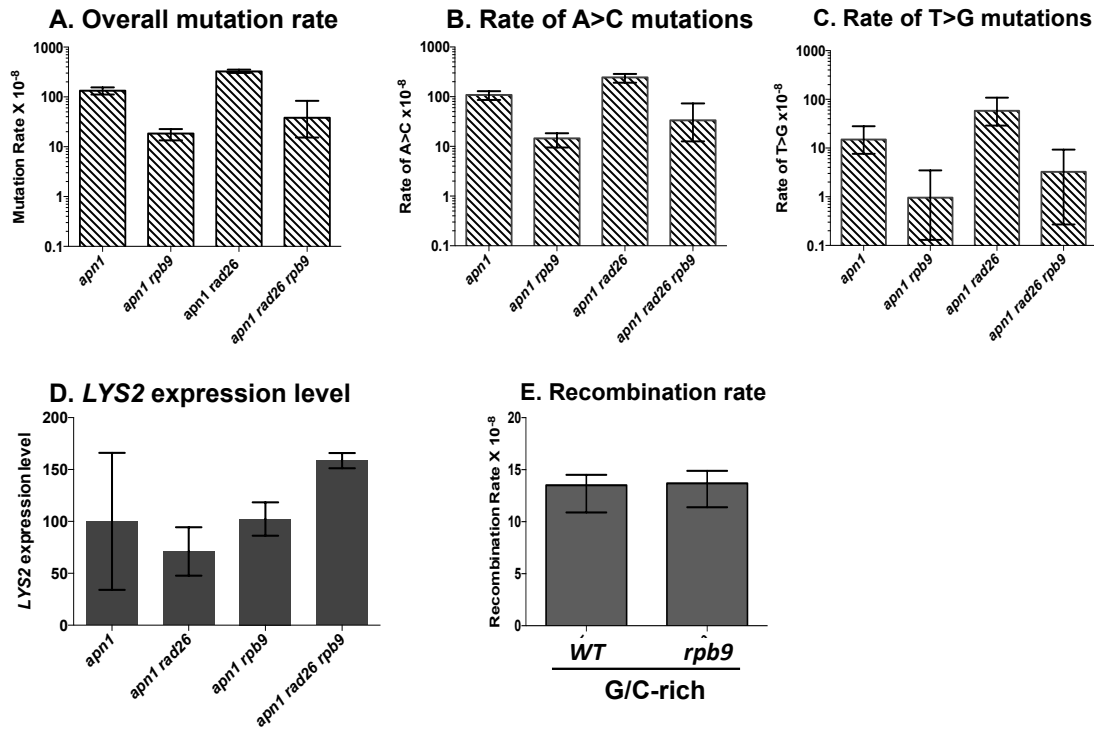
Figure 20: The effect of Rpb9 disruption on the rates of mutations and recombination

A) Overall rates of mutations at *pTET-lys2-TAA*. Error bars represents 95% confidence intervals. **B)** Rates of A>C mutations. **C)** Rates of T>G mutations. For the actual numbers see Table 7. **D)** *LYS2* expression levels calculated by RT-QPCR using *ACT1* as the reference gene. RNA was extracted from yeast cells with the indicated genetic strains. QRT-PCR was performed using primers targeting *LYS2*. For the actual numbers see Table 5. Error bars indicate the standard deviation calculated from three independent experiments. **E)** The rates of recombination at the *G/C-rich* cassette were determined using the method of the median. Error bars represent 95% confidence intervals. For the actual numbers see Table 6. Total, A>C, and T>G mutation rates in *apn1* and *apn1 rad26* backgrounds in this figure is identical to the rates shown in Figure 17.

Table 6: Recombination rates and confidence intervals

Relevant genotype	Recombination rate (x 10^{-8})
WT (G-rich)	20.3 (17.2 – 24.4)
WT (C-rich)	12.5 (10.4 – 15.6)
WT (G/C rich)	13.5 (10.9 – 14.9)
dst1 (G-rich)	44.8 (40.1 – 58.3)
dst1 (C-rich)	37.6 (28.3 – 46.6)
dst1 (G/C-rich)	27.5 (21.9 – 29.7)
rpb9 (G/C-rich)	13.7 (11.4 – 14.9)

* The recombination data are part of my first author publication however, the experiments were performed by Dr. Christopher Lopez. Dr. Lopez has authorized the use of these results in this dissertation.

4.2.6 Def1-disruption elevates mutations from AP lesions on the transcribed strand.

Def1 is recruited to the stalled and/or arrested RNAPII and promote the ubiquitination and degradation of RNAPII thereby clearing the way for the subsequent RNA polymerases and ensure the continued transcription as well as possibly allowing the access of repair proteins to DNA damage (Lee et al., 2002; Woudstra et al., 2002). Although previous studies have suggested that Def1 is not involved in TCR *per se* but functions as a last resort to allow for the resumption of transcription (Woudstra et al., 2002), other studies identified Def1 as a potential transcription-repair coupling factor by showing that Def1-disruption elevated the UV-sensitivity in the absence of GGR (Gaillard et al., 2009). I deleted *DEF1* in *apn1* and *apn1 rad26* backgrounds to determine whether it leads to a specific elevation in the A>C mutations indicative of further interruption in TCR. The overall mutation rate in *def1* and *apn1 def1* mutants showed no significant difference compared to WT and *apn1,* respectively (Fig. 21A and Table 7). However, the deletion of *DEF1* in *apn1 rad26* background led to a 1.7-fold increase in the overall mutation rate and a ~2-fold increase in the rate of A>C mutations (Fig. 21A and 21B). The rate of T>G mutations were generally lower in comparison to the A>C mutations and were slightly reduced upon the disruption of Def1 (Fig. 21C). This suggests that, in the absence of Rad26, the Def1-mediated degradation of the stalled RNAPII can alternatively allow the recruitment of repair enzymes to the AP lesions on the transcribed strand. However, when both Def1 and Rad26 are disrupted, the repair of the AP sites is partially compromised and mutations accumulate. The disruption of Def1 did not affect the T>G mutation rates in *apn1 rad14* background. The rate of A>C mutations in *apn1 rad26 def1* strains is significantly less than that in *apn1 rad14* strains (Fig. 21B; Table 7) where TCR as well as GGR are completely abolished suggesting that, even though both Def1 and Rad26 contribute to repairing the AP lesions on the transcribed strand, additional factor must be functioning to mediate TCR. The deletion of *RAD7,* which specifically disables GGR, did not affect the overall rate of mutagenesis in *apn1* or *apn1*

rad26 strains indicating that GGR is not significantly contributing to the repair of AP lesions. When I deleted *DEF1* in *apn1 rad7* background, I observed a 2.2-fold increase in the overall mutation rate further supporting our findings that Def1 contributes in repairing AP sites. Surprisingly, the overall mutation rate of *apn1 rad7 def1* strains was 2.1-fold higher compared to *apn1 def1* strains suggesting that Def1 and Rad7 might be operating in the same pathway to promote the repair of AP sites. Def1-disruption in *apn1 rad14* background led to a slight decrease in the mutation rate indicating that the degradation and removal of the stalled RNAPII might interfere with the efficient repair of AP sites by pathways other than NER. The attempt to generate *apn1 rad26 rad7 def1* strain by a consecutive one-step allele replacement approach was not successful. It is possible that this combination of deletion is lethal, although we lack the definitive proof since the construction of mutant diploids followed by the tetrad analysis was not carried out.

A. Overall mutation rate (Glvcerol/Ethanol)

B. Rate of A>C mutations

C. Rate of T>G mutations

Figure 21: Effects of Def1 disruption on the rates of mutations in YEPGE

A) Overall rates of mutations at *pTET-lys2-TAA*. Error bars represent 95% confidence intervals. **B)** Rates of A>C mutations. **C)** Rates of T>G mutations. *Zero mutation event of this class was detected and the mutation rate was calculated assuming 1 event. For the actual numbers see Table S2. Total, A>C, and T>G mutation rates in *apn1, apn1 rad14,* and *apn1 rad26* backgrounds in this figure is identical to the rates shown in Figure 2 and Figure 3.

4.2.7 The effect of Def1-disruption on mutations from AP lesions is dependent on the growth conditions.

For all of the data described above, I determined the reversion mutation rates of the *pTET-lys2-TAA* allele by growing the relevant yeast strains in the rich media supplemented with 2% glycerol and 2% ethanol as the primary carbon source (YEPGE). Because the *DEF1*-deletion strains showed a slow growth phenotype in YEPGE, I repeated the experiments in the rich medium supplemented with the fermentable sugar dextrose (YEPD). Although the growth conditions did not significantly affect the mutation rates for WT or *def1* strain, in *apn1* background, the rate of overall Lys+ reversion mutation was >20-fold lower when the cells are grown in YEPD than in YEPGE (Figs. 21A and 22A). In YEPD, in *apn1 def1* strain compared to *apn1* strain, there were ~15-fold and ~19-fold increases in the overall and A>C mutation rates, respectively. Compared to the modest 1.7- and 2-fold changes in the overall and A>C mutation rates observed in YEPGE, when analyzed in YEPD, *DEF1*-deletion in *apn1 rad26* background led to an overwhelming ~53-fold increase in the overall mutation rate and a ~60-fold increase in the A>C mutations. The effect of Def1-disruption on the rate of T>G mutations was nominal in all backgrounds. In YEPD, the rates of overall and A>C mutations in a*pn1 rad26 def1* strain were comparable to those in *apn1 rad14* strain where TCR is completely abolished (Fig. 22). I also determined the mutation rates in *apn1 rad7 def1* strains in YEPD. The deletion of *DEF1* in YEPD led to a 4.8-fold increase in mutagenesis in *apn1 rad7* strain background indicating the involvement of *DEF1* in the TCR repair of AP sites. However, unlike in YEPGE where I observed a 2.1-fold increase in the mutation rate in *apn1 rad7 def1* strain compared to *apn1 def1* strain, I observed a 2.3-fold reduction in the rate of mutagenesis. And, in *apn1 rad14* background, no further elevation in the overall mutation rate resulted from Def1-disruption. However, the rates of A>C mutations were slightly increased and the rates of T>G mutations were

101

reduced by two-fold further confirming that the effect of Def1 is specific to the AP lesions located on the transcribed strand.
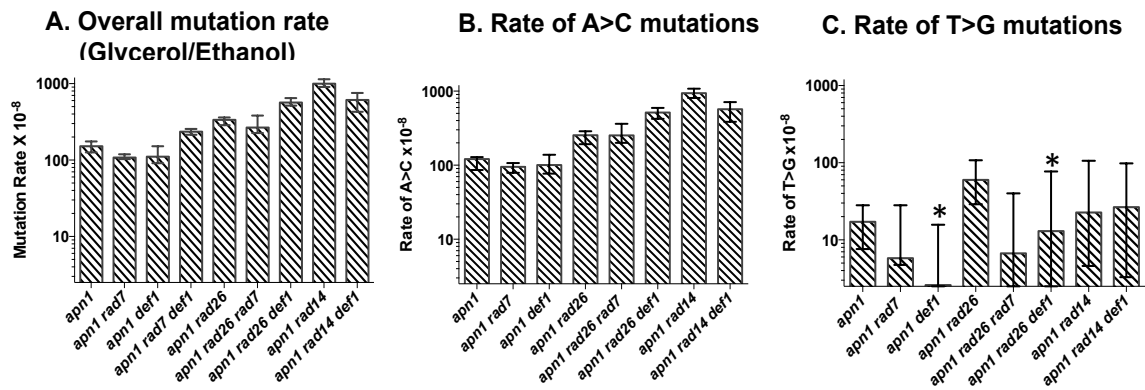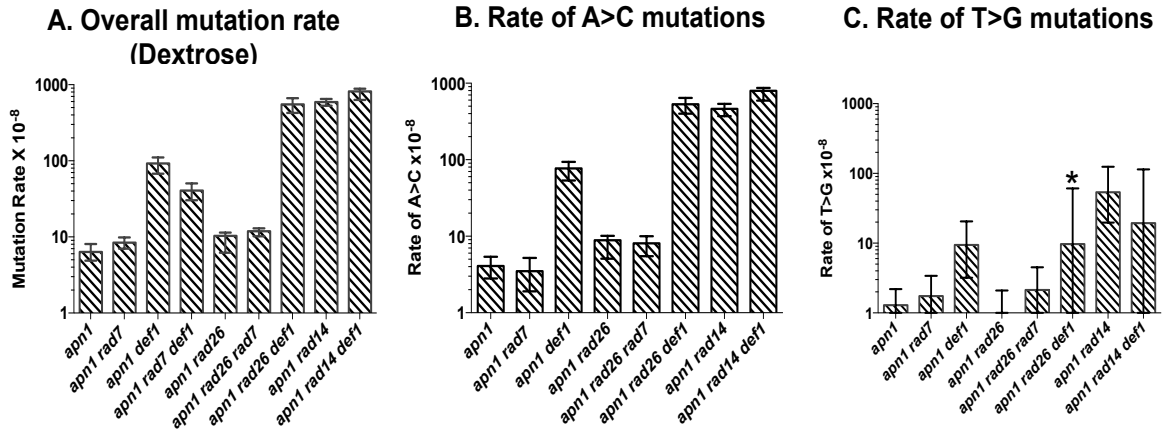
Figure 22: Effects of Def1 disruption on the rates of mutations in YEPD

A) Overall rates of mutations at *pTET-lys2-TAA*. Error bars represent 95% confidence intervals. **B)** Rates of A>C mutations. **C)** Rates of T>G mutations. *Zero mutation event of this class was detected and the mutation rate was calculated assuming 1 event. For the actual numbers see Table 8.

Table 7: Mutation rates and confidence intervals for pTET-lys2-TAA in glycerol/ethanol

pTET-lys2-TAA (Mutation rate x 10^{-8})

Relevant genotype	Total Mutation Rate (95%CI)	A>C rate (95%CI) (Fraction)	T>G rate (95%CI) (Fraction)	Other (Fraction)
WT	2.49 (1.22 – 3.42)	0.74 (0.32-1.17) (25/84)	0.18 (0.04-0.41) (6/84)	1.57 (53/84)
apn1	133 (112 – 155)	108 (86.1-129) (70/86)	15.4 (7.56-28.1) (10/86)	9.28 (6/86)
apn1 rad26	325 (302 – 354)	245 (192-288) (34/45)	57.8 (29.0-108) (8/45)	21.7 (3/45)
apn1 ntg1 ntg2	468 (367 – 1150)	55.0 (25.7-144) (11/94)	390 (296-959) (78/94)	24.9 (5/94)
apn1 rad14	997 (904 – 1140)	938 (811-1082) (80/88)	26.8 (4.58-106) (2/88)	68 (6/88)
apn1 srs2	145 (126 – 207)	142 (118-203) (45/46)	3.15 (0.13-18.9) (1/46)	<3.15 (0/46)
apn1 rad26 srs2	190 (179 – 287)	186 (163-281) (46/47)	<4.04 (0.2-24.7) (0/47)	4.04 (1/47)
apn1 sub1	89.4 (64 – 100)	67.5 (43.8-80.5) (34/45)	19.8 (9.09-33.2) (10/45)	1.99 (1/45)
apn1 rad26 sub1	322 (257 – 371)	299 (226-347) (40/43)	15 (2.25-54.9) (2/43)	7.5 (1/43)
apn1 dst1	39.3 (35 – 57)	36.0 (29.1-52.5) (42/48)	1.64 (0.28-6.0) (2/48)	3.3 (4/48)
apn1 rad26 dst1	53 (47 – 63)	51.5 (41.9-61.7) (38/41)	1.29 (0.06-7.41) (1/41)	2.6 (2/41)
apn1 ntg1 ntg2 dst1	74.9 (45 – 94)	25.0 (11.4-38.4) (16/48)	40.6 (21.2-54.7) (28/48)	9.4 (6/48)
apn1 rad14 dst1	168 (77 – 219)	151 (66.5-198) (43/48)	3.50 (0.31-21) (1/48)	14 (4/48)
apn1 rad51	111 (88.7 – 153)	N/D	N/D	N/D
apn1 rad51 dst1	42.1 (33 – 52.5)	N/D	N/D	N/D
apn1 rad7	108 (103 – 119)	95.0 (79.2-107) (37/42)	12.9 (4.75-28.2) (5/42)	<2.5 (1/42)

apn1 def1	111 (90.9 – 152)	101 (77-139) (39/43)	<2.6 (0.09-15.7) (0/43)	10.3 (4/43)
apn1 rad7 def1	235 (214 – 256)	N/D	N/D	N/D
apn1 rad26 def1	567 (516 – 646)	515 (424-595) (40/44)	<13 (0.63-77.2) (0/44)	51.5 (4/44)
apn1 rad26 rad7	267 (226 – 383)	253 (201-363) (38/40)	<6.68 (0.26-40.2) (0/40)	13.4 (2/40)
apn1 rad14 def1	611 (429 – 757)	571 (385-710) (43/46)	26.6 (3.26-98.2) (2/46)	13.3 (1/46)
apn1 rpb9	18.4 (13.4 – 22.6)	14.5 (9.48-18.4) (30/38)	0.96 (0.13-3.47) (2/38)	2.9 (6/38)
apn1 rad26 rpb9	38.3 (15.3 – 83.7)	33.6 (12.6-73.5) (36/41)	3.24 (0.27-9.33) (4/41)	0.93 (1/41)
srs2	0.86 (0.67 – 1.55)	N/D	N/D	N/D
dst1	0.8 (0.49 – 0.94)	0.15 (0.05-0.29) (7/37)	0.04(0.003-0.15) (2/37)	0.61 (28/37)
rad51	19.7 (16.6 – 34.1)	N/D	N/D	N/D

Table 8: Mutation rates and confidence intervals for *pTET-lys2-TAA* in dextrose

$PTET\text{-}lys2\text{-}TAA$ (Mutation rate x 10^{-8})

Relevant genotype	Total Mutation Rate (95%CI)	A>C rate (95%CI) (Fraction)	T>G rate (95%CI) (Fraction)	Other (Fraction)
WT	2.21 (1.75 – 2.42)	0.62 (0.26-1.11) (7/25)	0.09 (0.003-0.5) (1/25)	1.5 (17/25)
def1	2.53 (2.07 – 3.39)	0.633 (0.31-1.11) (10/40)	0.316 (0.12-0.72) (5/40)	1.58 (25/40)
apn1	6.34 (4.87 – 8.06)	4.08 (2.81-5.41) (38/59)	1.29 (0.63-2.17) (12/59)	0.97 (9/59)
apn1 rad26	10.3 (6.21 – 11.4)	8.85 (5.1-10.1) (43/50)	0.82 (0.21-2.05) (4/50)	0.62 (3/50)
apn1 rad7	8.43 (7 – 9.83)	3.49 (1.91-5.24) (12/29)	1.74 (0.71-3.39) (6/29)	3.2 (11/29)
apn1 def1	92.6 (67.7 – 111)	76.9 (53.4-93.8) (49/59)	9.42 (3.17-20.5) (6/59)	6.3 (4/59)
apn1 rad7 def1	40.8 (30.3 – 50.6)	N/D	N/D	N/D
apn1 rad26 rad7	11.9 (10.2 – 13)	8.1 (5.46-10.1) (19/28)	2.13 (0.8-4.54) (5/28)	1.7 (4/28)
apn1 rad26 def1	552 (427 – 666)	532 (401-643) (55/57)	<9.7 (0.24-60.9) (0/57)	19.4 (2/57)
apn1 rad14	592 (531 – 651)	462 (371-536) (43/55)	53.8 (19.7-124) (5/55)	75.3 (7/55)
apn1 rad14 def1	817 (632 – 885)	797 (593-866) (41/42)	19.4 (0.46-114) (1/42)	<19.4 (0/42)
apn1 srs2	6.10 (4.65 – 7.21)	N/D	N/D	N/D
apn1 rad26 srs2	10.6 (8.3 – 12.2)	N/D	N/D	N/D
apn1 sub1	8.64 (7.11 – 9.53)	N/D	N/D	N/D
apn1 rad26 sub1	10.1 (7.7 – 11.4)	N/D	N/D	N/D
apn1 dst1	4.4 (2.75 – 13.5)	N/D	N/D	N/D

apn1 rad26 dst1	15.3 (8.1 – 23.2)	N/D	N/D	N/D
srs2	2.64 (1.78 – 4.96)	N/D	N/D	N/D
sub1	1.78 (1.7 – 2.32)	N/D	N/D	N/D
dst1	1.36 (1.14 – 2.22)	N/D	N/D	N/D

4.3 Discussion

AP lesions, arguably the most prevalent type of spontaneously occurring DNA damage, are produced in cells through multiple pathways. Notably, AP lesions are obligatory intermediates in the repair of many different types of DNA damage including those produced by endogenously generated reactive oxygen species and by chemotherapeutic agents targeting thymidine synthesis pathway or directly damaging DNA by alkylation. AP sites present on the transcribed strand of a gene are particularly problematic since they not only endanger the genome integrity but also compromise efficient transcription by blocking RNA polymerases. It was previously reported that the TCR sub-pathway of NER can be recruited to AP lesions on the transcribed strand and removes this type of damage to a significant extent (Kim & Jinks-Robertson, 2010). A key aspect of the reporter assay used to demonstrate the involvement of Rad26 in TCR repair of AP lesions is that we can infer from the mutation spectra whether the relevant AP lesion was present on transcribed or non-transcribed strand. Exploiting this assay, it previously shown that, when TCR is compromised in *apn1 rad26* background, AP sites derived from uracil excision from DNA accumulate specifically on the transcribed strand. By expression of the engineered human glycosylase TDG that can excise undamaged thymine to generate AP sites, I demonstrate here that non-uracil derived AP sites are also processed and repaired *via* TCR.

AP lesions are poorly recognized by NER proteins but can be repaired by TCR by stalling RNAPII (Tornaletti et al., 2006), which subsequently interferes with transcription elongation and might be detrimental to the cell. In order to determine how multiple mechanisms involved in overcoming the obstacle to transcription might interact with the TCR repair of AP sites, I examined four potential regulators of stalled RNAPII. Sub1, Srs2, Dst1 and Def1 have been variously implicated in TCR repair of UV induced damages and/or transcription elongation (Awrey et al., 1997; Charlet-Berguerand et al., 2006; Epshtein et al., 2014; Reines & Mote, 1993; J. Y. Wang et al., 2004; Woudstra et al., 2002).

In the current report, I demonstrate that Srs2 and Sub1 are not involved in TCR repair of AP lesions since disruption of either proteins did not affect the spontaneous AP-associated A:T to C:G mutations at the *pTET-lys2-TAA* reporter in *apn1* or *apn1 rad26* background (Fig. 17). *DST1* deletion, on the other hand, resulted in up to a 10-fold reduction in mutation rates, which I showed was not due to the reduced transcription of the reporter gene. The reduction in mutation rate could be explained by more efficient repair in the absence of Dst1. I speculate that the action of Dst1 might be interfering with repair by encouraging transcriptional bypass of the AP lesion. In fact, Dst1 has been shown to induce RNAPII to bypass oxidative DNA damage (e.g. 8-oxoG) but not UV-induced damage (Kuraoka et al., 2007) leading to the suggestion that Dst1 might also promote transcriptional bypass of the AP lesions, which, like 8-xoG, is not helix-distorting. Interestingly, the effect of Dst1 disruption is not limited to the transcribed strand or TCR since mutations arising from AP lesions on both the transcribed (A>C) and the non-transcribed (T>G) DNA strands are significantly reduced and AP-associated mutagenesis reduced in both NER⁻ (*apn1 rad14)* and BER⁻ (*apn1 ntg1 ntg2*) strains*.* The elevation in homologous recombination upon the deletion of *DST1* suggests that, when present, Dst1 prevents AP lesion-repair by recombination through a mechanism yet to be defined. Disruption of *E. coli* GreA or GreB, which are the functional homologs of yeast Dst1, was recently shown to decrease mutations by increasing homologous recombination (C. Herman, Baylor College of Medicine, Houston, TX; Personal communications). In fact, in *E.coli*, *GreA* and *GreB,* which carry out similar functions as Dst1 during RNA polymerase stalling, promote transcription elongation at the expense of repair and repress recombination. This result is fully consistent with our data that recombination is enhanced when *DST1* is deleted (Figure 19D). To further understand the roles of Dst1, homologous recombination was disrupted by the deletion of *RAD51* (Table 7). In *apn1* or *apn1 dst1* background, *RAD51*-deletion did not affect the mutation rates indicating that recombination might be occurring in *RAD51*-independent manner (Pohl &

Nickoloff, 2008). It is also possible that the effect of Dst1-disruption on the AP-derived mutations is independent of its role in repressing homologous recombination. Further studies need to be performed to explore how yeast Dst1 or *E. Coli* GreA/GreB represses homologous recombination.

The degradation and clearance of RNAPII arrested at lesions require Def1, which mediates ubiquitination of the catalytic subunit of RNAPII (M. D. Wilson et al., 2013; Woudstra et al., 2002). Our effort to determine whether Def1 allows for the more efficient repair of AP lesions on the transcribed strand yielded a surprising result that the effect of Def1 disruption was highly dependent on the growth conditions (Fig. 21 and 22). When analyzed in media containing glycerol and ethanol as the major carbon source, we observed a modest (1.7-fold) increase in the mutation rate in *apn1 rad26* background as expected from previous findings that Rad26 normally inhibits the activity of Def1 and the degradation of RNAPII is used as a mechanism of last resort (Woudstra et al., 2002). However, in contrast to experiments in YEPGE, in dextrose-containing rich media (YEPD), Def1-disruption resulted in much greater increase in the mutation rate in both *apn1* (~15-fold) and *apn1 rad26* (~53-fold) backgrounds (Figs. 21A and 22A). During the fermentative growth in YEPD, the mutation rate of *apn1 rad26 def1* was similar to that of *apn1 rad14* suggesting that Def1 is involved in TCR repair of AP sites and that Rad26 and Def1 are sufficient in repairing all the AP sites in the transcribed strand without the need for additional factors when the cells are grown in dextrose. The effect of Def1 disruption in *apn1 rad14* background was also dependent on the growth conditions.

Yeast cells preferably grow *via* fermentation using dextrose as the carbon source. Using the non-fermentable carbon source such as ethanol and glycerol, yeast cells respire or grow oxidatively relying on the mitochondria. According to our data, the effect of cell metabolic state extends beyond the role of Def1; the overall mutation rates in *apn1* and *apn1 rad14* backgrounds showed significant differences depending on the growth condition

110

(Fig. 21A and 22A). I also observed that, although the effect of Srs2- or Sub1-disruption on the repair of AP lesion remained constant in YEPGE or YEPD, the effect of *DST1* deletion in *apn1 rad26* background on the AP-associated mutation rate was significantly dependent on the carbon source (Table 8). This difference is not due to different number of cell divisions during the culturing process since the starting and ending cell density during the experiments were not significantly different in these two growth conditions (data not shown). It has been previously shown in yeast that the metabolic state can be a determining factor in the types and numbers of endogenously-occurring DNA damage (Minesinger et al., 2006). There is a major difference in the accumulation of reactive oxygen species (ROS) in yeast cells growing by respiration (glycerol/ethanol) and by fermentation (dextrose) (Pan, 2011). The difference in the extent of the ROS-induced DNA damage in these two growth conditions is one potential explanation for the growth-condition-dependent effect of DEF1 deletion. However, the mutations in the *pTET-lys2-TAA* assay are originating not from oxidative damage but from uracil incorporated during DNA synthesis, leading us to suggest that the cell metabolic state and growth condition can also affect the various pathways involved in AP repair and/or transcriptional bypass. In *E. coli,* ppGpp, a small molecule mediator of the stringent response, was recently shown to play an important role in promoting TCR of helix-distorting DNA lesions (Kamarthapu et al., 2016). Stress response could similarly play an important role in shifting the balance among multiple pathways involved in the repair of AP lesions in yeast.

In summary, I report that Def1 and Dst1 play distinct yet significant function in how AP lesions in the transcribed genes are repaired (Fig. 23). I show that Def1, in addition to Rad26, can facilitate repair of AP lesions located on the transcribed strand by TCR as well as other repair pathways. I also describe a novel function of Dst1 in interfering with AP repair possibly by suppressing homologous recombination. Dst1-disruption reduces mutations resulting from the AP lesions located on both the transcribed and the non-

transcribed strands leading us to suggest that further study is needed to understand how the

role of Dst1 in AP lesion repair is mechanistically connected to its previously characterized

role in rescuing stalled transcription complexes.
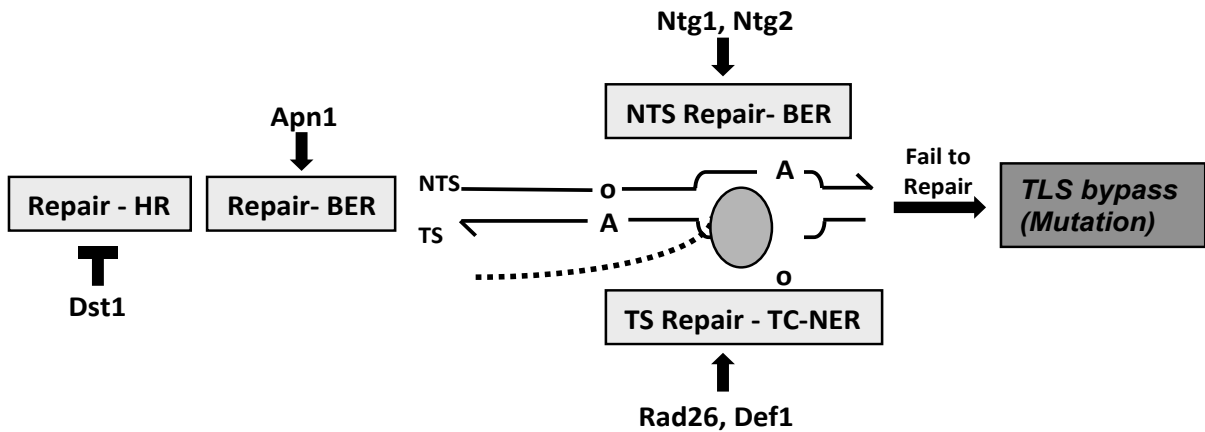
Figure 23: Proposed model of AP repair at transcribed regions

AP sites on both transcribed strand (TS) and non-transcribed strand (NTS) are repaired by either Apn1-mediated BER or homologous recombination pathway (HR). Repair by HR is suppressed by Dst1. While Ntg1/Ntg2-mediated BER repairs mostly the AP sites present on TS, Rad26- or Def1-mediated TC-NER repairs the AP sites on NTS. Those AP sites left unrepaired will invoke mutagenic translesion synthesis (TLS). Grey oval and dotted line represent RNA polymerase complex and nascent RNA chain, respectively. AP lesion is denoted by o.

Chapter V

Discussion and Future Directions

5.1 Summary

The goal of this work was to develop a better understanding of the role played by transcription in mutagenesis resulting from the incorporation of uracil into DNA. I show that high transcription rates correlate with high uracil density in the yeast genome, furthermore, I demonstrate that uracil incorporation at highly transcribed regions occurs during repair synthesis. In addition, I have characterized the roles played by two proteins, Dst1 and Def1, in the AP-induced transcription arrest.

In chapter 3, I showed that there are a greater number of uracil residues incorporated into DNA following elevation of the rate of transcription, using two complementary approaches: (i) a long-qPCR technique to quantifies the density of uracil at specific loci and (ii) the expression of a mutant glycosylase to further determine the contribution of uracil-density and uracil-glycosylase activity at highly transcribed genomic loci. In addition to these data, I provide evidence to support the hypothesis that uracil is incorporated into DNA via DNA repair synthesis. This was shown by using DNA damaging agents to induce damage and repair and assessing the effect on uracil dependent-mutagenesis and on uracil density in DNA.

In chapter 4, I examined the roles of Dst1 and Def1 in the transcription-coupled repair of abasic sites (AP sites) derived from uracil. First I described a novel role for Dst1 in promoting mutagenesis by interfering with the repair of AP lesions. Dst1 is presumably functioning by inhibiting homologous recombination. In addition, I demonstrated that Def1 participates in directing the transcription-coupled nucleotide excision repair (TCR) to those AP sites located in the highly transcribed regions. Surprisingly, its role in the regulation of TCR appears to be dependent on the yeast metabolic state. Overall, these studies contributed to the more comprehensive understanding of the mechanism underlying the transcription-associated mutagenesis resulting from the glycosylase-catalyzed excision of uracil bases from DNA.

5.2 Transcription levels and uracil DNA content

Previous studies in yeast using a tetracycline-regulatable promoter reversion assay found that there is a linear and proportional relationship between the level of transcription and the rate of uracil-derived mutagenesis (Kim et al., 2007; Kim & Jinks-Robertson, 2009, 2010). Specifically, mutations with a distinct signature, namely A:T to C:G transversions, were observed when base excision repair (BER) was disabled, as in *apn1* or *apn1 ntg1 ntg2* strains. These mutations were dramatically reduced when the uracil DNA glycosylase, Ung1, was disabled and also when *DUT1* was overexpressed. These results suggested that the mutagenesis observed was originating from the incorporation of uracil into DNA by DNA polymerases. The most surprising of these observations was that the mutations were almost completely suppressed when transcription of the reporter gene was repressed. These findings implied that the nucleotide composition in highly transcribed genomic regions is distinct from the other parts of the genome.

I tested the above hypothesis of transcription-dependent elevation of uracil residues in DNA by directly quantifying uracil residues at defined genomic loci under high- and low-transcription conditions using long-amplicon qPCR method. This method revealed a uracil 'hotspot' (i.e uracil enrichment at highly transcribed genomic regions). I observed a 2-fold increase in the uracil density at the *pTET-lys2-TAA* mutation reporter in *ung1* strains when transcription was activated (Fig. 8). In addition to the *pTET*-regulated reporter, the long amplicon qPCR method indicated that the uracil content was higher at the highly transcribed *TDH3* gene compared to the less transcribed *CAN1* gene. Additional uracil quantitation was performed following the disruption of Dcd1, which has been shown to elevate the genome-wide uracil incorporation (Bryan & Hesselberth, 2015) as well as uracil-dependent mutations (Fig. 7A). I showed a significant further increase in the uracil density at all three genomic locations tested when Dcd1 is disrupted (Fig.8). The ability to obtain equivalents results with two different primer sets targeting the region with *pTET-lys2-TAA* reporter and the elevated

116

uracil density measured in *dcd1* mutants consistent with the known role of Dcd1 protein reinforce the validity of the long qPCR method in quantitatively measuring the genomic uracil density in an accurate and reproducible manner.

The observation of general variation in uracil content in the genome has been previously implied in yeast (Bryan et al., 2014). It was demonstrated that replication origins are completely depleted of uracil residues and that altering nucleotide biosynthesis misregulates uracil-incorporation into DNA. To my knowledge, the findings described in this thesis are the first to directly highlight the correlation between the rate of transcription and the genomic uracil content. It is unknown whether the transcription-dependent regulation of uracil DNA content is conserved in humans and other organisms; however, due to the conservation of the transcription-associated mutagenesis (TAM) phenomenon, the preservation of major nucleotide biosynthesis enzymes between different species, and the importance of uracil DNA regulation, we suspect that this process would be conserved across most organisms.

Even though we observe a significant difference in uracil DNA content between highly transcribed genomic loci and loci transcribed at low rates, the uracil quantitation results do not directly correlate with the mutation rates results. For example, in *apn1* strains, the mutation rate at the *LYS2* reporter gene is elevated by 20-fold when the transcription is fully activated, while the uracil quantitation indicates a 2-fold difference in uracil content. This suggests that there are additional factors in addition to uracil DNA content, that could explain the increases in the mutation rates observed when transcription is activated. One of these factors could be a change in the activity of the uracil DNA glycosyase during transcription. To test this, I utilized a mutant glycosylase, CDG, which excises undamaged cytosines, in addition to uracil residues. The rationale behind the experiment is that if the glycosylase activity is not affected by transcription, then the rate of cytosine-derived mutations would remain the same whether the transcription of the reporter is activated or

117

repressed. However, if the glycosylase activity were enhanced by high transcription, then the number of cytosine-dependent mutations would increase when transcription of the reporter gene is highly transcribed. Surprisingly, the expression of this glycosylase led to only a small but still significant (~2 fold) transcription-dependent increase in the rate of cytosine-dependent mutations. This increase in cytosine-dependent mutations was much less than the 10-fold increase in uracil-dependent mutations that was observed. These observations suggest that the activity of the glycosylase is slightly enhanced during transcription; however, the difference observed in uracil-dependent mutagenesis with transcription is mainly a result of the higher uracil DNA content at these regions. Other potential explanations for the differences between the extents of elevation in mutagenesis and of increase in the uracil content include factors such as factors such as replication timing, position and orientation of the transcription machinery as well as the position of the replication fork that can modulate the incorporation of uracil into DNA.

Future studies should determine the correlation of uracil content and transcription genome-wide and further determine the location of genomic uracil genome-wide. This could be performed using two approaches 1) utilizing a biotinylated alkoxyamine that can label AP sites to perform a pull-down of uracil-enriched DNA followed by deep sequencing technique, and 2) directly identifying the positions of uracil in DNA by adding the sequence-specific linkers to the breaks generated from sites of uracil, followed by deep sequencing. Based on the results with the long-qPCR technique, I anticipate that highly transcribed genomic loci in yeast, such as rDNA and yeast transposon elements (TYs), would be significantly enriched with uracil.

5.3 Repair-dependent introduction of uracil into DNA

Using the long-amplicon qPCR and CDG expression, I have shown that there is increased uracil density at highly transcribed genomic loci. These uracil residues, if present

on the transcribed strand, have the potential to affect the binding of transcription factors, thereby affecting gene expression as previously shown (Luhnsdorf et al., 2014). The question of how transcription, a normal cellular process, can influence the density of a non-canonical nucleotide in DNA is yet to be elucidated. I hypothesize that the answer lies in the previous observations that highly transcribed genomic loci are more susceptible to spontaneous DNA damage (Gaillard & Aguilera, 2016). It is possible that the DNA synthesis that accompanies repair of transcription-induced damage occurs in regions with a high concentration of dUTP; this would lead to an increase in uracil incorporation into DNA. This could occur because eukaryotic DNA polymerases utilize dUTP as a substrate during DNA synthesis as efficiently as they use dTTP. This implies that the concentration of dUTP relative to dTTP in the nucleotide pool is the main limiting factor in the incorporation of uracil into DNA.

I showed that uracil incorporation into DNA can occur during repair of the damage introduced by DNA damaging agents such as 5-FU, 4NQO and CPT. Exogenously introducing damage (and thus subsequent repair) using 5-FU, CPT and 4NQO led to an increase in the uracil-dependent mutations at the reporter gene in the absence of BER (Fig. 9). Additionally, inducing damage using 5-FU and 4NQO led to an increase in the amount of uracil in DNA throughout the genome according to the experiments employing the chemical labeling of uracil-derived AP sites (Fig 12). I predict that this repair-dependent introduction of uracil into DNA is not only limited to damage introduced by 4NQO and CPT, but that it would occur following any type of DNA damage that activates repair, including those induced by high transcription.

Future studies should examine the effect of other DNA damaging agent besides 4NQO and CPT, which both induce recombination. It would be especially informative to utilize DNA damaging agents known to activate different repair pathways including BER, NER, Break-Induced Replication (BIR) and others. This would aid in further understanding

the specific repair synthesis pathways involved in the incorporation of uracil into DNA. In addition, it would be important to elucidate whether non-canonical, co-transcriptionally formed DNA structures, such as G4-DNA or R-loops, could induce damage to DNA and enhance incorporation of uracil into DNA. I anticipate that the degree of uracil incorporation would depend on the extent of the repair synthesis. For example, the DNA synthesis associated with recombination involves several kilobases of DNA surrounding the break and would lead to the higher uracil-incorporation in comparison to the synthesis associated with NER, which is only up to ~25bp per repair event. Interestingly, a high rate of mutagenesis, up to 2800-fold higher than spontaneous events, has been associated with BIR, a sub-pathway of homologous recombination that involves the repair of one-ended breaks and is associated with synthesis of up to hundreds of kilobases (Deem et al., 2011; Hicks et al., 2010; Kramara et al., 2018). These mutation events were shown to result from errors made by DNA polymerases including pol ζ and pol ε. It would be of interest to determine whether a significant amount of uracil is incorporated during the error-prone BIR-dependent repair synthesis. This could be explored by determining the effect of uracil-derived mutagenesis of the deletion of *POL32,* which encodes a non-essential subunit of pol δ required for BIR. These studies would further expand the understanding of the role played by the repair synthesis in the incorporation of uracil into DNA by elaborating on the specific repair pathways involved.

The mechanism that most likely explains how repair synthesis leads to the introduction of uracil into DNA lies in the cell-cycle specific regulation of dUTPase, which is highly expressed in S-phase (Cho et al., 1998; Ladner & Caradonna, 1997; Pardo & Gutierrez, 1990). dUTPase enzyme catalyzes dUTP into dUMP ensuring a low [dUTP]/[dTTP] in the nucleotide pool during replication (S-phase) and minimizes the incorporation of uracil during genome replication. For the DNA synthesis occurring outside of S-phase, such as that associated with repair, the available nucleotide pool has a relatively

higher [dUTP/dTTP] nucleotide pool. Therefore, the repair synthesis occurring during G1- or G2-phase of cell cycle has the higher risk of incorporating uracil into DNA. Reducing the [dUTP]/[dTTP] ratio by overexpressing dUTPase from the G1- or G2-specific promoters significantly reduced the mutation rates at the reporter under high transcription conditions (Fig. 14A and 14B). Interestingly, overexpressing dUTPase from the cell cycle specific promoters did not affect the mutation rates when transcription was lowered. These results indicate that the uracil incorporated into DNA during G1 or G2 is a main source of TAM (Fig.15). These data coupled with the fact that BIR is known to occur predominantly outside S-phase, specifically in G2/M phase, provide further impetus to determine whether this repair pathway is a major source of uracil incorporation into DNA (Kramara et al., 2018).

In summary, chapter 3 of this thesis describes a novel mechanism for introducing uracil into DNA during damage-induced repair synthesis occurring outside of S-phase. One major missing gap is the demonstration of induction of repair synthesis by transcription. The positive correlation between transcription and DNA damage has previously been shown, however, a direct correlation of transcription and unscheduled DNA synthesis is still lacking. To test this, bromodeoxyuridine immunoprecipitation (BrdU-IP) can be employed to quantify BrdU incorporated into DNA outside of S-phase. BrdU, a thymidine analog can be efficiently incorporated into newly synthesized DNA. Using the reporter system, the rates of BrdU incorporation into DNA at different cell cycles can be monitored. It is expected that enhanced BrdU incorporation into DNA would be observed at the mutation reporter under the high transcription conditions, which is associated with the accumulation of spontaneous damage, compared to the low-transcription conditions. Overall, I hypothesize that this role played by transcription in changing the nucleotide composition locally, would apply to other types of non-canonical DNA residues whose quantity in the genome has been shown to be affected by the rates of transcription. For example, I suspect that a similar mechanism could explain the elevated levels of ribonucleotides-dependent mutations, at highly transcribed

genomic regions (Kim, Huang, et al., 2011; Nick McElhinny et al., 2010). This is especially intriguing because similar to dUTPase, *RNR1,* the gene encoding for a subunit of ribonucleotide reductase is regulated in a cell-cycle dependent manner to ensure the maintenance of optimal [dNTP]/[rNTP] ratio during replication in S-phase (Elledge & Davis, 1990).

5.4 Role of Dst1 in transcription elongation

AP sites are necessary intermediates in the repair of many different types of DNA damage, including the excision of uracil from DNA. AP sites present on the transcribed DNA strand have been shown to block RNA polymerases; therefore, they interfere with transcription elongation. In addition to repair by BER, the TCR sub-pathway of NER can also remove AP sites that are located on the transcribed strand of DNA although the mechanism underlying this is poorly understood (Kim & Jinks-Robertson, 2010; Tornaletti et al., 2006). For example, it is unknown how different proteins interact with the stalled RNA polymerase at the AP lesion to allow for either the continued transcription elongation or the initiation of repair by TCR. In the TCR deficient *apn1 rad26* strains, uracil-derived AP sites accumulate specifically on the transcribed strand. However, the effect in the *apn1 rad26* strains is significantly less compared to the *apn1 rad14* strains where NER (and therefore TCR) is completely abolished. These findings suggest that factors additional to Rad26 exist to direct TCR to transcribed strand through interaction with the stalled RNA polymerases.

Dst1 has been implicated in TCR repair of several types of lesions, and it has been shown to be involved in transcription elongation (Awrey et al., 1997; Kuraoka et al., 2007). Dst1 restarts the backtracked RNA polymerase by stimulating the cleavage of the nascent RNA and realigning the 3' end of the transcript with the active site, further promoting elongation. In Chapter 4 of this thesis, I described the role of Dst1 in TCR. Surprisingly, the deletion of *DST1* resulted in an overall reduction in mutation rates both at the *pTET-lys2-*

TAA reporter gene and at the *CAN1* locus. This observation suggests that there is more efficient repair in the absence of Dst1; the presence of Dst1 interferes with the efficient repair of AP sites. Furthermore, the deletion of *DST1* led to a significant increase in homologous recombination (HR). One hypothesis to explain these results is that, Dst1, through its interaction with the stalled RNA polymerase complex, prevents AP lesion-repair *via* recombination. In *E. coli*, GreA, the functional homolog of Dst1, was recently shown to interfere with the DNA break repair by HR. It was shown that the accumulation of backtracked RNA polymerases in the absence of GreA allows for DNA damage to be repaired *via* recombination (Sivaramakrishnan et al., 2017). I suspect that the same mechanism applies in the case of Dst1 in yeast because I observe enhanced recombination and reduced mutagenesis in the absence of Dst1. A straightforward study to determine the role of Dst1 in inhibiting HR repair would be to determine the ability of *dst1* mutant to withstand double-strand break (DSB) damage. These mutants can be treated with DSB-inducing chemicals such as Bleomycin and the sensitivity of the cells examined. The cells will be highly resistant to these drugs if the absence of Dst1 enhances DSB repair by HR. Future studies should also incorporate biochemical assays to further assess the role of Dst1 in the AP site repair and recombination.

5.5 Role of Def1 in transcription elongation

Stalled/arrested RNA polymerase complex at lesions sites can be degraded to clear the way for the next round of transcription or for the repair proteins to assemble. In yeast, Def1 mediates the degradation of RNAPII in a ubiquitination and proteasome-dependent manner (M. D. Wilson et al., 2013; Woudstra et al., 2002). In an attempt to further understand the mechanisms of the TCR repair of AP sites, I questioned whether Def1 allows for more efficient repair of AP lesions on the transcribed DNA strand by degrading the RNAPII.

The role of Def1 in the repair of AP sites was highly dependent on the growth conditions (Figure 21 and 22). Def1-disruption experiments were performed in media containing either glycerol/ethanol or dextrose as the major carbon source. In both conditions, Def1-disruption resulted in a significant increase in the mutations arising from the unrepaired AP sites on the transcribed strand in *apn1 rad26* background, indicating that Def1's function is important in TCR repair of AP sites. However, the mutation rates were significantly greater when cells were grown in dextrose than in glycerol/ethanol (Fig. 21 and 22.)

In yeast, the metabolic state has been shown to affect both the type and amount of endogenous DNA damage (Minesinger et al., 2006). The accumulation of reactive oxygen species (ROS) in yeast cells growing by respiration (glycerol/ethanol) and by fermentation (dextrose) was shown to be very distinct (Pan, 2011). It is possible that the difference in mutation rates with the deletion of *DEF1* in these two growth conditions could be explained by the difference in the amounts of damages induced or alternatively, the difference in how the repair pathways might be regulated. Future studies should elucidate the effect of yeast metabolic state on mutagenesis including investigating the roles played by different stress inducers such as ROS or heat on not only the types and amount of DNA damage induced, but also the regulation of various DNA repair proteins.

5.6 Overall Conclusions.

This thesis has revealed a novel mechanism for introducing uracil into DNA and elucidated the role of Dst1 and Def1 in dealing with AP sites at highly transcribed loci. This is the first evidence for transcription-induced disproportionate distribution of uracil-DNA content. Overall, these studies further enhance our understanding of the mechanisms underlying the mutations associated with transcription, specifically those resulting from uracil misincorporation. If the correlation between the rate of transcription and the extent of uracil-

incorporation into DNA content is conserved in higher organisms such as humans, then these findings will further contribute in the designing of more targeted antimetabolite chemotherapy agents.

Bibliography

Adams, M. D., Celniker, S. E., Holt, R. A., Evans, C. A., Gocayne, J. D., Amanatides, P. G., Scherer, S. E., Li, P. W., Hoskins, R. A., Galle, R. F., George, R. A., Lewis, S. E., Richards, S., Ashburner, M., Henderson, S. N., Sutton, G. G., Wortman, J. R., Yandell, M. D., Zhang, Q., Chen, L. X., Brandon, R. C., Rogers, Y. H., Blazej, R. G., Champe, M., Pfeiffer, B. D., Wan, K. H., Doyle, C., Baxter, E. G., Helt, G., Nelson, C. R., Gabor, G. L., Abril, J. F., Agbayani, A., An, H. J., Andrews-Pfannkoch, C., Baldwin, D., Ballew, R. M., Basu, A., Baxendale, J., Bayraktaroglu, L., Beasley, E. M., Beeson, K. Y., Benos, P. V., Berman, B. P., Bhandari, D., Bolshakov, S., Borkova, D., Botchan, M. R., Bouck, J., Brokstein, P., Brottier, P., Burtis, K. C., Busam, D. A., Butler, H., Cadieu, E., Center, A., Chandra, I., Cherry, J. M., Cawley, S., Dahlke, C., Davenport, L. B., Davies, P., de Pablos, B., Delcher, A., Deng, Z., Mays, A. D., Dew, I., Dietz, S. M., Dodson, K., Doup, L. E., Downes, M., Dugan-Rocha, S., Dunkov, B. C., Dunn, P., Durbin, K. J., Evangelista, C. C., Ferraz, C., Ferriera, S., Fleischmann, W., Fosler, C., Gabrielian, A. E., Garg, N. S., Gelbart, W. M., Glasser, K., Glodek, A., Gong, F., Gorrell, J. H., Gu, Z., Guan, P., Harris, M., Harris, N. L., Harvey, D., Heiman, T. J., Hernandez, J. R., Houck, J., Hostin, D., Houston, K. A., Howland, T. J., Wei, M. H., Ibegwam, C., Jalali, M., Kalush, F., Karpen, G. H., Ke, Z., Kennison, J. A., Ketchum, K. A., Kimmel, B. E., Kodira, C. D., Kraft, C., Kravitz, S., Kulp, D., Lai, Z., Lasko, P., Lei, Y., Levitsky, A. A., Li, J., Li, Z., Liang, Y., Lin, X., Liu, X., Mattei, B., McIntosh, T. C., McLeod, M. P., McPherson, D., Merkulov, G., Milshina, N. V., Mobarry, C., Morris, J., Moshrefi, A., Mount, S. M., Moy, M., Murphy, B., Murphy, L., Muzny, D. M., Nelson, D. L., Nelson, D. R., Nelson, K. A., Nixon, K., Nusskern, D. R., Pacleb, J. M., Palazzolo, M., Pittman, G. S., Pan, S., Pollard, J., Puri, V., Reese, M. G., Reinert, K., Remington, K., Saunders, R. D., Scheeler, F., Shen, H., Shue, B. C., Siden-Kiamos, I., Simpson, M., Skupski, M. P.,

126

Smith, T., Spier, E., Spradling, A. C., Stapleton, M., Strong, R., Sun, E., Svirskas, R., Tector, C., Turner, R., Venter, E., Wang, A. H., Wang, X., Wang, Z. Y., Wassarman, D. A., Weinstock, G. M., Weissenbach, J., Williams, S. M., WoodageT, Worley, K. C., Wu, D., Yang, S., Yao, Q. A., Ye, J., Yeh, R. F., Zaveri, J. S., Zhan, M., Zhang, G., Zhao, Q., Zheng, L., Zheng, X. H., Zhong, F. N., Zhong, W., Zhou, X., Zhu, S., Zhu, X., Smith, H. O., Gibbs, R. A., Myers, E. W., Rubin, G. M., & Venter, J. C. (2000). The genome sequence of Drosophila melanogaster. *Science, 287*(5461), 2185-2195.

Aguilera, A. (2002). The connection between transcription and genomic instability. *EMBO J, 21*(3), 195-201.

Aguilera, A., & Garcia-Muse, T. (2012). R loops: from transcription byproducts to threats to genome stability. *Mol Cell, 46*(2), 115-124.

Alexandrov, L. B., Nik-Zainal, S., Wedge, D. C., Aparicio, S. A., Behjati, S., Biankin, A. V., Bignell, G. R., Bolli, N., Borg, A., Borresen-Dale, A. L., Boyault, S., Burkhardt, B., Butler, A. P., Caldas, C., Davies, H. R., Desmedt, C., Eils, R., Eyfjord, J. E., Foekens, J. A., Greaves, M., Hosoda, F., Hutter, B., Ilicic, T., Imbeaud, S., Imielinsk, M., Jager, N., Jones, D. T., Jones, D., Knappskog, S., Kool, M., Lakhani, S. R., Lopez-Otin, C., Martin, S., Munshi, N. C., Nakamura, H., Northcott, P. A., Pajic, M., Papaemmanuil, E., Paradiso, A., Pearson, J. V., Puente, X. S., Raine, K., Ramakrishna, M., Richardson, A. L., Richter, J., Rosenstiel, P., Schlesner, M., Schumacher, T. N., Span, P. N., Teague, J. W., Totoki, Y., Tutt, A. N., Valdes-Mas, R., van Buuren, M. M., van 't Veer, L., Vincent-Salomon, A., Waddell, N., Yates, L. R., Australian Pancreatic Cancer Genome, I., Consortium, I. B. C., Consortium, I. M.-S., PedBrain, I., Zucman-Rossi, J., Futreal, P. A., McDermott, U., Lichter, P., Meyerson, M., Grimmond, S. M., Siebert, R., Campo, E., Shibata, T., Pfister, S. M.,

Campbell, P. J., & Stratton, M. R. (2013). Signatures of mutational processes in human cancer. *Nature, 500*(7463), 415-421.

Andersen, S. L., Sloan, R. S., Petes, T. D., & Jinks-Robertson, S. (2015). Genome-destabilizing effects associated with top1 loss or accumulation of top1 cleavage complexes in yeast. *PLoS Genet, 11*(4), e1005098.

Auerbach, P., Bennett, R. A., Bailey, E. A., Krokan, H. E., & Demple, B. (2005). Mutagenic specificity of endogenously generated abasic sites in Saccharomyces cerevisiae chromosomal DNA. *Proc Natl Acad Sci U S A, 102*(49), 17711-17716.

Auerbach, P. A., & Demple, B. (2010). Roles of Rev1, Pol zeta, Pol32 and Pol eta in the bypass of chromosomal abasic sites in Saccharomyces cerevisiae. *Mutagenesis, 25*(1), 63-69.

Awrey, D. E., Weilbaecher, R. G., Hemming, S. A., Orlicky, S. M., Kane, C. M., & Edwards, A. M. (1997). Transcription elongation through DNA arrest sites. A multistep process involving both RNA polymerase II subunit RPB9 and TFIIS. *J Biol Chem, 272*(23), 14747-14754.

Balestrieri, E., Zanier, R., & Degrassi, F. (2001). Molecular characterisation of camptothecin-induced mutations at the hprt locus in Chinese hamster cells. *Mutat Res, 476*(1-2), 63-69.

Beaudenon, S. L., Huacani, M. R., Wang, G., McDonnell, D. P., & Huibregtse, J. M. (1999). Rsp5 ubiquitin-protein ligase mediates DNA damage-induced degradation of the large subunit of RNA polymerase II in Saccharomyces cerevisiae. *Mol Cell Biol, 19*(10), 6972-6979.

Beletskii, A., & Bhagwat, A. S. (1996). Transcription-induced mutations: increase in C to T mutations in the nontranscribed strand during transcription in Escherichia coli. *Proc Natl Acad Sci U S A, 93*(24), 13919-13924.

Belotserkovskii, B. P., Liu, R., Tornaletti, S., Krasilnikova, M. M., Mirkin, S. M., & Hanawalt, P. C. (2010). Mechanisms and implications of transcription blockage by guanine-rich DNA sequences. *Proc Natl Acad Sci U S A, 107*(29), 12816-12821.

Berset, C., Griac, P., Tempel, R., La Rue, J., Wittenberg, C., & Lanker, S. (2002). Transferable domain in the G(1) cyclin Cln2 sufficient to switch degradation of Sic1 from the E3 ubiquitin ligase SCF(Cdc4) to SCF(Grr1). *Mol Cell Biol, 22*(13), 4463-4476.

Bessman, M. J., Lehman, I. R., Adler, J., Zimmerman, S. B., Simms, E. S., & Kornberg, A. (1958). Enzymatic Synthesis of Deoxyribonucleic Acid. Iii. The Incorporation of Pyrimidine and Purine Analogues into Deoxyribonucleic Acid. *Proc Natl Acad Sci U S A, 44*(7), 633-640.

Bohr, V. A., Smith, C. A., Okumoto, D. S., & Hanawalt, P. C. (1985). DNA repair in an active gene: removal of pyrimidine dimers from the DHFR gene of CHO cells is much more efficient than in the genome overall. *Cell, 40*(2), 359-369.

Boiteux, S., & Guillet, M. (2004). Abasic sites in DNA: repair and biological consequences in Saccharomyces cerevisiae. *DNA Repair (Amst), 3*(1), 1-12.

Boiteux, S., & Jinks-Robertson, S. (2013). DNA repair mechanisms and the bypass of DNA damage in Saccharomyces cerevisiae. *Genetics, 193*(4), 1025-1064.

129

Bryan, D. S., & Hesselberth, J. R. (2015). Temporal regulation of dUTP biosynthesis limits uracil incorporation during early DNA replication. *bioRxiv, doi.org/10.1101/027508*.

Bryan, D. S., Ransom, M., Adane, B., York, K., & Hesselberth, J. R. (2014). High resolution mapping of modified DNA nucleobases using excision repair enzymes. *Genome Res, 24*(9), 1534-1542.

Brynolf, K., Eliasson, R., & Reichard, P. (1978). Formation of Okazaki fragments in polyoma DNA synthesis caused by misincorporation of uracil. *Cell, 13*(3), 573-580.

Cadet, J., Sage, E., & Douki, T. (2005). Ultraviolet radiation-mediated damage to cellular DNA. *Mutat Res, 571*(1-2), 3-17.

Caradonna, S. J., & Cheng, Y. C. (1980). The role of deoxyuridine triphosphate nucleotidohydrolase, uracil-DNA glycosylase, and DNA polymerase alpha in the metabolism of FUdR in human tumor cells. *Mol Pharmacol, 18*(3), 513-520.

Castellano-Pozo, M., Garcia-Muse, T., & Aguilera, A. (2012). R-loops cause replication impairment and genome instability during meiosis. *EMBO Rep, 13*(10), 923-929.

Castillo-Acosta, V. M., Estevez, A. M., Vidal, A. E., Ruiz-Perez, L. M., & Gonzalez-Pacanowska, D. (2008). Depletion of dimeric all-alpha dUTPase induces DNA strand breaks and impairs cell cycle progression in Trypanosoma brucei. *Int J Biochem Cell Biol, 40*(12), 2901-2913.

Cerritelli, S. M., & Crouch, R. J. (2009). Ribonuclease H: the enzymes in eukaryotes. *The FEBS journal, 276*(6), 1494-1505.

Chan, G. L., Doetsch, P. W., & Haseltine, W. A. (1985). Cyclobutane pyrimidine dimers and (6-4) photoproducts block polymerization by DNA polymerase I. *Biochemistry, 24*(21), 5723-5728.

Chan, Y. A., Aristizabal, M. J., Lu, P. Y., Luo, Z., Hamza, A., Kobor, M. S., Stirling, P. C., & Hieter, P. (2014). Genome-wide profiling of yeast DNA:RNA hybrid prone sites with DRIP-chip. *PLoS Genet, 10*(4), e1004288.

Charlet-Berguerand, N., Feuerhahn, S., Kong, S. E., Ziserman, H., Conaway, J. W., Conaway, R., & Egly, J. M. (2006). RNA polymerase II bypass of oxidative DNA damage is regulated by transcription elongation factors. *EMBO J, 25*(23), 5481-5491.

Chavez, S., & Aguilera, A. (1997). The yeast HPR1 gene has a functional role in transcriptional elongation that uncovers a novel source of genome instability. *Genes Dev, 11*(24), 3459-3470.

Chen, R., Wang, H., & Mansky, L. M. (2002). Roles of uracil-DNA glycosylase and dUTPase in virus replication. *J Gen Virol, 83*(Pt 10), 2339-2345.

Cho, R. J., Campbell, M. J., Winzeler, E. A., Steinmetz, L., Conway, A., Wodicka, L., Wolfsberg, T. G., Gabrielian, A. E., Landsman, D., Lockhart, D. J., & Davis, R. W. (1998). A genome-wide transcriptional analysis of the mitotic cell cycle. *Mol Cell, 2*(1), 65-73.

Chu, E., Callender, M. A., Farrell, M. P., & Schmitz, J. C. (2003). Thymidylate synthase inhibitors as anticancer agents: from bench to bedside. *Cancer Chemother Pharmacol, 52 Suppl 1*, S80-89.

Cordeiro-Stone, M., Zaritskaya, L. S., Price, L. K., & Kaufmann, W. K. (1997). Replication fork bypass of a pyrimidine dimer blocking leading strand DNA synthesis. *J Biol Chem, 272*(21), 13945-13954.

Costi, M. P., Ferrari, S., Venturelli, A., Calo, S., Tondi, D., & Barlocco, D. (2005). Thymidylate synthase structure, function and implication in drug discovery. *Curr Med Chem, 12*(19), 2241-2258.

Cox, M. M., Goodman, M. F., Kreuzer, K. N., Sherratt, D. J., Sandler, S. J., & Marians, K. J. (2000). The importance of repairing stalled replication forks. *Nature, 404*(6773), 37-41.

Crick, F. (1974). The double helix: a personal view. *Nature, 248*(5451), 766-769.

Danenberg, P. V., Heidelberger, C., Mulkins, M. A., & Peterson, A. R. (1981). The incorporation of 5-fluoro-2'-deoxyuridine into DNA of mammalian tumor cells. *Biochem Biophys Res Commun, 102*(2), 654-658.

Danenberg, P. V., & Lockshin, A. (1981). Fluorinated pyrimidines as tight-binding inhibitors of thymidylate synthetase. *Pharmacol Ther, 13*(1), 69-90.

Datta, A., & Jinks-Robertson, S. (1995). Association of increased spontaneous mutation rates with high levels of transcription in yeast. *Science, 268*(5217), 1616-1619.

De Bont, R., & van Larebeke, N. (2004). Endogenous DNA damage in humans: a review of quantitative data. *Mutagenesis, 19*(3), 169-185.

Deem, A., Keszthelyi, A., Blackgrove, T., Vayl, A., Coffey, B., Mathur, R., Chabes, A., & Malkova, A. (2011). Break-induced replication is highly inaccurate. *PLoS Biol, 9*(2), e1000594.

Demple, B., & Harrison, L. (1994). Repair of oxidative damage to DNA: enzymology and biology. *Annu Rev Biochem, 63*, 915-948.

Dengg, M., Garcia-Muse, T., Gill, S. G., Ashcroft, N., Boulton, S. J., & Nilsen, H. (2006). Abrogation of the CLK-2 checkpoint leads to tolerance to base-excision repair intermediates. *EMBO Rep, 7*(10), 1046-1051.

Dubois, E., Cordoba-Canero, D., Massot, S., Siaud, N., Gakiere, B., Domenichini, S., Guerard, F., Roldan-Arjona, T., & Doutriaux, M. P. (2011). Homologous recombination is stimulated by a decrease in dUTPase in Arabidopsis. *PLoS One, 6*(4), e18658.

Duquette, M. L., Handa, P., Vincent, J. A., Taylor, A. F., & Maizels, N. (2004). Intracellular transcription of G-rich DNAs induces formation of G-loops, novel structures containing G4 DNA. *Genes Dev, 18*(13), 1618-1629.

An early suggestion of DNA Repair. Effect os sublethal doses of monochromatic ultraviolet radiation on bacteria in liquid suspensions. By Alexander Hollaender and John T. Curtis. Proc Soc Exp Biol Med, 33,61-62(1935). (1975). *Basic Life Sci, 5A*, xi-xii.

El Hage, A., Webb, S., Kerr, A., & Tollervey, D. (2014). Genome-wide distribution of RNA-DNA hybrids identifies RNase H targets in tRNA genes, retrotransposons and mitochondria. *PLoS Genet, 10*(10), e1004716.

el-Hajj, H. H., Zhang, H., & Weiss, B. (1988). Lethality of a dut (deoxyuridine triphosphatase) mutation in Escherichia coli. *J Bacteriol, 170*(3), 1069-1075.

Elledge, S. J., & Davis, R. W. (1990). Two genes differentially regulated in the cell cycle and by DNA-damaging agents encode alternative regulatory subunits of ribonucleotide reductase. *Genes Dev, 4*(5), 740-751.

Epshtein, V., Kamarthapu, V., McGary, K., Svetlov, V., Ueberheide, B., Proshkin, S., Mironov, A., & Nudler, E. (2014). UvrD facilitates DNA repair by pulling RNA polymerase backwards. *Nature, 505*(7483), 372-377.

Farber, S., & Diamond, L. K. (1948). Temporary remissions in acute leukemia in children produced by folic acid antagonist, 4-aminopteroyl-glutamic acid. *N Engl J Med, 238*(23), 787-793.

Firbank, S. J., Wardle, J., Heslop, P., Lewis, R. J., & Connolly, B. A. (2008). Uracil recognition in archaeal DNA polymerases captured by X-ray crystallography. *J Mol Biol, 381*(3), 529-539.

Fix, D. F., & Glickman, B. W. (1987). Asymmetric cytosine deamination revealed by spontaneous mutational specificity in an Ung- strain of Escherichia coli. *Mol Gen Genet, 209*(1), 78-82.

Fogg, M. J., Pearl, L. H., & Connolly, B. A. (2002). Structural basis for uracil recognition by archaeal family B DNA polymerases. *Nat Struct Biol, 9*(12), 922-927.

Freireich, E. J. (1967). The management of acute leukemia. *Can Med Assoc J, 96*(25), 1605-1610.

Friedberg, E. C. (2008). A brief history of the DNA repair field. *Cell Res, 18*(1), 3-7.

Gadsden, M. H., McIntosh, E. M., Game, J. C., Wilson, P. J., & Haynes, R. H. (1993). dUTP pyrophosphatase is an essential enzyme in Saccharomyces cerevisiae. *The EMBO journal, 12*(11), 4425-4431.

Gaillard, H., & Aguilera, A. (2013). Transcription coupled repair at the interface between transcription elongation and mRNP biogenesis. *Biochim Biophys Acta, 1829*(1), 141-150.

Gaillard, H., & Aguilera, A. (2016). Transcription as a Threat to Genome Integrity. *Annu Rev Biochem, 85*, 291-317.

Gaillard, H., Tous, C., Botet, J., Gonzalez-Aguilera, C., Quintero, M. J., Viladevall, L., Garcia-Rubio, M. L., Rodriguez-Gil, A., Marin, A., Arino, J., Revuelta, J. L., Chavez, S., & Aguilera, A. (2009). Genome-wide analysis of factors affecting transcription elongation and DNA repair: a new role for PAF and Ccr4-not in transcription-coupled repair. *PLoS Genet, 5*(2), e1000364.

Garcia-Rubio, M., Huertas, P., Gonzalez-Barrera, S., & Aguilera, A. (2003). Recombinogenic effects of DNA-damaging agents are synergistically increased by transcription in Saccharomyces cerevisiae. New insights into transcription-associated recombination. *Genetics, 165*(2), 457-466.

Gellon, L., Barbey, R., Auffret van der Kemp, P., Thomas, D., & Boiteux, S. (2001). Synergism between base excision repair, mediated by the DNA glycosylases Ntg1 and Ntg2, and nucleotide excision repair in the removal of oxidatively damaged DNA bases in Saccharomyces cerevisiae. *Mol Genet Genomics, 265*(6), 1087-1096.

Gibbs, P. E., McDonald, J., Woodgate, R., & Lawrence, C. W. (2005). The relative roles in vivo of Saccharomyces cerevisiae Pol eta, Pol zeta, Rev1 protein and Pol32 in the

bypass and mutation induction of an abasic site, T-T (6-4) photoadduct and T-T cis-syn cyclobutane dimer. *Genetics, 169*(2), 575-582.

Glazer, R. I., & Lloyd, L. S. (1982). Association of cell lethality with incorporation of 5-fluorouracil and 5-fluorouridine into nuclear RNA in human colon carcinoma cells in culture. *Mol Pharmacol, 21*(2), 468-473.

Goldin, A., Venditti, J. M., Humphreys, S. R., Dennis, D., Mantel, N., & Greenhouse, S. W. (1955). A quantitative comparison of the antileukemic effectiveness of two folic acid antagonists in mice. *J Natl Cancer Inst, 15*(6), 1657-1664.

Gomez-Gonzalez, B., & Aguilera, A. (2007). Activation-induced cytidine deaminase action is strongly stimulated by mutations of the THO complex. *Proc Natl Acad Sci U S A, 104*(20), 8409-8414.

Gottipati, P., Cassel, T. N., Savolainen, L., & Helleday, T. (2008). Transcription-associated recombination is dependent on replication in Mammalian cells. *Mol Cell Biol, 28*(1), 154-164.

Greagg, M. A., Fogg, M. J., Panayotou, G., Evans, S. J., Connolly, B. A., & Pearl, L. H. (1999). A read-ahead function in archaeal DNA polymerases detects promutagenic template-strand uracil. *Proc Natl Acad Sci U S A, 96*(16), 9045-9050.

Gueldener, U., Heinisch, J., Koehler, G. J., Voss, D., & Hegemann, J. H. (2002). A second set of loxP marker cassettes for Cre-mediated multiple gene knockouts in budding yeast. *Nucleic Acids Res, 30*(6), e23.

Guillet, M., & Boiteux, S. (2003). Origin of endogenous DNA abasic sites in Saccharomyces cerevisiae. *Mol Cell Biol, 23*(22), 8386-8394.

Guillet, M., Van Der Kemp, P. A., & Boiteux, S. (2006). dUTPase activity is critical to maintain genetic stability in Saccharomyces cerevisiae. *Nucleic acids research, 34*(7), 2056-2066.

Guo, J., Hanawalt, P. C., & Spivak, G. (2013). Comet-FISH with strand-specific probes reveals transcription-coupled repair of 8-oxoGuanine in human cells. *Nucleic Acids Res, 41*(16), 7700-7712.

Guzder, S. N., Habraken, Y., Sung, P., Prakash, L., & Prakash, S. (1995). Reconstitution of yeast nucleotide excision repair with purified Rad proteins, replication protein A, and transcription factor TFIIH. *J Biol Chem, 270*(22), 12973-12976.

Guzder, S. N., Sung, P., Prakash, L., & Prakash, S. (1997). Yeast Rad7-Rad16 complex, specific for the nucleotide excision repair of the nontranscribed DNA strand, is an ATP-dependent DNA damage sensor. *J Biol Chem, 272*(35), 21665-21668.

Hagenkort, A., Paulin, C. B. J., Desroses, M., Sarno, A., Wiita, E., Mortusewicz, O., Koolmeister, T., Loseva, O., Jemth, A. S., Almlof, I., Homan, E., Lundback, T., Gustavsson, A. L., Scobie, M., & Helleday, T. (2017). dUTPase inhibition augments replication defects of 5-Fluorouracil. *Oncotarget, 8*(14), 23713-23726.

Hall, A. C., Ostrowski, L. A., Pietrobon, V., & Mekhail, K. (2017). Repetitive DNA loci and their modulation by the non-canonical nucleic acid structures R-loops and G-quadruplexes. *Nucleus, 8*(2), 162-181.

Hanawalt, P. C. (2002). Subpathways of nucleotide excision repair and their regulation. *Oncogene, 21*(58), 8949-8956.

Heidelberger, C., Chaudhuri, N. K., Danneberg, P., Mooren, D., Griesbach, L., Duschinsky, R., Schnitzer, R. J., Pleven, E., & Scheiner, J. (1957). Fluorinated pyrimidines, a new class of tumour-inhibitory compounds. *Nature, 179*(4561), 663-666.

Heintz, N. (1991). The regulation of histone gene expression during the cell cycle. *Biochim Biophys Acta, 1088*(3), 327-339.

Herman, R. K., & Dworkin, N. B. (1971). Effect of gene induction on the rate of mutagenesis by ICR-191 in Escherichia coli. *J Bacteriol, 106*(2), 543-550.

Herrera-Moyano, E., Mergui, X., Garcia-Rubio, M. L., Barroso, S., & Aguilera, A. (2014). The yeast and human FACT chromatin-reorganizing complexes solve R-loop-mediated transcription-replication conflicts. *Genes Dev, 28*(7), 735-748.

Herrick, D., & Kufe, D. W. (1984). Lethality associated with incorporation of 5-fluorouracil into preribosomal RNA. *Mol Pharmacol, 26*(1), 135-140.

Hicks, W. M., Kim, M., & Haber, J. E. (2010). Increased mutagenesis and unique mutation signature associated with mitotic gene conversion. *Science, 329*(5987), 82-85.

Hochhauser, S. J., & Weiss, B. (1978). Escherichia coli mutants deficient in deoxyuridine triphosphatase. *J Bacteriol, 134*(1), 157-166.

Horvath, A., Batki, J., Henn, L., Lukacsovich, T., Rona, G., Erdelyi, M., & Vertessy, B. G. (2015). dUTPase expression correlates with cell division potential in Drosophila melanogaster. *FEBS J, 282*(10), 1998-2013.

Huertas, P., & Aguilera, A. (2003). Cotranscriptionally formed DNA:RNA hybrids mediate transcription elongation impairment and transcription-associated recombination. *Mol Cell, 12*(3), 711-721.

Hunter, S. E., Jung, D., Di Giulio, R. T., & Meyer, J. N. (2010). The QPCR assay for analysis of mitochondrial DNA damage, repair, and relative copy number. *Methods, 51*(4), 444-451.

Ikenaga, M., Ishii, Y., Tada, M., Kakunaga, T., & Takebe, H. (1975). Excision-repair of 4-nitroquinolin-1-oxide damage responsible for killing, mutation, and cancer. *Basic Life Sci, 5B*, 763-771.

Jinks-Robertson, S., & Bhagwat, A. S. (2014). Transcription-associated mutagenesis. *Annu Rev Genet, 48*, 341-359.

Kamarthapu, V., Epshtein, V., Benjamin, B., Proshkin, S., Mironov, A., Cashel, M., & Nudler, E. (2016). ppGpp couples transcription to DNA repair in E. coli. *Science, 352*(6288), 993-996.

Kavli, B., Otterlei, M., Slupphaug, G., & Krokan, H. E. (2007). Uracil in DNA--general mutagen, but normal intermediate in acquired immunity. *DNA Repair (Amst), 6*(4), 505-516.

Kavli, B., Slupphaug, G., Mol, C. D., Arvai, A. S., Peterson, S. B., Tainer, J. A., & Krokan, H. E. (1996). Excision of cytosine and thymine from DNA by mutants of human uracil-DNA glycosylase. *EMBO J, 15*(13), 3442-3447.

Kawahara, A., Akagi, Y., Hattori, S., Mizobe, T., Shirouzu, K., Ono, M., Yanagawa, T., Kuwano, M., & Kage, M. (2009). Higher expression of deoxyuridine triphosphatase (dUTPase) may predict the metastasis potential of colorectal cancer. *J Clin Pathol, 62*(4), 364-369.

Keil, R. L., & Roeder, G. S. (1984). Cis-acting, recombination-stimulating activity in a fragment of the ribosomal DNA of S. cerevisiae. *Cell, 39*(2 Pt 1), 377-386.

Kelner, A. (1949). Effect of Visible Light on the Recovery of Streptomyces Griseus Conidia from Ultra-violet Irradiation Injury. *Proc Natl Acad Sci U S A, 35*(2), 73-79.

Kiljunen, S., Hakala, K., Pinta, E., Huttunen, S., Pluta, P., Gador, A., Lonnberg, H., & Skurnik, M. (2005). Yersiniophage phiR1-37 is a tailed bacteriophage having a 270 kb DNA genome with thymidine replaced by deoxyuridine. *Microbiology, 151*(Pt 12), 4093-4102.

Kim, N., Abdulovic, A. L., Gealy, R., Lippert, M. J., & Jinks-Robertson, S. (2007). Transcription-associated mutagenesis in yeast is directly proportional to the level of gene expression and influenced by the direction of DNA replication. *DNA Repair (Amst), 6*(9), 1285-1296.

Kim, N., Huang, S. Y., Williams, J. S., Li, Y. C., Clark, A. B., Cho, J. E., Kunkel, T. A., Pommier, Y., & Jinks-Robertson, S. (2011). Mutagenic processing of ribonucleotides in DNA by yeast topoisomerase I. *Science, 332*(6037), 1561-1564.

Kim, N., & Jinks-Robertson, S. (2009). dUTP incorporation into genomic DNA is linked to transcription in yeast. *Nature, 459*(7250), 1150-1153.

Kim, N., & Jinks-Robertson, S. (2010). Abasic sites in the transcribed strand of yeast DNA are removed by transcription-coupled nucleotide excision repair. *Mol Cell Biol, 30*(13), 3206-3215.

Kim, N., & Jinks-Robertson, S. (2011). Guanine repeat-containing sequences confer transcription-dependent instability in an orientation-specific manner in yeast. *DNA Repair (Amst)*.

Kim, N., & Jinks-Robertson, S. (2012). Transcription as a source of genome instability. *Nature reviews. Genetics, 13*(3), 204-214.

Kim, N., Mudrak, S. V., & Jinks-Robertson, S. (2011). The dCMP transferase activity of yeast Rev1 is biologically relevant during the bypass of endogenously generated AP sites. *DNA repair, 10*(12), 1262-1271.

Klapacz, J., & Bhagwat, A. S. (2002). Transcription-dependent increase in multiple classes of base substitution mutations in Escherichia coli. *J Bacteriol, 184*(24), 6866-6872.

Klapacz, J., & Bhagwat, A. S. (2005). Transcription promotes guanine to thymine mutations in the non-transcribed strand of an Escherichia coli gene. *DNA Repair (Amst), 4*(7), 806-813.

Klemm, L., Duy, C., Iacobucci, I., Kuchen, S., von Levetzow, G., Feldhahn, N., Henke, N., Li, Z., Hoffmann, T. K., Kim, Y. M., Hofmann, W. K., Jumaa, H., Groffen, J., Heisterkamp, N., Martinelli, G., Lieber, M. R., Casellas, R., & Muschen, M. (2009). The B cell mutator AID promotes B lymphoid blast crisis and drug resistance in chronic myeloid leukemia. *Cancer Cell, 16*(3), 232-245.

Koehler, S. E., & Ladner, R. D. (2004). Small interfering RNA-mediated suppression of dUTPase sensitizes cancer cell lines to thymidylate synthase inhibition. *Mol Pharmacol, 66*(3), 620-626.

Kohalmi, S. E., Glattke, M., McIntosh, E. M., & Kunz, B. A. (1991). Mutational specificity of DNA precursor pool imbalances in yeast arising from deoxycytidylate deaminase deficiency or treatment with thymidylate. *J Mol Biol, 220*(4), 933-946.

Kow, Y. W., Bao, G., Minesinger, B., Jinks-Robertson, S., Siede, W., Jiang, Y. L., & Greenberg, M. M. (2005). Mutagenic effects of abasic and oxidized abasic lesions in Saccharomyces cerevisiae. *Nucleic Acids Res, 33*(19), 6196-6202.

Kramara, J., Osia, B., & Malkova, A. (2018). Break-Induced Replication: The Where, The Why, and The How. *Trends Genet*.

Krejci, L., Van Komen, S., Li, Y., Villemain, J., Reddy, M. S., Klein, H., Ellenberger, T., & Sung, P. (2003). DNA helicase Srs2 disrupts the Rad51 presynaptic filament. *Nature, 423*(6937), 305-309.

Krokan, H. E., & Bjoras, M. (2013). Base excision repair. *Cold Spring Harb Perspect Biol, 5*(4), a012583.

Kumura, K., & Sekiguchi, M. (1984). Identification of the uvrD gene product of Escherichia coli as DNA helicase II and its induction by DNA-damaging agents. *J Biol Chem, 259*(3), 1560-1565.

Kuraoka, I., Suzuki, K., Ito, S., Hayashida, M., Kwei, J. S., Ikegami, T., Handa, H., Nakabeppu, Y., & Tanaka, K. (2007). RNA polymerase II bypasses 8-oxoguanine in the presence of transcription elongation factor TFIIS. *DNA Repair (Amst), 6*(6), 841-851.

Ladner, R. D., & Caradonna, S. J. (1997). The human dUTPase gene encodes both nuclear and mitochondrial isoforms. Differential expression of the isoforms and

characterization of a cDNA encoding the mitochondrial species. *J Biol Chem, 272*(30), 19072-19080.

Lari, S. U., Chen, C. Y., Vertessy, B. G., Morre, J., & Bennett, S. E. (2006). Quantitative determination of uracil residues in Escherichia coli DNA: Contribution of ung, dug, and dut genes to uracil avoidance. *DNA Repair (Amst), 5*(12), 1407-1420.

Lasken, R. S., Schuster, D. M., & Rashtchian, A. (1996). Archaebacterial DNA polymerases tightly bind uracil-containing DNA. *J Biol Chem, 271*(30), 17692-17696.

Lee, K. B., Wang, D., Lippard, S. J., & Sharp, P. A. (2002). Transcription-coupled and DNA damage-dependent ubiquitination of RNA polymerase II in vitro. *Proc Natl Acad Sci U S A, 99*(7), 4239-4244.

Li, S., & Smerdon, M. J. (2002). Rpb4 and Rpb9 mediate subpathways of transcription-coupled DNA repair in Saccharomyces cerevisiae. *EMBO J, 21*(21), 5921-5929.

Li, Z., Woo, C. J., Iglesias-Ussel, M. D., Ronai, D., & Scharff, M. D. (2004). The generation of antibody diversity through somatic hypermutation and class switch recombination. *Genes Dev, 18*(1), 1-11.

Lindahl, T. (1979). DNA glycosylases, endonucleases for apurinic/apyrimidinic sites, and base excision-repair. *Prog Nucleic Acid Res Mol Biol, 22*, 135-192.

Lindahl, T. (1993). Instability and decay of the primary structure of DNA. *Nature, 362*(6422), 709-715.

Lindahl, T., & Barnes, D. E. (2000). Repair of endogenous DNA damage. *Cold Spring Harb Symp Quant Biol, 65*, 127-133.

Lindahl, T., Modrich, P., & Sancar, A. (2016). The 2015 Nobel Prize in Chemistry The Discovery of Essential Mechanisms that Repair DNA Damage. *J Assoc Genet Technol, 42*(1), 37-41.

Lindahl, T., & Nyberg, B. (1972). Rate of depurination of native deoxyribonucleic acid. *Biochemistry, 11*(19), 3610-3618.

Lindahl, T., & Nyberg, B. (1974). Heat-induced deamination of cytosine residues in deoxyribonucleic acid. *Biochemistry, 13*(16), 3405-3410.

Liu, B., & Alberts, B. M. (1995). Head-on collision between a DNA replication apparatus and RNA polymerase transcription complex. *Science, 267*(5201), 1131-1137.

Longerich, S., Basu, U., Alt, F., & Storb, U. (2006). AID in somatic hypermutation and class switch recombination. *Curr Opin Immunol, 18*(2), 164-174.

Longley, D., Harkin, D., & Johnston, P. (2003). 5-fluorouracil: mechanisms of action and clinical strategies. *Nat Rev Cancer, 3*(7f2646d2-fa60-c923-c514-51b88a113482), 330-338.

Luhnsdorf, B., Epe, B., & Khobta, A. (2014). Excision of uracil from transcribed DNA negatively affects gene expression. *J Biol Chem, 289*(32), 22008-22018.

Maizels, N. (2005). Immunoglobulin gene diversification. *Annu Rev Genet, 39*, 23-46.

Marini, V., & Krejci, L. (2010). Srs2: the "Odd-Job Man" in DNA repair. *DNA Repair (Amst), 9*(3), 268-275.

Maul, R. W., & Gearhart, P. J. (2010). Controlling somatic hypermutation in immunoglobulin variable and switch regions. *Immunol Res, 47*(1-3), 113-122.

Maul, R. W., Saribasak, H., Martomo, S. A., McClure, R. L., Yang, W., Vaisman, A., Gramlich, H. S., Schatz, D. G., Woodgate, R., Wilson, D. M., 3rd, & Gearhart, P. J. (2011). Uracil residues dependent on the deaminase AID in immunoglobulin gene variable and switch regions. *Nature immunology, 12*(1), 70-76.

McIntosh, E. M., Gadsden, M. H., & Haynes, R. H. (1986). Transcription of genes encoding enzymes involved in DNA synthesis during the cell cycle of Saccharomyces cerevisiae. *Mol Gen Genet, 204*(3), 363-366.

McIntosh, E. M., & Haynes, R. H. (1984). Isolation of a Saccharomyces cerevisiae mutant strain deficient in deoxycytidylate deaminase activity and partial characterization of the enzyme. *J Bacteriol, 158*(2), 644-649.

Meadows, K. L., Song, B., & Doetsch, P. W. (2003). Characterization of AP lyase activities of Saccharomyces cerevisiae Ntg1p and Ntg2p: implications for biological function. *Nucleic Acids Res, 31*(19), 5560-5567.

Mellon, I., Bohr, V. A., Smith, C. A., & Hanawalt, P. C. (1986). Preferential DNA repair of an active gene in human cells. *Proc Natl Acad Sci U S A, 83*(23), 8878-8882.

Minesinger, B. K., Abdulovic, A. L., Ou, T. M., & Jinks-Robertson, S. (2006). The effect of oxidative metabolism on spontaneous Pol zeta-dependent translesion synthesis in Saccharomyces cerevisiae. *DNA Repair (Amst), 5*(2), 226-234.

Miyahara, S., Miyakoshi, H., Yokogawa, T., Chong, K. T., Taguchi, J., Muto, T., Endoh, K., Yano, W., Wakasa, T., Ueno, H., Takao, Y., Fujioka, A., Hashimoto, A., Itou, K., Yamamura, K., Nomura, M., Nagasawa, H., Shuto, S., & Fukuoka, M. (2012). Discovery of highly potent human deoxyuridine triphosphatase inhibitors based on the conformation restriction strategy. *J Med Chem, 55*(11), 5483-5496.

Morey, N. J., Greene, C. N., & Jinks-Robertson, S. (2000). Genetic analysis of transcription-associated mutation in Saccharomyces cerevisiae. *Genetics, 154*(1), 109-120.

Muha, V., Horvath, A., Bekesi, A., Pukancsik, M., Hodoscsek, B., Merenyi, G., Rona, G., Batki, J., Kiss, I., Jankovics, F., Vilmos, P., Erdelyi, M., & Vertessy, B. G. (2012). Uracil-containing DNA in Drosophila: stability, stage-specific accumulation, and developmental involvement. *PLoS Genet, 8*(6), e1002738.

Muramatsu, M., Sankaranand, V. S., Anant, S., Sugai, M., Kinoshita, K., Davidson, N. O., & Honjo, T. (1999). Specific expression of activation-induced cytidine deaminase (AID), a novel member of the RNA-editing deaminase family in germinal center B cells. *J Biol Chem, 274*(26), 18470-18476.

Nelson, J. R., Lawrence, C. W., & Hinkle, D. C. (1996). Deoxycytidyl transferase activity of yeast REV1 protein. *Nature, 382*(6593), 729-731.

Newcombe, R. G. (1998). Two-sided confidence intervals for the single proportion: comparison of seven methods. *Stat Med, 17*(8), 857-872.

Nick McElhinny, S. A., Kumar, D., Clark, A. B., Watt, D. L., Watts, B. E., Lundstrom, E. B., Johansson, E., Chabes, A., & Kunkel, T. A. (2010). Genome instability due to ribonucleotide incorporation into DNA. *Nat Chem Biol, 6*(10), 774-781.

Otsuka, C., Loakes, D., & Negishi, K. (2002). The role of deoxycytidyl transferase activity of yeast Rev1 protein in the bypass of abasic sites. *Nucleic Acids Res Suppl*(2), 87-88.

Owiti, N., Lopez, C., Singh, S., Stephenson, A., & Kim, N. (2017). Def1 and Dst1 play distinct roles in repair of AP lesions in highly transcribed genomic regions. *DNA Repair (Amst), 55*, 31-39.

Ozlu, T., & Bulbul, Y. (2005). Smoking and lung cancer. *Tuberk Toraks, 53*(2), 200-209.

Pallan, P. S., & Egli, M. (2008). Insights into RNA/DNA hybrid recognition and processing by RNase H from the crystal structure of a non-specific enzyme-dsDNA complex. *Cell Cycle, 7*(16), 2562-2569.

Pan, Y. (2011). Mitochondria, reactive oxygen species, and chronological aging: a message from yeast. *Exp Gerontol, 46*(11), 847-852.

Pardo, E. G., & Gutierrez, C. (1990). Cell cycle- and differentiation stage-dependent variation of dUTPase activity in higher plant cells. *Exp Cell Res, 186*(1), 90-98.

Percival, K. J., Klein, M. B., & Burgers, P. M. (1989). Molecular cloning and primary structure of the uracil-DNA-glycosylase gene from Saccharomyces cerevisiae. *J Biol Chem, 264*(5), 2593-2598.

Petersen-Mahrt, S. K., Harris, R. S., & Neuberger, M. S. (2002). AID mutates E. coli suggesting a DNA deamination mechanism for antibody diversification. *Nature, 418*(6893), 99-103.

Pettersen, H. S., Galashevskaya, A., Doseth, B., Sousa, M. M., Sarno, A., Visnes, T., Aas, P. A., Liabakk, N. B., Slupphaug, G., Saetrom, P., Kavli, B., & Krokan, H. E. (2015). AID expression in B-cell lymphomas causes accumulation of genomic uracil and a distinct AID mutational signature. *DNA Repair (Amst), 25*, 60-71.

Pfeifer, G. P. (1997). Formation and processing of UV photoproducts: effects of DNA sequence and chromatin environment. *Photochem Photobiol, 65*(2), 270-283.

Pohl, T. J., & Nickoloff, J. A. (2008). Rad51-independent interchromosomal double-strand break repair by gene conversion requires Rad52 but not Rad55, Rad57, or Dmc1. *Mol Cell Biol, 28*(3), 897-906.

Popoff, S. C., Spira, A. I., Johnson, A. W., & Demple, B. (1990). Yeast structural gene (APN1) for the major apurinic endonuclease: homology to Escherichia coli endonuclease IV. *Proc Natl Acad Sci U S A, 87*(11), 4193-4197.

Postow, L., Crisona, N. J., Peter, B. J., Hardy, C. D., & Cozzarelli, N. R. (2001). Topological challenges to DNA replication: conformations at the fork. *Proc Natl Acad Sci U S A, 98*(15), 8219-8226.

Postow, L., Ullsperger, C., Keller, R. W., Bustamante, C., Vologodskii, A. V., & Cozzarelli, N. R. (2001). Positive torsional strain causes the formation of a four-way junction at replication forks. *J Biol Chem, 276*(4), 2790-2796.

Prado, F., & Aguilera, A. (2005). Impairment of replication fork progression mediates RNA polII transcription-associated recombination. *EMBO J, 24*(6), 1267-1276.

Prado, F., Piruat, J. I., & Aguilera, A. (1997). Recombination between DNA repeats in yeast hpr1delta cells is linked to transcription elongation. *EMBO J, 16*(10), 2826-2835.

Pryor, J. M., & Washington, M. T. (2011). Pre-steady state kinetic studies show that an abasic site is a cognate lesion for the yeast Rev1 protein. *DNA Repair (Amst), 10*(11), 1138-1144.

Reines, D., & Mote, J., Jr. (1993). Elongation factor SII-dependent transcription by RNA polymerase II through a sequence-specific DNA-binding protein. *Proc Natl Acad Sci U S A, 90*(5), 1917-1921.

Rosebrock, A. P. (2017). Synchronization and Arrest of the Budding Yeast Cell Cycle Using Chemical and Genetic Methods. *Cold Spring Harb Protoc, 2017*(1), pdb prot088724.

Rudolph, C. J., Dhillon, P., Moore, T., & Lloyd, R. G. (2007). Avoiding and resolving conflicts between DNA replication and transcription. *DNA Repair (Amst), 6*(7), 981-993.

Ryu, J., Youn, J., Kim, Y., Kwon, O., Song, Y., Kim, H., Cho, K., & Chang, I. (1999). Mutation spectrum of 4-nitroquinoline N-oxide in the lacI transgenic Big Blue Rat2 cell line. *Mutat Res, 445*(1), 127-135.

Saito, K., Nagashima, H., Noguchi, K., Yoshisue, K., Yokogawa, T., Matsushima, E., Tahara, T., & Takagi, S. (2014). First-in-human, phase I dose-escalation study of single and multiple doses of a first-in-class enhancer of fluoropyrimidines, a dUTPase inhibitor (TAS-114) in healthy male volunteers. *Cancer Chemother Pharmacol, 73*(3), 577-583.

Sallmyr, A., Fan, J., & Rassool, F. V. (2008). Genomic instability in myeloid malignancies: increased reactive oxygen species (ROS), DNA double strand breaks (DSBs) and error-prone repair. *Cancer Lett, 270*(1), 1-9.

Sanchez, A., Sharma, S., Rozenzhak, S., Roguev, A., Krogan, N. J., Chabes, A., & Russell, P. (2012). Replication fork collapse and genome instability in a deoxycytidylate deaminase mutant. *Mol Cell Biol, 32*(21), 4445-4454.

Santos, J. H., Meyer, J. N., Mandavilli, B. S., & Van Houten, B. (2006). Quantitative PCR-based measurement of nuclear and mitochondrial DNA damage and repair in mammalian cells. *Methods Mol Biol, 314*, 183-199.

Savic, D. J., & Kanazir, D. T. (1972). The effect of a histidine operator-constitutive mutation on UV-induced mutability within the histidine operon of Salmonella typhimurium. *Mol Gen Genet, 118*(1), 45-50.

Scott, A. D., Neishabury, M., Jones, D. H., Reed, S. H., Boiteux, S., & Waters, R. (1999). Spontaneous mutation, oxidative DNA damage, and the roles of base and nucleotide excision repair in the yeast Saccharomyces cerevisiae. *Yeast, 15*(3), 205-218.

Seiple, L., Jaruga, P., Dizdaroglu, M., & Stivers, J. T. (2006). Linking uracil base excision repair and 5-fluorouracil toxicity in yeast. *Nucleic Acids Res, 34*(1), 140-151.

Senturker, S., Auffret van der Kemp, P., You, H. J., Doetsch, P. W., Dizdaroglu, M., & Boiteux, S. (1998). Substrate specificities of the ntg1 and ntg2 proteins of Saccharomyces cerevisiae for oxidized DNA bases are not identical. *Nucleic Acids Res, 26*(23), 5270-5276.

Shlomai, J., & Kornberg, A. (1978). Deoxyuridine triphosphatase of Escherichia coli. Purification, properties, and use as a reagent to reduce uracil incorporation into DNA. *J Biol Chem, 253*(9), 3305-3312.

Sivaramakrishnan, P., Sepulveda, L. A., Halliday, J. A., Liu, J., Nunez, M. A. B., Golding, I., Rosenberg, S. M., & Herman, C. (2017). The transcription fidelity factor GreA impedes DNA break repair. *Nature, 550*(7675), 214-218.

Skandalis, A., Ford, B. N., & Glickman, B. W. (1994). Strand bias in mutation involving 5-methylcytosine deamination in the human hprt gene. *Mutat Res, 314*(1), 21-26.

Skurnik, M., Hyytiainen, H. J., Happonen, L. J., Kiljunen, S., Datta, N., Mattinen, L., Williamson, K., Kristo, P., Szeliga, M., Kalin-Manttari, L., Ahola-Iivarinen, E.,

Kalkkinen, N., & Butcher, S. J. (2012). Characterization of the genome, proteome, and structure of yersiniophage varphiR1-37. *J Virol, 86*(23), 12625-12642.

Spell, R. M., & Jinks-Robertson, S. (2004). Determination of mitotic recombination rates by fluctuation analysis in Saccharomyces cerevisiae. *Methods Mol Biol, 262*, 3-12.

Storms, R. K., Ord, R. W., Greenwood, M. T., Mirdamadi, B., Chu, F. K., & Belfort, M. (1984). Cell cycle-dependent expression of thymidylate synthase in Saccharomyces cerevisiae. *Mol Cell Biol, 4*(12), 2858-2864.

Sugasawa, K., Okamoto, T., Shimizu, Y., Masutani, C., Iwai, S., & Hanaoka, F. (2001). A multistep damage recognition mechanism for global genomic nucleotide excision repair. *Genes Dev, 15*(5), 507-521.

Swanson, R. L., Morey, N. J., Doetsch, P. W., & Jinks-Robertson, S. (1999). Overlapping specificities of base excision repair, nucleotide excision repair, recombination, and translesion synthesis pathways for DNA base damage in Saccharomyces cerevisiae. *Mol Cell Biol, 19*(4), 2929-2935.

Sweder, K. S., & Hanawalt, P. C. (1992). Preferential repair of cyclobutane pyrimidine dimers in the transcribed strand of a gene in yeast chromosomes and plasmids is dependent on transcription. *Proc Natl Acad Sci U S A, 89*(22), 10696-10700.

Swenberg, J. A., Lu, K., Moeller, B. C., Gao, L., Upton, P. B., Nakamura, J., & Starr, T. B. (2011). Endogenous versus exogenous DNA adducts: their role in carcinogenesis, epidemiology, and risk assessment. *Toxicol Sci, 120 Suppl 1*, S130-145.

Takeuchi, Y., Horiuchi, T., & Kobayashi, T. (2003). Transcription-dependent recombination and the role of fork collision in yeast rDNA. *Genes Dev, 17*(12), 1497-1506.

Terleth, C., Waters, R., Brouwer, J., & van de Putte, P. (1990). Differential repair of UV damage in Saccharomyces cerevisiae is cell cycle dependent. *EMBO J, 9*(9), 2899-2904.

Tinkelenberg, B. A., Hansbury, M. J., & Ladner, R. D. (2002). dUTPase and uracil-DNA glycosylase are central modulators of antifolate toxicity in Saccharomyces cerevisiae. *Cancer Res, 62*(17), 4909-4915.

Tornaletti, S., Maeda, L. S., & Hanawalt, P. C. (2006). Transcription arrest at an abasic site in the transcribed strand of template DNA. *Chem Res Toxicol, 19*(9), 1215-1220.

Torres-Ramos, C. A., Johnson, R. E., Prakash, L., & Prakash, S. (2000). Evidence for the involvement of nucleotide excision repair in the removal of abasic sites in yeast. *Mol Cell Biol, 20*(10), 3522-3528.

Traut, T. W. (1994). Physiological concentrations of purines and pyrimidines. *Mol Cell Biochem, 140*(1), 1-22.

Tsai, A. G., Lu, H., Raghavan, S. C., Muschen, M., Hsieh, C. L., & Lieber, M. R. (2008). Human chromosomal translocations at CpG sites and a theoretical basis for their lineage and stage specificity. *Cell, 135*(6), 1130-1142.

Tye, B. K., Nyman, P. O., Lehman, I. R., Hochhauser, S., & Weiss, B. (1977). Transient accumulation of Okazaki fragments as a result of uracil incorporation into nascent DNA. *Proc Natl Acad Sci U S A, 74*(1), 154-157.

Van Houten, B., & McCullough, A. (1994). Nucleotide excision repair in E. coli. *Ann N Y Acad Sci, 726*, 236-251.

Van Komen, S., Reddy, M. S., Krejci, L., Klein, H., & Sung, P. (2003). ATPase and DNA helicase activities of the Saccharomyces cerevisiae anti-recombinase Srs2. *J Biol Chem, 278*(45), 44331-44337.

Verhage, R. A., van Gool, A. J., de Groot, N., Hoeijmakers, J. H., van de Putte, P., & Brouwer, J. (1996). Double mutants of Saccharomyces cerevisiae with alterations in global genome and transcription-coupled repair. *Mol Cell Biol, 16*(2), 496-502.

Vertessy, B. G., & Toth, J. (2009). Keeping uracil out of DNA: physiological role, structure and catalytic mechanism of dUTPases. *Acc Chem Res, 42*(1), 97-106.

Voelkel-Meiman, K., Keil, R. L., & Roeder, G. S. (1987). Recombination-stimulating sequences in yeast ribosomal DNA correspond to sequences regulating transcription by RNA polymerase I. *Cell, 48*(6), 1071-1079.

Wahba, L., Costantino, L., Tan, F. J., Zimmer, A., & Koshland, D. (2016). S1-DRIP-seq identifies high expression and polyA tracts as major contributors to R-loop formation. *Genes Dev, 30*(11), 1327-1338.

Wang, J. C. (2002). Cellular roles of DNA topoisomerases: a molecular perspective. *Nat Rev Mol Cell Biol, 3*(6), 430-440.

Wang, J. Y., Sarker, A. H., Cooper, P. K., & Volkert, M. R. (2004). The single-strand DNA binding activity of human PC4 prevents mutagenesis and killing by oxidative DNA damage. *Mol Cell Biol, 24*(13), 6084-6093.

Wang, L., & Weiss, B. (1992). dcd (dCTP deaminase) gene of Escherichia coli: mapping, cloning, sequencing, and identification as a locus of suppressors of lethal dut (dUTPase) mutations. *J Bacteriol, 174*(17), 5647-5653.

Wardle, J., Burgers, P. M., Cann, I. K., Darley, K., Heslop, P., Johansson, E., Lin, L. J., McGlynn, P., Sanvoisin, J., Stith, C. M., & Connolly, B. A. (2008). Uracil recognition by replicative DNA polymerases is limited to the archaea, not occurring with bacteria and eukarya. *Nucleic Acids Res, 36*(3), 705-711.

Warner, H. R., & Duncan, B. K. (1978). In vivo synthesis and properties of uracil-containing DNA. *Nature, 272*(5648), 32-34.

Warner, H. R., Duncan, B. K., Garrett, C., & Neuhard, J. (1981). Synthesis and metabolism of uracil-containing deoxyribonucleic acid in Escherichia coli. *J Bacteriol, 145*(2), 687-695.

Waters, L. S., Minesinger, B. K., Wiltrout, M. E., D'Souza, S., Woodruff, R. V., & Walker, G. C. (2009). Eukaryotic translesion polymerases and their roles and regulation in DNA damage tolerance. *Microbiol Mol Biol Rev, 73*(1), 134-154.

Wei, S., Shalhout, S., Ahn, Y. H., & Bhagwat, A. S. (2015). A versatile new tool to quantify abasic sites in DNA and inhibit base excision repair. *DNA Repair (Amst), 27*, 9-18.

Wellinger, R. E., Prado, F., & Aguilera, A. (2006). Replication fork progression is impaired by transcription in hyperrecombinant yeast cells lacking a functional THO complex. *Mol Cell Biol, 26*(8), 3327-3334.

Wilson, D. M., 3rd, & Barsky, D. (2001). The major human abasic endonuclease: formation, consequences and repair of abasic lesions in DNA. *Mutat Res, 485*(4), 283-307.

Wilson, M. D., Harreman, M., Taschner, M., Reid, J., Walker, J., Erdjument-Bromage, H., Tempst, P., & Svejstrup, J. Q. (2013). Proteasome-mediated processing of Def1, a critical step in the cellular response to transcription stress. *Cell, 154*(5), 983-995.

Wilson, P. M., Danenberg, P. V., Johnston, P. G., Lenz, H. J., & Ladner, R. D. (2014). Standing the test of time: targeting thymidylate biosynthesis in cancer therapy. *Nat Rev Clin Oncol, 11*(5), 282-298.

Wilson, P. M., LaBonte, M. J., Lenz, H. J., Mack, P. C., & Ladner, R. D. (2012). Inhibition of dUTPase induces synthetic lethality with thymidylate synthase-targeted therapies in non-small cell lung cancer. *Mol Cancer Ther, 11*(3), 616-628.

Woudstra, E. C., Gilbert, C., Fellows, J., Jansen, L., Brouwer, J., Erdjument-Bromage, H., Tempst, P., & Svejstrup, J. Q. (2002). A Rad26-Def1 complex coordinates repair and RNA pol II proteolysis in response to DNA damage. *Nature, 415*(6874), 929-933.

Wu, H. Y., Shyy, S. H., Wang, J. C., & Liu, L. F. (1988). Transcription generates positively and negatively supercoiled domains in the template. *Cell, 53*(3), 433-440.

Yadav, P., Harcy, V., Argueso, J. L., Dominska, M., Jinks-Robertson, S., & Kim, N. (2014). Topoisomerase I plays a critical role in suppressing genome instability at a highly transcribed G-quadruplex-forming sequence. *PLoS Genet, 10*(12), e1004839.

Zhang, Y., Gostissa, M., Hildebrand, D. G., Becker, M. S., Boboila, C., Chiarle, R., Lewis, S., & Alt, F. W. (2010). The role of mechanistic factors in promoting chromosomal translocations found in lymphoid and other cancers. *Adv Immunol, 106*, 93-133.

Zhao, X. N., & Usdin, K. (2015). The Repeat Expansion Diseases: The dark side of DNA repair. *DNA Repair (Amst), 32*, 96-105.

VITA

Norah Auma Owiti was born in Homabay, Kenya to Lister Boaz Owiti and Perez Aoko Okatch. She, along with her siblings Edwin, June, Selma "Bibi" and Lydia "Lido" were raised in Kenya. After graduating from Lambwe Secondary School in November 2006, through the Zawadi Africa Scholarship Program, she applied and enrolled at Simpson College, in Indianola, Iowa in August 2008. Owiti received her Bachelor of Art degree in 2012 with a double major in Biochemistry and Chemistry (Cum Laude). In August of 2012, she joined the University of Texas Graduate School of Biomedical Sciences at Houston to pursue a doctorate degree.