Electronic Thesis and Dissertation Repository

4-17-2015 12:00 AM

Translesion synthesis and mutations: On the mutagenic properties of the two DNA lesions, 8-oxo-G and Pt-GG, and the functions of Y-family DNA polymerases and Rev3L on the bypass of each of the DNA lesions in mammalian cells

Lizhen Guo, *The University of Western Ontario*

Supervisor: Dr. Hong Ling, *The University of Western Ontario*
A thesis submitted in partial fulfillment of the requirements for the Master of Science degree in Biochemistry
© Lizhen Guo 2015

Recommended Citation

TRANSLESION SYNTHESIS AND MUTATIONS: ON THE MUTAGENIC PROPERTIES OF THE TWO DNA LESIONS 8-OXO-G AND PT-GG, AND THE FUNCTIONS OF Y-FAMILY DNA POLYMERASES AND REV3L ON BYPASSING EACH OF THE DNA LESIONS IN MAMMALIAN CELLS

(Thesis format: Monograph)

by

Lizhen Guo

Graduate Program in Biochemistry

A thesis submitted in partial fulfillment
of the requirements for the degree of
A Master of Science

The School of Graduate and Postdoctoral Studies
The University of Western Ontario
London, Ontario, Canada

Abstract

I studied the capabilities of the two DNA lesions 8-oxo-guanine and cisplatin intrastrand crosslinked 1,2-d(GpG) or Pt-GG to cause mutations in mammalian cells. Using isogenic cell lines generated from mice with selective gene knockouts of distinct DNA polymerases as models, I deduced the biological functions of the translesion DNA polymerases Pol eta, Pol kappa, Pol iota, Rev1 and Rev3L on bypassing each of the lesions 8-oxo-G and Pt-GG. My study takes advantage of the Next Generation Sequencing (NGS) technology to determine mutagenic effects of the DNA lesions *in vivo* and effects of translesion DNA polymerases on bypassing the lesions. Through adapting a translesion synthesis (TLS) assay into a high-throughput TLS assay which uses NGS, I eliminated steps in the traditional TLS assay that involves *E. coli*. My study reveals that 8-oxo-guanine caused a high frequency of C->A substitutions, and that significantly, Pol eta correctly bypassed 8-oxo-G in cells. The high mutation rate of 8-oxo-G underlines the importance of DNA repair to remove this lesion before it is replicated. By contrast, Pt-GG caused very few mutations *in vivo*: the mutation rate of Pt-GG was only two times the background error rate in my experiment. The five translesion DNA polymerases studied bypassed Pt-GG in an error-prone manner; specifically, Rev3L and Rev1 DNA polymerases can each bypass the lesion and Pol iota might share the function of Pol kappa or enhance that of Pol eta to bypass the lesion. Since Pt-GG does not appear to block DNA replication more than 8-oxo-G in mammalian cells, the cytotoxicity of the drug cisplatin that produces the cisplatin-DNA adduct is likely attributed to other causes, such as inter-strand cisplatin-DNA adducts that are more potent for stalling DNA replication.

Keywords

Acknowledgments

Table of Contents

List of Tables

List of Figures

List of Appendices

List of Abbreviations

8-oxo-G	8-oxo-guanine
ATP	adenosine triphosphate
BER	base excision repair
BH adjusted	Benjamini Hochberg method of adjusting P-values
BL	Burkitt's lymphoma
BL2 cell line	Burkitt's lymphoma 2 cell line, B lymphocytes, tumorgenic
bp	base pair
BsaI	restriction endonuclease cloned from *Bacillus stearothermophilus* 6-55
BstXI	restriction endonuclease cloned from *stearothermophilus* X1
cisplatin	cis-diamine-dichloroplatinum (II)
clr	centered log ratio transformation
cm	chloramphenicol
CPD	cyclobutane pyrimidine dimer
Ct	cycle threshold number in Q-PCR reactions
dCMP	deoxycytidine monophosphate
dCTP	deoxycytidine triphosphate
d(GpG)	deoxyguanidyl-deoxyguanosine dinucleotides
DKO	double knock out
DNA	deoxyribonucleic acid
DpnI	restriction enzyme cloned from *Diplococus pneumonia* G41
E. coli	*Escherichia coli*
EDTA	ethylenediaminetetraacetic acid
GFP	green fluorescent protein
HeLa cell	a cell line derived from cervical cancer cells taken from the patient Henrietta Lacks (thus 'HeLa') in Feb 1951
HMG protein	High Mobility Group-family of protein
IU	international unit, a standardized measure of a drug's biological activity
kan	kanamycin
kb	kilobase

m	minutes(s)
MEF	mouse embryonic fibroblast
NER	nucleotide excision repair
NGS	Next Generation Sequencing
nt	nucleotide
OGG1	the gene encoding 8-oxo-guanine DNA glycosylase, also known as MUTM
PAGE	polyacrylamide gel electrophoresis
PCA	Principal Component Analysis
PCNA	Proliferating Cell Nuclear Antigen
PBS	phosphate buffer saline
Pol	DNA polymerase
POLH	the gene encoding DNA polymerase eta/ Pol eta/ Pol η
POLI	the gene encoding DNA polymerase iota/ Pol iota/ Pol ι
POLK	the gene encoding DNA polymerase kappa/ Pol kappa/ Pol κ
Pol ζ	DNA polymerase zeta, a B-family DNA polymerase
Pt-GG	cisplatin crosslinked intrastrand 1,2-d(GpG)
Q-PCR	quantitative PCR
REV	'Reversionless', describing the phenotype of yeast cells in response to UV light when a *REV* gene is knocked out
RNA	ribonucleic acid
ROS	reactive oxygen species
rpm	revolutions per minute
s	second(s)
SDS	sodium dodecyl sulphate
TAE	Tris-Acetate-EDTA buffer
TKO	triple knock out
T_m	melting temperature
TLS	translesion synthesis
UV light	ultraviolet light
vs.	versus

Preface

This dissertation is original, unpublished, independent work by the author, L. Guo, with funding and support provided by Dr. Ling. All the figures except Figures 2 and 3 are produced by the author.

Concept formation of this project was initiated by Dr. Ling. To the author's knowledge, the concept was, at various times, consulted with Dr. Edgell, Dr. Gloor, and Guangxing Xing. The author was involved in concept formation, planning and execution of the project.

This work relies on a translesion synthesis assay that requires construction of double stranded gap plasmids. Even though the method of constructing the plasmids has been reported, when performing actual experiments the author used her ingenuity to solve several major problems, including gel purification, electro-elution, desalting DNA samples and scaling down transfection experiments. Although the traditional TLS assay described herein is similar in principal to what has been published by others, it may not be apparent from reading the thesis that the actual work is different in many aspects from those experiments carried out by others. The author in her own ways produced pure and highly transfectable gap plasmids that is an absolute requirement for obtaining high quality data. This should be noted as one of the achievements of the thesis project.

The author is particularly interested in mutagenesis as a phenomenon of the living cell and as a random event that disturbs, makes disorder and creates variation in DNA sequence. Once a mutation has come to be through translesion synthesis, even if the initial effect is imperceptible, the mutation will not be erased but continue to develop and be influenced by the living condition of the cell and its environment. A mutation produced by DNA lesion is not merely a substitution of DNA bases, but is always a reflection of what is outside of the cell. From quantitative results presented in the thesis, the author became interested in thinking qualitatively about changes or other mutations brought by those produced from translesion synthesis in the cells. The author found Chaos theory an attractive model (but unfortunately, it was not defensible) to explain the nonlinear dynamics of mutations in relation to qualitative changes that occur in the cancer cell.

Chapter 1

1 Introduction

1.1 Y-family DNA polymerases are prone to making mutations

Information encoded by genomic DNA is not static. The biological content of the genome adapts to the environment and is shaped by outside forces over time. One way genomic DNA sequences are changed is from chemical modifications of nucleotides. Y-family DNA polymerases bypass modified nucleotides or DNA lesions in a process called translesion synthesis (TLS) (Yang and Woodgate, 2007; Sale et al., 2012). Compared to DNA polymerases that accurately copy the whole genome, Y-family polymerases are prone to making mutations (Ohmori et al., 2001).

Mutation is a natural phenomenon and a driving force of evolution. One way that mutations can arise is when DNA polymerases bypass lesions that block normal DNA replication. A stalled replication fork if not rescued by translesion DNA polymerases would fall apart—proteins of the replication machinery disassemble and a double stranded break forms. It is generally accepted that Y-family polymerases maintain genome stability by allowing DNA replication to continue despite DNA lesions that are ubiquitously present. This way, translesion synthesis ensures the cell's survival and tolerates DNA damage; on the other hand, it increases the likelihood that the base inserted opposite to the DNA lesion would be changed.

Translesion synthesis is one of the ways whereby DNA mutations occur. It's a major source of point mutation that involves change in the nucleotide opposite to the DNA lesion. Spontaneous mutations without facilitation of DNA lesions are rare but can arise in cells. At the gene level, these mutations are base substitution, insertion and deletion of one or several nucleotides given rise to by replication errors and spontaneous loss of DNA bases. There are also copy number DNA mutations such as expansion of trinucleotide DNA repeats exemplified in a genetic disease called fragile X syndrome,

whereby thousands of nucleotides were duplicated from CGG repeats. DNA mutations that affect production and folding of proteins are deleterious and could lead to increased susceptibility to illness and disease.
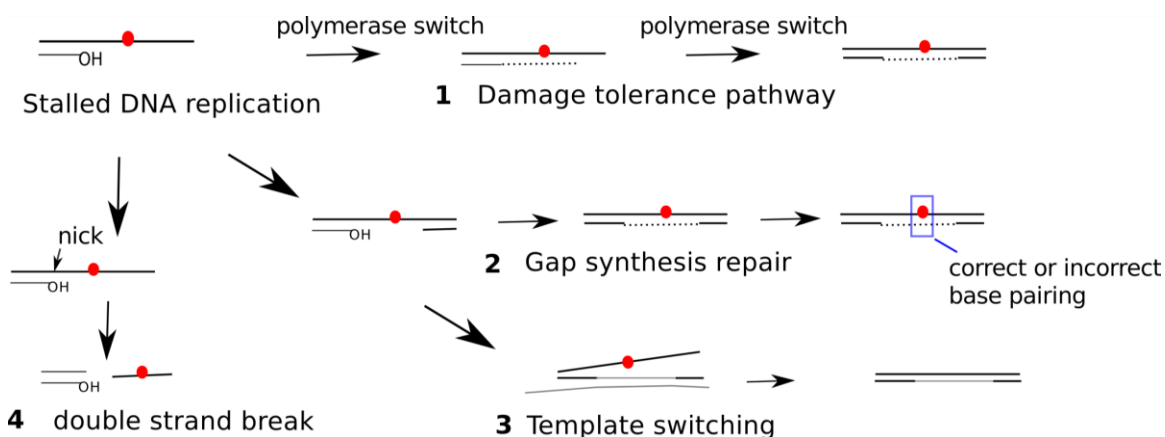
Figure 1. Possible outcomes for when DNA replication is stalled by a DNA lesion in the template

When DNA replication is stalled by a DNA lesion (indicated by the red circle), the cell can use the damage tolerance pathway to bypass the lesion (case 1). This involves two polymerase-switching events whereby in the first event, a replicative DNA polymerase is replaced by a translesion DNA polymerase which inserts a nucleotide across from the lesion, and in the second event, the replicative DNA polymerase is recruited back to continue normal DNA replication. Alternatively, the DNA replication machinery can skip the DNA lesion in what is called the damage avoidance pathway (cases 2 and 3) whereby the replicative DNA polymerase reinitiates DNA replication downstream from the lesion, leaving a small single stranded gap (Jansen, 2006). This gap is filled by translesion DNA polymerases later in the S or G2 phase of cell cycle (Daigaku et al., 2010; Karras and Jentsch, 2010; Elvers et al., 2011), and the nucleotide incorporated opposite to the lesion could either form the correct or mismatched base pair with the lesion (case 2). The single stranded gap can also be repaired by homologous recombination mediated by template switching (case 3). In this scenario, the DNA segment containing the lesion is not used as a template for DNA replication, thus the mutagenic effect of the DNA lesion is completely avoided. Failing the damage tolerance and the damage avoidance measures, a stalled DNA replication fork can be nicked by a single stranded endonuclease to form a double stranded break (case 4) whereupon DNA replication is terminated.

Translesion synthesis is a major source of point mutations or variations in the genome. These variations—if they are not strongly deleterious, are passed on for as long as a cell's lineage is kept alive, without causing harmful effects to daughter cells. The purpose of studying translesion synthesis is to look at the initial conditions that perturb the genome. Similar to the butterfly effect (Lorenz, 1963; Lorenz, E.N., 1972), small and random changes that are regarded as errors or noises in the genome can have a remarkable effect on determining the cell's fate. There is an element of chance and randomness in what shapes and what continues to shape the human genome. While the rate of mutation is tightly controlled in normal cells, a high rate of mutation is characteristic of tumor cells (Jackson and Loeb, 1998). Accumulation of point mutations increases the risk of developing cancer, especially if it is accompanied by defects in the cell's DNA repair and cell cycle control mechanisms (Hartwell and Kastan, 1994; Lodish et al., 2000). Understanding the regulation of non-stringent DNA polymerases, such as those of the Y-family, and how their regulations intercept DNA repair and cell cycle control pathways, are key to understanding the role of translesion synthesis in the formation of cancer (Lange et al., 2011; Makridakis and Reichardt, 2012).

While translesion synthesis works toward subtly changing the sequence of genomic DNA, not all mutations are detrimental and sometimes new functions are created, such as in generating antibody diversity. Somatic hypermutation in activated B cells is an active mutagenic process carried out by Y-family DNA polymerases on normal DNA templates (Faili et al., 2002; Delbos et al., 2005). The DNA polymerases produce mutations in the variable regions of immunoglobulin genes (V_H/V_L), thus giving rise to antibodies recognizing different antigens (McDonald et al., 2003; Delbos et al., 2005; Martomo et al., 2005; Faili et al., 2002).

The behaviors of Y-family DNA polymerases have been studied in the contexts of disease and of adaptive immunity. One way to elucidate the functions of Y-family DNA polymerases is to find out the DNA lesions they are adept at bypassing. A good example of a translesion DNA polymerase that can correctly bypass DNA lesions and thus contribute to reducing the incidence of a disease is the human Pol eta.

1.2   A brief history of the study of Y-family DNA polymerases

The human Pol eta gene (*POLH* or *XPV*) is defective in some patients with an inherited skin disorder called Xeroderma Pigmentosum-Variant (XP-V). Patients with this disease are predisposed to sunlight induced skin cancer and their skins retain high levels of damage due to DNA lesions, such as *cis-syn* thymine-thymine dimers arising from UV irradiation. Cells of these patients have normal nucleotide excision repair to remove these DNA lesions, but lack specialized DNA polymerases to bypass the lesions in DNA replication.

The XPV protein first isolated in HeLa cells is the human Pol eta (Masutani et al., 1999), its amino acid sequence shares 20% identity with the *Saccharomyces cerevisiae* RAD30 protein that encodes the yeast Pol eta (Johnson et al., 1999). Cell extracts of XP-V patients cannot bypass thymine dimers, however, when purified recombinant XPV proteins were added there was bypass activity. It was established from this experiment that the human Pol eta is able to bypass thymine-thymine dimers in cells and that a defect in the *POLH* gene causes skin disease.

According to their phylogenetic relationships to yeast *RAD30*, *POLH* along with three other human DNA polymerase genes (*POLK*, *POLI* and *REV1*) belong to the Y-family (Burgers et al., 2001; Ohmori et al., 2001). Over 200 orthologs of Y-family DNA polymerase genes have been identified in all three kingdoms of life (Yeiser et al., 2002). We would not have learned about those in higher eukaryotes without first identifying their homologous genes in yeast and in bacteria—for example, yeast *RAD30* was discovered by its homology to *DinB* and *UmuC* genes in *Escherichia coli* that encode DNA Polymerases IV and V, respectively (McDonald et al., 1997). The study of translesion synthesis began in the 1970s (Radman et al., 1974), and using *E. coli*, a two-polymerase model was developed to describe translesion synthesis as a process that occurs in two steps—insertion and extension, each effected by a different error-prone DNA polymerase (Radman et al., 1974). The two-polymerase model is also widely

adopted as the mechanism of translesion synthesis in higher eukaryotes (Friedberg et al., 2005; Shachar et al., 2009).

The two-polymerase model gained support partly because of the regulation of Y-family DNA polymerases by a ring-structured protein called Proliferating Cell Nuclear Antigen (PCNA) which functions as a sliding clamp at DNA replication fork (Tsurimoto, 1999; Maga and Hübscher, 2003; Naryzhny, 2008). Human PCNA is a homotrimer for which each of its subunits can bind to a Y-family DNA polymerase (Kannouche et al., 2004; Bienko et al., 2005; Haracska et al., 2005; Guo et al., 2006; Jansen et al., 2007). Recruitment of more than one Y-family DNA polymerases by PCNA favored the two-polymerase model.

The model also gained support with the evidence that Y-family DNA polymerases co-localize to replication foci in the nuclei when cells are induced by DNA damage. This suggests that the bypass of DNA lesions can be coordinated among the DNA polymerases (Kannouche et al., 2003). Furthermore, mutations are observed not only opposite to but also adjacent to DNA lesions in cellular studies done using a translesion synthesis assay (Avkin and Livneh, 2002; Shachar et al., 2009; Taggart et al., 2013) and these mutations are believed to be made by different error-prone DNA polymerases. In my work, I will call into question the two-polymerase model and argue that mutations adjacent to DNA lesions do not result immediately from translesion synthesis.

1.3   Common features of Y-family DNA polymerases

The overall structure of a Y-family DNA polymerase resembles a 'right-hand' with palm, finger and thumb domains that form the protein's active site (Friedberg et al., 2001). In addition, a Y-family DNA polymerase has a domain called the 'Little Finger' (Ling et al., 2001); its amino acid sequence is divergent among Y-family members (Boudsocq et al., 2004). Tethered to the thumb domain, the 'Little Finger' domain binds DNA. It was thought that movement of the 'Little Finger' domain modulates substrate specificity of Y-family DNA polymerases, so that each of them is unique in the type of DNA lesions it

can bypass (Boudsocq et al., 2004). However, my results show that there are functional overlaps among Y-family DNA polymerases in the cell to bypass DNA lesions.

Y-family DNA polymerases are tolerant of unusual local structures in DNA caused by a damaged nucleotide. A major reason is because their active sites have a larger space than replicative DNA polymerases, so they can accommodate a base pair mismatch, Hoogsteen base pairing or a base flip-out in DNA template caused by a DNA lesion (Friedberg et al., 2001). This is not to say that Y-family DNA polymerases do not select for incoming nucleotides, but rather, they tend to make flexible choices. Another reason that Y-family DNA polymerases can replicate through DNA lesions is that they lack a 3'-5' exonuclease or proofreading activity (Friedberg et al., 2001). While this ensures the complete bypass of a DNA lesion in that the enzyme does not track backwards to remove a nucleotide it has just inserted, the lack of proofreading activity makes the DNA polymerases error-prone. The error frequency of a Y-family DNA polymerase on normal DNA template is one mistake in every 10^2 to 10^4 nucleotides replicated (Ollivierre et al., 2011; Tissier et al., 2000) compared to one mistake in every 10^6 to 10^8 nucleotides synthesized by a replicative DNA polymerase that has intrinsic proofreading activity (Kunkel, 2004). Unlike replicative DNA polymerases that can incorporate thousands of nucleotides in a stretch, Y-family DNA polymerases disengage from the template after adding several nucleotides (Ohashi et al., 2000; Friedberg et al., 2001; Boudsocq et al., 2004). Y-family DNA polymerases are also slower than replicative DNA polymerases in catalyzing polymerization reactions (Beard et al., 2002).

1.4   Distinct functions of translesion DNA polymerases

Each Y-family DNA polymerase was thought to have unique structural properties that enable it to bypass different DNA lesions, but Rev1 is an exception because it was first discovered as a protein that has nucleotidyl transferase activity in *S. cerevisiae* and it preferentially inserts dCMP opposite abasic site (Nelson et al., 1996). There are at least two ways Rev1 can function in translesion synthesis depending on whether its catalytic domain is involved.

First, it was recently shown that Rev1 can bypass a DNA lesion directly, the Rev1 dCMP transferase domain is required to bypass 1,N6-ethenoadenine (Zhou et al., 2010). Second, Rev1 can mediate translesion synthesis indirectly by acting as a scaffold, which does not require the protein's catalytic domain to recruit other DNA polymerases, such as Pol kappa, Pol eta, Pol iota and Pol zeta (Guo et al., 2003; Ohashi et al., 2004; Tissier et al., 2004).

Pol zeta is a B-family DNA polymerase that plays an important role in error-prone translesion synthesis (Gan et al., 2008). The Rev3L protein and the Rev7 protein are the catalytic subunit and the accessory subunit of Pol zeta, respectively. The *REV3L* gene is essential for mouse embryonic development (Wittschieben et al., 2006). Cells with targeted disruption of *REV3L* have reduced numbers of point mutations and are very sensitive to UV light and some DNA damaging agents (Gan et al., 2008; Wittschieben et al., 2006). Therefore, Rev3L is essential for bypassing various types of DNA lesions and in error-prone translesion synthesis (Jansen et al., 2009). It has been reported that Rev3 adopts the function of an extender DNA polymerase, acting sequentially after an insertor DNA polymerase to extend from a lesion site (Acharya et al., 2006; Lee et al., 2014). Rev1 and Rev3L cooperate in translesion synthesis through an interaction mediated by Rev7 (Murakumo et al., 2001; Kikuchi et al., 2012). Rev1 and Rev3 also contribute to the process of homologous recombination to repair double strand breaks induced in cells by ionizing radiation (Sharma et al., 2011; Wittschieben et al., 2010) and post-replication gaps resulting from DNA lesions (Jansen et al., 2009). Regulations of Rev1 and Rev3 are governed by PCNA and the consequences of Rev1 and Rev3's interactions are dependent on the states of post-translational modifications of PCNA (Hoege et al., 2002; Stelter and Ulrich, 2003; Andersen et al., 2008). Mono-ubiquitinlation of PCNA controls the timing of translesion synthesis and regulates the access of the Rev3L protein and Y-family DNA polymerases to DNA templates (Vidal et al., 2004; Kannouche et al., 2004; Garg et al., 2005; Bi et al., 2006; Guo et al., 2006; Chun and Jin, 2010).

Pol iota is considered the most error prone among the four human Y-family DNA polymerases (Tissier et al., 2000; Zhang et al., 2000a; Makarova and Kulbachinskiy, 2012) because it has not yet been shown what kind of DNA lesions Pol iota can bypass

correctly. The pol iota gene, arising from an ancestral gene duplication of Pol eta, is not essential for survival. Mice with homozygous knockout of the Pol iota gene (*POLI*) have normal phenotypes (McDonald et al., 2003).

On the other hand, mice with a knockout of the Pol eta gene (*POLH*) are susceptible to developing skin tumors after UV irradiation (Lin et al., 2006). Thus, Pol eta protects mammalian cells from DNA damage caused by UV light.

Pol kappa can replicate past several bulky DNA adducts *in vitro*, including benzo(a)pyrene (Zhang et al., 2000b; Ogi et al., 2002; Ogi and Lehmann, 2006). Mouse and chicken cells with complete disruption of the *POLK* gene grow normally but show slightly increased UV sensitivity (Ogi et al., 2002; Okada et al., 2002). Pol kappa on its own cannot bypass major UV lesions such as *cis-syn* T-T dimers *in vitro*, but it is able to extend from a nucleotide inserted to the 3'T of the dimer by another DNA polymerase (Carpio et al., 2011; Zhang et al., 2000b). Pol kappa has been shown to act as an extender DNA polymerase in translesion synthesis as well as having a role in nucleotide excision repair (NER) (Ogi and Lehmann, 2006).

1.5   Project

The initial aim of my project is to investigate the cellular function of Pol iota on bypassing the DNA lesion 8-oxo-guanine. X-ray crystal structures of Pol iota solved by former members in the Ling Lab have shown that the DNA polymerase can bypass 8-oxo-G accurately *in vitro* (Kirouac and Ling, 2011a). Therefore, by extension, I hypothesized that Pol iota can accurately bypass 8-oxo-G *in vivo*. If this function of Pol iota could be confirmed in the cell and that no other Y-family DNA polymerases shared the same function, Pol iota would have the indispensable role of protecting cells from oxidative stress that produces 8-oxo-G.

My project expanded to include the study of another DNA lesion, the cisplatin instrastrand crosslinked 1,2-d(GpG) or Pt-GG for short. I obtained isogenic mouse embryonic fibroblast (MEF) cell lines that allowed me to study translesion synthesis of

each of the 8-oxo-G and the Pt-GG lesion by all four of the human Y-family DNA polymerases (Pol iota, Pol kappa, Pol eta, Rev1) and Rev3L. 8-oxo-G and Pt-GG would make for valuable comparisons because I assumed, as were implicated in literatures, that 8-oxo-G, a relatively small DNA lesion does not block DNA replication, and that Pt-GG, a purine dimer does. In this project, I found that Pt-GG and 8-oxo-G had similar blocking effects to DNA replication in mammalian cells.

From the perspective of generating mutations, interactions between DNA lesions and translesion DNA polymerases contribute to shaping the mutational landscape of the cell. My work has added knowledge to the mutagenicity of the two DNA lesions 8-oxo-G and Pt-GG and to the biological functions of translesion DNA polymerases on bypassing these two lesions.

1.6   8-oxo-guanine

8-oxo-guanine (8-oxo-G) is a very common DNA lesion formed by the chemical reaction between a guanine base and reactive oxygen species (Figure 2).
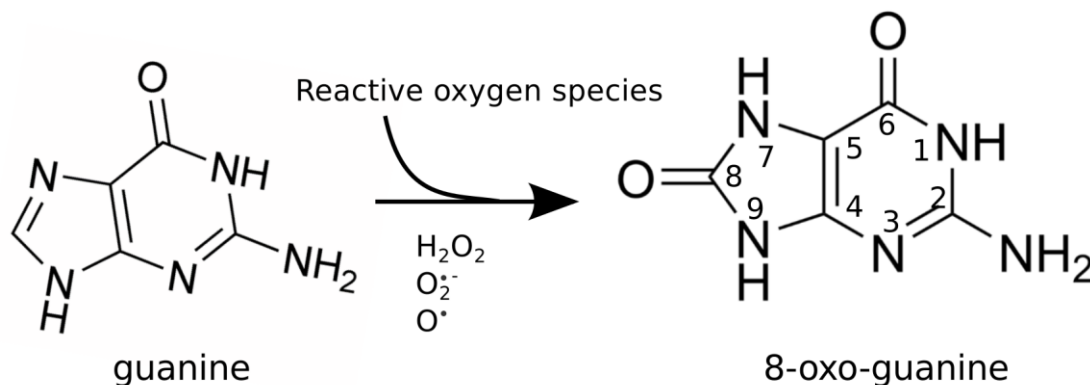
Figure 2 How 8-oxo-guanine is formed.

Reactive oxygen species such as superoxide radical ($O_2 \cdot^-$), hydrogen peroxide (H_2O_2) and hydroxyl radical ($\cdot OH$) are generated in the body as byproducts of oxygen metabolism and are found in the environment. Guanine is the base most prone to DNA oxidation (Weimann et al., 2002). The steady state concentration of 8-oxo-G in cells remains in

debate because there is no reliable method to accurately measure its concentration (Marnett, 2000). The level of 8-oxo-G in human tissues has been reported by several labs and the numbers differ widely, ranging from one adduct in 10^3 nucleotides to one adduct in 10^7 nucleotides (Cadet et al., 1997). While the abundance of 8-oxo-G in genomic DNA remains controversial (Marnett, 2000), it is generally considered pretty high (Bohr et al., 2002; van Loon et al., 2010). Free 8-oxo-G in the cellular nucleotide pool can be incorporated into DNA (Satou et al., 2009) and deoxyguanosine in DNA can be directly converted to 8-oxo-G. The DNA lesion is removed by DNA glycosylase OGG1 in base excision repair pathway (Boiteux and Radicella, 2000; Hazra et al., 2001), and that which is not removed causes a high frequency of C->A transversion mutation when replicated (Cheng et al., 1992). My study shows that the mutation frequency of 8-oxo-G in mammalian cells is very high, in fact higher than all previously reported, and the cell lines used should have normal *OGG1* genes. This finding underlines the importance of DNA repairs, perhaps in addition to the action of DNA glycosylase OGG1, to remove 8-oxo-G before its replication.

1.7 Repair of 8-oxo-G lesions by base excision repair

Base excision repair is the predominant pathway for removing 8-oxo-G and other DNA damages caused by free radicals and reactive species generated by metabolism (Clancy, 2008). In mammalian cells, 8-oxoguanine-DNA glycosylase (OGG1) and endonuclease III homolog 1 (NTH1) are two enzymes that recognize and cleave oxidized guanine bases from DNA (Hailer et al., 2005). First, DNA glycosylase releases a damaged base and thus creates an abasic site by cleaving a covalent bond called the N-glycosidic bond between the DNA base and ribose sugar. Then, apurinic/apyrimidinic (AP) endonucleases cleaves the phosphodiester bond in DNA at the abasic site and makes a single stranded break. The break is repaired by either DNA synthesis with strand displacement (long-patch pathway) or by another pathway (called short patch) that removes the 5' deoxyribose phosphate group at the abasic site by enzymes with 5' deoxyribose phosphatase lyase activity followed by DNA synthesis. The final step of both pathways is sealing of the single stranded gap by DNA ligase. *OGG1* knockout mice accumulate more 8-oxo-G

lesions in hepatocytes but not in splenocytes, spermatocytes and kidney cells (Minowa et al., 2000; Osterod et al., 2001). *OGG1*-deficient fibroblast culture does not accumulate more 8-oxo-G lesions than the control wild type fibroblast culture. *NTH1* knockout mice are healthy (Ocampo et al., 2002). Both *OGG1*-deficient mice and *NTH1*-deficient mice do not have association with increased incidence of cancer, suggesting that there are other base excision repair enzymes in mammalian cells that recognize and remove 8-oxo-G lesions (Dou et al., 2003; Hailer et al., 2005). MUYTH is a DNA glycosylase in human that excises an adenine base mispaired with 8-oxo-G. Mutations in the MUTYH gene are associated with heritable predisposition to colorectal cancer.

1.8   Cisplatin intrastrand crosslinked 1,2-d(GpG)

Cisplatin is an anti-tumor agent used for treating a variety of cancers. Pt-GG is a major species (65%) of the various types of interstrand and intrastrand cisplatin-DNA adducts formed in cancer patients (Fichtinger-Schepman et al., 1987; Jung and Lippard, 2007; Cepeda et al., 2007). It is believed that cisplatin kills rapidly dividing tumor cells partly because Pt-GG blocks DNA replication (Vaisman et al., 2000; Chijiwa et al., 2010).
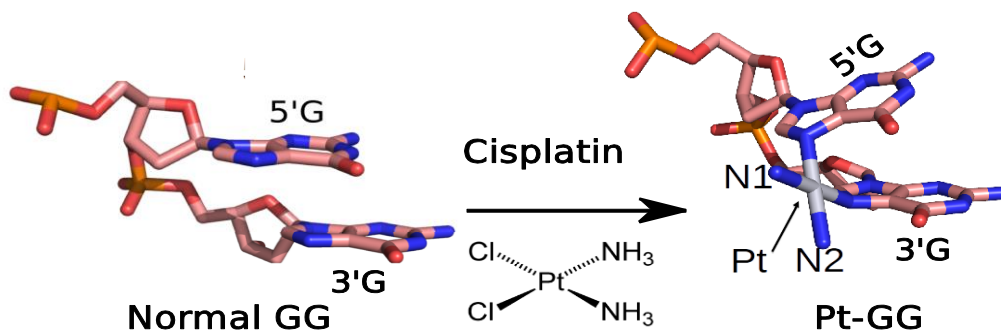
Figure 3 How Pt-GG is formed.

Cisplatin crosslinks the N7 atoms of adjacent guanosine nucleotides within the same DNA strand to form Pt-GG (image was provided by colleague Chuanbing Bian).

Following this rationale, the tumor cell develops resistance to cisplatin when it has acquired the ability to bypass cisplatin-DNA lesions and particularly, when the bypass of the lesion inserts the correct nucleotides. Researchers studying translesion synthesis of Pt-GG saw DNA polymerases with the ability to bypass the lesion as potential drug targets in chemotherapy. They postulated that drugs inactivating these DNA polymerases would reduce cisplatin resistance in cancer treatment (Zhao et al., 2012).

In my study, I found that Pt-GG was bypassed correctly 99% of the time in the mammalian cell due to DNA polymerase(s) other than those of the Y-family and other than Rev3L. The translesion DNA polymerases bypassed Pt-GG in an error-prone manner, and these enzymes were responsible for cisplatin resistance to the extent that they could bypass the DNA lesion. This result implies that the drug mechanism of cisplatin was likely not due to the direct effect of Pt-GG on blocking DNA replication, because the lesion was bypassed relatively efficiently and by several DNA polymerases in the cell. On the other hand, cisplatin also generates interstrand DNA adducts, and it is conceivable that these DNA lesions would be highly blocking to DNA replication and may therefore be responsible for the drug's cytotoxicity.

1.9   Repair of Pt-GG lesions by nucleotide excision repair

Nucleotide excision repair (NER) is the predominant pathway for removing bulky and dinucleotide DNA lesions, such as cyclobutane-pyrimidine dimers and 6-4 photoproducts induced by UV radiation. Nucleotide excision repair is also thought to be the major pathway for removing Pt-GG lesions (Nouspikel, 2009; Wang et al., 2011). In mammalian cells, there are at least 18 protein complexes involved in nucleotide excision repair (Clancy, 2008), and defects in some of the proteins are associated with the diseases Xeroderma Pigmentosum and Cockayne Syndrome (Nouspikel, 2009). In nucleotide excision repair, distortion in local DNA structure caused by Pt-GG is recognized by proteins XPA, RPA and the XPC-hHR23B complex, then the two strands of DNA around the lesion are pried apart by the DNA helicase TFIIH in an ATP-dependent manner. The strand containing the lesion is cleaved at the 3' end by XPG nuclease and at the 5' end by

the ERCC1-XPF nuclease, thereby releasing a short DNA fragment that is 24-32 nucleotides long (Wood, 1997). Synthesis across the short gap is carried out by the holoenzymes of DNA polymerase δ or ε, mediated by PCNA and finally, the single stranded gap is sealed by DNA ligase (Wood, 1997). It was also reported that mismatch repair could be involved in repairing Pt-GG lesions because the human mismatch repair protein hMSH2 binds to Pt-GG (Duckett et al., 1996; Mello et al., 1996). The mechanisms for removing Pt-GG lesions are different from that for removing 8-oxo-G lesions.

1.10   Translesion synthesis assays

To study mutations arising from DNA lesions in mammalian cells, I used a quantitative translesion synthesis (TLS) assay (Ziv et al., 2012). To deduce the functions of translesion DNA polymerases on bypassing the lesions *in vivo*, I obtained pairs of isogenic mouse embryonic fibroblast cell lines that were generated from knockout mice, where the mutant cell lines have specific knockouts of translesion DNA polymerases. I could infer the biological function of a DNA polymerase by comparing a mutant cell line to its isogenic wild type cell line. In the TLS assay, double stranded plasmids containing a small gap with and without a site-specific DNA lesion are co-transfected in equal amounts into mammalian cells, inside which the plasmids do not replicate but their gaps are filled. Then the plasmid DNAs are retrieved from the cell by the alkaline lysis method, which ensures that only the fully double stranded plasmids are retained and any unfilled or unrepaired gap plasmids are removed. Then I focused on two questions: First, what are the nucleotides that have been incorporated against the DNA lesions and second, what is the efficiency at which the lesion gap plasmid is repaired relative to the control gap plasmid, in other words, what is the efficiency at which the DNA lesion is bypassed relative to a regular nucleotide. I used two different approaches to address these questions: the traditional TLS assay and the high-throughput TLS assay. The traditional TLS assay uses *E. coli* to amplify the plasmids and Sanger Sequencing to determine the plasmid sequences (Ziv et al., 2012). The high-throughput TLS assay, which is one of the innovations in my project, uses PCR to amplify the gap regions of the plasmids and Next

Generation Sequencing to determine their sequences. The two approaches generated very different results that would make for an interesting comparison of the immediate effect versus the long term effect of a DNA lesion in the cell.

One of the advantages of the translesion synthesis assay is that the lesion gap plasmid is introduced into the cell along with the control gap plasmid that does not have a DNA lesion. Thus, the blocking effect of a DNA lesion can be measured by comparing the gap filling efficiency of the lesion plasmid relative to that of the control plasmid.

There are two major assumptions of the translesion synthesis assay. First, a DNA lesion in the gap plasmid is not removed in the cell prior to gap-filling. Second, a mismatched base pair resulting from translesion synthesis is preserved and unaltered when the plasmid is incubated in the mammalian cell. It has been shown that base excision repair is not active in the seven-hour period following transfection of the gap plasmids into the mammalian cell (Ziv et al., 2012).

An alternative substrate to use in the TLS assay is a single stranded gap plasmid that replicates in the mammalian cell (Karata et al., 2009; Pandya and Moriya, 1996; Watt et al., 2007). Some investigators have argued that this is a better model than the double stranded gap plasmid model to study translesion synthesis in DNA replication because the latter simulates gap-filling in DNA repair processes that occur in a different cell cycle stage. It has been shown, however, that translesion DNA polymerases are active in the late S and G2 phases of the cell cycle when the double stranded gap plasmid is being repaired (Daigaku et al., 2010; Karras and Jentsch, 2010; Diamant et al., 2011). Therefore, the double stranded gap plasmid is an appropriate substrate for studying mutations produced by translesion DNA polymerases.

The traditional TLS assay has several drawbacks. First, to study the spectrum of mutations caused by a DNA lesion, the traditional TLS assay uses *E. coli* to amplify and separate the control plasmids and lesion plasmids prior to Sanger sequencing. This makes the results questionable because the DNA sequences can undergo additional changes in bacteria as bacterial cells divide and multiply. Second, to study the bypass efficiency of a DNA lesion, the traditional TLS assay takes the ratio of the number of bacterial colonies

arising on a kanamycin plate, which selects for the lesion plasmids, to that arising on a chloramphenicol plate, which selects for the control plasmids. Since bacteria have different growth preferences in culturing media with different antibiotics, this method of measuring the translesion synthesis efficiency is flawed. Third, the traditional TLS assay is inefficient, and thus limits the amount of sequence information that can be obtained from each transfection experiment; any conclusions drawn from a small number of samples are misleading.

In adapting the traditional TLS assay into a high-throughput TLS assay, I replaced the steps in the traditional TLS assay that used *E. coli* with PCR, so instead of amplifying the whole plasmid in bacteria, a small region on the plasmid was amplified in a test tube. I designed a multiplexed study of the two DNA lesions 8-oxo-G and Pt-GG, which were acted upon by each of the five translesion DNA polymerases (Pol iota, Pol kappa, Pol eta, Rev1 and Rev3). I used barcoded primers in PCR reactions to keep track of DNA samples of different transfection experiments and pooled all the PCR products together to do Next Generation Sequencing (NGS). For positive control, I used fully double stranded plasmids with known sequence in PCR reactions to monitor experimental errors. For negative control, I incubated gap plasmid mixture with mammalian cells in the absence of transfection reagent to show that only plasmids that had been repaired in cells and were double stranded would be preserved (not destroyed by alkaline lysis) and sequenced. A control that is missing is to use the lesion gap plasmid as template in PCR reaction to check the effect of recombinant DNA polymerase used in PCR. The high-throughput TLS assay generated thousands and tens of thousands of sequences for each transfection experiment, so the NGS technology greatly increased the sample size and the statistical power of my study. Before implementing the high-throughput TLS assay, however, I obtained some results using the traditional TLS assay. I will reconcile the difference in findings of the two methods in the Discussion section.

1.11   Objectives and hypotheses

This thesis can be read with a focus on any of the following three dimensions:

1)   The mutagenic properties of the two DNA lesions 8-oxo-G and Pt-GG in terms of their mutation frequency (Section 3.1 & 3.6), mutation spectrum (Section 3.2 & 3.7) and bypass efficiency (Section 3.12). Comparisons were made between the control plasmids and the lesion plasmids to elucidate the effect of a DNA lesion. Properties of 8-oxo-G lesion and Pt-GG lesion were compared in the Discussion (Section 4.3).

2)   The roles of the translesion DNA polymerases Pol iota, Pol eta, Pol kappa, Rev1 and Rev3L on the bypass of either 8-oxo-G (Section 3.4 & 3.5) or Pt-GG (Section 3.8-3.11). Comparisons were made between a cell line that is wild type for a particular translesion DNA polymerase gene and its isogenic mutant cell line that is deficient for that gene to deduce the *in vivo* functions of the translesion DNA polymerase (Appendix E).

3)   A comparative and critical analyses of the results obtained from the traditional TLS assay and those obtained from the high-throughput TLS assay. In the discussion section, results of the traditional TLS assay and those of the high-throughput TLS assay are compared (Section 4.1) and reconciled (Section 4.7 to 4.9).

In the early phase of this project, I formulated hypotheses based on the reasoning that if a DNA polymerase has the ability to bypass a particular DNA lesion correctly *in vitro*, it would also have a similar function *in vivo*. My first hypothesis is that Pol iota can accurately bypass 8-oxo-G lesions in mammalian cells (Section 3.4). My second hypothesis is that Pol eta can accurately bypass Pt-GG lesions in mammalian cells (Section 3.8). Because the *POLI* gene is phylogenetically related to the *POLH* gene, Pol iota and Pol eta might have similar functions; I further hypothesized that Pol iota, in addition to Pol eta can correctly bypass Pt-GG in cells (Section 3.9). This last hypothesis sought to complement the work of Chuanbing Bian, my colleague who was working on determining the kinetics and solving the X-ray crystal structures of Pol iota on bypassing the Pt-GG lesion. All three of the hypotheses were first investigated by the traditional TLS assay and later by the high-throughput TLS assay.

The high-throughput TLS assay extended the scope of my project beyond the initial hypotheses. I used the assay to determine the mutagenic properties of 8-oxo-G and Pt-GG

and to explore which of the translesion DNA polymerases (Pol iota, Pol eta, Pol kappa, Rev1 and Rev3) bypassed the DNA lesions in mammalian cells. Data of the high-throughput TLS assay could not be analyzed in the same way that data of the traditional TLS assay were analyzed. At first, I thought results of the high-throughput TLS assay would validate those of the traditional TLS assay, but it turned out not to be the case.

1.12   Significance

A change in methodology from the traditional TLS assay to the high-throughput TLS assay led to new understandings about the properties of the two DNA lesions 8-oxo-G and Pt-GG and about translesion synthesis in the mammalian cell. My project generated new hypotheses about the *in vivo* functions of translesion DNA polymerases, overturned previous assumptions on the blocking effects of the two DNA lesions, and called into question the two-polymerase model.

Results of the traditional TLS assay are totally different from those of the high-throughput TLS assay even though they are all produced from the same initial condition, which is a single DNA lesion in the gap plasmid. The high-throughput TLS assay captured a still photograph and the traditional TLS assay captured a time-lapsed photograph of mutations arising from a DNA lesion. The results of the two methods provided insights into the immediate versus the long-term effect of a DNA lesion, where the latter had been influenced by the selective pressure of the cell.

Chapter 2

2 Materials and methods

2.1 Plasmids

Two plasmids, pSKSL-Cm^R (3618 bp) and pSKSL-Kan^R (3365 bp), were obtained from Omer Ziv from Dr. Zvi Livneh's laboratory (Weizmann Institute of Science). The sequences of the two plasmids are almost identical except for their antibiotic markers. The pSKSL-Cm^R plasmid with a chloroamphenicol (Cm) selectable marker was used to make the control gap plasmid that does not have a DNA lesion. The pSKSL-Kan^R plasmid with a kanamycin (Kan) selectable marker was used to make the lesion gap plasmid that has a site-specific DNA lesion, either 8-oxo-G or Pt-GG.

The sequence of the pSKSL-Kan^R plasmid was modified twice by QuikChange$^®$ PCR (Stratagene, 2005). After the first modification, the plasmid was used to construct the lesion gap plasmid containing 8-oxo-G. Based on the first modification, additional changes were made in the sequence of the pSKSL-Kan^R plasmid by a second QuikChange PCR reaction. The resulting plasmid was used to construct the lesion gap plasmid containing Pt-GG.

2.2 QuikChange$^®$ PCR

Primers were designed for each QuikChange$^®$ PCR reaction to make three nucleotide substitutions in the pSKSL-Kan^R plasmid. Sequences of the primers for the two QuikChange$^®$ PCR reactions are listed in Table 1. QuikChange$^®$ PCR reactions were carried out as previously described (Stratagene, 2005). Nucleotide substitutions in the plasmids were confirmed by sequencing.

Table 1   Primers used in QuikChange PCR reactions to make three nucleotide substitutions in the pSKSL-Kan^R plasmid

The first QuikChange® PCR reaction	
Template:	pSKSL-Kan^R plasmid DNA as it was received
Forward primer: 1LG136-oligo 1	GCCCGTCGTGTAGATAACTA**ACG**TACGGGAGGGCTTACCATC (42 bases, GC content 55%, % mismatch to the template: 3/42 bases, T_m calculated by the Stratagene formula: 81°C)
Reverse primer: 1LG136-oligo 2	GATGGTAAGCCCTCCCGTA**CGT**TAGTTATCTACACGACGGGC (42 bases, GC content 55%, % mismatch to the template: 3/42 bases, T_m calculated by the Stratagene formula: 81°C)
Note:	'CAG' in the template was modified to 'ACG' as bolded and underlined in the primer sequences.
The second QuikChange® PCR reaction	
Template:	pSKSL-Kan^R plasmid that has been modified by the first QuikChange® PCR
Forward primer: 4LG94-L	ACTGCCCGTCGTGTAGATAA**GTG**ACGTACGGGAGGGCTTACC (42 bases, GC content 57%, % mismatch to the template: 3/42 bases, T_m calculated by the Stratagene formula: 82°C)
Reverse primer: 4LG94-R	GGTAAGCCCTCCCGTACGT**CAC**TTATCTACACGACGGGCAGT (42 bases, GC content 57%, % mismatch to the template: 3/42 bases, T_m calculated by the Stratagene formula: 82°C).
Note:	'CTA' in the template was modified to 'GTG' as bolded and underlined in the primer sequences. Right after 'GTG' is 'ACG' from the first modification.

2.3   Oligonucleotides with site-specific DNA lesions

54mer oligos containing 8-oxo-G were synthesized at the W. M. Keck Oligonucleotide Synthesis Facility (Yale). The sequence of the 54mer oligo is ACCGCAACGAAGTGATTCCGGCATGC**X**TCCTACCTGGCTACTTGAACCAGACCG, where X denotes 8-oxo-G and the sequence of the single stranded gap region is underlined. The oligos were purified in our lab by anion exchange chromatography using

a SOURCE 15Q column (GE Healthcare) on ÄKTAFPLC system (GE Healthcare). 54mer oligos containing Pt-GG were made by ligating three shorter oligos (5'-end 21mer, 3'-end 21mer and 12mer) together in the presence of a 34mer scaffold DNA (Table 2). The sequence of the 54mer oligo is ACCGCAACGAAGTGATTCCGG<u>CTTCCTCT**GG**CC</u>CTGGCTACTTGAACCAGACCG, where GG denotes Pt-GG and the sequence of the single stranded gap region is underlined.

12mer oligos with Pt-GG were made by Chuanbing Bian in our lab following an established protocol (Zhao et al., 2012). 1.8nmol of the 5'-end 21mer oligos, 1.8nmol of the 3'-end 21mer oligos and 1.5nmol of the 12mer oligos were 5'-phosphorylated using T4 polynucleotide kinase (NEB). The three phosphorylated oligos were combined, annealed to 1.8nmol of the 34mer scaffold oligos, and then ligated using T4 DNA ligase (NEB) in the presence of 1mM ATP. The 54mer ligated products were separated from any unligated oligos and scaffold DNAs on an 8M urea 12% polyacrylamide gel. The gel was ran under constant power of 30-50 Watts for one hour (Figure 5). The 54mer products were extracted from the gel, purified and re-suspended in a final concentration of 15pmol/μl.

Table 2 Sequences of the shorter oligos for making the 54mer oligo containing Pt-GG

5'-end 21mer	ACCGCAACGAAGTGATTCCGG
12mer with Pt-GG (red)	CTTCCTCT**GG**CC
3'-end 21mer	CTGGCTACTTGAACCAGACCG
34mer scaffold	AAGTAGCCGGGGCCAGAGGAAGCCGGAATCACTT

2.4 Construction of double stranded gap plasmids

150pmol of the 54mer oligos, 240pmol of the upstream 15mer primer and 240pmol of the downstream 15mer primer were 5'-phosphorylated. They were annealed to make gap duplexes. 150pmol of the gap duplexes were ligated into 100pmol of linearized DNA

vectors in a large-scale ligation reaction using 60,000 units of T4 DNA ligase in a total final volume of 12ml. The DNA vectors were prepared by digesting 600μg of the pSKSL-Cm^R or p-SKSL-Kan^R plasmid DNAs with 300 units each of the *BsaI* and *BstXI* restriction enzymes, and isolating the larger DNA fragment (around 3kb) by gel purification. After ligation, DNA was concentrated by ethanol precipitation. The singly ligated gap plasmids were separated from any unligated DNA vectors and multiply ligated DNA products by gel purification and electroeluted. The isolated gap plasmids were concentrated by ethanol precipitation and desalted using an Amicon Ultra-0.5mL Microcon$^®$ column (Ultracel 3K Membrane, Millipore). The concentration of the purified gap plasmids was measured by NanoDrop.

Table 3 Sequences of the two 15mer oligos annealing to the 54mer oligo for making the gap duplex

Upstream 15mer primer	GGAATCACTTCGTTG
Downstream 15mer primer	CTGGTTCAAGTAGCC

2.5   Cell lines and cell culture

Isogenic pairs of immortalized mouse embryonic fibroblast (MEF) cell lines with knockout(s) of one or a few translesion DNA polymerase genes were obtained (Table 4). A pair of Burkitt's lymphoma (BL2) cell lines, one wild type and one *POLI*-deficient, was also obtained (Table 4).

Table 4 Mammalian cell lines with various knockouts of translesion DNA polymerase genes

Cell line and genotype	Nick-name	Sender / Laboratory	Reference
MEF *POLI* (+/+)	724-1SV	Dr. Woodgate (National Institute of Child Health and Human Development, USA)	(McDonald et al., 2003)
MEF *POLI* (-/-)	725-3SC	Dr. Woodgate	(McDonald et al., 2003)
BL2 *POLI*(+/+)		Dr. Woodgate	(Faili et al., 2002)
BL2 *POLI*(-/-)	267	Dr. Woodgate	(Faili et al., 2002)
MEF *POLH*(+/+)	B3	K. Hashimoto, Moriya Lab (Stony Brook University, New York)	(Ohkumo et al., 2006)
MEF *POLH*(-/-)		K. Hashimoto, Moriya Lab	(Ohkumo et al., 2006)
MEF *POLK*(+/+)		K. Hashimoto, Moriya Lab	(Ogi et al., 2002)
MEF *POLK*(-/-)		K. Hashimoto, Moriya Lab	(Ogi et al., 2002)
MEF *POLH*(-/-) *POLI*(-/-)		K. Hashimoto, Moriya Lab	(Ohkumo et al., 2006)
MEF *POLK*(-/-) *POLI*(-/-)		K. Hashimoto, Moriya Lab	(Hashimoto et al., 2012)
MEF *POLH*(-/-) *POLI*(-/-) *POLK*(-/-)	2091	K. Hashimoto, Moriya Lab	(Hashimoto et al., 2012)
MEF *POLH*(-/-) *POLI*(-/-) *POLK*(-/-)	2095	K. Hashimoto, Moriya Lab	(Hashimoto et al., 2012)
MEF *REV1*(+/+)		Dr. Niels de Wind (Leiden University Medical Center, Netherlands)	(Jansen, 2006)
MEF *REV1*(-/-)		Dr. Niels de Wind	(Jansen, 2006)
MEF *REV3L* (+/lox) Cre	3(+)6	L. Sabine, Wood Lab (MD Madison Cancer	(Lange et al., 2012)

		Center, Texas)	
MEF *REV3L* (-/lox) Cre	4(-)5	L. Sabine, Wood Lab	(Lange et al., 2012)
MEF *REV3L* (-/lox) Cre	4(-)11	L. Sabine, Wood Lab	(Lange et al., 2012)
MEF *REV3L*(+/+) *P53*(-/-)	B2	L. Sabine, Wood Lab	(Wittschieben et al., 2006)
MEF *REV3L*(-/-) *P53*(-/-)	B4-9	L. Sabine, Wood Lab	(Wittschieben et al., 2006)
MEF *REV3L*(-/-) *P53*(-/-)	B4-18	L. Sabine, Wood Lab	(Wittschieben et al., 2006)

Mouse embryonic fibroblast (MEF) cell lines were cultured in Dulbecco's Modified Eagle's Medium (DMEM) with 4.5g/L glucose and sodium pyruvate, adding 1x GlutaMAX (Invitrogen), 10% new born calf serum (Wisent), and penicillin (100 IU) and streptomycin (100ug/ml). Cells were passaged with 0.25% trypsin 2.21mM EDTA solution. Burkitt's lymphoma (BL2) cells grow in suspension. BL2 cell lines were cultured in RPMI 1640 medium with L-glutamine and sodium bicarbonate (Wisent), adding 10% fetal bovine serum (Wisent), and penicillin (100IU) and streptomycin (100ug/ml). The cell lines were maintained at 37°C in 5% CO_2 environment chamber.

2.6   Transient Transfection

Cells were seeded at a density of $2x10^5$ cells per 35mm dish the day before transfection. 33ng of the control gap plasmids, 33ng of the lesion gap plasmids and 2μg of the carrier plasmids (pBlueScript) were transfected into cells using jetPRIME® reagent (Polyplus Transfection^{TM}), as per manufacturer's recommendations. Cells were incubated for seven hours.

2.7   Traditional TLS assay

Mammalian cells that had been transfected with gap plasmids were harvested and washed three times with PBS. The plasmids were extracted by the alkaline lysis method using a

plasmid Miniprep kit (Invitrogen). Plasmid DNAs were transformed into *E. coli* (CSH117 strain) by electroporation. Bacterial cells were plated in triplicates on chloramphenicol (30μg/ml) and kanamycine (50μg/ml) plates to separate the control plasmids and the lesion plasmids, respectively. Translesion synthesis efficiency was determined by taking the ratio of the average number of bacterial colonies arising on the kanamycin plates to that arising on the chloramphenicol plates. Isolated bacterial colonies were inoculated into LB starter cultures and grown overnight. Plasmid DNAs were extracted and their sequences were determined by Sanger Sequencing.

2.8 Statistical analyses of traditional TLS assay

To determine if the POLI gene has an effect on bypassing 8-oxo-G, two-tailed Fisher's Exact Test (Fisher, 1922) was used to determine if the number of sequences with mutation across from a DNA lesion of the POLI wild type cell line is significantly different from that of an isogenic POLI-deficient cell line. A two by two contingency table was set up for computing the exact P-value (Weisstein), whereby the two groups compared are POLI wild type cell line and its isogenic POLI(-/-) cell line and the two outcomes are "no mutation" and "with mutation" across from the DNA lesion. If the computed P-value is less than α of 0.05, the difference in observations between the two cell lines is deemed statistically significant. The same method was used to determine if the POLH gene and the POLI gene each had an effect on bypassing Pt-GG.

2.9 Quantitative PCR

Q-PCR reactions were used to measure copy numbers of the control plasmids and lesion plasmids extracted from mammalian cells. The ratio of the number of copies of the lesion plasmids to that of the control plasmids (Kan/Cm) is the translesion synthesis efficiency. A 72 bp fragment of the chloramphenicol resistant gene of the control plasmids and a 71 bp fragment of the kanamycin resistant gene of the lesion plasmids were amplified using the primers listed in Table 5. Melting curve analyses showed that the primers were

specific for amplifying their respective targets. Standard curves were plotted using known concentrations of each of the control plasmids and lesion plasmids from 5ng/µl to 5×10^{-6}ng/µl in ten-fold serial dilutions. Concentrations of unknown samples were interpolated from the standard curves.

Table 5 Sequences of Q-PCR primers used to measure copy numbers of the control plasmids and the lesion plasmids extracted from mammalian cells

Primers amplifying from the control plasmids	
Left	GGGAAATAGGCCAGGTTTTC (20bp, Tm=59.8°C, GC content 50%)
Right	CGATTTCCGGCAGTTTCTAC (20bp, Tm=59.7°C, GC content 50%)
Primers amplifying from the lesion plasmids	
Left	AGAAACAACTCTGGCGCATC (20bp, Tm=60.4°C, GC content 50%)
Right	GATAATGTCGGGCAATCAGG (20bp, Tm=60.3°C, GC content 50%)

In an earlier experiment, four DNA samples (3LG188.1, 3LG188.2, 3LG102.5 and 3LG102.6) were measured. A typical Q-PCR reaction consisted of 2µl left primer (1µM stock), 2µl right primer (1µM stock), 1µl DNA extracted from mammalian cells or 1µl DNA standard, and 5µl 2x EvaGreen qPCR Master Mix. Reactions were set up in replicates and ran on Eppendorf RealPlex Mastercycler. The thermocycling condition was 95°C 2 min, 35 cycles of 95°C 20 s, 57°C 20 s and 72°C 35 s, then 72°C 10 min.

In a later experiment, the translesion synthesis efficiencies of 8-oxo-G and Pt-GG were determined from each of the nine MEF cell lines with the following genotypes: *POLI*(+/+), *POLI*(-/-), *POLH*(+/+), *POLH*(-/-), *POLK*(+/+), *POLK*(-/-), *POLH*(-/-) and *POLI*(-/-), *POLK*(-/-) and *POLI*(-/-), and 2091 *POLH*(-/-) *POLK*(-/-) and *POLI*(-/-). For each cell line, three biological replicates were used and each biological replicate was measured in triplicates. Q-PCR reactions were set up on a 384-well plate using SYBR Select Master Mix (Invitrogen) and ran on the ViiA7™ Real-Time PCR System. The thermocycling condition was 50°C 2 min, 95°C 2 min, 95°C 15 s, 40 cycles of 95°C 15 s, 57°C 15 s and 72°C 1 min.

2.10   High-throughput TLS assay

The gap regions of the plasmids extracted from mammalian cells were amplified by PCR using primers that contain barcodes and Illumina adapter sequences. Twenty-four barcodes were embedded in left primers and right primers, respectively. The primers were ordered from Integrated DNA Technologies in 25nmole scale.

In the 8-oxo-G study, the control plasmids and lesion plasmids of a transfection experiment were amplified in the same PCR reaction. They would be separated computationally based on a three nucleotide difference in their sequences. The sequence of the plasmid-specific portion of the left primer is GACTGCCCGTCGTGTAGATAA, and that of the plasmid-specific portion of the right primer is AGCGGTCTGGTTCAAGTAGC. A PCR reaction was set up using 11.8µl water, 4µl 5x HF Phusion buffer (Thermo Scientific), 0.4µl 10mM dNTP, 0.8µl left primer (5uM stock), 0.8µl right primer (5uM stock), 2µl plasmid DNA extracted from mammalian cells, and 0.2µl Phusion Hot Start DNA Polymerase (2U/µl, Thermo Scientific). The thermocycling condition was 98°C 30 s, 25 cycles of 98°C 10 s, 66°C 20 s and 72°C 15 s, and lastly 72°C 5 min. There are 192 transfection experiments in the 8-oxo-G study, including nine biological replicates for each of the twenty cell lines and positive and negative controls.

In the Pt-GG study, the control plasmids and the lesion plasmids of a transfection experiment were amplified in two separate PCR reactions using different left primers and a common right primer. The sequence of the plasmid-specific portion of the left primer used to amplify the control plasmid is CCCGTCGTGTAGATAACTACGA, and that used to amplify the lesion plasmid is GCCCGTCGTGTAGATAAGTGA. The sequence of the right primer is the same as that used in the 8-oxo-G study. A typical PCR reaction consisted of 8.2µl water, 4µl 5x HF Phusion buffer, 0.6µl DMSO, 0.4µl 10mM dNTP, 0.8µl left primer (5uM stock), 0.8µl right primer (5uM stock), 5µl plasmid DNA extracted from mammalian cells and 0.2µl Phusion Hot Start DNA Polymerase (2U/µl). Two rounds of DNA amplifications were carried out for a total of 55 PCR cycles. The first PCR reaction was ran 98°C 30 s, 25 cycles of 98°C 10 s and 72°C 15 s, and ended at 72°C 5 min. DNA products were purified and then used as templates in the second PCR

reaction. There are 384 barcoded samples in the Pt-GG study. They came from 192 transfection experiments, including nine biological replicates for each of the twenty cell lines and positive and negative controls.

PCR products of the 8-oxo-G and Pt-GG studies were purified, and their concentrations were measured using a Qubit® dsDNA High Sensitivity Assay Kit (Invitrogen). 10ng of each of the 576 barcoded DNA samples were combined and further concentrated, then the DNA sample was sequenced using 2 by 150 Illumina MiSeq (Robarts Institute).

2.11   Processing Next Generation Sequencing raw data

Two FASTQ files were generated from the MiSeq run (Appendix 4), one file contains reads sequenced in the forward direction (R1) and the other file contains reads sequenced in the reverse direction (R2). Each file was processed independently following the same procedures. Briefly, reads of the PhiX genome were removed using BOWTIE2, a read alignment tool (Langmead and Salzberg, 2012). FASTQ/A Short-reads Pre-processing Tools (Hannon Lab, Cold Spring Harbor Laboratory) were used to trim, clip and split data. Reads with the same barcodes were grouped and then further split into control plasmids and lesion plasmids, and those that don't match to barcodes were discarded. All the sequences were trimmed to 25 bases including the 12-nucleotide gap region. Base substitutions of every nucleotide in the gap region were tallied using a 'NGS Position Counter' program (Taggart et al., 2013). To count the number of reads of different permutations of dinucleotides, such as those opposite the Pt-GG lesion, a Perl script given to me by Greg Gloor was used to parse the data.

2.12   Statistical analyses of Next Generation Sequencing data

The same care was taken to analyze reads of the control plasmids and reads of the lesion plasmids. The overall mutation frequencies across from 8-oxo-G or Pt-GG were calculated based on the sum total of reads obtained for each DNA lesion, regardless of the genotype of cell lines. The mutation frequency of the lesion plasmids was compared

to that of the control plasmids to elucidate the effect of a DNA lesion. To understand how the mutation spectrums differ between the control plasmids and the lesion plasmids, principal component analysis (Aitchison, 2003) was applied. The CoDaPack Software 2.01 (Comas-Cufi and Thió-Henestrosa, 2011) and the compositions package in R (van den Boogaart and Tolosana-Delgado, 2008) were used to generate biplots. A biplot shows the correlations among variables between the control plasmids and the lesion plasmids, and the relative proportions of changes in the variables effected by a DNA lesion.

To infer the biological function of a translesion DNA polymerase on the bypass of a certain DNA lesion, results from two isogenic cell lines were compared. See Appendix E for a list of all the cell lines compared, both for the control plasmids and for the lesion plasmids in the high-throughput TLS assay. For each cell line and for each of its nine biological replicates, the within-sample ratios were calculated and log2 transformed. In the 8-oxo-G study, the within-sample variables are log2(A/C), log2(T/C), log2(G/C) and log2(deletion/C). In the Pt-GG study, the within-sample variables are log2(CT/CC), log2(AT/CC), log2(CG/CC), log2(AG/CC), log2(C/CC), log2(TC/CC), log2(GC/CC), log2(AC/CC), log2(A/CC), log2(CA/CC), log2(AA/CC), log2(double deletions/CC). Wilcoxon Rank Sum Test was used to compare each of the variables between a pair of isogenic cell lines (N=9) and to calculate P-values. P-values were adjusted by the Benjamini-Hochberg method (Benjamini and Hochberg, 1995). Where the adjusted P-values were less than 0.05 (α), the effect sizes (Cohen's d) and their 95% confidence intervals (C.I.) were calculated and the biological functions of the translesion DNA polymerases were deduced based on the genotypes of the cell lines being compared.

Chapter 3

3   Results

3.1   Construction of double stranded gap plasmids containing an 8-oxo-G or Pt-GG lesion

To study mutations in mammalian cells caused by a DNA lesion, I constructed double stranded gap plasmids containing a site-specific 8-oxo-G or Pt-GG lesion. The gap plasmids were transfected into mammalian cells as the first step in the translesion synthesis assay. To produce double stranded gap plasmids, gapped duplexes were ligated into double stranded DNA vectors as shown in Figure 4. Each gap duplex was formed by annealing two short primers to a 54 mer oligonucleotide containing a DNA lesion. The 54mer oligo was produced by bridging together three shorter oligos that had been 5'-phosphorylated, in the presence of a 34mer scaffold DNA, and the final product was purified by denaturing PAGE (Figure 5). DNA vectors were prepared by digesting 600µg of the pSKSL-Cm^R or p-SKSL-Kan^R plasmid DNAs with 300 units each of the *BsaI* and *BstXI* restriction enzymes and isolating the larger DNA fragment (around 3kb) by gel purification. 150pmol of the gap duplexes were ligated into 100pmol of linearized DNA vectors in a large-scale ligation reaction using 60,000 units of T4 DNA ligase in a total final volume of 12ml. After ligation, DNA was concentrated by ethanol precipitation and then the singly ligated gap plasmids were separated from any unligated DNA vectors and multiply ligated DNA products by gel purification (Figure 6), and electroeluted using dialysis bags. The isolated gap plasmids were concentrated by ethanol precipitation and desalted using an Amicon Ultra-0.5mL Microcon® column (Ultracel 3K Membrane, Millipore). The gap plasmids prepared transfected with high efficiency into mouse embryonic fibroblast cells. Control gap plasmids without DNA lesions were produced in a similar way.
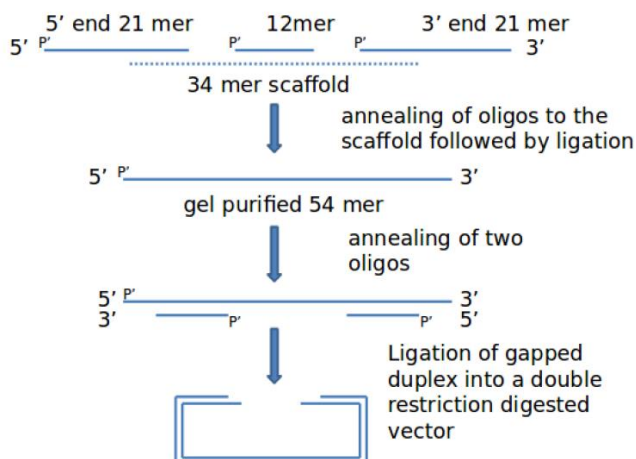
Figure 4 Gap plasmids were made by ligating gap duplexes into linear DNA vectors

A 54mer oligonucleotide with or without a site-specific DNA lesion was produced by ligating three shorter oligonucleotides that were phosphorylated at the 5'-end and that were annealed to a 34mer scaffold oligo. The 54mer oligo was gel purified (Figure 5) and then annealed to two primers to form a gap duplex, which was then ligated into a DNA vector to form a double stranded plasmid with a small single stranded gap (Figure 6).
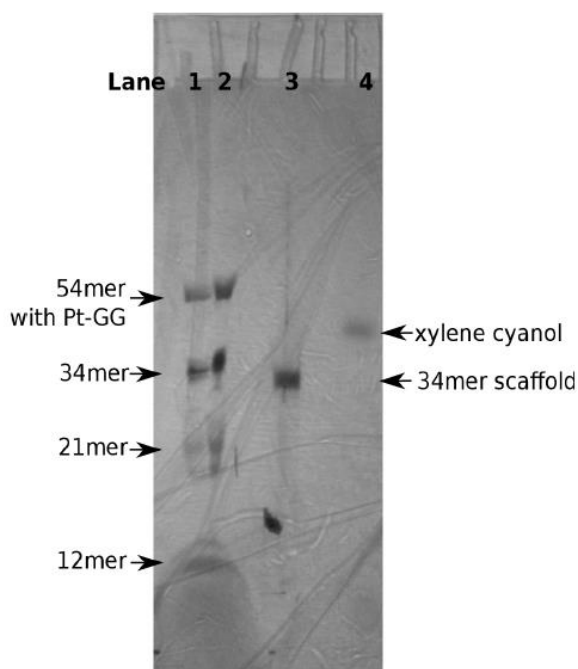
Figure 5   Isolating 54mer oligos containing a site-specific DNA lesion by gel purification

8M urea 12% PAGE. Lane 1 and 2 were loaded with the ligation reaction—ligation of three oligonucleotides annealed to 34mer scaffold oligos. Lane 3 was loaded with the scaffold DNA only. Lane 4 was loaded with a tracking dye.

* marks the gap plasmid to isolate
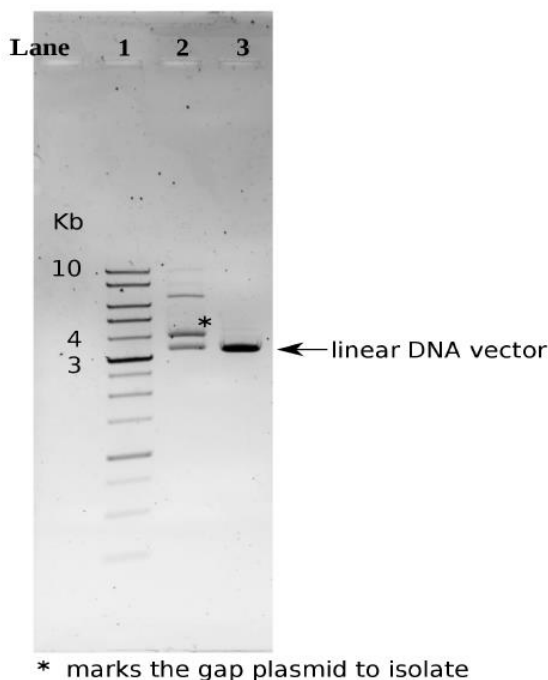
Figure 6 Gap plasmids produced from a large-scale ligation reaction

Singly ligated gap plasmids (marked by *) ran at 4kb in 0.8% agarose gel made with 1xTAE buffer. Lane 2: an aliquot of the large-scale ligation reaction—ligation of gapped duplexes into DNA vectors; Lane 3: DNA vectors only. In gel purification of the large-scale ligation reaction, DNA bands stained with ethidium bromide were visualized under short wavelength UV light. Exposure to UV light was kept to minimum to prevent formation of pyrimidine dimers.

3.2   8-oxo-G caused a high frequency of mutation

To determine the mutation frequency of an 8-oxo-G lesion, gap plasmids containing the lesion were co-transfected with control gap plasmids without the lesion into mouse embryonic fibroblast cells and BL2 cells. The mutation frequency of 8-oxo-G was determined by the traditional TLS assay and by the high-throughput TLS assay.

In the traditional TLS assay, gap plasmids that had been repaired were retrieved from mammalian cells and then transformed into *E. coli* for amplification. The control plasmids and lesion plasmids were separated by plating the transformed bacterial cells on chloramphenicol plates and on kanamycin plates, respectively. Isolated bacterial colonies

arising on the plates were grew in starter cultures and the plasmid DNAs were extracted and sent for sequencing. In the high-throughput TLS assay, the gap regions of the plasmids retrieved from mammalian cells were amplified by PCR and then sequenced; the control plasmids and lesion plasmids were separated based on a three-nucleotide difference in their sequence.

In the traditional TLS assay, the mutation probability of 8-oxo-G was 64 out of 174 sequences analyzed (36.8%) for MEF cells, and 41 out of 182 sequences analyzed (22.5%) for BL2 cells. There were no mutations across from a regular G nucleotide in the control plasmids, where 60 control plasmid sequences were analyzed for the MEF cell type and for the BL2 cell type, respectively.

In the high-throughput TLS assay, the mutation frequency of 8-oxo-G or that of a regular G nucleotide was the same between MEF cells and BL2 cells. The mutation frequency of 8-oxo-G was 91% from a total of 319,777 reads analyzed (279,650 reads for MEF cells and 40,127 reads for BL2 cells), and that of a regular G nucleotide was 1.6% from a total of 541,965 reads analyzed (461,711 reads for MEF cells and 80,254 reads for BL2 cells). The background error rate determined as the average of three positive control experiments where a fully double stranded plasmid with known sequence was used as template in PCR reactions was 0.09%.

3.3   Almost all the mutations caused by 8-oxo-G were C->A substitutions

The mutation spectrum of 8-oxo-G was determined by the traditional TLS assay and by the high-throughput TLS assay. Most of the mutations formed across from 8-oxo-G were C->A substitutions. In the traditional TLS assay, 93% of mutations (60/64 sequences) opposite to the lesion transfected into MEF cells were C->A substitutions, 5% (3/64 sequences) were C->G substitutions and 2% (1/64 sequences) were C->T substitutions; all of the mutations opposite to the lesion transfected into BL2 cells were C->A substitutions (38/38 sequences).

In the high-throughput TLS assay, 99% of all the mutations across from 8-oxo-G were C->A substitutions and only 1% were C->G substitutions (Figure 7). Furthermore, principal component analysis of the NGS data shows that, first of all, reads of the control plasmids and reads of the lesion plasmids are highly separable based on their nucleotide compositions at the position of the DNA lesion (Figure 8); second, the major difference between the control plasmid and the lesion plasmid is in the frequency of C->A substitutions; third, change in the variable C (inserted against a regular nucleotide G) in the control plasmids is linearly correlated to changes in the variables A and G (inserted opposite to 8-oxo-G) in the lesion plasmids at different proportions (Figure 8). The variable C is weakly correlated to the variable T and uncorrelated to the variable deletion (Figure 8). Principal component analysis indicates that 8-oxo-G caused purine bases, mostly A to be incorporated thus producing transverion mutations.
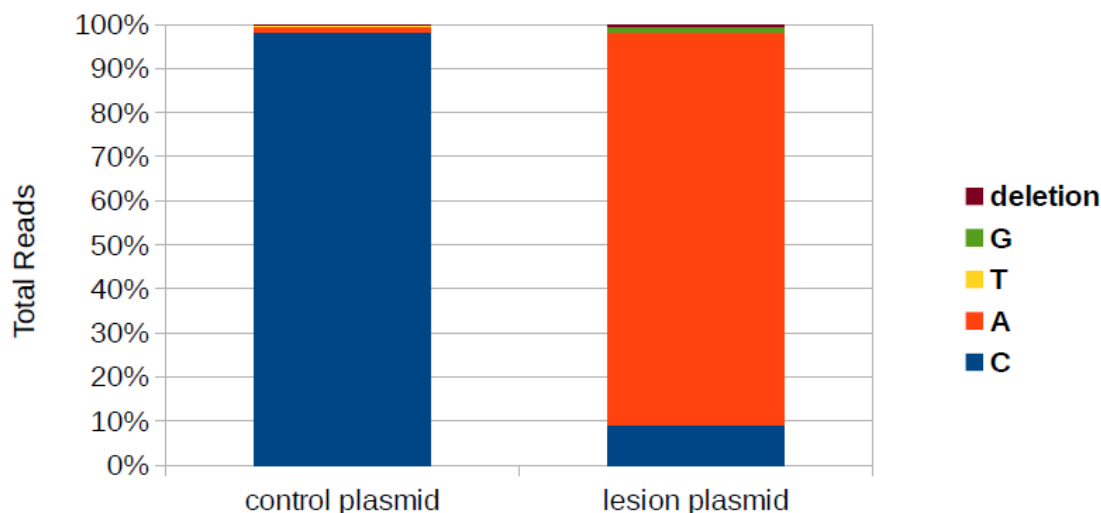
Figure 7 Mutation frequency and mutation profile of 8-oxo-G in mammalian cells.

Results of the high-throughput TLS assay show that while C was correctly inserted to a regular G nucleotide in the control plasmids, A was misincorporated against 8-oxo-G in the lesion plasmids and the mutation frequency of 8-oxo-G was 91%. 541,965 reads were analyzed for the control plasmids and 319,777 reads were analyzed for the lesion plasmids.
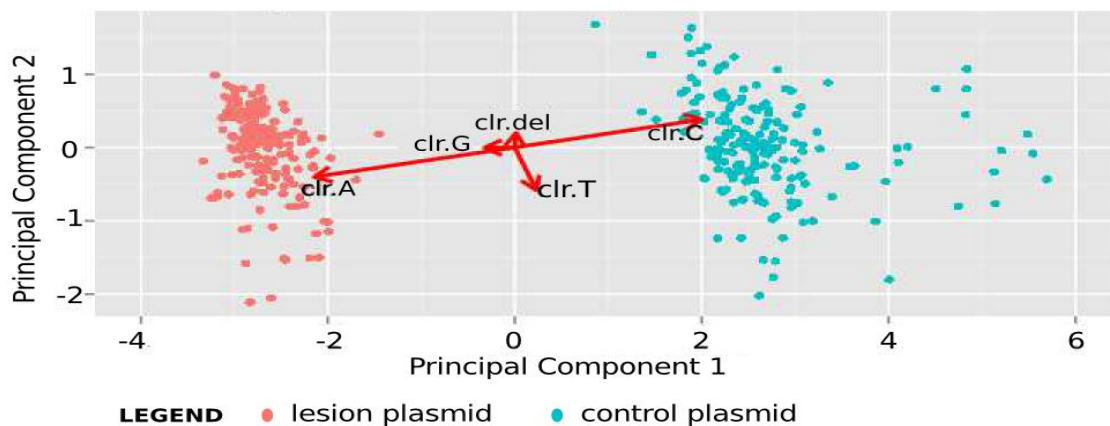
Figure 8 Biplot generated from principal component analysis of the 8-oxo-G data

Principal component analysis (PCA) reduces a high dimensional or multivariate dataset into a smaller number of linearly uncorrelated variables called principal components through orthogonal transformation of the data. The transformation is defined in such a way that the first principal component contains the largest possible variance and subsequent principal components contain the largest possible variance under the constraint that they are uncorrelated to the previous principal components (Aitchison, 2003). Through PCA the data is placed in a new dimension and is deconstructed into eigenvectors and eigenvalues that form the principal components. Eigenvectors and values exist in pairs with the eigenvector indicating the direction and the eigenvalue indicating the spread of the data in that direction. In simpler terms, PCA reveals the underlying structure of the data and divides the variance of the data into directions that are perpendicular or principal components that are uncorrelated to each other. A biplot displays experimental observations and variables of an experiment on the same plot. Each point in the biplot is data of a transfection experiment. The raw experimental data was converted into compositional data by first taking log-ratios and then applying center log transformation (clr). PCA does not lose the relative scaling of the original variables. In the 8-oxo-G study, reads of the control plasmids are well separated from reads of the lesion plasmids along principal component one that accounts for 88.4±2.76% of total variance of the data. This proves that constructions of the control and the lesion double stranded gap plasmids were successful. While the magnitude of composition vectors shows the relative proportion of variables, the direction of composition vectors shows the correlation among different variables. Data points for the control plasmids (blue dots) are largely driven by C incorporated across from a regular G nucleotide and those for the lesion plasmids (red dots) are driven by A incorporated against 8-oxo-G. Because a change in the variable C is highly correlated to changes in the variables A and G, 8-oxo-G caused transversion mutations.

3.4 There was no apparent biological function of Pol iota in the bypass of 8-oxo-G

Because 8-oxo-G is a very abundant and highly mutagenic DNA lesion (van Loon et al., 2010), DNA polymerases that can correctly bypass the lesion are biologically important. Pol iota is one of the most error-prone DNA polymerases in eukaryotes (Tissier et al., 2000; Zhang et al., 2000a; Kirouac and Ling, 2011b; Makarova and Kulbachinskiy, 2012). Previous structural findings demonstrated that Pol iota correctly bypasses 8-oxo-G *in vitro*, thus pointing to a plausible biological function of the DNA polymerase (Kirouac and Ling, 2011a).

To investigate the function of Pol iota in the bypass of 8-oxo-G in mammalian cells, I used both the traditional TLS assay and the high-throughput TLS assay. For each assay the mutation frequency of 8-oxo-G of a *POLI* wild type cell line was compared to that of a *POLI*-deficient cell line to deduce Pol iota's function. In the traditional TLS assay, the mutation probability of 8-oxo-G was 33 out of 89 sequences analyzed (37%) for the MEF *POLI* wild type cell line and 31 out of 85 sequences analyzed (37%) for the MEF *POLI*(-/-) cell line, so Pol iota did not have an effect on bypassing the lesion in MEF cells (P-value=1, Fisher's Exact Test, α=0.05). Also in the traditional TLS assay, the mutation probability of 8-oxo-G was 17 out of 93 sequences analyzed (18%) in the BL2 wild type cell line and 24 out of 89 sequences analyzed (27%) in the BL2 *POLI*(-/-) cell line. Pol iota also did not play a role in the correct bypass of the lesion in BL2 cells (P-value=0.2139, Fisher's Exact Test, α=0.05).

In the high-throughput TLS assay, there was no difference in the mutation frequency of 8-oxo-G between the MEF *POLI*(+/+) wild type cell line and the MEF *POLI*(-/-) cell line nor between the BL2 wild type cell line and the BL2 *POLI*(-/-) cell line. Therefore, it was not likely that Pol iota bypassed 8-oxo-G in mammalian cells. The reason for this result could be that the function of Pol iota was redundant with those of other DNA polymerases and that Pol iota's activity was minor compared to those of other DNA polymerases so its effect could not be detected in mammalian cells.

3.5   Pol eta correctly bypassed 8-oxo-G *in vivo*

To determine the functions of Pol iota, Pol eta, Pol kappa, Rev1 and Rev3L on the bypass of an 8-oxo-G lesion in mammalian cells, I used the high-throughput TLS assay that incorporated the use of Next Generation Sequencing. The high-throughput TLS assay shows that Pol eta, among the five translesion DNA polymerases tested correctly bypassed 8-oxo-G *in vivo*.

To deduce the function of Pol eta, I compared the mutation frequency of 8-oxo-G between the MEF *POLH*(+/+) cell line and the MEF *POLH*(-/-) cell line, where there were nine transfection experiments set up for each cell line. Similarly, to deduce the function of Pol kappa, I compared between the MEF *POLK*(+/+) cell line and the MEF *POLK*(-/-) cell line; to deduce the function of Rev1, I compared between the MEF *REV1*(+/+) cell line and the MEF *REV1*(-/-) cell line. To deduce the function of Rev3L, I compared the 3(+)6 *REV3L*(+/lox) cell line to each of the 4(-)5 and 4(-)11 cell lines that have the genotype *REV3L*(-/lox); and compared the B2 *REV3L*(+/+) *P53*(-/-) cell line to each of the B4(-)9 and B4(-)18 cell lines that have the genotype *REV3L*(-/-) *P53*(-/-). To uncover synergistic effects of the translesion DNA polymerases, I compared each of the *POLI*(+/+), *POLH*(+/+) and *POLK*(+/+) wild type cell lines to the two double knockout cell lines with the genotypes *POLH*(-/-)*POLI*(-/-) and *POLK*(-/-) *POLI*(-/-), respectively. Moreover, I compared each of the *POLI*(-/-), *POLH*(-/-) and *POLK*(-/-) single knockout cell lines to the two triple knockout cell lines named 2091 and 2095, respectively, that have the genotype *POLH*(-/-)*POLI*(-/-)*POLK*(-/-). The reader may refer to Appendix E for a list of all the cell lines compared in the high-throughput TLS assay.

I focused on analyzing the frequency of C->A substitutions across from 8-oxo-G because the frequencies of C->G and C->T substitutions were negligible. When processing the NGS data, I took the within-sample ratio A to C—the proportion of reads with A to those with C for a given transfection experiment. This normalized the data to one and cancelled out any experimental errors. Then the A/C ratio was log2 transformed to express the proportion of reads with A to those with C as fold changes. The log2 transformed A to C ratios of nine biological replicates of a cell line were compared to those of the nine biological replicates of the other cell line by the Wilcoxon Rank Sum Test. I used this

statistical test because it does not make assumption about the data distribution. The Wilcox Rank Sum Test evaluates the likelihood that the median of the nine log2(A/C) ratios of a wild type cell line and that of the nine log2(A/C) ratios of its isogenic mutant cell line were from two different populations. In my case, the test is interpreted as the probability that the translesion DNA polymerase(s) in question, which is the only difference between the two isogenic cell lines compared, affected the mutation frequency of 8-oxo-G. P-values calculated by Wilcox Rank Sum Tests were adjusted by the Benjamini-Hochberg method (Benjamini and Hochberg, 1995) to keep Type I error in check. The chance of falsely rejecting a null hypothesis is one in twenty when alpha of P-value is set at 0.05, but when the number of statistical tests performed increases, so is the likelihood of making false positive inferences. Benjamini-Hochberg (BH) method controls for false positive findings of a study that involves multiple comparisons by scaling P-values, so the overall false discovery rate (the ratio of the false positives to true positives) of the study is kept low, say at 5% when alpha of the BH-adjusted P-value is chosen at 0.05. Where the adjusted P-value was less than 0.05, the difference between two isogenic cell lines was deemed statistically significant and the effect size (Cohen's d) and its 95% confidence interval (C. I.) were calculated. Cohen's d expressed the difference in mutation frequency of 8-oxo-G between two isogenic cell lines in units of standard deviation. An effect size of 0.6 or greater is considered large by convention.

The 8-oxo-G study and bypass of the DNA lesion by five translesion DNA polymerases took up 9% of the total sequencing capacity. The number of reads obtained for each transfection experiment is shown in Figure 9. The most significant finding of the high-throughput study is that Pol eta correctly bypassed 8-oxo-G. Compared to the MEF *POLH*(+/+) cell line, knocking out the *POLH* gene alone or knocking out the *POLH* gene along with the *POLI* gene led to more frequent C->A substitutions across from an 8-oxo-G lesion but not across from a regular G nucleotide (Figure 10).
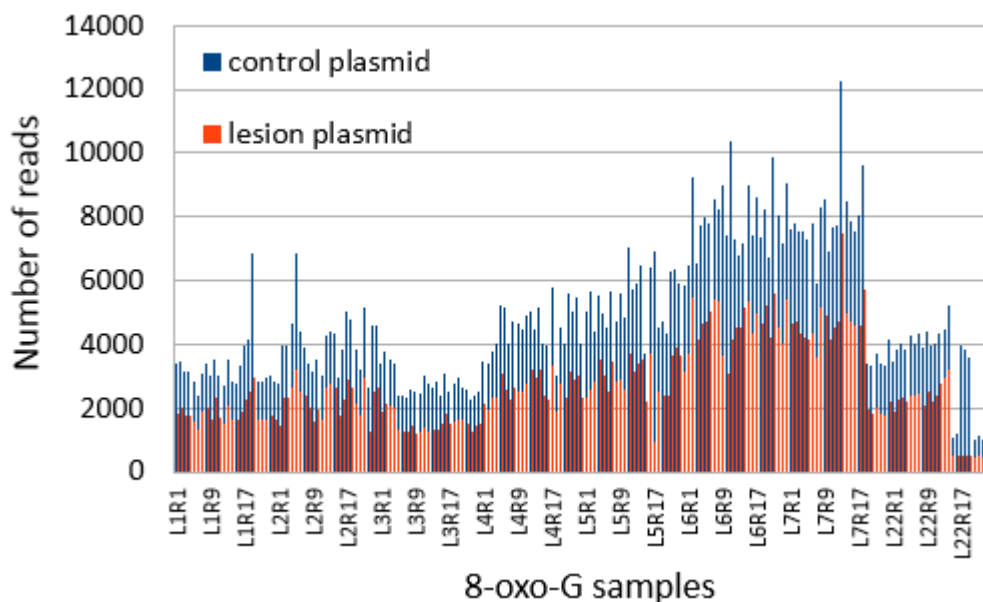
Figure 9 Reads obtained for every transfection experiment in the 8-oxo-G study

The number of reads for the control plasmid and that for the lesion plasmid are shown for every transfection experiment in the 8-oxo-G study (see Appendix F for raw data). For a given DNA sample extracted from mammalian cells, the control plasmids and the lesion plasmids were amplified in the same PCR reaction (for a total of 25 cycles) and thus their reads were labelled with the same barcodes; the reads were split based on a three nucleotide difference in their sequences close to the gap region. Data for a total of 192 transfection experiments are shown. They include nine biological replicates for each of the twenty cell lines listed in Table 4, and positive and negative controls done in triplicates. Positive control PCR reactions used the control plasmid (without gap) as template. There are two types of negative control experiments: in the first type, no transfection reagent was added, as a result none of the gap plasmids entered into mammalian cells and during alkaline lysis, the gap plasmids were denatured and eliminated; in the second type, no DNA template was added in PCR reactions. The purpose of this graph is to present raw data for the 8-oxo-G study and to show how the NGS capacity was divided among transfection experiments. There are more reads for the control plasmid than for the lesion plasmid, indicating that the control plasmid is filled more efficiently than the lesion plasmid in mammalian cells; this point will be examined in more detail in Section 3.12.
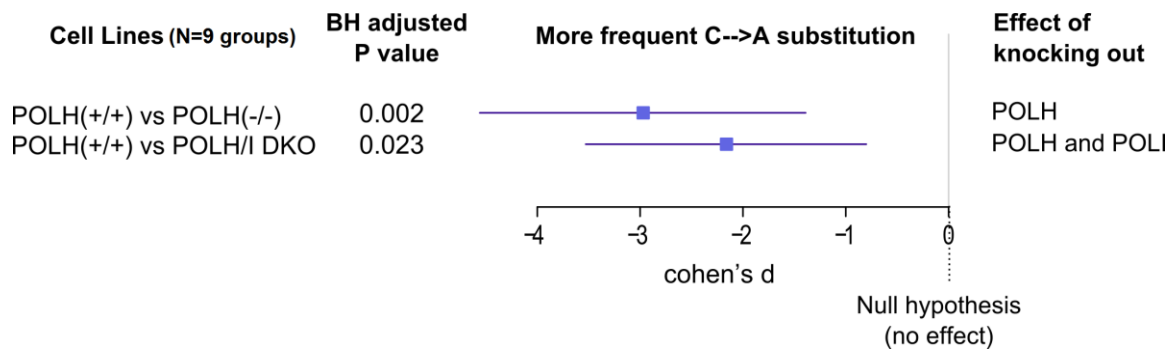
Figure 10 The role of *POLH* on the bypass of 8-oxo-G

Knocking out the *POLH* gene alone or knocking out the *POLH* gene along with the *POLI* gene led to more frequent C->A substitutions. The forest plot shows the effect size and the Benjamini-Hochberg (BH) adjusted P-value for comparisons made between the *POLH*(+/+) wild type cell line and each of the *POLH*(-/-) single knockout and the *POLH*(-/-) *POLI*(-/-) double knockout (DKO) cell lines.

Besides Pol eta, another translesion DNA polymerase Rev3L correctly bypassed 8-oxo-G lesion in one of the four pairs of *REV3L* cell lines analyzed. Knocking out the *REV3L* gene in a *P53* null genetic background resulted in more frequent C->A substitutions against an 8-oxo-G lesion but not against a regular G nucleotide (Table 6).

Table 6 The effect of *REV3L* on the bypass of 8-oxo-G

Cell lines (N=9 groups)	Adjusted P-value	Cohen's d	95% C. I.
REV3L(+/+) *P53*(-/-) versus *REV3L*(-/-) *P53*(-/-) Nick name: B2 versus B4(-)18	0.023	-2.05	-3.38 to -0.71

Pol kappa and Rev1 had no effects on bypassing 8-oxo-G. Results were validated by the NGS data sequenced from the opposite direction (see Appendix A).

Given that the mutation frequency of 8-oxo-G was found to be 91% in the high-throughput TLS assay, the finding that Pol eta was the major DNA polymerase in mouse embryonic fibroblasts to correctly bypass the lesion is biologically important. Efficient and accurate replication of 8-oxo-G by the yeast Pol eta and the human Pol eta have been

reported, whereas the role of Rev3L in the tolerance of 8-oxo-G has not been characterized.

3.6 The mutation frequency of Pt-GG was similar to that of regular GG dinucleotides

To determine the mutation rate of Pt-GG I used both the traditional TLS assay and the high-throughput TLS assay. Gap plasmids containing Pt-GG lesions were co-transfected with control gap plasmids without lesions into mouse embryonic fibroblast cells. In the traditional TLS assay, gap plasmids that had been repaired in mammalian cells were amplified by *E. coli* and then sequenced. The average mutation probability of Pt-GG was 19 out of 110 sequences analyzed (17.3%) for the first insertion step, 6 out of 110 sequences analyzed (5.4%) for the second insertion step, and overall, 25 out of 110 sequences analyzed (22.7%). There were no mutations across from regular GG dinucleotides in the control plasmids from a total of 35 sequences analyzed.

In the high-throughput TLS assay, the gap region of the plasmids that had been repaired in mammalian cells were amplified by PCR. The mutation frequency of Pt-GG was 21,520 reads out of 4,491,510 reads analyzed (4.79×10^{-3}) and that of regular GG dinucleotides was 38,212 reads out of 5,901,558 reads analyzed (6.47×10^{-3}). The experimental error rates were 2.05×10^{-3} for the lesion plasmids (Table 7) and 2.65×10^{-3} for the control plasmids (Table 8) determined from positive control experiments. The mutation frequency of Pt-GG was similar to that of a regular GG dinucleotides and was equal to two times the background error rate of the experiment.

Table 7 Determining the background error rate for amplifying Pt-GG lesion plasmids.

Positive PCR reactions for Pt-GG control plasmids	Total reads obtained for the sample	Number of reads containing errors across from GG dinucleotides
L24R16	25996	65
L24R17	20242	36
L24R18	25642	48
Average error rate	0.00205	

Table 8 Determining the background error rate for amplifying Pt-GG control plasmids.

Positive PCR reactions for Pt-GG control plasmids	Total reads obtained for the sample	No. of reads containing errors across from GG dinucleotides
L23R16	46450	127
L23R17	52032	144
L23R18	47291	116
Average error rate	0.002652	

3.7   Mutation spectrum of Pt-GG

To determine what nucleotides were incorporated across from Pt-GG, I used the traditional TLS assay and the high-throughput TLS assay. The most abundant mutation arising from Pt-GG lesion was CC->AC substitution. In the traditional TLS assay, most of the mutations were CC->AC substitutions: 19 out of 25 sequences that have mutations had A mis-inserted at the first insertion step to the DNA lesion, the other 6 sequences have mutation at the second insertion step (see Table C-2, Figure C-1 and C-2 in Appendix C). In the high-throughput TLS assay, among the 21,520 reads that have mutations across from the Pt-GG lesion, 50% are CC->AC substitutions, 12% are CC->TC substitutions (Figure 11). Interestingly, in the control plasmids among the 38,212 reads that have mutations, close to 70% have a single nucleotide deletion at one of the two 'G' positions; while in the lesion plasmids, only 15% of the reads with mutations have a single nucleotide deletion (Figure 11). Compare the mutation spectrum of Pt-GG to that of GG, there are large increases in the proportions of reads with CC->AC, AT, AA substitutions, small increases in the proportions of reads with CC->TC, CT, CA substitutions (Figure 11) and a small decrease in the proportion of reads with double base deletions.
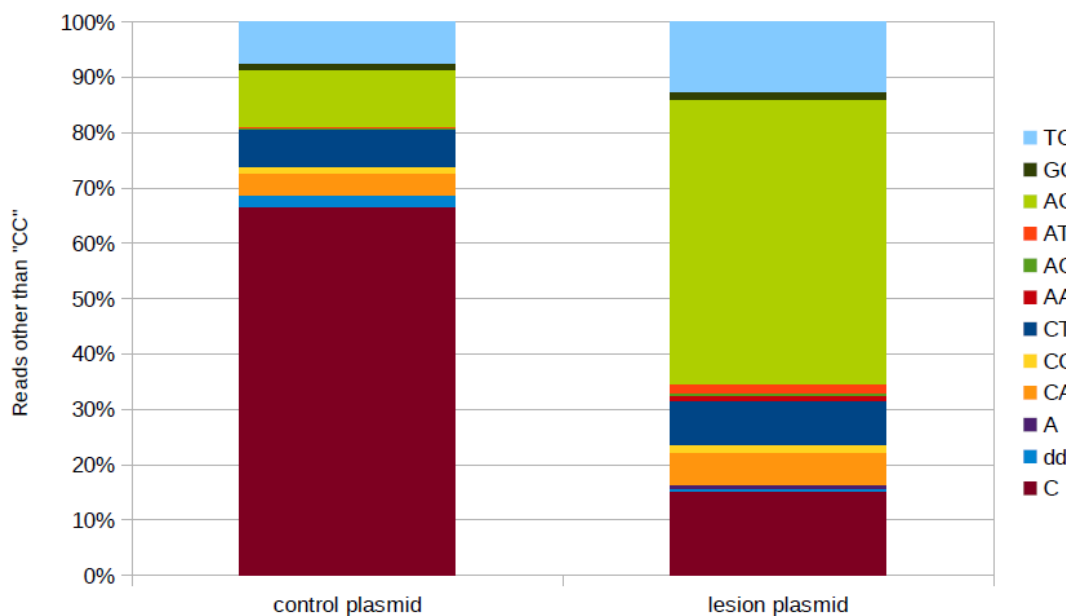
Figure 11 Mutation spectrums across from Pt-GG lesion and GG dinucleotides in mammalian cells

Mutations across from GG dinucleotides in the control plasmids and Pt-GG in the lesion plasmids were measured by the high-throughput TLS assay. The mutation spectrums are derived from 38,212 reads of the control plasmids and 21,520 reads of the lesion plasmids. In the legend, both 'A' and 'C' denote a single nucleotide deletion at one of the two nucleotide positions of GG or Pt-GG. 'dd' stands for double base deletion. 70% of the mutations found opposite to GG in the control plasmids are a single nucleotide deletion. 65% of the mutations caused by Pt-GG are mis-incorporations of A at the first insertion step with or without a mutation at the second insertion step. Incorporations of two purine bases to the Pt-GG lesion are rare; CC->AG and CC->AA substitutions across from the lesion are much less favored compared to other types of mutations and CC->GA and CC->GG substitutions are not found.

Despite their similar mutation frequencies, Pt-GG and the regular GG dinucleotides have different mutation patterns. Principal component analysis of the NGS data shows that the mutation spectrum of GG in the control plasmids and that of Pt-GG in the lesion plasmids are separable (Figure 12A). The length of a composition vector in Figure 12B indicates the relative proportion of change in a variable. For example, compared to the GG

dinucleotides, Pt-GG causes a large decrease in the frequency of deletions (both a single nucleotide deletion and a double base deletion) and large increases in the frequencies of mutations CC->AC, AT, AA and AG. The variables AC, AT and AA that form a cluster in Figure 12B could be interpreted as being produced by the same molecular mechanism. In other words, the mutations CC->AC, AT and AA could be produced by the same DNA polymerase or DNA polymerases recruited by a common pathway.

Figure 12 Biplot generated from principal component analysis of the Pt-GG data

Principal component analysis (PCA) reveals the underlying structure of a compositional multivariate dataset by reducing its complexity. PCA highlights those variables that explain the most variance in the data. Principal Component 1 of the biplot, the x-axis explains 43±1.52% of the total variance in the data and Principal Component 2, the y-axis explains 10±0.73% of the total variance in the data. A) Each point represents data from a transfection experiment. Mutation profiles opposite to 'GG' (blue dots) are separable from those opposite to Pt-GG (red dots). This indicates that constructions of the control plasmids and the lesion plasmids for the Pt-GG study were successful. B) The same biplot as shown in A) with composition vectors only. Changes in the variables 'C' and 'dd' of the control plasmids are correlated to changes in the variables 'AC', 'AT', 'AA' and 'AG' of the lesion plasmids. The AC, AT, AA variants are highly correlated because the distance between the tips of the rays is small. 'Clr.' stands for centered log-ratio transformation. 'dd' stands for double deletion mutation. The biplots were generated by the CoDaPack software (Comas-Cufi and Thió-Henestrosa, 2011).

3.8   The role of Pol eta by itself in the bypass of Pt-GG *in vivo*

Pt-GG is a major type of cisplatin-DNA adducts found in cancer patients treated with the drug cisplatin (Cepeda et al., 2007; Fichtinger-Schepman et al., 1987; Jung and Lippard, 2007). Pt-GG is considered a bulky DNA lesion because it involves crosslinking of two neighboring purines. Due to its large size Pt-GG is thought to be highly blocking to DNA replication and is thought to be largely responsible for the drug action of cisplatin (Vaisman et al., 2000; Chijiwa et al., 2010). DNA polymerases with the abilities to bypass Pt-GG are undesirable in chemotherapy, because their up-regulations are believed to increase drug resistance to cisplatin. My aim is to identify translesion DNA polymerases that can bypass the Pt-GG lesion because presumably, inactivating these DNA polymerases could benefit cancer treatment.

Recently, it has been shown that Pol eta correctly bypasses Pt-GG *in vitro* (Ummat et al., 2012; Zhao et al., 2012). To investigate whether Pol eta has a similar function in mammalian cells, I used the traditional TLS assay and the high-throughput TLS assay to deduce the biological function of Pol eta. Gap plasmids containing a site-specific Pt-GG lesion were co-transfected with control gap plasmids without the lesion into a pair of isogenic MEF cell lines with the genotypes *POLH*(+/+) and *POLH*(-/-), respectively.

In the traditional TLS assay, the mutation probability of Pt-GG was 4 out of 29 sequences analyzed (14%) in the MEF *POLH*(+/+) cell line and 10 out of 28 sequences analyzed (36%) in the MEF *POLH*(-/-) cell line. Notably, the incidence of C->A substitution at the first insertion step increased by three times when the *POLH* gene was knocked out and that at the second insertion step was unaffected (Appendix C, Figures C-6). This suggests that Pol eta correctly bypassed Pt-GG at the first insertion step (P-value=0.0411, Fisher's Exact Test, α=0.05).

In the high-throughput TLS assay, the mutation frequency of Pt-GG was the same in the MEF *POLH*(+/+) cell line and in the MEF *POLH*(-/-) cell line, so Pol eta's effect on bypassing Pt-GG was not detected. Using the double and triple knockout cell lines, however, it could be inferred that Pol eta and Pol iota together promoted CC->AC and

CC->AT substitutions across from the lesion (Section 3.11), so Pol eta bypassed Pt-GG in an error-prone manner *in vivo*.

The function of Pol eta deduced from the traditional TLS assay was supportive of recent structural reports (Ummat et al., 2012; Zhao et al., 2012), but it was not confirmed by the high-throughput TLS assay. However, an earlier structural study has shown that Pol eta incorporates the correct dCTP at the first insertion step to Pt-GG but it cannot incorporate any nucleotides in the second insertion step to the lesion *in vitro*; therefore, Pol eta cannot carry out translesion synthesis of Pt-GG (Chijiwa et al., 2010). This conclusion is consistent with my finding of Pol eta's cellular function in the high-throughput TLS assay.

3.9   The role of Pol iota by itself in the bypass of Pt-GG *in vivo*

The *POLI* gene arises from an ancestral gene duplication of the *POLH* gene. To determine if Pol iota could bypass Pt-GG *in vivo*, I used the traditional TLS assay and the high-throughput TLS assay to deduce the biological function of Pol iota in mammalian cells. Lesion gap plasmids containing a Pt-GG lesion were co-transfected with control gap plasmids without the lesion into a pair of isogenic MEF cell lines with the genotypes *POLI*(+/+) and *POLI*(-/-), respectively.

In the traditional TLS assay, the mutation probability of Pt-GG was 7 out of 29 sequences analyzed (24%) in the MEF *POLI*(+/+) cell line and 4 out of 24 sequences analyzed (16.5%) in the MEF *POLI*(-/-) cell line. Knocking out the *POLI* gene did not affect the probability of mutation at the first insertion step (P-value=1, Fisher's Exact Test, α=0.05), nor that at the second insertion step (P-value=0.49, Fisher's Exact Test, α=0.05). Therefore, Pol iota did not have any effect on bypassing the Pt-GG lesion.

In the high-throughput TLS assay, I also could not detect the effect of Pol iota between the MEF *POLI*(+/+) cell line and the MEF *POLI*(-/-) cell line.

In the traditional TLS assay, while Pol eta inserted the correct C at the first insertion step, Pol iota had no effect. In the high-throughput assay, I could not detect the biological function of Pol iota nor that of Pol eta by comparing their wild type cell line to their respective isogenic single knockout cell line. In Section 3.11, I detected effects of knocking out both the *POLI* gene and the *POLH* gene using double- and triple-knockout MEF cell lines (Figure 15).

3.10   Rev3L and Rev1 bypassed Pt-GG in mammalian cells

Rev3L is the catalytic subunit of Pol zeta, a B-family DNA polymerase. It has been shown that Rev3L and Rev1 functionally interact in translesion synthesis with Rev1 acting as a scaffold protein to recruit Rev3L via an interaction with Rev7, the accessory subunit of Pol zeta (Murakumo et al., 2001; Kikuchi et al., 2012). Knocking down *REV3L* or *REV1* sensitizes mammalian cells to cisplatin (Doles et al., 2010; Xie et al., 2010). To determine if Rev3L and Rev1 played any roles in the bypass of the Pt-GG lesion in mammalian cells, I used the high-throughput TLS assay.

The high-throughput TLS assay showed that Rev3L and Rev1 bypassed Pt-GG *in vivo* by producing mutations. Lesion gap plasmids containing a Pt-GG lesion were co-transfected with control gap plasmids without the lesion into pairs of isogenic MEF cell lines. To deduce the cellular function of Rev3L, the frequencies of CC->AC, CC->TC, CC->AA and CC->AT substitutions across from the Pt-GG lesion were compared between the 3(+)6 *REV3L*(+/lox) Cre cell line and each of the 4(-)5 and 4(-)11 cell lines with the genotype *REV3L*(-/lox) Cre, and between the B2 *REV3L*(+/+) *P53*(-/-) cell line and each of the B4(-)9 and B4(-)18 cell lines with the genotype *REV3L*(-/-) *P53*(-/-). To deduce the function of Rev1, mutation frequencies of Pt-GG of the MEF *REV1*(+/+) cell line were compared to those of the MEF *REV1*(-/-) cell line. Nine transfection experiments were set up for each cell line.

Gap regions of the control plasmids and lesion plasmids recovered from mammalian cells were amplified in separate PCR reactions and sequenced by Next Generation Sequencing.

In data analysis, the within-sample ratios AC/CC, TC/CC, AT/CC and AA/CC were taken and then log 2 transformed. Each of the ratios log2(AC/CC), log2(TC/CC), log2(AT/CC), log2(AA/CC) of a cell line (with nine biological replicates) was compared to that of the other cell line (also with nine biological replicates) by the Wilcoxon Rank Sum Test. Wilcox Rank Sum Test evaluates the likelihood that the medians of the ratios from the two cell lines are from two different populations. P-values calculated by Wilcox Rank Sum Tests were adjusted by the Benjamini-Hochberg method to control the false discovery rate of the study. Where the adjusted P-value was less than 0.05, meaning that the probability of falsely detecting an effect for the study is 5% at most, the difference between the cell lines was deemed statistically significant and the effect size (Cohen's d) and its 95% confidence interval were calculated. An effect size of 0.6 or greater is considered large by convention.

The Pt-GG study and bypass of the lesion by five translesion DNA polymerases took up 86% of the total sequencing capacity. Numbers of reads (for the control and for the lesion plasmids) obtained for every transfection experiment are shown in Figure 13.
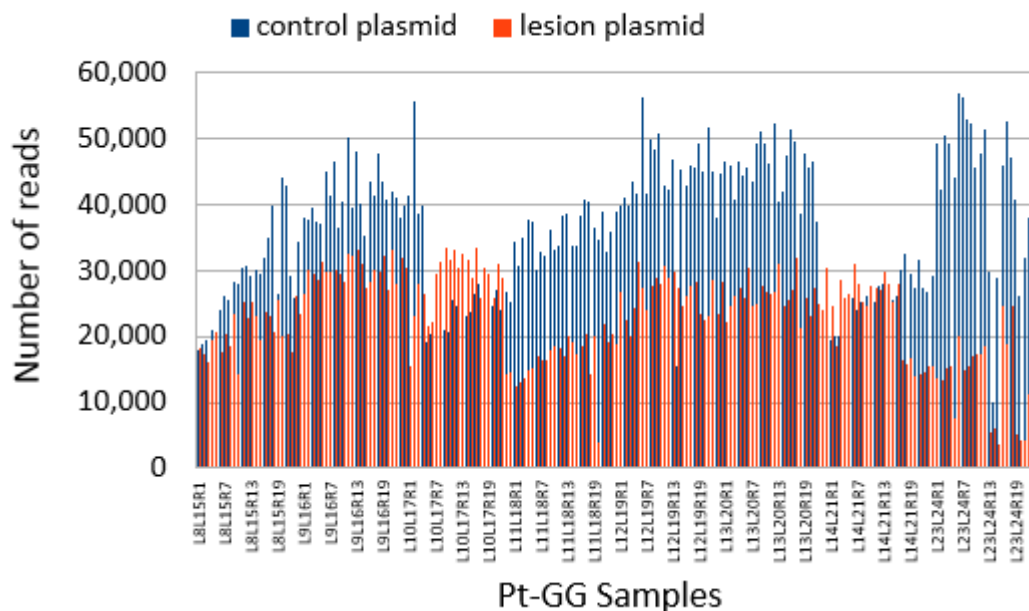
Figure 13 Reads obtained for every transfection experiment in the Pt-GG study

Number of reads for the control plasmid and number of reads for the lesion plasmid for every transfection experiment in the Pt-GG study (see Appendix I for raw data). The control plasmids and the lesion plasmids of a transfection experiment were amplified in separate PCR reactions for a total of 55 cycles and their reads were sorted by barcodes. There are 192 transfection experiments, including nine biological replicates for each of the twenty cell lines listed in Table 4 and positive and negative control reactions done in triplicates. Positive control PCR reactions used the control plasmid without the gap or the lesion plasmid without the lesion and without the gap as DNA template. There are two types of negative control experiments: in the first type, no transfection reagent was added, so gap plasmids did not enter into mammalian cells. The purpose of this negative control experiment is to check that the gap plasmids were effectively removed by alkaline lysis. In the second type of negative control, no DNA template was added in the PCR reactions.

The mutation frequency of Pt-GG was in the order of 10^{-3} and was similar to that of a regular GG dinucleotides. Rev3L and Rev1 had no effects on bypassing GG in the control plasmids but produced mutations opposite to Pt-GG in the lesion plasmids where 50% of the mutations were CC->AC substitutions. The effect of Rev3L was validated in more than one pair of cell lines. Knocking out *REV3L* or *REV1* resulted in less frequent CC->AC substitutions (Figure 14). In addition to CC->AC substitutions, knocking out

REV3L in a *P53* null genetic background resulted in less frequent CC->TC substitutions (Table 9). The same results were obtained from NGS data sequenced in the reverse direction (Appendix B).

Genotypes	Cell Lines (N=9 groups)	BH-adjusted P value	Less frequent CC->AC substitution	Effect of Knocking out
Rev3L(+/Δ) vs. Rev3L(-/Δ)	3(+)6 vs. 4(−)5	0.002		Rev3L
	3(+)6 vs. 4(−)11	0.002		
Rev3L(+/+) vs. Rev3L(-/-) P53(-/-) P53(-/-)	B2 vs. B4(−)9	0.015		Rev3L
	B2 vs. B4(−)18	0.002		
	Rev1 wild type vs. Rev1−/−	0.002		Rev1

cohen's d

null hypothesis (no effect)
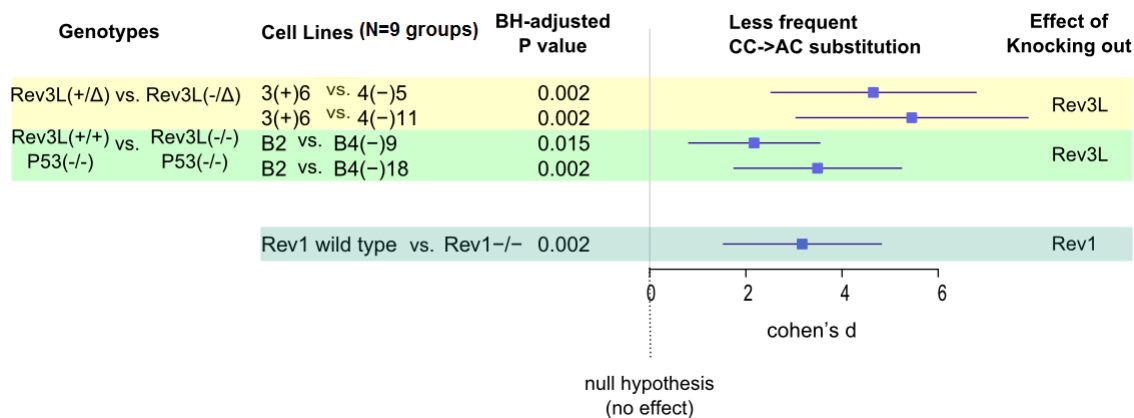
Figure 14   Knocking out *REV3L* or *REV1* resulted in less frequent CC->AC substitutions across from Pt-GG

Knocking out the *REV3L* gene conditionally or in a *P53* null genetic background resulted in less frequent CC->AC substitutions across from Pt-GG; each result was confirmed in two pairs of isogenic cell lines. Knocking out the *REV1* gene produced a similar effect as knocking out the *REV3L* gene. Effect sizes and Benjamini Hochberg (BH) adjusted P-values are shown for the comparisons between isogenic cell lines.

Table 9   Knocking out *REV3L* resulted in less frequent substitutions of CC->TC across from Pt-GG

Genotypes of cell lines compared (N=9 groups)	Adjusted P-value	Cohen's d	95% C. I.
REV3L(+/+) *P53*(-/-) versus *REV3L*(-/-) *P53*(-/-) Nicknames: B2 versus B4(-)18	0.0018	2.28	0.88-3.67

There is strong support for Rev3L functioning as a major DNA polymerase to bypass Pt-GG in mammalian cells. It has been reported that cell lines lacking a functional *REV3L* gene are hypersensitive to cisplatin treatment (Zander and Bemark, 2004). A two-polymerase model based on protein-protein interaction studies suggested that the complete bypass of Pt-GG requires Rev3L to act as an extender DNA polymerase, preceded by another DNA polymerase inserting a nucleotide at the first insertion step. Here I show that knocking out *REV3L* caused either A or T to be mis-incorporated at the first insertion step to a Pt-GG lesion, hence Rev3L can function on its own to bypass the lesion.

Knocking out *REV1* produced a similar effect as knocking out *REV3L*. Rev1 may bypass Pt-GG indirectly by acting as a scaffold protein to recruit Rev3L or other translesion DNA polymerases. If REV1 and REV3L proteins require each other to function properly in translesion synthesis, disrupting either the *REV1* or the *REV3L* gene could entail a completely different DNA polymerase to bypass Pt-GG. The ability of Rev1 to bypass Pt-GG directly has not been studied *in vitro*.

3.11 Each of Pol kappa and Pol eta associated with Pol iota in the bypass of Pt-GG *in vivo*

To determine the effects of Pol kappa, Pol iota and Pol eta on the bypass of the Pt-GG lesion, I used the high-throughput TLS assay. Although recently structural explanations were provided for the correct bypass of Pt-GG at the first insertion step by Pol eta (Ummat et al., 2012; Zhao et al., 2012), I did not find any effects of Pol eta, at least not by itself, on bypassing the DNA lesion *in vivo*. My results show that Pol kappa, rather than Pol eta bypassed Pt-GG and the bypass was error-prone (Figure 15).

To deduce the function of Pol kappa, isogenic cell lines with and without the *POLK* gene were compared. The *POLK* wild type cell line was compared to the *POLK*(-/-) single knockout cell line; the *POLI*(-/-) cell line was compared to the *POLK*(-/-) *POLI*(-/-) double knockout cell line; the *POLH*(-/-) *POLI*(-/-) cell line (with functional *POLK* gene) was compared to each of the *POLH*(-/-) *POLI*(-/-) *POLK*(-/-) triple knockout cell lines

2091 and 2095, respectively (Figure 15). Although *in vitro* studies have shown that Pol kappa cannot bypass a Pt-GG adduct in DNA template (Gerlach et al., 2001), work done by Vikash Jha in our lab showed that Pol kappa had some activities in bypassing the lesion *in vitro* (unpublished data). The ability of Pol kappa to bypass Pt-GG was functionally redundant to that of Pol iota, because the effect of knocking out Pol kappa was only detected in the absence of Pol iota, and vice versa (Figure 15). Knocking out Pol kappa and Pol iota together resulted in less frequent CC->AC and CC->AA substitutions across from the lesion (Figure 15, Table 10).

Table 10   Knocking out *POLK* and *POLI* together resulted in less frequent substitutions of CC->AA across from Pt-GG

Genotypes of cell lines compared (N=9 groups)	Adjusted P-value	Cohen's d	95% C. I.
POLK wild type vs. *POLK*(-/-) *POLI*(-/-) DKO	0.038	2.13	0.77-3.49
POLI wild type vs. *POLK*(-/-) *POLI*(-/-) DKO	0.0018	2.27	0.88-3.66
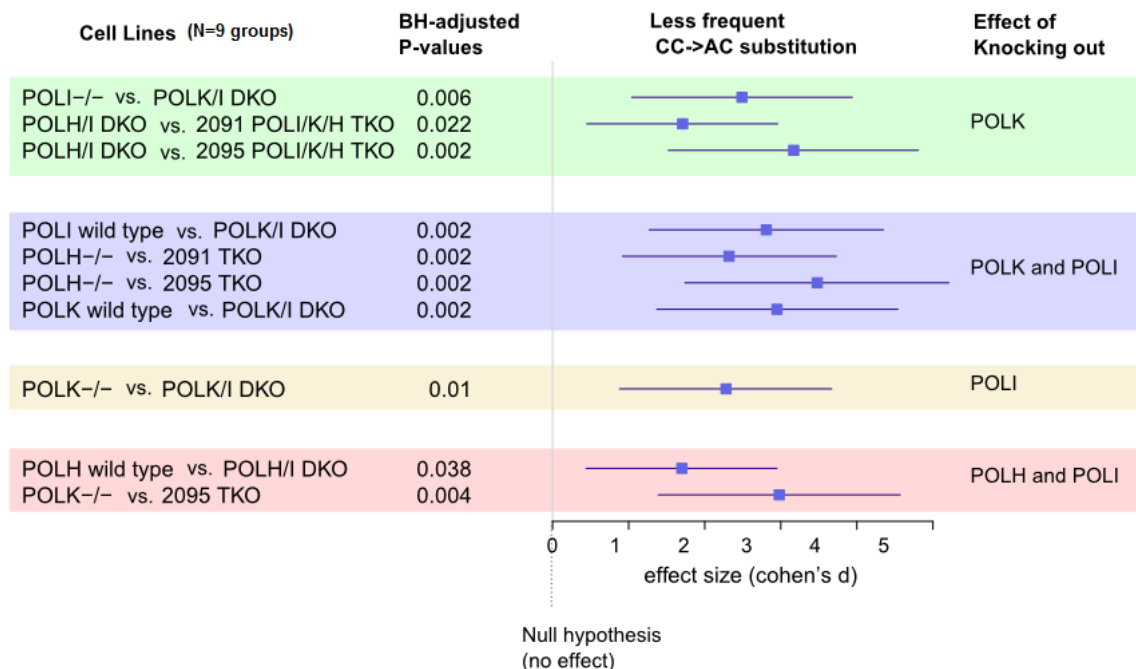
Figure 15 Knocking out *POLK* in the absence of *POLI* or knocking out *POLH* in the absence of *POLI* resulted in less frequent CC->AC substitutions across from Pt-GG

While the effect of Pol kappa could not be detected between the *POLK*(+/+) wild type cell line and the *POLK*(-/-) single knockout cell line, knocking out the *POLK* gene in the absence of the *POLI* gene or vice versa resulted in less frequent CC->AC substitutions. Knocking out the *POLK* gene and the *POLI* gene together produced less frequent CC->AC substitutions. Knocking out the *POLH* gene and the *POLI* gene together also resulted in less frequent CC->AC substitutions. The function of Pol iota was associated with those of Pol kappa and Pol eta in the bypass of Pt-GG.

To deduce the function of Pol eta, isogenic MEF cell lines with and without the *POLH* gene were compared. The *POLH*(+/+) cell line was compared to the *POLH*(-/-) single knockout cell line and to the *POLH*(-/-) *POLI*(-/-) double knockout cell line, respectively. The *POLK*(-/-) cell line (with functional *POLH* and *POLI* genes) was compared to each of the *POLK*(-/-) *POLI*(-/-) *POLH*(-/-) triple knockout cell lines 2091 and 2095, respectively. My result shows that Pol eta and Pol iota together bypassed Pt-GG and made CC->AC and CC->AT substitutions (Figure 15, Table 11), suggesting that the two DNA polymerases cooperated to bypass the Pt-GG lesion. It has been shown that

localization of Pol iota to replication foci is largely dependent on Pol eta and that the C-terminal 224 amino acids of Pol iota are sufficient for binding to Pol eta (Kannouche et al., 2003). It has also been reported that physical and functional interactions of Pol eta and Pol iota are mediated by ubiquitin and are dependent on ubiquitination of either of the DNA polymerases (McIntyre et al., 2013).

Table 11   Knocking out *POLH* and *POLI* together resulted in less frequent substitutions of CC->AT across from Pt-GG

Genotypes of cell lines compared (N=9 groups)	Adjusted P-value	Cohen's d	95% C. I.
POLH wild type vs. *POLH*(-/-) *POLI*(-/-) DKO	0.038	1.85	0.56-3.14

Despite the low mutation rate of Pt-GG, which was similar to that of a regular GG dinucleotides, several translesion DNA polymerases could be recruited, possibly via different pathways to bypass the lesion in mammalian cells. I do not know yet which DNA polymerase(s) correctly bypassed Pt-GG in the cell, but it would be interesting to find out.

Since Pt-GG is implicated in blocking DNA replication (Vaisman et al., 2000; Chijiwa et al., 2010; Zhao et al., 2012), I would like to determine the bypass efficiency of the DNA lesion. In the next section, I present results of the traditional TLS assay, Q-PCR and the high-throughput TLS assay for studying translesion synthesis efficiencies of Pt-GG and 8-oxo-G and in other words, measuring the blocking effects of the DNA lesions.

3.12 Pt-GG and 8-oxo-G had similar blocking effects to DNA replication

In addition to mutagenesis, DNA lesions are replicated less efficiently than regular nucleotides. The delay in replicating a DNA lesion relative to a normal nucleotide allows recruitment of translesion DNA polymerases to bypass damaged DNA. I speculated that the blocking effect of a DNA lesion is directly related to its size. Pt-GG is believed to be highly blocking to DNA replication because it involves crosslinking of two adjacent

purine bases (Vaisman et al., 2000; Chijiwa et al., 2010); by contrast, 8-oxo-G is believed to be less obstructive to DNA replication. (Shibutani et al., 1991; Einolf and Guengerich, 2001; Avkin and Livneh, 2002; Markkanen et al., 2012) I hypothesized that Pt-GG has a larger blocking effect than 8-oxo-G *in vivo*.

Studying the blocking effect of a DNA lesion is important because the difference in the bypass efficiency of Pt-GG and 8-oxo-G could entail different mechanisms of recruiting translesion DNA polymerases. The widespread presence of Pt-GG in patients who are treated with cisplatin is thought to kill tumor cells by preventing DNA replication.

To investigate translesion synthesis efficiencies of 8-oxo-G and Pt-GG in mammalian cells, I first used the traditional TLS assay and Q-PCR to study bypass efficiency of 8-oxo-G in four DNA samples. There I found a discrepancy in the results of the two methods (Table 12 and Table 13). Then I used Q-PCR to study the bypass efficiencies of 8-oxo-G and Pt-GG, respectively, in nine MEF cell lines and validated the results using data of Next Generation Sequencing (Table 14 and Table 15).

In the traditional TLS assay, the translesion synthesis efficiency is estimated by taking the ratio of the number of bacterial colonies arising on a kanamycin plate (selecting for the lesion plasmids), to that arising on a chloramphenicol plate (selecting for the control plasmids). Four DNA samples A, B, C, D extracted from the four cell lines MEF *POLI*(+/+), MEF *POLI*(-/-), BL2 *POLI*(+/+), and BL2 *POLI*(-/-), respectively, were transformed into CSH117 electrocompetent cells in triplicates. Following each transformation, undiluted cells were plated in triplicates on kanamycin plates and on chloramphenicol plates, and the number of bacterial cells arising on the plates were counted. The Kan/Cm ratio was calculated using average numbers of bacterial colonies from the three kanamycin plates (Kan) and from the three chloramphenicol plates (Cm). Results of the translesion synthesis efficiencies of 8-oxo-G measured from the four DNA samples with the traditional TLS assay are shown in Table 12.

Using the same four DNA samples as above, I measured the number of copies of lesion plasmids and that of the control plasmids by Q-PCR and then calculated the Kan/Cm ratio for each DNA sample (Table 13). DNA standards and unknown samples were

measured in replicates. Q-PCR reactions were set up using 2x EvaGreen qPCR Master Mix and ran on the Eppendorf RealPlex Mastercycler instrument.

Table 12 Bypass efficiencies of 8-oxo-G estimated by the traditional TLS assay
Gap plasmids transfected into mammalian cells were repaired and then recovered by the Alkaline Lysis method. The DNA was transformed, in three independently transformation experiments into bacteria (CSH117 strain). In the table are numbers of bacterial colonies arising on three kanamycin (Kan) plates and on three chloramphenicol (Cm) plates when undiluted bacterial cells from each transformation experiment were spread. The bypass efficiency of 8-oxo-G (Kan/Cm) is calculated as the ratio of the average number of bacterial colonies on the Kan plates to that on the Cm plates.

		A	B	C	D
	Plate	3LG188.1 MEF *POLI*(+/+)	3LG188.2 MEF *POLI*(-/-)	3LG102.5 BL2 *POLI*(+/+)	3LG102.6 BL2 *POLI*(-/-)
The 1st transformation experiment	Kan	22, 32, 24	26, 28, 24		
	Cm	18, 20, 18	12, 26, 18		
	Kan/Cm	1.42	1.37		
The 2nd transformation experiment	Kan	54, 70, 55	52, 69, 171	26, 24, 20	10, 5, 8
	Cm	25, 24, 38	30, 22, 24	24, 12, 28	4, 5, 8
	Kan/Cm	2.07	2.34	1.09	1.33
The 3rd transformation experiment	Kan	30, 43, 29	79, 86, 86	49, 36, 35	25, 37, 23
	Cm	28, 35, 27	62, 53, 63	41, 24, 49	19, 26, 19
	Kan/Cm	1.13	1.4	1.05	1.32

Table 13 Bypass efficiencies of 8-oxo-G of the same DNA samples as in Table 12 measured by Q-PCR

Gap plasmids transfected into mammalian cells were repaired and then recovered by the Alkaline Lysis method that effectively eliminated any unrepaired plasmids. From the DNA sample, the number of copies of control plasmids and the number of copies of lesion plasmids were measured by quantitative PCR using standard curves. Ct is the cycle threshold number set above the background in a Q-PCR reaction; it is used to determine the concentration or the copy number of the plasmids.

	A		B		C		D	
	3LG188.1 MEF *POLI*(+/+)		3LG188.2 MEF *POLI*(-/-)		3LG102.5 BL2 *POLI*(+/+)		3LG102.6 BL2 *POLI*(-/-)	
replicates	1	2	1	2	1	2	1	2
Ct value, p-Kan	10.74	10.54	14.95	15.19	11.59	11.6	11.75	11.72
Ct value, p-Cm	11.93	11.55	17.46	17.56	12.96	12.86	12.65	13
Copy no. p-Kan	4.55E+8	5.34E+8	1.54E+7	1.27E+7	2.30E+8	2.28E+8	2.02E+8	2.07E+8
Copy no. p-Cm	3.57E+9	4.59E+9	9.39E+7	8.79E+7	1.81E+9	1.94E+9	2.22E+9	1.77E+9
Kan/Cm ratio	0.13	0.12	0.16	0.14	0.13	0.12	0.09	0.12

The bypass efficiencies of 8-oxo-G measured by the traditional TLS assay exceed one and show large variations among repeat transformation experiments (

Table 12). Given that the lesion gap plasmids and the control gap plasmids were co-transfected in equal molar amounts, a Kan/Cm ratio greater than one would indicate that the lesion plasmids were repaired more efficiently than the control plasmids and that 8-oxo-G was replicated more rapidly than a regular G—a result contrary to my expectation. Using Q-PCR, however, I found that for the same DNA samples the bypass efficiencies of 8-oxo-G were around 0.12 (Table 13), which indicates that for every ten control gap plasmids that were filled, one lesion gap plasmid was repaired, or that 8-oxo-G was bypassed at about one tenth the rate of a regular nucleotide in cells. Results of the Q-PCR

experiments are more reproducible. I believe Q-PCR is a more reliable and accurate method than the traditional TLS assay in estimating a DNA lesion's bypass efficiency.

Next, I used Q-PCR to determine the translesion synthesis efficiencies of 8-oxo-G and Pt-GG, respectively, from nine MEF cell lines, and validated the results using the high-throughput data. I wanted to know how the bypass efficiencies of 8-oxo-G and Pt-GG differ. Furthermore, I asked if the translesion synthesis efficiencies of the lesions were influenced by knocking out one, two or three of the DNA polymerase genes *POLI*, *POLK* and *POLH*. If Pol iota or Pol kappa or Pol eta was the only DNA polymerase in cells able to bypass either of the DNA lesion, inactivating the DNA polymerase should greatly reduce the bypass efficiency for that DNA lesion.

Q-PCR reactions were set up on a 384-well plate using Syber Select Master Mix and ran in the ViiA7 instrument as described in Methods. The four DNA samples measured previously were not used in this experiment. For each cell line, samples of three transfection experiments were measured (Table 14) and three technical replicates were set up to measure each sample. Translesion synthesis efficiencies of the same samples in Table 14 were also determined and validated using results of the high-throughput TLS assay (Table 15), based on the ratio of the number of reads obtained for the lesion plasmids to that obtained for the control plasmids for a given transfection experiment (Table 15).

Table 14 Bypass efficiencies of 8-oxo-G and Pt-GG in nine MEF cell lines measured by Q-PCR

Each Kan/Cm ratio in the table is the number of copies of lesion plasmids to that of control plasmids for a transfection experiment and the number shown is the average of three technical replicates.

Genotypes of MEF cell lines	8-oxo-G Kan/Cm ratios of three biological replicates			Pt-GG Kan/Cm ratios of three biological replicates		
	1	2	3	1	2	3
POLI wild type	0.68	0.62	0.9	0.8	0.7	0.67
POLI(-/-)	2.34	0.9	0.83	0.45	0.68	0.63
POLH wild type	0.81	0.65	0.65	0.57	0.53	0.51
POLH(-/-)	0.53	0.67	0.68	0.56	0.53	0.49
POLK wild type	0.52	0.46	0.50	0.72	0.61	0.57
POLK(-/-)	0.65	1.20	0.51	0.68	0.85	0.69
POLH/I DKO	0.54	0.71	0.60	0.56	0.56	0.64
POLK/I DKO	0.58	0.59	0.54	0.57	0.53	0.68
2091 *POLH/K/I* TKO	0.74	0.83	0.72	1.20	0.86	0.85

Table 15 Bypass efficiencies of 8-oxo-G and Pt-GG of the same samples as in Table 14 determined from the NGS data

Each number is the ratio of the number of reads obtained for the lesion plasmids to that obtained for the control plasmids for a transfection experiment.

Genotypes of MEF cell lines	8-oxo-G Kan/Cm ratios of three biological replicates			Pt-GG Kan/Cm ratios of three biological replicates		
	1	2	3	1	2	3
POLI wild type	0.54	0.61	0.61	1.09	0.95	0.90
POLI(-/-)	0.75	0.59	0.55	0.54	0.90	0.79
POLH wild type	0.67	0.63	0.59	0.74	0.80	0.83
POLH(-/-)	0.58	0.59	0.57	0.74	0.91	0.82
POLK wild type	0.47	0.47	0.42	1.83	1.61	1.73
POLK(-/-)	0.59	0.57	0.65	1.37	1.02	1.21
POLH/I DKO	0.56	0.59	0.55	0.60	0.11	0.61
POLK/I DKO	0.55	0.59	0.60	0.69	0.66	0.67
2091 *POLH/K/I* TKO	0.60	0.65	0.66	0.60	0.80	0.52

Translesion synthesis efficiencies of 8-oxo-G and Pt-GG determined by Q-PCR and by the high-throughput TLS assay are in good agreement (Table 14 and Table 15) despite a few outliers with values greater than 1. All the data considered there is no statistically significant difference between the bypass efficiencies of 8-oxo-G and Pt-GG; the average Kan/Cm ratio of either of the two DNA lesions is about 0.6. Knocking out the genes *POLI*, *POLK* and *POLH* did not have any effects on the bypass efficiency of 8-oxo-G or the bypass efficiency of Pt-GG.

Even though Pt-GG has been shown to be highly blocking to DNA replication under well-defined *in vitro* conditions, it was not the case *in vivo*. My results show that the blocking effect of Pt-GG was the same as that of 8-oxo-G in mammalian cells. It can be reasoned that DNA polymerases accommodated a single base of Pt-GG rather than the

cross-linked dimer in catalysis. Since the genotypes of the cell lines did not have any effects on the TLS efficiencies of either of the two DNA lesions, the functions of Pol iota, Pol kappa and Pol eta were not obligatory for the bypass of 8-oxo-G or Pt-GG in mammalian cells. Other DNA polymerases could fill the gap plasmids and bypass the DNA lesions.

Even though Pt-GG presents itself as a major roadblock for certain DNA polymerases *in vitro*, cells have capabilities to bypass this lesion efficiently because multiple DNA polymerases could carry out this function *in vivo*. Pt-GG hindered DNA replication, but surprisingly, its blocking effect was comparable to that of 8-oxo-G and was not as severe as previously implicated. This suggests that the presence of Pt-GG in DNA is not a major contributing factor to the anti-tumor property of cisplatin. On the other hand, 8-oxo-G was replicated less efficiently compared to a regular G nucleotide and the lesion's blocking effect was greater than previously considered. The finding that 8-oxo-G and Pt-GG lesions were bypassed with similar efficiencies *in vivo* was not expected judging by the relative size of the two lesions.

Chapter 4

4   Discussions

4.1   Traditional TLS assay and high-throughput TLS assay revealed totally different results

Mutation frequencies of 8-oxo-G and Pt-GG were 37% (64/174 sequences) and 22% (25/110 sequences), respectively, in MEF cells estimated by the traditional TLS assay, and 91% (254,481/279,650 reads) and less than 1% (21,520/4,491,510 reads), respectively, in MEF cells estimated by the high-throughput TLS assay. Whereas the traditional TLS assay was inconclusive about the role of Pol iota in bypassing 8-oxo-G, the high-throughput TLS assay showed that Pol iota did not have an effect. Whereas the traditional TLS assay showed that Pol eta correctly bypassed Pt-GG in the first insertion step and that Pol iota had no effect in bypassing the DNA lesion, the high-throughput TLS assay showed that neither Pol eta nor Pol iota had any effects, but that Pol kappa bypassed Pt-GG and the bypass was error-prone. Results of the traditional TLS assay show that 8-oxo-G was replicated more efficiently than a regular G nucleotide, which is contrary to my expectation. Results of the high-throughput TLS assay on the other hand, show that the bypass efficiency of 8-oxo-G and that of Pt-GG were about 0.6 relative to regular nucleotides measured by Q-PCR and Next Generation Sequencing. This means that each of 8-oxo-G and Pt-GG was replicated at 0.6 times the rate of regular nucleotides. I believe the differences in results between the traditional TLS assay and the high-throughput TLS assay are due to the use of *E. coli* in the traditional TLS assay. The high-throughput TLS assay is better able to capture mutations made inside mammalian cells caused by DNA lesions.

The high-throughput TLS assay produced more consistent results across transfection experiments than the traditional TLS assay. Next Generation Sequencing generated a much larger sample size than the traditional TLS assay for each transfection experiment. The traditional TLS assay revealed that the mutation frequency of 8-oxo-G was 33% in both the MEF *POLI*(+/+) cell line (33/89 sequences) and the MEF *POLI*(-/-) cell line (31/85 sequences), and was 18% in the BL2 *POLI*(+/+) cell line (17/93 sequences)and

27% in the BL2 *POLI*(-/-) cell line (24/89 sequences). The high-throughput TLS assay showed that the mutation frequency of 8-oxo-G was 91% in all four of the cell lines where thousands of sequences were analyzed for each cell line. The difference in mutation frequency of a DNA lesion was very subtle in the high-throughput TLS assay across all the cell lines, less than 1% for sure. As a result, the high-throughput data must be analyzed differently from the data of the traditional TLS assay. The high-throughput data were treated as compositional and log-ratio analyses were used to detect differences in mutation frequencies of a DNA lesion in pairs of isogenic cell lines.

The use of *E. coli* in the traditional TLS assay not only had impact on the mutation frequencies of 8-oxo-G and Pt-GG, but also resulted in additional mutations arising in the gap region close to the DNA lesion (Appendix C). These mutations were not observed in the high-throughput TLS assay so they must have arisen in bacteria. In the high-throughput assay, a mutation was observed only where across from a DNA lesion. Observations from the traditional TLS assay demonstrate that a damaged nucleotide in a DNA sequence could, over the course of many rounds of DNA replications and cell divisions in bacteria, cause mutations of its neighboring bases. Not realizing that mutations in the traditional TLS assay result from multiple events, observations of mutations caused by a DNA lesion in bacteria led to the development of a two-polymerase model for translesion synthesis. In this model, translesion synthesis is divided into two steps, insertion to the lesion and extension with each step requiring a different error-prone DNA polymerase, because mutations across from and next to a DNA lesion were thought to be made by different error-prone DNA polymerases (Prakash and Prakash, 2002; Friedberg et al., 2005; McCulloch and Kunkel, 2008; Shachar et al., 2009). Results of the high-throughput TLS assay show that mutations that were not directly opposite to DNA lesions did not result immediately from translesion synthesis, thus calling into question the two-polymerase model.

Results of the high-throughput TLS assay are consistent with polymerase switching whereby a translesion DNA polymerase is recruited to specifically insert a nucleotide opposite to the lesion. The results are also consistent with having just one translesion DNA polymerase filling the entire gap region, because the error rate of a translesion

DNA polymerase is 10^{-2} to 10^{-4} and the gap region spans only 12 nucleotides. Therefore, translesion DNA polymerases play important roles in synthesizing short gaps in DNA repairs.

4.2   Several DNA polymerases can bypass 8-oxo-G and Pt-GG in cells

We predicted that Pol iota is the only DNA polymerase to correctly bypass 8-oxo-G *in vivo*, but found instead Pol eta carried this function in mammalian cells. Effects of Pol iota were detected in the absence of Pol eta or Pol kappa; therefore, Pol iota together with Pol eta and Pol iota together with Pol kappa bypassed 8-oxo-G correctly. In one of the four pairs of *REV3L* cell lines, Rev3L was found to bypass 8-oxo-G correctly.

Rev3L, Rev1 and Pol kappa bypassed Pt-GG in an error-prone manner in mammalian cells. The functions of Rev3L and Rev1 may be associated because Rev1 can act as a scaffold protein to recruit Rev3L (Acharya et al., 2006; Hashimoto et al., 2012; Kikuchi et al., 2012). The biological function of Pol kappa was redundant with that of Pol iota because Pol kappa's effect was detected only in the absence of Pol iota, and vice versa. Among Pol kappa, Pol iota and Pol eta, Pol kappa was the major DNA polymerase to bypass Pt-GG, and contrary to previous reports (Ummat et al., 2012; Zhao et al., 2012), Pol eta did not have a major role in bypassing the Pt-GG lesion. It was also reported that Pol eta could not carry out translesion synthesis of Pt-GG (Chijiwa et al., 2010), because in the structural study Pol eta could not incorporate nucleotides at the second insertion step of the DNA lesion.

4.3   Properties of 8-oxo-G and Pt-GG in mammalian cells

In the beginning of the study, I thought that Pt-GG but not 8-oxo-G blocks DNA replication because Pt-GG is a more bulky DNA lesion. While previous studies show that Pt-GG cannot be readily bypassed by many DNA polymerases *in vitro* (Vaisman et al.,

2000; Chijiwa et al., 2010), my results show that Pt-GG was bypassed with the same efficiency as 8-oxo-G in mammalian cells.

The mutation rate of 8-oxo-G (9.1×10^{-1}) was two orders of magnitudes higher than that of Pt-GG (4.7×10^{-3}) *in vivo*. The large difference in mutation rate reflected the difference in intrinsic mutagenic properties of the DNA lesions. Pt-GG is less prone to base mis-insertions because DNA polymerases allow some flexibility in DNA backbone and are able to traverse the slight bent introduced by the intrastrand cisplatin crosslink when incorporating nucleotides to Pt-GG. For the 8-oxo-G lesion, an additional oxygen atom in a guanine base alters its charge and therefore resulted in a high frequency of A mis-incorporation against 8-oxo-G. The large difference in mutation rate of the two DNA lesions could also be due to different DNA repair mechanisms involved in removing the DNA lesions. Perhaps the mechanism for removing Pt-GG lesions acted sooner than that for 8-oxo-G lesions. While 8-oxo-G is largely removed by base excision repair pathway through recognition of the lesion by the DNA glycosylase enzyme OGG1, Pt-GG is largely removed by nucleotide excision pathway that recognizes the distortion in local DNA structure caused by the DNA lesion. The basis for recognizing the 8-oxo-G lesion relies less on DNA geometry and possibly, more on charge and that for recognizing the Pt-GG lesion rests on deviation in its local DNA structure. This could entail different mechanisms (recruitment of different protein ensembles), thus different timings at which the lesions were repaired. Clearly, 8-oxo-G is the more harmful of the two DNA lesions. The high incidence of transversion mutations arising from 8-oxo-G underlines the importance of DNA repair to remove this lesion before DNA replication.

Pt-GG caused very few mutations: its mutation rate was only two times the background error rate in the NGS study. All of Pol iota, Pol kappa, Pol eta, Rev1 and Rev3L were error-prone to the lesion. I do not know which DNA polymerase(s) bypassed Pt-GG correctly *in vivo*, but it would be important to find out, because inactivating those DNA polymerases should reduce drug resistance to cisplatin in chemotherapy.

Then a question arises: what makes the cisplatin drug effective to kill cancerous cells since Pt-GG, a major species of cisplatin-DNA adducts, did not block DNA replication

more than 8-oxo-G, and 8-oxo-G is an abundant DNA lesion. I argue that cisplatin's drug effect does not depend on Pt-GG to block DNA replication, at least not by Pt-GG alone, and the toxicity of cisplatin to tumor cells have other explanations (Cepeda et al., 2007; Jung and Lippard, 2007). For one, it is conceivable that inter-strand cisplatin DNA adducts would stall normal DNA replication. Also, cisplatin-DNA adducts attract binding of the nonhistone High Mobility Group (HMG)-family of proteins that alter chromatin structures (Rabik and Dolan, 2007); HMG proteins bound to Pt-GG lesions or other cisplatin-DNA adducts could stall DNA replication (Rabik and Dolan, 2007). Cisplatin also modifies amino acid residues of cytoplasmic and nuclear proteins (Woźniak and Błasiak, 2002) and HMG proteins bound to cisplatin modified proteins could disrupt the cell's internal architecture (Woźniak and Błasiak, 2002).

4.4 Limitations on using cell lines as models in the high-throughput TLS assay to study protein functions

There are limitations on using mammalian cell lines with specific gene knockouts to study functions of translesion DNA polymerases. First, even with the Next Generation Sequencing technology that produces thousands of reads for each cell line, the biological effect of a single gene is elusive because only subtle differences in mutation frequencies could be detected between isogenic cell lines. Second, results of the high-throughput TLS assay are indirect as I do not know which DNA polymerase(s) actually fill the gap plasmids in cells. The best I could do is to acknowledge a probability that the translesion DNA polymerase in question has taken action. I do not know if the proteins are expressed in wild type cells, since I could not detect Pol iota, Pol eta and Pol kappa in Western Blotting experiments using monoclonal primary antibodies (results not shown). Labs that generated the knockout cell lines used in this project (Table 4) have confirmed the genotypes of the cell lines by PCR and checked the expressions of the DNA polymerases by Western Blotting. Third, I overloaded the cells or stressed the system with exogenous DNA in transfection, which might cause the cells to not behave as they would under normal physiological conditions. Forth, in analyzing the NGS data I made exhaustive comparisons between cell lines to deduce the functions of translesion DNA polymerases.

The many t-tests performed greatly increases the chance of committing Type 1 error—detecting an effect for a translesion DNA polymerase when in fact it does not have any. Type 1 error cannot be eliminated even with P-value adjustment: the Benjamini Hochberg method was chosen because it is less conservative than the Bonferroni correction. For all of these reasons, my findings on the biological functions of translesion DNA polymerases in the high-throughput TLS assay are preliminary and would need to be validated by other means, such as *in vitro* assays using purified proteins.

4.5 The high-throughput TLS assay generated new hypotheses for further testing

The high-throughput TLS assay is a valuable tool for generating new hypotheses. From my results, Pol eta had a major role in correctly bypassing 8-oxo-G in mammalian cells. This finding is supported by other reports that study the kinetics (Haracska et al., 2000; Washington et al., 2001), X-ray crystal structures (Silverstein et al., 2010), and genetics (Rodriguez et al., 2013) of the yeast Pol eta in bypassing 8-oxo-G.

Rev3L bypassed Pt-GG in all four pairs of the *REV3L* cell lines analyzed (Figure 14). This finding is supported by previous reports that showed that Rev3L confers chemo-resistance to cisplatin (Wang et al., 2009) and *REV3L*-deficient cells are sensitive to cisplatin (Sonoda et al., 2003; Zander and Bemark, 2004; Doles et al., 2010). Knocking out *REV1* produced a similar effect as knocking out *REV3L*. Rev1 bypassed Pt-GG, and Rev1 may function on its own or recruit Rev3L to bypass the DNA lesion. Compared to Pol eta and Pol iota, Pol kappa had a greater activity in bypassing Pt-GG *in vivo*. Pol kappa's biological function was detected in the absence of Pol iota, and vice versa, suggesting their functions were either complementary or redundant. Pol eta and Pol iota together bypassed Pt-GG, suggesting a link between the two proteins. The functions of Pol kappa, Pol iota and Pol eta in cells were not mutually exclusive; they could bypass the same type of DNA lesions such as Pt-GG and could potentially interact when carrying out translesion synthesis in mammalian cells.

4.6   Other applications of the high-throughput TLS assay

The high-throughput TLS assay is a very useful method to determine the mutation rate and the mutation spectrum of a DNA lesion in mammalian cells. The assay can be conveniently adapted to test many DNA lesions in a few cell lines to pinpoint those lesions that have the most mutagenic effects for further study. Researchers can choose to use different cell types and change the sequence of the gap region of double stranded gap plasmids in their experiments.

The high-throughput TLS assay can also be used to study the long-term effect of a DNA lesion. A time-series study can be carried out to monitor changes in the mutation frequency of a DNA lesion and changes in the mutation spectrum of the gap region of the double stranded gap plasmids over the course of hours or days in mammalian cells.

4.7   A problem with using *E. coli* in the traditional TLS assay

The traditional TLS assay is not a quantitative method to study mutations made by DNA lesions precisely because the method uses *E.coli*. The mutation frequency of a DNA lesion observed after bacterial cells have grown overnight does not reflect the mutation frequency of the lesion at the time of bacterial transformation.

To reconcile the results obtained from the traditional TLS assay and those obtained from the high-throughput TLS assay, we must understand the difference in dynamics of using *E. coli* versus PCR to amplify DNA. In both cases, the number of DNA sequences increases exponentially, but a PCR reaction is a more stable and predictable system than the bacterial cell. DNA amplification in PCR reactions is only limited by the concentration of DNA templates and the number of PCR cycles. The Phusion$^{®}$ High Fidelity DNA polymerase used in PCR reaction is assumed to not make any mutations. DNA amplification in *E. coli* on the other hand, is not constant and is under selective pressure—reflected by the change in growth rate of bacteria that may have a qualitative effect on producing additional mutations from the one caused by DNA lesion. The mutation rate of a DNA lesion conveys an impression of a linear relationship between

mutations and the number of nucleotides replicated. Accumulation of mutations in bacteria, however, is nonlinear and is subject to influences of bacteria's growth conditions that change with environment. Mutations observed in bacteria due to a DNA lesion are results of a series of events starting with translesion synthesis and those observed in mammalian cells are results of translesion synthesis only. Difference in time produced different observations of mutations that arise from a single DNA lesion.

4.8   Accumulation of mutation is related to the cell's growth condition

My study stumbles upon seeing mutations at different times, from within one generation in mammalian cells—as in a still photograph, to across many generations in bacterial cells—as in a time-lapsed photograph. The high-throughput TLS assay captures mutations caused by a DNA lesion in the short term—the time frame is seven hours in cells that double every 24 hours. The traditional TLS assay captures mutations caused by a DNA lesion in the long term—the time frame is 16 hours in bacterial cells that double every 20 minutes. In the high-throughput assay, we saw only mutation directly opposite to the DNA lesion; in the traditional assay, however, we saw accumulation of mutations at positions adjacent to the DNA lesion (Appendix C). The difference in results of the two methods suggests that the long-term effect of a DNA lesion is closely related to the cell's growth condition, and that a single mutation caused by a DNA lesion can lead to other mutations in the nearby DNA sequence under selective pressure of the cell. Observations from the traditional TLS assay are outcomes from a series of events led by translesion synthesis of a DNA lesion.

4.9   Significance of translesion synthesis in survival

The long-term effects of a DNA lesion in mammalian cells after they have divided many times may come to resemble the mutation spectrums observed in bacteria, with increased likelihood of mutations for nucleotides adjacent to the DNA lesion. This would be the

case especially if the growth condition of mammalian cells is limiting, such as under cell culture environment.

In bacteria, a DNA lesion may have long been removed from DNA sequence as cells continue to divide, but the effect of the DNA lesion, such as mutations resulting immediately from the lesion can quickly lead to other changes in its surrounding nucleotides. Cells have DNA repairs and other mechanisms to maintain the genome's equilibrium so mutations cannot increase uncontrollably without qualitative changes in the cell.

Cultured mammalian cells share the same growth dynamic as cultured bacterial cells, but an important distinction is that mammalian cells divide every 24 hours instead of every 20 minutes so mutations in mammalian cells accumulate much more slowly. Mammalian cells have more stable dynamics than bacterial cells due to their slower growth rates. This implies that properties of mammalian cells change less than those of bacteria in the long term and that mammalian cells can better resist changes to DNA sequence arising from a DNA lesion. To further explain, it would be much harder for translesion synthesis to cause mutations in nucleotides adjacent to DNA lesions in mammalian cells unless there are also qualitative changes that speed up the cells' growth rates. The longer time it takes for mammalian cells to divide also leaves more chance for DNA repairs to remove a DNA lesion before it is replicated. I argue that mutations resulting from DNA lesions observed in bacteria forecast those that will be seen in malignantly transformed mammalian cells with uncontrolled growth rate.

Studying translesion synthesis is looking at a very fundamental level how mutations arise spontaneously in response to the environment. Accumulation of mutation is closely linked to the cell's growth condition. While the high-throughput TLS assay is useful for determining the mutation frequency and the mutation spectrum of a DNA lesion resulting immediately from translesion synthesis in the mammalian cell, the traditional TLS assay is useful for determining the long-term effect of the lesion in causing additional changes to a DNA sequence.

The genome is constantly under pressure from DNA lesions and translesion synthesis to change its information content. There is an element of chance in where changes in DNA sequences would occur. Though translesion synthesis causes subtle variations in the genome and is thus easily dismissed, small changes in initial conditions can have far-reaching consequences. DNA lesions and translesion synthesis are potent driving forces of evolution, and similar to Lorenz's butterfly effect (Lorenz, E.N., 1972) small perturbations occurring by chance in the sequence of the genome can, over a long time, have a large effect in determining the cell's fate.

References

Acharya, N., Johnson, R.E., Prakash, S., and Prakash, L. (2006). Complex formation with Rev1 enhances the proficiency of Saccharomyces cerevisiae DNA polymerase zeta for mismatch extension and for extension opposite from DNA lesions. Mol. Cell. Biol. 26, 9555–9563.

Aitchison, J. (2003). A concise guide to compositional data analysis.

Andersen, P.L., Xu, F., and Xiao, W. (2008). Eukaryotic DNA damage tolerance and translesion synthesis through covalent modifications of PCNA. Cell Res 18, 162–173.

Avkin, S., and Livneh, Z. (2002). Efficiency, specificity and DNA polymerase-dependence of translesion replication across the oxidative DNA lesion 8-oxoguanine in human cells. Mutat. Res. 510, 81–90.

Beard, W.A., Shock, D.D., Vande Berg, B.J., and Wilson, S.H. (2002). Efficiency of correct nucleotide insertion governs DNA polymerase fidelity. J. Biol. Chem. 277, 47393–47398.

Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. J. R. Stat. Soc. 57, 289–300.

Bi, X., Barkley, L.R., Slater, D.M., Tateishi, S., Yamaizumi, M., Ohmori, H., and Vaziri, C. (2006). Rad18 regulates DNA polymerase kappa and is required for recovery from S-phase checkpoint-mediated arrest. Mol. Cell. Biol. 26, 3527–3540.

Bienko, M., Green, C.M., Crosetto, N., Rudolf, F., Zapart, G., Coull, B., Kannouche, P., Wider, G., Peter, M., Lehmann, A.R., et al. (2005). Ubiquitin-binding domains in Y-family polymerases regulate translesion synthesis. Science 310, 1821–1824.

Bohr, V.A., Stevnsner, T., and de Souza-Pinto, N.C. (2002). Mitochondrial DNA repair of oxidative damage in mammalian cells. Gene 286, 127–134.

Boiteux, S., and Radicella, J.P. (2000). The human OGG1 gene: structure, functions, and its implication in the process of carcinogenesis. Arch. Biochem. Biophys. 377, 1–8.

Van den Boogaart, K.G., and Tolosana-Delgado, R. (2008). "compositions": A unified R package to analyze compositional data. Comput. Geosci. 34, 320–338.

Boudsocq, F., Kokoska, R.J., Plosky, B.S., Vaisman, A., Ling, H., Kunkel, T.A., Yang, W., and Woodgate, R. (2004). Investigating the role of the little finger domain of Y-family DNA polymerases in low fidelity synthesis and translesion replication. J. Biol. Chem. 279, 32932–32940.

Burgers, P.M., Koonin, E.V., Bruford, E., Blanco, L., Burtis, K.C., Christman, M.F., Copeland, W.C., Friedberg, E.C., Hanaoka, F., Hinkle, D.C., et al. (2001). Eukaryotic DNA polymerases: proposal for a revised nomenclature. J. Biol. Chem. 276, 43487–43490.

Cadet, J., Douki, T., and Ravanat, J.L. (1997). Artifacts associated with the measurement of oxidized DNA bases. Environ. Health Perspect. 105, 1034–1039.

Carpio, R.V.-D., Silverstein, T.D., Lone, S., Johnson, R.E., Prakash, L., Prakash, S., and Aggarwal, A.K. (2011). Role of human DNA polymerase κ in extension opposite from a cis-syn thymine dimer. J. Mol. Biol. 408, 252–261.

Cepeda, V., Fuertes, M.A., Castilla, J., Alonso, C., Quevedo, C., and Pérez, J.M. (2007). Biochemical mechanisms of cisplatin cytotoxicity. Anticancer Agents Med. Chem. 7, 3–18.

Cheng, K.C., Cahill, D.S., Kasai, H., Nishimura, S., and Loeb, L.A. (1992). 8-Hydroxyguanine, an abundant form of oxidative DNA damage, causes G----T and A----C substitutions. J. Biol. Chem. 267, 166–172.

Chijiwa, S., Masutani, C., Hanaoka, F., Iwai, S., and Kuraoka, I. (2010). Polymerization by DNA polymerase η is blocked by cis-diamminedichloroplatinum(II) 1,3-d(GpTpG) cross-link: implications for cytotoxic effects in nucleotide excision repair-negative tumor cells. Carcinogenesis 31, 388–393.

Chun, A.C.S., and Jin, D.-Y. (2010). Ubiquitin-dependent regulation of translesion polymerases. Biochem. Soc. Trans. 38, 110–115.

Clancy, S. (2008). DNA damage & repair: mechanisms for maintaining DNA integrity. Nat. Educ. 1, 103.

Comas-Cufi, M., and Thió-Henestrosa, S. (2011). CoDaPack 2.0: a stand-alone, multi-platform compositional software. In 4th International Workshop on Compositional Data Analysis, (Sant Feliu de Guíxols),.

Daigaku, Y., Davies, A.A., and Ulrich, H.D. (2010). Ubiquitin-dependent DNA damage bypass is separable from genome replication. Nature 465, 951–955.

Delbos, F., De Smet, A., Faili, A., Aoufouchi, S., Weill, J.-C., and Reynaud, C.-A. (2005). Contribution of DNA polymerase η to immunoglobulin gene hypermutation in the mouse. J. Exp. Med. 201, 1191–1196.

Diamant, N., Hendel, A., Vered, I., Carell, T., Reißner, T., de Wind, N., Geacinov, N., and Livneh, Z. (2011). DNA damage bypass operates in the S and G2 phases of the cell cycle and exhibits differential mutagenicity. Nucleic Acids Res.

Doles, J., Oliver, T.G., Cameron, E.R., Hsu, G., Jacks, T., Walker, G.C., and Hemann, M.T. (2010). Suppression of Rev3, the catalytic subunit of Pol{zeta}, sensitizes drug-

resistant lung tumors to chemotherapy. Proc. Natl. Acad. Sci. U. S. A. 107, 20786–20791.

Dou, H., Mitra, S., and Hazra, T.K. (2003). Repair of Oxidized Bases in DNA Bubble Structures by Human DNA Glycosylases NEIL1 and NEIL2. J. Biol. Chem. 278, 49679–49684.

Duckett, D.R., Drummond, J.T., Murchie, A.I., Reardon, J.T., Sancar, A., Lilley, D.M., and Modrich, P. (1996). Human MutSalpha recognizes damaged DNA base pairs containing O6-methylguanine, O4-methylthymine, or the cisplatin-d(GpG) adduct. Proc. Natl. Acad. Sci. 93, 6443–6447.

Einolf, H.J., and Guengerich, F.P. (2001). Fidelity of nucleotide insertion at 8-oxo-7,8-dihydroguanine by mammalian DNA polymerase delta. Steady-state and pre-steady-state kinetic analysis. J. Biol. Chem. 276, 3764–3771.

Elvers, I., Johansson, F., Groth, P., Erixon, K., and Helleday, T. (2011). UV stalled replication forks restart by re-priming in human fibroblasts. Nucleic Acids Res. 39, 7049–7057.

Faili, A., Aoufouchi, S., Flatter, E., Guéranger, Q., Reynaud, C.-A., and Weill, J.-C. (2002). Induction of somatic hypermutation in immunoglobulin genes is dependent on DNA polymerase iota. Nature 419, 944–947.

Fichtinger-Schepman, A.M., van Oosterom, A.T., Lohman, P.H., and Berends, F. (1987). cis-Diamminedichloroplatinum(II)-induced DNA adducts in peripheral leukocytes from seven cancer patients: quantitative immunochemical detection of the adduct induction and removal after a single dose of cis-diamminedichloroplatinum(II). Cancer Res. 47, 3000–3004.

Friedberg, E.C., Fischhaber, P.L., and Kisker, C. (2001). Error-prone DNA polymerases: novel structures and the benefits of infidelity. Cell 107, 9–12.

Friedberg, E.C., Lehmann, A.R., and Fuchs, R.P.P. (2005). Trading places: how do DNA polymerases switch during translesion DNA synthesis? Mol. Cell 18, 499–505.

Gan, G.N., Wittschieben, J.P., Wittschieben, B.Ø., and Wood, R.D. (2008). DNA polymerase zeta (pol zeta) in higher eukaryotes. Cell Res. 18, 174–183.

Garg, P., Stith, C.M., Majka, J., and Burgers, P.M.J. (2005). Proliferating cell nuclear antigen promotes translesion synthesis by DNA polymerase zeta. J. Biol. Chem. 280, 23446–23450.

Gerlach, V.L., Feaver, W.J., Fischhaber, P.L., and Friedberg, E.C. (2001). Purification and characterization of pol kappa, a DNA polymerase encoded by the human DINB1 gene. J. Biol. Chem. 276, 92–98.

Guo, C., Fischhaber, P.L., Luk-Paszyc, M.J., Masuda, Y., Zhou, J., Kamiya, K., Kisker, C., and Friedberg, E.C. (2003). Mouse Rev1 protein interacts with multiple DNA polymerases involved in translesion DNA synthesis. EMBO J. 22, 6621–6630.

Guo, C., Tang, T.-S., Bienko, M., Parker, J.L., Bielen, A.B., Sonoda, E., Takeda, S., Ulrich, H.D., Dikic, I., and Friedberg, E.C. (2006). Ubiquitin-binding motifs in REV1 protein are required for its role in the tolerance of DNA damage. Mol. Cell. Biol. 26, 8892–8900.

Hailer, M.K., Slade, P.G., Martin, B.D., Rosenquist, T.A., and Sugden, K.D. (2005). Recognition of the oxidized lesions spiroiminodihydantoin and guanidinohydantoin in DNA by the mammalian base excision repair glycosylases NEIL1 and NEIL2. DNA Repair 4, 41–50.

Haracska, L., Yu, S.L., Johnson, R.E., Prakash, L., and Prakash, S. (2000). Efficient and accurate replication in the presence of 7,8-dihydro-8-oxoguanine by DNA polymerase eta. Nat. Genet. 25, 458–461.

Haracska, L., Acharya, N., Unk, I., Johnson, R.E., Hurwitz, J., Prakash, L., and Prakash, S. (2005). A Single Domain in Human DNA Polymerase ι Mediates Interaction with PCNA: Implications for Translesion DNA Synthesis. Mol. Cell. Biol. 25, 1183–1190.

Hartwell, L., and Kastan, M. (1994). Cell cycle control and cancer. Science 266, 1821–1828.

Hashimoto, K., Cho, Y., Yang, I.-Y., Akagi, J., Ohashi, E., Tateishi, S., de Wind, N., Hanaoka, F., Ohmori, H., and Moriya, M. (2012). The Vital Role of Polymerase ζ and REV1 in Mutagenic, but Not Correct, DNA Synthesis across Benzo[a]pyrene-dG and Recruitment of Polymerase ζ by REV1 to Replication-stalled Site. J. Biol. Chem. 287, 9613–9622.

Hazra, T.K., Hill, J.W., Izumi, T., and Mitra, S. (2001). Multiple DNA glycosylases for repair of 8-oxoguanine and their potential in vivo functions. Prog. Nucleic Acid Res. Mol. Biol. 68, 193–205.

Hoege, C., Pfander, B., Moldovan, G.-L., Pyrowolakis, G., and Jentsch, S. (2002). RAD6-dependent DNA repair is linked to modification of PCNA by ubiquitin and SUMO. Nature 419, 135–141.

Jackson, A.L., and Loeb, L.A. (1998). The Mutation Rate and Cancer. Genetics 148, 1483–1490.

Jansen, J.G. (2006). Strand-biased defect in C/G transversions in hypermutating immunoglobulin genes in Rev1-deficient mice. J. Exp. Med. 203, 319–323.

Jansen, J.G., Fousteri, M.I., and de Wind, N. (2007). Send in the Clamps: Control of DNA Translesion Synthesis in Eukaryotes. Mol. Cell 28, 522–529.

Jansen, J.G., Tsaalbi-Shtylik, A., Hendriks, G., Verspuy, J., Gali, H., Haracska, L., and de Wind, N. (2009). Mammalian polymerase zeta is essential for post-replication repair of UV-induced DNA lesions. DNA Repair 8, 1444–1451.

Johnson, R.E., Kondratick, C.M., Prakash, S., and Prakash, L. (1999). hRAD30 mutations in the variant form of xeroderma pigmentosum. Science 285, 263–265.

Jung, Y., and Lippard, S.J. (2007). Direct cellular responses to platinum-induced DNA damage. Chem. Rev. 107, 1387–1407.

Kannouche, P., Fernández de Henestrosa, A.R., Coull, B., Vidal, A.E., Gray, C., Zicha, D., Woodgate, R., and Lehmann, A.R. (2003). Localization of DNA polymerases eta and iota to the replication machinery is tightly co-ordinated in human cells. EMBO J. 22, 1223–1233.

Kannouche, P.L., Wing, J., and Lehmann, A.R. (2004). Interaction of human DNA polymerase eta with monoubiquitinated PCNA: a possible mechanism for the polymerase switch in response to DNA damage. Mol. Cell 14, 491–500.

Karata, K., Vidal, A.E., and Woodgate, R. (2009). Construction of a circular single-stranded DNA template containing a defined lesion. DNA Repair 8, 852–856.

Karras, G.I., and Jentsch, S. (2010). The RAD6 DNA damage tolerance pathway operates uncoupled from the replication fork and is functional beyond S phase. Cell 141, 255–267.

Kikuchi, S., Hara, K., Shimizu, T., Sato, M., and Hashimoto, H. (2012). Structural basis of recruitment of DNA polymerase ζ by interaction between REV1 and REV7 proteins. J. Biol. Chem. 287, 33847–33852.

Kirouac, K.N., and Ling, H. (2011a). Unique active site promotes error-free replication opposite an 8-oxo-guanine lesion by human DNA polymerase iota. Proc. Natl. Acad. Sci. U. S. A. 108, 3210–3215.

Kirouac, K.N., and Ling, H. (2011b). Polι: Shining light on repair of oxidative DNA lesions and mutations. Cell Cycle Georget. Tex 10, 1520–1521.

Kunkel, T.A. (2004). DNA Replication Fidelity. J. Biol. Chem. 279, 16895–16898.

Lange, S.S., Takata, K., and Wood, R.D. (2011). DNA polymerases and cancer. Nat. Rev. Cancer 11, 96–110.

Lange, S.S., Wittschieben, J.P., and Wood, R.D. (2012). DNA polymerase zeta is required for proliferation of normal mammalian cells. Nucleic Acids Res. 40, 4473–4482.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat Meth 9, 357–359.

Lee, Y.-S., Gregory, M.T., and Yang, W. (2014). Human Pol ζ purified with accessory subunits is active in translesion DNA synthesis and complements Pol η in cisplatin bypass. Proc. Natl. Acad. Sci. U. S. A. 111, 2954–2959.

Lin, Q., Clark, A.B., McCulloch, S.D., Yuan, T., Bronson, R.T., Kunkel, T.A., and Kucherlapati, R. (2006). Increased susceptibility to UV-induced skin carcinogenesis in polymerase eta-deficient mice. Cancer Res. 66, 87–94.

Ling, H., Boudsocq, F., Woodgate, R., and Yang, W. (2001). Crystal structure of a Y-family DNA polymerase in action: a mechanism for error-prone and lesion-bypass replication. Cell 107, 91–102.

Lodish, H., Berk, A., and Zipursky, S.L. (2000). Mutations affecting genome stability. In Molecular Cell Biology, (New York: W. H. Freeman), p. Section 24.5.

Van Loon, B., Markkanen, E., and Hübscher, U. (2010). Oxygen as a friend and enemy: How to combat the mutational potential of 8-oxo-guanine. DNA Repair 9, 604–616.

Lorenz, E.N. (1963). Deterministic nonperiodic flow. J. Atmospheric Sci. 20, 130–141.

Lorenz, E.N. (1972). Predictability: does the flap of a butterfly's wings in Brazil set off a tornado in Texas?

Maga, G., and Hübscher, U. (2003). Proliferating cell nuclear antigen (PCNA): a dancer with many partners. J. Cell Sci. 116, 3051–3060.

Makarova, A.V., and Kulbachinskiy, A.V. (2012). Structure of human DNA polymerase iota and the mechanism of DNA synthesis. Biochem. Biokhimii︠a︡ 77, 547–561.

Makridakis, N.M., and Reichardt, J.K.V. (2012). Translesion DNA Polymerases and Cancer. Front. Genet. 3, 174.

Markkanen, E., Castrec, B., Villani, G., and Hübscher, U. (2012). A switch between DNA polymerases δ and λ promotes error-free bypass of 8-oxo-G lesions. Proc. Natl. Acad. Sci. U. S. A. 109, 20401–20406.

Marnett, L.J. (2000). Oxyradicals and DNA damage. Carcinogenesis 21, 361–370.

Martomo, S.A., Yang, W.W., Wersto, R.P., Ohkumo, T., Kondo, Y., Yokoi, M., Masutani, C., Hanaoka, F., and Gearhart, P.J. (2005). Different mutation signature in DNA polymerase eta- and MSH6-deficient mice suggest separate roles in antibody diversification. Proc Natl Acad Sci U A 102.

Masutani, C., Kusumoto, R., Yamada, A., Dohmae, N., Yokoi, M., Yuasa, M., Araki, M., Iwai, S., Takio, K., and Hanaoka, F. (1999). The XPV (xeroderma pigmentosum variant) gene encodes human DNA polymerase eta. Nature 399, 700–704.

May, R.M. (1976). Simple mathematical models with very complicated dynamics. Nature 261, 459–467.

Mello, J.A., Acharya, S., Fishel, R., and Essigmann, J.M. (1996). The mismatch-repair protein hMSH2 binds selectively to DNA adducts of the anticancer drug cisplatin. Chem. Biol. 3, 579–589.

Minowa, O., Arai, T., Hirano, M., Monden, Y., Nakai, S., Fukuda, M., Itoh, M., Takano, H., Hippou, Y., Aburatani, H., et al. (2000). Mmh/Ogg1 gene inactivation results in accumulation of 8-hydroxyguanine in mice. Proc. Natl. Acad. Sci. 97, 4156–4161.

McCulloch, S.D., and Kunkel, T.A. (2008). The fidelity of DNA synthesis by eukaryotic replicative and translesion synthesis polymerases. Cell Res. 18, 148–161.

McDonald, J.P., Levine, A.S., and Woodgate, R. (1997). The Saccharomyces cerevisiae RAD30 gene, a homologue of Escherichia coli dinB and umuC, is DNA damage inducible and functions in a novel error-free postreplication repair mechanism. Genetics 147, 1557–1568.

McDonald, J.P., Frank, E.G., Plosky, B.S., Rogozin, I.B., Masutani, C., Hanaoka, F., Woodgate, R., and Gearhart, P.J. (2003). 129-derived Strains of Mice Are Deficient in DNA Polymerase and Have Normal Immunoglobulin Hypermutation. J. Exp. Med. 198, 635–643.

McIntyre, J., Vidal, A.E., McLenigan, M.P., Bomar, M.G., Curti, E., McDonald, J.P., Plosky, B.S., Ohashi, E., and Woodgate, R. (2013). Ubiquitin mediates the physical and functional interaction between human DNA polymerases η and ι. Nucleic Acids Res. 41, 1649–1660.

Murakumo, Y., Ogura, Y., Ishii, H., Numata, S., Ichihara, M., Croce, C.M., Fishel, R., and Takahashi, M. (2001). Interactions in the error-prone postreplication repair proteins hREV1, hREV3, and hREV7. J. Biol. Chem. 276, 35644–35651.

Naryzhny, S.N. (2008). Proliferating cell nuclear antigen: a proteomics view. Cell. Mol. Life Sci. CMLS 65, 3789–3808.

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

Nouspikel, T. (2009). DNA Repair in Mammalian Cells. Cell. Mol. Life Sci. 66, 994–1009.

Ocampo, M.T.A., Chaung, W., Marenstein, D.R., Chan, M.K., Altamirano, A., Basu, A.K., Boorstein, R.J., Cunningham, R.P., and Teebor, G.W. (2002). Targeted Deletion of mNth1 Reveals a Novel DNA Repair Enzyme Activity. Mol. Cell. Biol. 22, 6111–6121.

Ogi, T., and Lehmann, A.R. (2006). The Y-family DNA polymerase κ (pol κ) functions in mammalian nucleotide-excision repair. Nat. Cell Biol. 8, 640–642.

Ogi, T., Shinkai, Y., Tanaka, K., and Ohmori, H. (2002). Polκ protects mammalian cells against the lethal and mutagenic effects of benzo [a] pyrene. Proc. Natl. Acad. Sci. 99, 15548–15553.

Ohashi, E., Bebenek, K., Matsuda, T., Feaver, W.J., Gerlach, V.L., Friedberg, E.C., Ohmori, H., and Kunkel, T.A. (2000). Fidelity and processivity of DNA synthesis by DNA polymerase kappa, the product of the human DINB1 gene. J. Biol. Chem. 275, 39678–39684.

Ohashi, E., Murakumo, Y., Kanjo, N., Akagi, J.-I., Masutani, C., Hanaoka, F., and Ohmori, H. (2004). Interaction of hREV1 with three human Y-family DNA polymerases. Genes Cells Devoted Mol. Cell. Mech. 9, 523–531.

Ohkumo, T., Kondo, Y., Yokoi, M., Tsukamoto, T., Yamada, A., Sugimoto, T., Kanao, R., Higashi, Y., Kondoh, H., Tatematsu, M., et al. (2006). UV-B Radiation Induces Epithelial Tumors in Mice Lacking DNA Polymerase and Mesenchymal Tumors in Mice Deficient for DNA Polymerase. Mol. Cell. Biol. 26, 7696–7706.

Ohmori, H., Friedberg, E.C., Fuchs, R.P., Goodman, M.F., Hanaoka, F., Hinkle, D., Kunkel, T.A., Lawrence, C.W., Livneh, Z., Nohmi, T., et al. (2001). The Y-family of DNA polymerases. Mol. Cell 8, 7–8.

Okada, T., Sonoda, E., Yamashita, Y.M., Koyoshi, S., Tateishi, S., Yamaizumi, M., Takata, M., Ogawa, O., and Takeda, S. (2002). Involvement of vertebrate polkappa in Rad18-independent postreplication repair of UV damage. J. Biol. Chem. 277, 48690–48695.

Ollivierre, J., Silva, M., Sefcikova, J., and Beuning, P. (2011). Polymerase Switching in Response to DNA Damage. In Biophysics of DNA-Protein Interactions, M.C. Williams, and L.J. Maher III, eds. (Springer New York), pp. 241–292.

Osterod, M., Hollenbach, S., Hengstler, J.G., Barnes, D.E., Lindahl, T., and Epe, B. (2001). Age-related and tissue-specific accumulation of oxidative DNA base damage in 7,8-dihydro-8-oxoguanine-DNA glycosylase (Ogg1) deficient mice. Carcinogenesis 22, 1459–1463.

Pandya, G.A., and Moriya, M. (1996). 1,N6-ethenodeoxyadenosine, a DNA adduct highly mutagenic in mammalian cells. Biochemistry (Mosc.) 35, 11487–11492.

Prakash, S., and Prakash, L. (2002). Translesion DNA synthesis in eukaryotes: a one- or two-polymerase affair. Genes Dev. 16, 1872–1883.

Rabik, C.A., and Dolan, M.E. (2007). Molecular mechanisms of resistance and toxicity associated with platinating agents. Cancer Treat. Rev. 33, 9–23.

Radman, M., Prakash, L., Sherman, F., Miller, M.W., Lawrence, C.W., and Tabor, H.W. (1974). Phenomenology of an inducible mutgenic DNA repair pathway in Escherichia coli: SOS repair hypothesis. Mol. Environ. Asp. Mutagen. 128–142.

Rodriguez, G.P., Song, J.B., and Crouse, G.F. (2013). In vivo bypass of 8-oxodG. PLoS Genet. 9, e1003682.

Sale, J.E., Lehmann, A.R., and Woodgate, R. (2012). Y-family DNA polymerases and their role in tolerance of cellular DNA damage. Nat. Rev. Mol. Cell Biol. 13, 141–152.

Satou, K., Hori, M., Kawai, K., Kasai, H., Harashima, H., and Kamiya, H. (2009). Involvement of specialized DNA polymerases in mutagenesis by 8-hydroxy-dGTP in human cells. DNA Repair 8, 637–642.

Shachar, S., Ziv, O., Avkin, S., Adar, S., Wittschieben, J., Reissner, T., Chaney, S., Friedberg, E.C., Wang, Z., Carell, T., et al. (2009). Two-polymerase mechanisms dictate error-free and error-prone translesion DNA synthesis in mammals. EMBO J. 28, 383–393.

Sharma, S., Hicks, J.K., Chute, C.L., Brennan, J.R., Ahn, J.-Y., Glover, T.W., and Canman, C.E. (2011). REV1 and polymerase ζ facilitate homologous recombination repair. Nucleic Acids Res.

Shibutani, S., Takeshita, M., and Grollman, A.P. (1991). Insertion of specific bases during DNA synthesis past the oxidation-damaged base 8-oxodG. Nature 349, 431–434.

Silverstein, T.D., Jain, R., Johnson, R.E., Prakash, L., Prakash, S., and Aggarwal, A.K. (2010). Structural basis for error-free replication of oxidatively damaged DNA by yeast DNA polymerase η. Struct. Lond. Engl. 1993 18, 1463–1470.

Sonoda, E., Okada, T., Zhao, G.Y., Tateishi, S., Araki, K., Yamaizumi, M., Yagi, T., Verkaik, N.S., van Gent, D.C., Takata, M., et al. (2003). Multiple roles of Rev3, the catalytic subunit of polzeta in maintaining genome stability in vertebrates. EMBO J. 22, 3188–3197.

Stelter, P., and Ulrich, H.D. (2003). Control of spontaneous and damage-induced mutagenesis by SUMO and ubiquitin conjugation. Nature 425, 188–191.

Stratagene (2005). QuikChange Site-directed mutagenesis kit (Instruction Manual).

Strogatz, S.H. (2001). Nonlinear dynamics and chaos: with applications to physics, biology, chemistry and engineering (Westview Press).

Taggart, D.J., Camerlengo, T.L., Harrison, J.K., Sherrer, S.M., Kshetry, A.K., Taylor, J.-S., Huang, K., and Suo, Z. (2013). A high-throughput and quantitative method to assess the mutagenic potential of translesion DNA synthesis. Nucleic Acids Res. 41, e96.

Tissier, A., McDonald, J.P., Frank, E.G., and Woodgate, R. (2000). poliota, a remarkably error-prone human DNA polymerase. Genes Dev. 14, 1642–1650.

Tissier, A., Kannouche, P., Reck, M.-P., Lehmann, A.R., Fuchs, R.P.P., and Cordonnier, A. (2004). Co-localization in replication foci and interaction of human Y-family members, DNA polymerase pol eta and REVl protein. DNA Repair 3, 1503–1514.

Tsurimoto, T. (1999). PCNA binding proteins. Front. Biosci. J. Virtual Libr. 4, D849–D858.

Ummat, A., Rechkoblit, O., Jain, R., Roy Choudhury, J., Johnson, R.E., Silverstein, T.D., Buku, A., Lone, S., Prakash, L., Prakash, S., et al. (2012). Structural basis for cisplatin DNA damage tolerance by human polymerase η during cancer chemotherapy. Nat Struct Mol Biol 19, 628–632.

Vaisman, A., Masutani, C., Hanaoka, F., and Chaney, S.G. (2000). Efficient translesion replication past oxaliplatin and cisplatin GpG adducts by human DNA polymerase eta. Biochemistry (Mosc.) 39, 4575–4580.

Vidal, A.E., Kannouche, P., Podust, V.N., Yang, W., Lehmann, A.R., and Woodgate, R. (2004). Proliferating Cell Nuclear Antigen-dependent Coordination of the Biological Functions of Human DNA Polymerase ι. J. Biol. Chem. 279, 48360–48368.

Wang, H., Zhang, S.-Y., Wang, S., Lu, J., Wu, W., Weng, L., Chen, D., Zhang, Y., Lu, Z., Yang, J., et al. (2009). REV3L confers chemoresistance to cisplatin in human gliomas: The potential of its RNAi for synergistic therapy. Neuro-Oncol. 11, 790–802.

Wang, Q.E., Milum, K., Han, C., Huang, Y.W., Wani, G., Thomale, J., and Wani, A.A. (2011). Differential contributory roles of nucleotide excision and homologous recombination repair for enhancing cisplatin sensitivity in human ovarian cancer cells. Mol. Cancer 10, 24–24.

Washington, M.T., Johnson, R.E., Prakash, L., and Prakash, S. (2001). Accuracy of lesion bypass by yeast and human DNA polymerase eta. Proc. Natl. Acad. Sci. U. S. A. 98, 8355–8360.

Watt, D.L., Utzat, C.D., Hilario, P., and Basu, A.K. (2007). Mutagenicity of the 1-nitropyrene-DNA adduct N-(deoxyguanosin-8-yl)-1-aminopyrene in mammalian cells. Chem. Res. Toxicol. 20, 1658–1664.

Weimann, A., Belling, D., and Poulsen, H.E. (2002). Quantification of 8-oxo-guanine and guanine as the nucleobase, nucleoside and deoxynucleoside forms in human urine by high-performance liquid chromatography–electrospray tandem mass spectrometry. Nucleic Acids Res. 30, e7–e7.

Wittschieben, J.P., Reshmi, S.C., Gollin, S.M., and Wood, R.D. (2006). Loss of DNA polymerase zeta causes chromosomal instability in mammalian cells. Cancer Res. 66, 134–142.

Wittschieben, J.P., Patil, V., Glushets, V., Robinson, L.J., Kusewitt, D.F., and Wood, R.D. (2010). Loss of DNA polymerase zeta enhances spontaneous tumorigenesis. Cancer Res. 70, 2770–2778.

Wood, R.D. (1997). Nucleotide Excision Repair in Mammalian Cells. J. Biol. Chem. 272, 23465–23468.

Woźniak, K., and Błasiak, J. (2002). Recognition and repair of DNA-cisplatin adducts. Acta Biochim. Pol. 49, 583–596.

Xie, K., Doles, J., Hemann, M.T., and Walker, G.C. (2010). Error-prone translesion synthesis mediates acquired chemoresistance. Proc. Natl. Acad. Sci. 107, 20792–20797.

Yang, W., and Woodgate, R. (2007). What a difference a decade makes: Insights into translesion DNA synthesis. Proc. Natl. Acad. Sci. U. S. A. 104, 15591–15598.

Yeiser, B., Pepper, E.D., Goodman, M.F., and Finkel, S.E. (2002). SOS-induced DNA polymerases enhance long-term survival and evolutionary fitness. Proc. Natl. Acad. Sci. 99, 8737–8741.

Zander, L., and Bemark, M. (2004). Immortalized mouse cell lines that lack a functional Rev3 gene are hypersensitive to UV irradiation and cisplatin treatment. DNA Repair 3, 743–752.

Zhang, Y., Yuan, F., Wu, X., and Wang, Z. (2000a). Preferential incorporation of G opposite template T by the low-fidelity human DNA polymerase iota. Mol. Cell. Biol. 20, 7099–7108.

Zhang, Y., Yuan, F., Wu, X., Wang, M., Rechkoblit, O., Taylor, J.S., Geacintov, N.E., and Wang, Z. (2000b). Error-free and error-prone lesion bypass by human DNA polymerase kappa in vitro. Nucleic Acids Res. 28, 4138–4146.

Zhao, Y., Biertümpfel, C., Gregory, M.T., Hua, Y.-J., Hanaoka, F., and Yang, W. (2012). Structural basis of human DNA polymerase η-mediated chemoresistance to cisplatin. Proc. Natl. Acad. Sci. U. S. A. 109, 7269–7274.

Zhou, Y., Wang, J., Zhang, Y., and Wang, Z. (2010). The catalytic function of the Rev1 dCMP transferase is required in a lesion-specific manner for translesion synthesis and base damage-induced mutagenesis. Nucleic Acids Res. 38, 5036–5046.

Ziv, O., Diamant, N., Shachar, S., Hendel, A., and Livneh, Z. (2012). Quantitative measurement of translesion DNA synthesis in mammalian cells. Methods Mol. Biol. Clifton NJ 920, 529–542.

Appendix A   Effects of translesion DNA polymerases in the 8-oxo-G study

Table A-1   Findings from the control plasmids about the frequency of C->A substitution across from a regular G nucleotide

	Analyzed from reads sequenced in the FORWARD direction			Analyzed from reads sequenced in the REVERSE direction		
Genotypes of the MEF cell lines compared (N= 9 groups)	Adj. P-value	Cohen's d	95% C. I.	Adj. P-value	Cohen's d	95% C. I.
POLI(-/-) versus *POLK*(-/-) *POLI*(-/-) DKO	0.0029	-2.46	-3.9 to -1.02	0.027	-1.6	-2.83 to -0.36
POLK wild type versus *POLK*(-/-) *POLI*(-/-) DKO	0.0024	-2.11	-3.47 to -0.76	0.0015	-21.32	-29.86 to -12.78
POLK(-/-) versus *POLK*(-/-) *POLI*(-/-) DKO	0.0024	-2.59	-4.06 to -1.11	0.0015	-3.15	-4.79 to -1.51
POLK(-/-) versus 2091 *POLI/H/K* TKO	0.0143	-2.34	-3.75 to -0.93	0.0015	-2.96	-4.54 to -1.37
POLK(-/-) versus 2095 *POLI/H/K* TKO	0.023	-1.97	-3.3 to -0.65	0.0054	-2.36	-3.78 to -0.94

Table A-2   Findings from the lesion plasmids about the frequency of C->A substitution across from 8-oxo-G

	Analyzed from reads sequenced in the FORWARD direction			Analyzed from reads sequenced in the REVERSE direction		
Genotypes of the MEF cell lines compared (N=9 groups)	Adj. P-value	Cohen's d	95% C. I.	Adj. P-value	Cohen's d	95% C. I.
REV3L(+/+) *P53*(-/-) vs. *REV3L*(-/-) *P53*(-/-) Nick name: B2 vs. B4(-)18	0.023	-2.05	-3.38 to -0.71	0.015	-1.87	-3.16 to -0.57
POLH wild type versus *POLH*(-/-)	0.0024	-2.97	-4.56 to -1.39	0.0026	-3.03	-4.63 to -1.42
POLH wild type versus *POLH*(-/-) *POLI*(-/-) DKO	0.023	-2.16	-3.53 to -0.80	0.027	-2.07	-3.42 to -0.73

Appendix B   Effects of translesion DNA polymerases in the Pt-GG study

Translesion DNA polymerases did not make any mutations in the control plasmids.

Table B-1 Effects of translesion DNA polymerases on causing CC->AC substitution across from Pt-GG

Genotypes of the MEF cell lines compared (N=9 groups)	Analyzed from reads sequenced in the FORWARD direction			Analyzed from reads sequenced in the REVERSE direction		
	Adj. P-value	Cohen's d	95% C. I.	Adj. P-value	Cohen's d	95% C. I.
REV1 wild type vs. *REV1*(-/-)	0.0024	3.17	1.53-4.82	0.0041	3.22	1.56-4.88
REV3L(+/+) vs. *REV3L*(-/Δ) Nickname: 3(+)6 vs. 4(-)5	0.0018	4.65	2.52-6.79	0.0041	4.86	2.66-7.06
REV3L(+/+) vs. *REV3L*(-/Δ) Nickname: 3(+)6 vs. 4(-)11	0.0018	5.45	3.04-7.87	0.0041	5.12	2.82-7.42
REV3L(+/+) *P53*(-/-) vs. *REV3L*(-/-) *P53*(-/-) Nick name: B2 vs. B4(-)18	0.015	2.17	0.81-3.54	0.037	1.79	0.51-3.06
REV3L(+/+) *P53*(-/-) vs. *REV3L*(-/-) *P53*(-/-) Nick name: B2 vs. B4(-)18	0.0018	3.49	1.75-5.24	0.0041	2.92	1.35-4.49
POLI wild type vs. *POLK*(-/-) *POLI*(-/-) DKO	0.0024	2.81	1.27-4.35	0.0097	2.64	1.15-4.13
POLI(-/-) vs. *POLK*(-/-) *POLI*(-/-) DKO	0.006	2.49	1.04-3.94	0.016	2.45	1.01-3.9
POLI(-/-) vs. 2091 *POLI*(-/-) *POLK*(-/-) *POLH*(-/-) TKO	0.029	1.86	0.57-3.15	0.042	1.97	0.65-3.28
POLI(-/-) vs. 2095 *POLI*(-/-) *POLK*(-/-) *POLH*(-/-) TKO	0.0024	3.08	1.46-4.69	0.0041	3.31	1.62-5.0

Table B- continues	Analyzed from reads sequenced in the FORWARD direction			Analyzed from reads sequenced in the REVERSE direction		
Genotypes of the MEF cell lines compared (N=9 groups)	Adj. P-value	Cohen's d	95% C. I.	Adj. P-value	Cohen's d	95% C. I.
POLH wild type vs. *POLH*(-/-) *POLI*(-/-) DKO	0.038	1.7	0.44-2.95	0.049	1.76	0.49-3.03
POLH(-/-) vs. 2091 *POLI*(-/-) *POLK*(-/-) *POLH*(-/-) TKO	0.0024	2.32	0.92-3.73	0.0041	2.36	0.95-3.78
POLH(-/-) vs. 2095 *POLI*(-/-) *POLK*(-/-) *POLH*(-/-) TKO	0.0018	3.48	1.74-5.22	0.0041	3.59	1.81-5.37
POLK wild type vs. *POLK*(-/-) *POLI*(-/-) DKO	0.0024	2.95	1.37-4.54	0.0041	2.96	1.37-4.54
POLK(-/-) vs. *POLK*(-/-) *POLI*(-/-) DKO	0.01	2.28	0.88-3.67	0.042	1.96	0.64-3.28
POLK(-/-) vs. 2095 *POLI*(-/-) *POLK*(-/-) *POLH*(-/-) TKO	0.004	2.98	1.39-4.57	0.0041	3.93	2.04-5.81
POLH(-/-) *POLI*(-/-) DKO vs. 2091 *POLI*(-/-) *POLK*(-/-) *POLH*(-/-) TKO	0.022	1.71	0.45-2.96	0.06	1.48	0.27-2.69
POLH(-/-) *POLI*(-/-) DKO vs. 2095 *POLI*(-/-) *POLK*(-/-) *POLH*(-/-) TKO	0.0024	3.17	1.52-4.81	0.0041	2.86	1.31-4.42

Table B-2 Effect of Rev3L on causing CC->TC substitution across from Pt-GG

	Analyzed from reads sequenced in the FORWARD direction			Analyzed from reads sequenced in the REVERSE direction		
Genotypes of the MEF cell lines compared (N=9 groups)	Adj. P-value	Cohen's d	95% C. I.	Adj. P-value	Cohen's d	95% C. I.
REV3L(+/+) *P53*(-/-) vs. *REV3L*(-/-) *P53*(-/-) Nick name: B2 vs. B4(-)18	0.0018	2.28	0.88-3.67	0.049	1.64	0.4-2.89

Table B-3 Effect of knocking out both Pol kappa and Pol iota on causing CC->AA substitution across Pt-GG

	Analyzed from reads sequenced in the FORWARD direction			Analyzed from reads sequenced in the REVERSE direction		
Genotypes of the MEF cell lines compared (N=9 groups)	Adj. P-value	Cohen's d	95% C. I.	Adj. P-value	Cohen's d	95% C. I.
POLK wild type vs. *POLK*(-/-) *POLI*(-/-) DKO	0.038	2.13	0.77-3.49	0.042	2.17	0.8-3.53
POLI wild type vs. *POLK*(-/-) *POLI*(-/-) DKO	0.0018	2.27	0.88-3.66	0.0041	2.3	0.9-3.71

Table B-4 Effect of knocking out Pol eta and Pol iota together on causing CC->AT substitution

	Analyzed from reads sequenced in the FORWARD direction			Analyzed from reads sequenced in the REVERSE direction		
Genotypes of the MEF cell lines compared (N=9 groups)	Adj. P-value	Cohen's d	95% C. I.	Adj. P-value	Cohen's d	95% C. I.
POLH wild type vs. *POLH*(-/-) *POLI*(-/-) DKO	0.038	1.85	0.56-3.14	0.042	1.97	0.65-3.28

Appendix C   Mutations arising in the gap region of lesion gap plasmids from the traditional TLS assay

For the 8-oxo-G lesion

Figure C-1 Mutations in the gap regions of 8-oxo-G lesion plasmids from a pair of MEF cell lines

89 sequences were analyzed for the MEF *POLI*(+/+) cell line and 85 sequences were analyzed for the MEF *POLI*(-/-) cell line. The gap sequence is flanked by two dashes. The direction of gap-filling is from left to right or 5' to 3' of the synthesized strand. The 'C' with the highest mutation rate is opposite to 8-oxo-G.

Figure C-2 Mutations in the gap regions of 8-oxo-G lesion plasmid from a pair of BL2 cell lines

93 sequences were analyzed for the BL2 *POLI*(+/+) cell line and 89 sequences were analyzed for the BL2 *POLI*(-/-) cell line. The gap sequence is flanked by two dashes. The direction of gap-filling is from left to right or 5' to 3' of the synthesized strand. The 'C' with the highest mutation rate is opposite to 8-oxo-G.

Table C-1 Numbers of reads with mutations in the gap region of 8-oxo-G lesion plasmids in a pair of MEF cell lines for the *POLI* gene and in a pair of BL2 cell lines for the *POLI* gene. The "C" colored in red is across from the 8-oxo-G lesion. The 12-nucleotide gap region is highlighted in yellow.

Cell line	Total reads	C	C	A	G	G	T	A	G	G	A	C	G	C	A	T	G	C	C	G	G	A	A	T
MEF *POLI* WT	89	1			1				2	1		33	4			1	2							
MEF *POLI*(-/-)	85	1			1							31			1		1							1
BL2 *POLI* WT	93								3			17	4			4	8				4		4	
BL2 *POLI*(-/-)	89								6	1		24	2			7	9				10			

For the Pt-GG lesion
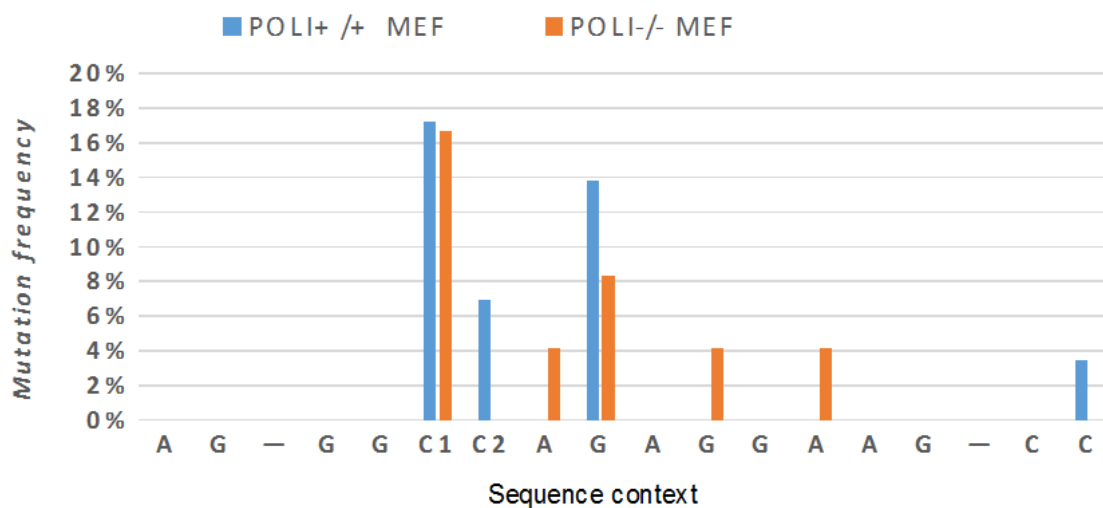
Figure C-3 Mutations in the gap regions of Pt-GG lesion plasmids from a pair of MEF cell lines for studying the effect of Pol iota

29 sequences for the MEF *POLI*(+/+) cell line and 24 sequences were analyzed for the *POLI*(-/-) cell line. The gap sequence is flanked by two dashes. The direction of gap-filling is from left to right of the given sequence or 5' to 3' of the synthesized strand. C1 denotes the first G of Pt-GG bypassed by a DNA polymerase. C2 denotes the second G of Pt-GG bypassed.
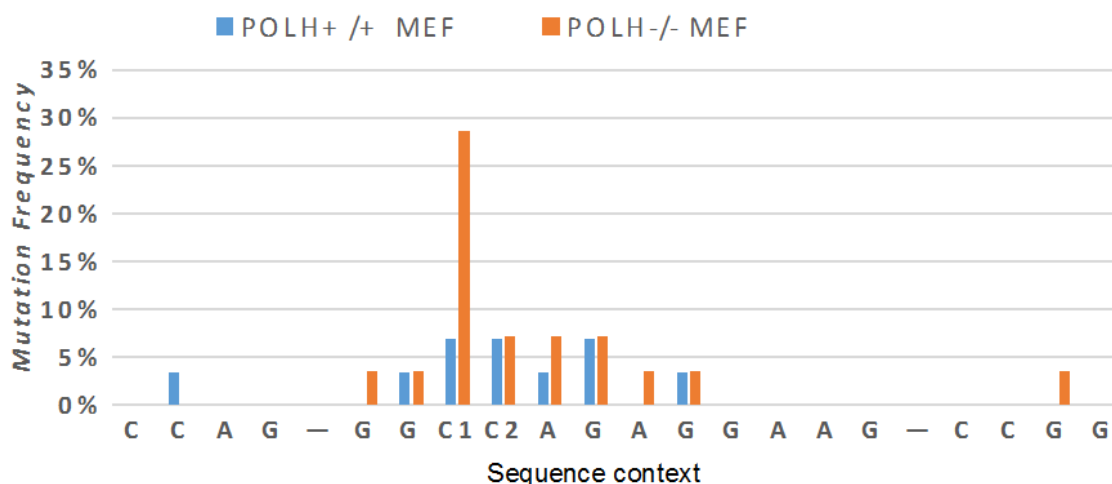
Figure C-4 Mutations in the gap region of Pt-GG lesion plasmids for a pair of MEF cell lines for studying the effect of Pol iota.

29 sequences were analyzed for the MEF *POLH*(+/+) cell line and 28 sequences were analyzed for the MEF *POLH*(-/-) cell line. The gap sequence is flanked by two dashes. The direction of gap-filling is from left to right of the given sequence or 5' to 3' of the synthesized strand. C1 denotes the first G of Pt-GG bypassed by a DNA polymerase. C2 denotes the second G of Pt-GG bypassed.

Table C-2 Numbers of reads with mutations in the gap region of Pt-GG lesion plasmids in a pair of MEF cell lines for the *POLI* gene and in a pair of MEF cell lines for the *POLH* gene. The "CC" colored in red is across from the Pt-GG lesion with C^1 denoting the first insertion step and C^2 denoting the second insertion step to the lesion. The 12 nucleotide gap region is highlighted in yellow.

Cell line	Total reads	C	C	A	G	G	G	C^1	C^2	A	G	A	G	G	A	A	G	C	C	G	G	A	A	T
MEF *POLI* WT	29							5	2	4									1					
MEF *POLI* (-/-)	24							4		1	2		1		1									
MEF *POLH* WT	29		1				1	2	2	1	2		1											
MEF *POLH*(-/-)	28					1	1	8	2	2	2	1	1								1			

Figure C-5 Mutation spectrums of the gap region of the Pt-GG lesion plasmids in a pair of MEF cell lines for the *POLI* gene.

C^1 and C^2 denote the first and the second insertion step to the Pt-GG lesion. The gap region of the lesion plasmid is highlighted in yellow. Each symbol above and below the sequence shown in black is one observation of the mutation. Δ is deletion.

Figure C-6 Mutation spectrums of the gap region of the Pt-GG lesion plasmids in a pair of MEF cell lines for the *POLH* gene.

C^1 and C^2 denote the first and the second insertion step to the Pt-GG lesion. The gap region of the lesion plasmid is highlighted in yellow. Each symbol above and below the sequence shown in black is one observation of the mutation. Δ is deletion. "AC" with the same shading appears in the same sequence.

Appendix D　Summary statistics of the MiSeq run

Table D-1 Summary statistics of the bi-directional 150 bp MiSeq run

	Cycles	Yield	Yield Perfect	Yield <=3 errors	% Aligned to PhiX
Read 1	150	2.8 gigabase	258.0 M	305.6M	10.95
Read 2	150	2.8 gigabase	231.8 M	293.7M	10.74
Total	300	5.6 gigabase	489.8 M	599.2 M	10.84

	% Perfect [num usable cycles]	Error rate[1] (%)	Intensity Cycle	% Intensity Cycle 20	% >= Q30 [2]
Read 1	83.3 [149]	0.29	139	20.3	97.5
Read 2	76.3 [149]	0.48	120	18.0	96.0
Total	79.8	0.39	129	19.1	96.8

Note 1: The error rate was determined from a spiked-in PhiX control sample.

Note 2: Q score, also known as the Phred score was calculated as $-10 log 10(P)$, where P is the error probability. A Q score of 30 means that there is one mistake per every thousand nucleotides sequenced.
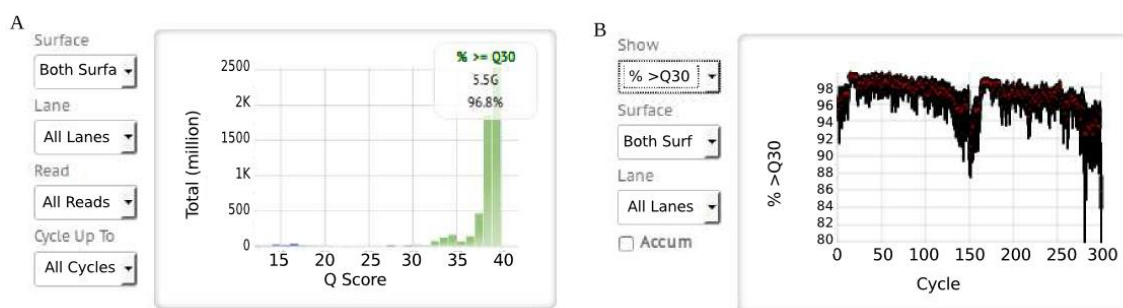
Figure D-1 Q score distributions of the MiSeq run

　　(A) By total reads. (B) By cycle. 96.8% of 5.5 gigabases have Q-scores equal to or greater than 30.

Appendix E Isogenic cell lines compared when deducing effects of translesion DNA polymerases on bypassing 8-oxo-G or Pt-GG

Table E-1 A list of all the isogenic cell lines that were compared

Each cell line was used to set up nine transfection experiments or nine biological replicates for the study of 8-oxo-G and Pt-GG lesions (Appendices G and J). For each pair of cell lines, comparisons were made for the control plasmids and for the lesion plasmids, respectively.

Effect of the translesion DNA polymerase(s) being deduced	Genotypes of the cell lines compared (each cell line has nine biological replicates)	
	Cell line (with the gene of the DNA polymerase in question)	Cell line (deficient for the gene of the DNA polymerase in question)
Pol iota	MEF *POLI*(+/+)	MEF *POL*(-/-)
	BL2 *POLI*(+/+)	BL2 *POLI*(-/-)
Pol eta	MEF *POLH*(+/+)	MEF *POLH*(-/-)
	MEF *POLK*(-/-) *POLI*(-/-)	2091 *POLH*(-/-) *POLI*(-/-) *POLK*(-/-)
	MEF *POLK*(-/-) *POLI*(-/-)	2095 *POLH*(-/-) *POLI*(-/-) *POLK*(-/-)
Pol kappa	MEF *POLK*(+/+)	MEF *POLK*(-/-)
	MEF *POLH*(-/-) *POLI*(-/-)	2091 *POLH*(-/-) *POLI*(-/-) *POLK*(-/-)
	MEF *POLH*(-/-) *POLI*(-/-)	2095 *POLH*(-/-) *POLI*(-/-) *POLK*(-/-)
Rev1	MEF *REV1*(+/+)	MEF *REV1*(-/-)
Rev3	MEF *REV3L*(+/+) *P53*(-/-) or B2	MEF *REV3L*(-/-) *P53*(-/-) or B4(-)9
	MEF *REV3L*(+/+) *P53*(-/-) or B2	MEF *REV3L*(-/-) *P53*(-/-) or B4(-)18
	MEF *REV3L*(+/lox) Cre or 3(+)6	MEF *REV3L*(-/lox) Cre or 4(-)5
	MEF *REV3L*(+/lox) Cre or 3(+)6	MEF *REV3L*(-/lox) Cre or 4(-)11
Pol iota and Pol eta	MEF *POLI*(+/+)	MEF *POLH*(-/-) *POLI*(-/-)
	MEF *POLH*(+/+)	MEF *POLH*(-/-) *POLI*(-/-)
	MEF *POLK*(-/-)	2091 *POLH*(-/-) *POLI*(-/-) *POLK*(-/-)
	MEF *POLK*(-/-)	2095 *POLH*(-/-) *POLI*(-/-) *POLK*(-/-)
Pol iota and Pol kappa	MEF *POLI*(+/+)	MEF *POLI*(-/-) *POLK*(-/-)
	MEF *POLK*(+/+)	MEF *POLI*(-/-) *POLK*(-/-)
	MEF *POLH*(-/-)	2091 *POLH*(-/-) *POLI*(-/-) *POLK*(-/-)
	MEF *POLH*(-/-)	2095 *POLH*(-/-) *POLI*(-/-) *POLK*(-/-)
Pol eta and Pol kappa	MEF *POLI*(-/-)	2091 *POLH*(-/-) *POLI*(-/-) *POLK*(-/-)
	MEF *POLI*(-/-)	2095 *POLH*(-/-) *POLI*(-/-) *POLK*(-/-)

Appendix F No. of reads for DNA samples in the 8-oxo-G study

Transfection Experiment	Total Reads	Reads of Control Plasmids	Reads of Lesion Plasmids	Transfection Experiment	Total Reads	Reads of Control Plasmids	Reads of Lesion Plasmids
L1R1	5193	3394	1799	L2R1	4229	2767	1462
L1R2	5523	3485	2038	L2R2	6307	3982	2325
L1R3	4881	3121	1760	L2R3	6257	3954	2303
L1R4	4938	3161	1777	L2R4	7310	4648	2662
L1R5	4441	2854	1587	L2R5	10006	6828	3178
L1R6	3722	2382	1340	L2R6	6933	4394	2539
L1R7	4963	3086	1877	L2R7	6245	3877	2368
L1R8	5413	3392	2021	L2R8	5365	3370	1995
L1R9	4619	2994	1625	L2R9	4736	3137	1599
L1R10	5820	3490	2330	L2R10	5485	3504	1981
L1R11	4724	3046	1678	L2R11	4675	3020	1655
L1R12	4256	2735	1521	L2R12	6966	4298	2668
L1R13	5626	3538	2088	L2R13	7147	4401	2746
L1R14	4451	2822	1629	L2R14	6961	4332	2629
L1R15	4418	2786	1632	L2R15	4672	2927	1745
L1R16	5236	3332	1904	L2R16	6050	3809	2241
L1R17	6182	3936	2246	L2R17	7911	5009	2902
L1R18	6631	4137	2494	L2R18	7432	4782	2650
L1R19	9834	6870	2964	L2R19	5998	3858	2140
L1R20	4456	2820	1636	L2R20	5008	3221	1787
L1R21	4446	2824	1622	L2R21	8145	5177	2968
L1R22	4543	2932	1611	L2R22	3894	2625	1269
L1R23	4770	3032	1738	L2R23	7058	4571	2487
L1R24	4439	2811	1628	L2R24	7232	4569	2663

Transfection Experiment	Total Reads	Reads of Control Plasmids	Reads of Lesion Plasmids	Transfection Experiment	Total Reads	Reads of Control Plasmids	Reads of Lesion Plasmids
L3R1	5280	3366	1914	L4R1	5567	3442	2125
L3R2	5888	3750	2138	L4R2	5355	3410	1945
L3R3	5577	3500	2077	L4R3	6128	3795	2333
L3R4	5392	3378	2014	L4R4	6345	4000	2345
L3R5	3687	2361	1326	L4R5	8271	5220	3051
L3R6	3656	2376	1280	L4R6	7694	5142	2552
L3R7	3592	2314	1278	L4R7	6325	4051	2274
L3R8	4043	2584	1459	L4R8	7357	4734	2623
L3R9	3740	2543	1197	L4R9	7111	4623	2488
L3R10	3680	2447	1233	L4R10	6945	4442	2503
L3R11	4387	2992	1395	L4R11	7669	4920	2749
L3R12	3997	2767	1230	L4R12	8223	5000	3223
L3R13	3974	2667	1307	L4R13	7401	4459	2942
L3R14	4160	2816	1344	L4R14	8379	5170	3209
L3R15	3915	2403	1512	L4R15	6419	4017	2402
L3R16	4949	3107	1842	L4R16	6199	3934	2265

Transfection Experiment	Total Reads	Reads of Control Plasmids	Reads of Lesion Plasmids
L3R17	4010	2522	1488
L3R18	4354	2784	1570
L3R19	4610	2958	1652
L3R20	4327	2667	1660
L3R21	4031	2547	1484
L3R22	3548	2260	1288
L3R23	3821	2378	1443
L3R24	3995	2493	1502

Transfection Experiment	Total Reads	Reads of Control Plasmids	Reads of Lesion Plasmids
L4R17	9094	5758	3336
L4R18	4918	3027	1891
L4R19	7306	4534	2772
L4R20	6398	4051	2347
L4R21	8753	5605	3148
L4R22	7910	5026	2884
L4R23	8483	5463	3020
L4R24	6358	4013	2345

Transfection Experiment	Total Reads	Reads of Control Plasmids	Reads of Lesion Plasmids
L5R1	7351	4998	2353
L5R2	8251	5682	2569
L5R3	7224	4420	2804
L5R4	9021	5513	3508
L5R5	7985	4954	3031
L5R6	7033	4497	2536
L5R7	9087	5648	3439
L5R8	7589	4745	2844
L5R9	8477	5594	2883
L5R10	7374	4816	2558
L5R11	10736	7024	3712
L5R12	8844	5689	3155
L5R13	9324	5913	3411
L5R14	9991	6476	3515
L5R15	5919	3739	2180
L5R16	10110	6401	3709
L5R17	7882	6912	970
L5R18	7006	4505	2501
L5R19	7136	4722	2414
L5R20	6720	4344	2376
L5R21	9912	6279	3633
L5R22	10269	6376	3893
L5R23	9522	5901	3621
L5R24	9009	5856	3153

Transfection Experiment	Total Reads	Reads of Control Plasmids	Reads of Lesion Plasmids
L6R1	10167	6464	3703
L6R2	14701	9212	5489
L6R3	10725	6556	4169
L6R4	12411	7734	4677
L6R5	12677	7970	4707
L6R6	12833	7822	5011
L6R7	13942	8520	5422
L6R8	13615	8249	5366
L6R9	12618	8989	3629
L6R10	10513	7405	3108
L6R11	14519	10347	4172
L6R12	11808	7271	4537
L6R13	11314	6809	4505
L6R14	12280	7137	5143
L6R15	14372	9002	5370
L6R16	11779	7439	4340
L6R17	13535	8581	4954
L6R18	12023	7359	4664
L6R19	13418	8231	5187
L6R20	10946	6750	4196
L6R21	15456	9860	5596
L6R22	12554	8020	4534
L6R23	11146	7134	4012
L6R24	14447	9048	5399

Transfection Experiment	Total Reads	Reads of Control Plasmids	Reads of Lesion Plasmids
L7R1	12256	7607	4649
L7R2	12487	7794	4693
L7R3	11907	7548	4359
L7R4	11752	7557	4195
L7R5	11474	7306	4168
L7R6	12120	7797	4323

Transfection Experiment	Total Reads	Reads of Control Plasmids	Reads of Lesion Plasmids
L22R1	5312	3437	1875
L22R2	6086	3844	2242
L22R3	6395	4053	2342
L22R4	5998	3822	2176
L22R5	6670	4260	2410
L22R6	6360	3992	2368

	Total Reads	Reads of Control Plasmids	Reads of Lesion Plasmids		Total Reads	Reads of Control Plasmids	Reads of Lesion Plasmids
L7R7	9487	5913	3574	L22R7	6808	4340	2468
L7R8	13455	8305	5150	L22R8	5962	3874	2088
L7R9	13440	8535	4905	L22R9	6888	4372	2516
L7R10	11052	6926	4126	L22R10	6118	3948	2170
L7R11	12217	7663	4554	L22R11	6427	4047	2380
L7R12	12457	7756	4701	L22R12	7113	4330	2783
L7R13	19743	12271	7472	L22R13	7428	4489	2939
L7R14	13451	8474	4977	L22R14	8403	5204	3199
L7R15	12535	7849	4686	L22R15	1535	1058	477
L7R16	12181	7568	4613	L22R16	1700	1191	509
L7R17	12666	8063	4603	L22R17	4440	3932	508
L7R18	15353	9618	5735	L22R18	4322	3841	481
L7R19	5311	3378	1933	L22R19	4087	3605	482
L7R20	5174	3359	1815	L22R20	1419	996	423
L7R21	5684	3694	1990	L22R21	1628	1120	508
L7R22	5236	3390	1846	L22R22	1415	991	424
L7R23	5091	3320	1771	L22R23	1367	960	407
L7R24	6371	4144	2227	L22R24	1602	1144	458

Note:

L22R15 and L22R16 are no DNA negative PCR controls whereby no DNA was added in PCR reactions.

L22R17, L22R18 and L22R19 are positive PCR controls whereby intact double stranded control plasmids were added in PCR reactions for determining experimental errors.

L22R20, L22R21 and L22R22 are negative controls for transfection whereby control gap plasmids were added to MEF cells without transfection reagent.

L22R23 and L22R24 are negative controls for transfection whereby control gap plasmids were added to BL2 cells without transfection reagent.

Appendix G Cell lines used for transfection experiments in the 8-oxo-G study

Cell line	Transfection experiments
MEF *POLI*(+/+)	L1R1 to L1R9
MEF *POLI*(-/-)	L1R10 to L1R18
BL2 *POLI*(+/+)	L1R19 to L1R24, L2R1 to L2R4
BL2 *POLI*(-/-)	L2R5 to L2R14
MEF *POLH*(+/+)	L2R15 to L2R23
MEF *POLH*(-/-)	L2R24, L3R1 to L3R8
MEF *POLK*(+/+)	L3R9 to L3R17
MEF *POLK*(-/-)	L3R18 to L3R24, L4R1, L4R2
MEF *REV1*(+/+)	L4R3 to L4R11
MEF *REV1*(-/-)	L4R12 to L4R20
MEF *POLH*(-/-) *POLI*(-/-)	L4R21 to L4R24, L5R1 to L5R5
MEF *POLH*(-/-) *POLI*(-/-) *POLK*(-/-), 2091	L5R6 to L5R14
MEF *POLH*(-/-) *POLI*(-/-) *POLK*(-/-), 2095	L5R15 to L5R23
MEF *POLK*(-/-) *POLI*(-/-), 2092	L5R24, L6R1 to L6R8
MEF *REV3L*(+/Δ), nickname: 3(+)6	L6R9 to L6R17
MEF *REV3L*(-/Δ), nickname: 4(-)5	L6R18 to L6R24, L7R1, L7R2
MEF *REV3L*(-/Δ), nickname: 4(-)11	L7R3 to L7R11
MEF *REV3L*(+/+), P53(-/-), nickname: B2	L7R12 to L7R20
MEF *REV3L*(+/+), P53(-/-), nickname: B4-9	L7R21 to L7R24, L22R1 to L22R5
MEF *REV3L*(+/+), P53(-/-), nickname: B4-18	L22R6 to L22R14

Appendix H   Mutation frequencies of control plasmids and lesion plasmids in the 8-oxo-G study for each cell line

Cell line	Average mutation frequency across from a regular G in control plasmids (N=9 samples)	Average mutation frequency across from an 8-oxo-G in lesion plasmids (N=9 samples)
MEF *POLI*(+/+)	0.019265±0.003387	0.901709±0.020447
MEF *POLI*(-/-)	0.013826±0.002641	0.921682±0.025595
BL2 *POLI*(+/+)	0.010293±0.007336	0.918242±0.043828
BL2 *POLI*(-/-)	0.015254±0.005925	0.903451±0.031048
MEF *POLH*(+/+)	0.012623±0.004703	0.868926±0.029094
MEF *POLH*(-/-)	0.014634±0.006322	0.929683±0.010711
MEF *POLK*(+/+)	0.008282±0.003878	0.927235±0.024269
MEF *POLK*(-/-)	0.007575±0.005674	0.932431±0.014121
MEF *POLH*(-/-) *POLI*(-/-)	0.018589±0.005959	0.919158±0.016402
MEF *POLH*(-/-) *POLI*(-/-) *POLK*(-/-), 2091	0.023589±0.004759	0.913149±0.004392
MEF *POLH*(-/-) *POLI*(-/-) *POLK*(-/-), 2095	0.018639±0.002599	0.919177±0.015477
MEF *POLK*(-/-) *POLI*(-/-), 2092	0.01898±0.006801	0.931155±0.009086
MEF *REV1*(+/+)	0.019328±0.005146	0.906739±0.011112
MEF *REV1*(-/-)	0.01682±0.00477	0.914846±0.011736
MEF *REV3L*(+/Δ), nickname: 3(+)6	0.020843±0.003397	0.885401±0.020573
MEF *REV3L*(-/Δ), nickname: 4(-)5	0.019108±0.001546	0.908886±0.009407
MEF *REV3L*(-/Δ), nickname: 4(-)11	0.020931±0.004855	0.906337±0.007816
MEF *REV3L*(+/+), P53(-/-), nickname: B2	0.025063±0.008512	0.896716±0.020351
MEF *REV3L*(+/+), P53(-/-), nickname: B4-9	0.025855±0.01121	0.897937±0.03203
MEF *REV3L*(+/+), P53(-/-), nickname: B4-18	0.016891±0.005043	0.926601±0.011012

Appendix I No. of reads for DNA samples in the Pt-GG study

Transfection Experiment	Total Reads	Reads of Control Plasmids	Reads of Lesion Plasmids	Transfection Experiment	Total Reads	Reads of Control Plasmids	Reads of Lesion Plasmids
L8L15R1	36024	17863	18161	L9L16R1	64482	37988	26494
L8L15R2	36207	18928	17279	L9L16R2	67825	37640	30185
L8L15R3	35457	19361	16096	L9L16R3	68797	39394	29403
L8L15R4	40351	20797	19554	L9L16R4	65804	37270	28534
L8L15R5	40306	19659	20647	L9L16R5	68370	37120	31250
L8L15R6	41563	23831	17732	L9L16R6	74466	44853	29613
L8L15R7	46500	26165	20335	L9L16R7	71140	41356	29784
L8L15R8	44003	25446	18557	L9L16R8	76563	46366	30197
L8L15R9	51500	28085	23415	L9L16R9	65740	36375	29365
L8L15R10	42392	28024	14368	L9L16R10	68844	40462	28382
L8L15R11	55695	30419	25276	L9L16R11	82529	49950	32579
L8L15R12	53285	30590	22695	L9L16R12	71837	39618	32219
L8L15R13	54559	29274	25285	L9L16R13	81171	47970	33201
L8L15R14	53104	30179	22925	L9L16R14	70830	40003	30827
L8L15R15	48927	29555	19372	L9L16R15	62436	35253	27183
L8L15R16	55594	32009	23585	L9L16R16	71543	43289	28254
L8L15R17	58006	34832	23174	L9L16R17	71266	41221	30045
L8L15R18	60290	39704	20586	L9L16R18	77326	47557	29769
L8L15R19	51834	26436	25398	L9L16R19	75568	43406	32162
L8L15R20	64230	44079	20151	L9L16R20	67533	40635	26898
L8L15R21	63154	42904	20250	L9L16R21	74950	41790	33160
L8L15R22	46743	29133	17610	L9L16R22	68953	40884	28069
L8L15R23	51651	25681	25970	L9L16R23	69949	37927	32022
L8L15R24	57671	34374	23297	L9L16R24	70154	39769	30385

Transfection Experiment	Total Reads	Reads of Control Plasmids	Reads of Lesion Plasmids	Transfection Experiment	Total Reads	Reads of Control Plasmids	Reads of Lesion Plasmids
L10L17R1	56715	41330	15385	L11L18R1	46782	34241	12541
L10L17R2	78852	55635	23217	L11L18R2	43657	30693	12964
L10L17R3	66577	38615	27962	L11L18R3	48643	35021	13622
L10L17R4	66259	39749	26510	L11L18R4	52460	37590	14870
L10L17R5	40689	19015	21674	L11L18R5	52404	37337	15067
L10L17R6	42661	20383	22278	L11L18R6	46875	29987	16888
L10L17R7	47258	17672	29586	L11L18R7	49028	32675	16353
L10L17R8	52330	21168	31162	L11L18R8	48714	32294	16420
L10L17R9	54341	21079	33262	L11L18R9	54014	36173	17841
L10L17R10	52122	20580	31542	L11L18R10	51793	33239	18554
L10L17R11	58771	25643	33128	L11L18R11	51801	33664	18137
L10L17R12	55036	24683	30353	L11L18R12	55341	38338	17003
L10L17R13	59710	27163	32547	L11L18R13	58658	38585	20073
L10L17R14	54427	22966	31461	L11L18R14	52695	33694	19001
L10L17R15	52433	23587	28846	L11L18R15	50965	33755	17210
L10L17R16	59790	26321	33469	L11L18R16	56964	38407	18557

Transfection Experiment	Total Reads	Reads of Control Plasmids	Reads of Lesion Plasmids
L10L17R17	53811	27940	25871
L10L17R18	57204	26774	30430
L10L17R19	57464	27886	29578
L10L17R20	50519	24620	25899
L10L17R21	58053	27008	31045
L10L17R22	52774	23931	28843
L10L17R23	41139	26785	14354
L10L17R24	39741	25207	14534

Transfection Experiment	Total Reads	Reads of Control Plasmids	Reads of Lesion Plasmids
L11L18R17	61006	40607	20399
L11L18R18	54630	40289	14341
L11L18R19	56546	36449	20097
L11L18R20	38641	34745	3896
L11L18R21	60615	38777	21838
L11L18R22	51836	32653	19183
L11L18R23	56257	35966	20291
L11L18R24	57467	38795	18672

Transfection Experiment	Total Reads	Reads of Control Plasmids	Reads of Lesion Plasmids
L12L19R1	66628	39811	26817
L12L19R2	63410	41054	22356
L12L19R3	59615	39642	19973
L12L19R4	67930	43535	24395
L12L19R5	72946	41709	31237
L12L19R6	83461	56198	27263
L12L19R7	65460	41530	23930
L12L19R8	77384	49889	27495
L12L19R9	77126	48229	28897
L12L19R10	78838	50848	27990
L12L19R11	73492	42849	30643
L12L19R12	71107	42338	28769
L12L19R13	76450	46819	29631
L12L19R14	42905	15582	27323
L12L19R15	69864	45212	24652
L12L19R16	68766	42805	25961
L12L19R17	73579	45959	27620
L12L19R18	73856	45505	28351
L12L19R19	72559	49140	23419
L12L19R20	67281	44924	22357
L12L19R21	74570	51622	22948
L12L19R22	73241	44819	28422
L12L19R23	61487	38023	23464
L12L19R24	73060	44784	28276

Transfection Experiment	Total Reads	Reads of Control Plasmids	Reads of Lesion Plasmids
L13L20R1	68368	46307	22061
L13L20R2	70656	45948	24708
L13L20R3	66731	40607	26124
L13L20R4	73834	46485	27349
L13L20R5	70185	44272	25913
L13L20R6	75937	45527	30410
L13L20R7	68043	43317	24726
L13L20R8	73955	49054	24901
L13L20R9	78625	51148	27477
L13L20R10	75918	49228	26690
L13L20R11	72757	46242	26515
L13L20R12	78867	52278	26589
L13L20R13	71363	40508	30855
L13L20R14	66372	41795	24577
L13L20R15	72626	47232	25394
L13L20R16	78395	51335	27060
L13L20R17	81469	49601	31868
L13L20R18	59811	38608	21203
L13L20R19	73654	47798	25856
L13L20R20	68656	45474	23182
L13L20R21	73762	46316	27446
L13L20R22	62216	37289	24927
L13L20R23	43166	19251	23915
L13L20R24	52392	21890	30502

Transfection Experiment	Total Reads	Reads of Control Plasmids	Reads of Lesion Plasmids
L14L21R1	43780	19325	24455
L14L21R2	38717	20072	18645
L14L21R3	48448	19932	28516
L14L21R4	47107	21231	25876
L14L21R5	48792	22446	26346
L14L21R6	56677	25800	30877

Transfection Experiment	Total Reads	Reads of Control Plasmids	Reads of Lesion Plasmids
L23L24R1	62856	49312	13544
L23L24R2	55453	42078	13375
L23L24R3	65605	50502	15103
L23L24R4	64659	49128	15531
L23L24R5	51577	44095	7482
L23L24R6	76682	56799	19883

Transfection Experiment	Total Reads	Reads of Control Plasmids	Reads of Lesion Plasmids	Transfection Experiment	Total Reads	Reads of Control Plasmids	Reads of Lesion Plasmids
L14L21R7	51910	23832	28078	L23L24R7	70990	56094	14896
L14L21R8	50361	25094	25267	L23L24R8	68275	52921	15354
L14L21R9	50758	26100	24658	L23L24R9	69109	52149	16960
L14L21R10	52261	24623	27638	L23L24R10	62660	45472	17188
L14L21R11	52510	25159	27351	L23L24R11	64814	47544	17270
L14L21R12	54589	27700	26889	L23L24R12	69844	51431	18413
L14L21R13	57830	28064	29766	L23L24R13	35231	29816	5415
L14L21R14	45336	17483	27853	L23L24R14	16229	10119	6110
L14L21R15	50429	25363	25066	L23L24R15	32519	28990	3529
L14L21R16	54078	26037	28041	L23L24R16	70670	45932	24738
L14L21R17	46502	30175	16327	L23L24R17	71456	52494	18962
L14L21R18	48119	32340	15779	L23L24R18	71659	47168	24491
L14L21R19	46204	29511	16693	L23L24R19	45740	40544	5196
L14L21R20	41429	27453	13976	L23L24R20	30318	26105	4213
L14L21R21	45934	31675	14259	L23L24R21	35904	31754	4150
L14L21R22	41911	27266	14645	L23L24R22	49326	38090	11236
L14L21R23	42233	26767	15466	L23L24R23	31946	26151	5795
L14L21R24	44787	29171	15616	L23L24R24	44769	34364	10405

Note:

L23/L24R13, L23/L24R14 and L23/L24R15 are no DNA negative PCR controls whereby no DNA was added in PCR reactions.

L23R16, L23R17 and L23R18 are positive control PCR reactions whereby intact and fully double stranded control plasmids were added in PCR reactions to determine experimental errors. Similarly, L24R16, L24R17 and L24R18 are positive control PCR reactions whereby intact and fully double stranded lesion plasmids (without Pt-GG crosslinks) were added in PCR reactions.

L23/L24R19, L23/L24R20 and L23/L24R21 are negative control reactions for transfection where control and lesion gap plasmid mixtures were added and incubated with MEF cells without using transfection reagent.

Similarly, L23/L24R22, L23/L24R23, L23/L24R24 are negative control reactions for transfection where control and lesion gap plasmid mixtures were added and incubated with BL2 cells without using transfection reagent.

Appendix J   Cell lines used for transfection experiments in the Pt-GG study

Cell line	Transfection experiments
MEF *POLI*(+/+)	L8L15R1 to L8L15R9
MEF *POLI*(-/-)	L8L15R10 to L8L15R18
BL2 *POLI*(+/+)	L8L15R19 to L8L15R24, L9L16R1 to L9L16R3
BL2 *POLI*(-/-)	L9L16R4 to L9L16R12
MEF *POLH*(+/+)	L9L16R13 to L9L16R21
MEF *POLH*(-/-)	L9L16R22 to L9L16R24, L10L17R1 to L10L17R6
MEF *POLK*(+/+)	L10L17R7 to L10L17R15
MEF *POLK*(-/-)	L10L17R16 to L10L17R24
MEF *REV1*(+/+)	L11L18R1 to L11L18R9
MEF *REV1*(-/-)	L11L18R10 to L11L18R18
MEF *POLH*(-/-) *POLI*(-/-)	L11L18R19 to L11L18R24 L12L19R1 to L12L19R3
MEF *POLH*(-/-) *POLI*(-/-) *POLK*(-/-), 2091	L12L19R4 to L12L19R12
MEF *POLH*(-/-) *POLI*(-/-) *POLK*(-/-), 2095	L12L19R13 to L12L19R21
MEF *POLK*(-/-) *POLI*(-/-), 2092	L12L19R22 to L12L19R24 L13L20R1 to L13L20R6
MEF *REV3L*(+/Δ), nickname: 3(+)6	L13L20R7 to L13L20R15
MEF *REV3L*(-/Δ), nickname: 4(-)5	L13L20R16 to L13L20R24
MEF *REV3L*(-/Δ), nickname: 4(-)11	L14L21R1 to L14L21R9
MEF *REV3L*(+/+), P53(-/-), nickname: B2	L14L21R10 to L14L21R18
MEF *REV3L*(+/+), P53(-/-), nickname: B4-9	L14L21R19 to L14R21R24 L23L24R1 to L23R24R3
MEF *REV3L*(+/+), P53(-/-), nickname: B4-18	L23L24R4 to L23L24R12

Appendix K Mutation frequencies of control plasmids and lesion plasmids in the Pt-GG study for each cell line

Cell lines	Average mutation frequencies of **control plasmids** across from GG (N=9 samples)		Average mutation frequencies of **lesion plasmids** across from Pt-GG (N=9 samples)	
	The 1st insertion step	The 2nd insertion step	The 1st insertion step	The 2nd insertion step
MEF *POLI*(+/+)	0.00633582 ±0.00060603	0.00211604 ±0.00042729	0.00797014 ±0.00235359	0.00325844 ±0.00081405
MEF *POLI*(-/-)	0.00623971 ±0.00026943	0.00210229 ±0.00033156	0.00676418 ±0.00178312	0.00335374 ±0.00046932
BL2 *POLI*(+/+)	0.00672665 ±0.00037447	0.00211498 ±0.00050253	0.00357752 ±0.00104655	0.00274966 ±0.0008659
BL2 *POLI*(-/-)	0.00697197 ±0.0004663	0.00213745 ±0.0002034	0.00453663 ±0.00094203	0.00373258 ±0.00065256
MEF *POLH*(+/+)	0.00671445 ±0.00063946	0.00213002 ±0.00023077	0.01264681 ±0.00534483	0.00498744 ±0.00091888
MEF *POLH*(-/-)	0.00659034 ±0.00070424	0.00223973 ±0.0002914	0.00967711 ±0.00290917	0.00426218 ±0.0004818
MEF *POLK*(+/+)	0.00631168 ±0.00053691	0.00197384 ±0.00028981	0.00817649 ±0.00205302	0.00405642 ±0.00065313
MEF *POLK*(-/-)	0.00647793 ±0.00061005	0.00205539 ±0.00054222	0.00692088 ±0.00098325	0.00403449 ±0.0004382
MEF *POLH*(-/-) *POLI*(-/-)	0.00669208 ±0.00032507	0.00218667 ±0.00035575	0.00747425 ±0.00113811	0.0040025 ±0.00033642
MEF *POLK*(-/-) *POLI*(-/-), 2092	0.00651591 ±0.00044	0.00215294 ±0.00036743	0.00506356 ±0.0009081	0.00361153 ±0.00053964
MEF *POLH*(-/-) *POLI*(-/-) *POLK*(-/-), 2091	0.00665362 ±0.00021903	0.00221712 ±0.00022728	0.00501292 ±0.00076192	0.00345792 ±0.00056011
Table continues in the next page …				

Cell line	Average mutation frequencies of **control plasmids** across from GG (N=9 samples)		Average mutation frequencies of **lesion plasmids** across from Pt-GG (N=9 samples)	
	The 1st insertion step (C^1)	The 2nd insertion step (C^2)	The 1st insertion step (C^1)	The 2nd insertion step (C^2)
MEF *POLH*(-/-) *POLI*(-/-) *POLK*(-/-), 2095	0.00657383 ±0.00041924	0.00227921 ±0.0005008	0.00409557 ±0.00071602	0.00307922 ±0.00065915
MEF *REV1*(+/+)	0.00685436 ±0.00052085	0.00213745 ±0.00030857	0.00756895 ±0.00153204	0.00431359 ±0.00063232
MEF *REV1*(-/-)	0.00652176 ±0.00024146	0.00214813 ±0.00035795	0.00435796 ±0.00056597	0.00338934 ±0.0007335
MEF *REV3L*(+/Δ), 3(+)6	0.00658979 ±0.00043858	0.00206947 ±0.00023479	0.01030133 ±0.00235666	0.00440965 ±0.00076919
MEF *REV3L*(-/Δ), 4(-)5	0.00678478 ±0.00053557	0.00217725 ±0.00046891	0.00490431 ±0.00095976	0.00389058 ±0.00105612
MEF *REV3L*(-/Δ), 4(-)11	0.00678595 ±0.00036631	0.0020206 ±0.00028687	0.00413251 ±0.00091898	0.00370398 ±0.00064044
MEF *REV3L*(+/+) P53(-/-), B2	0.00659605 ±0.00034117	0.00205079 ±0.00025514	0.00651752 ±0.00073177	0.00400263 ±0.00060623
MEF *REV3L*(+/+) P53(-/-), B4-9	0.00665444 ±0.0050896	0.00217355 ±0.00023279	0.00498804 ±0.00057864	0.00395544 ±0.00076486
MEF *REV3L*(+/+) P53(-/-), B4-18	0.00656054 ±0.00034216	0.00227396 ±0.00027169	0.00457931 ±0.00109652	0.00427594 ±0.00105971

Appendix K continues …

Curriculum Vitae

Lizhen Guo

Education

2012 - 2015   Master of Science in Biochemistry, Western University

2003 – 2008   Honours Bachelor of Science, University of Toronto

Research Experience

2012 – 2015   Research Assistant
Dr. Hong Ling's Lab, Dept of Biochemistry
Schulich School of Medicine and Dentistry, Western U.

2009-2012   Research Scientist
Cytogenomics Inc. 150-700 Collip Circle, London, Ontario

2006-2008   Laboratory Assistant
Cytogenetics Laboratory, the Hospital for Sick Children

Others

2013 – 2015   Councilor, Society of Graduate Students (SOGS), Western U.

2014 – 2015   Chairperson of the Academic Committee, SOGS, Western U.

2012 – 2013   Member of the Outdoors Club, Western U.

Scholarships

2013 – 2014   Western Graduate Research Scholarship

2013   Graduate Thesis Research Award