Functional DNA Repair Capacity Assays: A Focus On Base Excision Repair

By

Isaac Alexander Chaim

B.S. Biology
Universidad Simón Bolívar - Caracas, Venezuela - 2008

Submitted to the Department of Biological Engineering
in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2016

Signature of Author..... Signature redacted
Department of Biological Engineering
February 19th, 2016

Certified by ..... Signature redacted .....
Leona D. Samson
Professor of Biological Engineering and Biology
Thesis Supervisor

Accepted by ..... Signature redacted .....
Forest White
Professor of Biological Engineering
Chair, Biological Engineering Graduate Committee

Functional DNA Repair Capacity Assays: A Focus On Base Excision Repair

By

Isaac Alexander Chaim

Abstract

The integrity of our DNA is challenged by roughly 100,000 lesions per cell on a daily basis. Failure to cope with DNA damage can lead to cancer, immunodeficiency and degenerative disease. Quantitating and understanding an individual's DNA repair capacity may enable us to predict and prevent disease in a personalized manner. Base Excision Repair (BER) is known for the recognition and repair of endogenous and exogenous mutagenic non-helix-distorting lesions produced by DNA base alkylation, deamination and oxidation. BER is initiated by the action of one of eleven DNA glycosylases known-to-date. Many studies have shown that levels of these glycosylases can vary between individuals, suggesting a basis for inter-individual differences in DNA repair capacity. Moreover, the methods for measuring DNA repair capacity used so far are cumbersome, time consuming, low throughput and only allow for the analysis of one glycosylase at a time.

We have taken a fluorescence-based multiplex flow-cytometric host cell reactivation assay wherein the activity of several DNA glycosylases and their immediate downstream endonuclease (APE1) can be tested simultaneously, at single-cell resolution, under physiological conditions. Taking advantage of the transcriptional properties of several DNA lesions we have designed and engineered specific fluorescent reporter plasmids for OGG1, AAG, MUTYH, UNG and APE1. Inter-individual differences in DNA repair capacity of a panel of cell lines derived from healthy individuals have been measured. Regression models that incorporate these measurements have been developed in order to predict cellular sensitivity to the chemotherapeutic and DNA damaging agents 5-FU, H_2O_2 and MMS, with the interest of understanding the contributions that these differences can have on personalized disease prevention and treatment. Finally, we have conducted a pilot population study with 56 healthy subjects where we implemented all the methods developed in order to determine the feasibility of measuring DNA repair capacity variations in a healthy human population. Additionally, we report the discovery of a novel *in vivo* role of the TC-NER pathway in the repair of the lipid-peroxidation product, $3,N^4$-etheno-cytosine.

Thesis Supervisor: Leona D. Samson
Title: Professor of Biological Engineering and Biology

ב"ה

Acknowledgments

First I would like to thank Leona Samson, who has been an outstanding advisor for my scientific career and for life in general. I will always be grateful to her for taking me into her lab and believing in me, and for giving me the scientific freedom needed to make mistakes and learn from them.

I am grateful to the members of my thesis committee Bevin Engelward, John Essigmann and Pete Dedon who were never short on good advice and whose doors were always open for me.

I am very appreciative of Doug Lauffenburger, whose wise advice shapes the careers of more scientists that he will ever imagine.

This journey would have been impossible without current and former members of the Samson Lab. I was lucky enough to find a lab where everyone was willing to share and help, and simply be themselves. In particular I want to thank Jennifer Calvo, Nona Abolhassani, Dragony Fu, Carrie Margulies and José McFaline who were great friends, colleagues and scientific advisors.

Thank you to the "Pioneer Team" in the Samson Lab, Zachary Nagel and Patrizia Mazucatto, with whom I shared nearly every day of my time in the Samson lab. The lessons I take from you and our work will help me be a better scientist.

To so many BE and MIT employees that have helped me on my journey: Dalia, Sue, Aran, Alexis, Lindsay, Julie, Amanda, Sophea, Glenn, Mike, Catherine, Zhuli.

To my classmates, who certainly provided a great deal of memories throughout these years.

To those friends from home and around the globe that are simply there, no matter what, giving me their support and knowing they have mine.

I am grateful for the infinite support of all my loving family, my parents Margot and Ernesto who have provided everything a son can ask for, who have open all the possible doors they thought were good for me, but always let me chose my own way.

To my sister Esther, who always finds time and a way to be there for me and with me, for always looking up to and after me when I should be the one looking up to and after her. To her husband David for loving her every day.

To my grandma Esther, the matriarch of the family, the one that brought us all here and the one that keeps us all together. Her infinite love and her genuine words will never cease to amaze me and make me a better human being.

My aunts and uncles that have shaped me to be the person I am today: Mireya, Reina, Ramón, Walter, Isabel and all my cousins. My newly extended family who has supported me since the day they met me: Ellen, Michael, Arlene, Kurt, Paul, Bruce, Bonnie and Tom and all the cousins.

Finally, I am thankful for my beautiful wife Jordan, who has always been there for me, through the ups and downs, on good days when I'm saving the world from cancer and in bad days when I can't believe I graduated from kindergarten. I am grateful for her endless love, for always being herself, for believing in me.

For my beautiful Venezuela ... better times are coming your way ...

Table of Contents

List of Tables

List of Figures

CHAPTERS

Chapter I: Inter-individual variation in DNA repair capacity: a need for multi-pathway functional assays to promote translational DNA repair research.

Zachary D. Nagel, Isaac. A. Chaim, Leona D. Samson

Published as:

Nagel, Z.D., I.A. Chaim, and L.D. Samson, *Inter-individual variation in DNA repair capacity: A need for multi-pathway functional assays to promote translational DNA repair research.* DNA repair, 2014. **19**: p. 199–213.

Table of Contents

Overview of the Presented Study

In the presented work, we develop functional DNA repair assays for the assessment of inter-individual differences in DNA repair capacity, with a special focus on the Base Excision Repair (BER) pathway. The method seeks to fill a void in the need for multi-pathway functional assays to promote translational DNA repair research (Chapter I). The assay itself is a multiplexed fluorescent adaptation of the original host cell reactivation assay that has allowed the simultaneous measurement of HR, NER, NHEJ, MGMT, MMR and OGG1 as a representative of BER (Chapter II). A review on the current methodologies for the assessment of BER in the human population (Chapter III) puts into context the relevance of the methods developed to simultaneously measure inter-individual differences in the activities of several DNA glycosylases and APE1 *in vivo* (Chapter IV). A major breakthrough in the development of these assays was the exploitation of previously unknown *in vivo* transcriptional mutagenesis properties of several substrates of the BER pathway. The adaptability of the method developed for the assessment of virtually any site-specific DNA lesion allowed for the discovery of a novel *in vivo* role of the TC-NER pathway in the repair of etheno-cytosine (Chapter V). Finally, a pilot population study involving DRC reporters for all the major DNA repair pathways evaluated inter- and intra-individual variability in lymphocytes isolated from a group of 56 healthy subjects (Chapter VI).

Inter-individual variation in DNA repair capacity: a need for multi-pathway functional assays to promote translational DNA repair research.

Zachary D. Nagel[1,2], Isaac. A. Chaim[1,2], Leona D. Samson[1,2,3,4,*]

[1]Department of Biological Engineering, [2]Center for Environmental Health Sciences, [3]Department of Biology, [4]The David H. Koch Institute for Integrative Cancer Research, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

*To whom correspondence should be addressed: lsamson@mit.edu

Key Words: DNA repair capacity, multiplex assays, personalized disease prevention and treatment

Abstract

Why does a constant barrage of DNA damage lead to disease in some individuals, while others remain healthy? This article surveys current work addressing the implications of inter-individual variation in DNA repair capacity for human health, and discusses the status of DNA repair assays as potential clinical tools for personalized prevention or treatment of disease. In particular, we highlight research showing that there are significant inter-individual variations in DNA Repair Capacity (DRC), and that measuring these differences provides important biological insight regarding disease susceptibility and cancer treatment efficacy. We emphasize work showing that it is important to measure repair capacity in multiple pathways, and that functional assays are required to fill a gap left by genome wide association studies, global gene expression and proteomics.

Finally, we discuss research that will be needed to overcome barriers that currently limit the use of DNA repair assays in the clinic.

1. Introduction

During the time it takes to read this sentence, it can be estimated that the reader's DNA will incur on the order of 10 trillion DNA lesions. Left unrepaired DNA damage has the potential to lead to mutant cells, dead cells and ensuing disease (**Figure 1.1**). The precise number and type of DNA lesions formed varies from one individual to the next in part because of differences in exposure and lifestyle, and also because of variation in metabolism and other cellular processes. Many types of DNA damage, such as abasic sites, alkylation damage, oxidative damage, mismatches, single and double strand breaks, result from normal metabolic processes. Others are induced upon exposure to environmental agents. Among the environmentally induced lesions are bulky DNA adducts, including heterocyclic amines induced by compounds in cooked foods, cyclobutane pyrimidine dimers induced by sunlight, alkylation damage from nitroso compounds in combustion products, and oxidative damage and DNA strand breaks induced by ionizing radiation from cosmic rays and radionuclides such as Radon gas. In addition, some environmental exposures such as arsenic do not directly induce DNA damage, but are thought to increase DNA damage levels both by inducing inflammation and by disrupting DNA repair (1-3)(4, 5).

Fortunately, human cells mount a robust response to DNA damage that includes at least 7 major DNA repair pathways that specialize in the repair of subsets of DNA lesions, namely direct reversal (DR), mismatch repair (MMR), nucleotide excision repair (NER), homologous recombination (HR), base excision repair (BER), single strand break repair (SSBR), ,non-homologous end joining (NHEJ), and Fanconi Anemia DNA crosslink repair (FANC) (**Table 1.1**). The relationship between DNA damage and DNA repair is complex; no single pathway efficiently repairs all types of DNA lesions, some lesions are substrates for more than one pathway, and evidence for extensive interactions among

proteins involved in distinct pathways continues to emerge (6-11). Mutations in DNA repair genes can have profound consequences for disease risk. The classic example is that individuals with the disease Xeroderma Pigmentosum (XP) are highly prone to skin cancer because they have mutations in genes required for nucleotide excision repair (NER), which repairs bulky lesions such as those induced by UV light. These individuals are at a 2000-fold higher risk of skin cancer in sun-exposed skin (12). A variety of other diseases including neurological, developmental and immunological disorders, as well as premature aging, are associated with aberrant DNA repair in humans (**Table 1.1**) (13). Thus, it is clear that defective DNA repair caused by mutations in repair genes represents a major disease risk factor, and genetic tests are now available for the most common disease-associated mutations in DNA repair genes (14).

Intuitively, one might expect that DRC in a given pathway should vary even among individuals who do not have rare disease-associated mutations in key DNA repair genes, perhaps due to common sequence variants and epigenetic heterogeneity across populations. DRC might thus adopt a normal distribution among individuals with disease-associated DNA repair defects (red curve in **Figure 1.2A**), as well as among apparently healthy individuals in the general population (black curve in **Figure 1.2A**) (15). One might further hypothesize, based simply on interpolation (**Figure 1.2B**), that those in the general population falling to the left of the distribution would be at higher than average risk for disease, and that they might be candidates for personalized prevention schemes. It has been over two decades since these ideas were articulated by Hsu (16), and by Grossman and Wei (15), and although a wealth of evidence for interindividual DRC differences has since emerged from multiple laboratories using various methods (**Table 1.2**), it seems fair to say that the intervening studies have not yet resulted in personalized prevention efforts. Other possible relationships between DRC and risk of disease must also be considered (**Figures 1.2C-2F**), and these will be discussed below. Inter-individual variation in DRC might also account for differing tolerance among cancer patients for cancer therapy with DNA damaging agents. Moreover, the

sometimes dramatic changes in DRC in cancer cells versus non-cancer cells might be exploited for individualized treatment.

Over the last two decades, significant efforts have focused on testing the idea that measuring DRC has the potential to inform medical and clinical practice. In this review, we discuss major lines of evidence supporting the notion that there are significant interindividual differences in DRC, and further, supporting the claim that such variations are indeed associated with disease risk. We survey the experimental approaches used to measure DRC, and discuss future work that will be needed for clinical translation of functional DRC measurements.

2. Evidence for inter-individual differences in DRC from indirect measurements

Genetics

XP was the first human cancer-susceptibility disease found to be associated with a DNA repair defect, namely NER; direct *in vivo* DRC measurements in cells isolated from XP patients provided the critical insight that a DNA repair defect was the cause of the disease (17). Complementation studies identified numerous genes responsible for XP, providing the foundation for predicting NER defects and associated disease indirectly from genotype analysis. Thus DNA sequencing of well-characterized mutations in XP genes can be used to predict impaired DRC and increased disease susceptibility. Subsequent research has identified numerous other disease-associated rare gene mutations that cause severe defects in the MMR, NER, HR, BER, SSBR, NHEJ, and FANC pathways (**Table 1.1**), as well as defects in DNA damage surveillance (18) and tolerance pathways (19).

Common single nucleotide polymorphisms (SNPs) that are associated with disease have been identified in genes in the DR, BER, MMR, NER, HR, and NHEJ pathways (recently reviewed in (20)). In candidate gene association studies, SNPs in DNA repair genes have been associated with increased or

decreased risk of many cancers including lung, colorectal, gallbladder, oral, breast, prostate, liver, ovarian, and laryngeal cancer, as well as lymphoma and squamous cell carcinoma (20). Genome wide association studies (GWAS) have revealed many additional lower penetrance disease-associated sequence variants using unbiased computational approaches (21, 22), but surprisingly few of these turn out to be DNA repair genes. This may be explained in part by the observation that the variants identified so far explain only a small portion of disease heritability. As yet unidentified DNA repair variants may contribute to the missing heritability if they are relatively rare but confer a relatively large risk increment. Variants in DNA repair genes that confer risk could also be missed if they represent copy number variants or they have relatively small effects; further, gene-gene interactions involving DNA repair gene variants may also be missed in GWAS studies due to low statistical power (23). Moreover, most GWAS-identified variants are not located in genic regions, but rather in intergenic regions that are presumably involved in gene regulation. Increased sample sizes,better accounting for rare variants and structural variants, and better understanding of the role of regulatory variants will likely increase the ability of DNA sequence-based assessments to identify individuals with elevated disease risk. In **section 3**, we will discuss in detail functional assays that may complement DNA sequence-based predictors of DRC defects.

In addition to disease prevention, genome profiling for sequence variants in DNA repair genes has the potential to enable personalized disease treatment (24); it is already clear that SNPs in DNA repair genes can play a role in assessing a prognosis for patients being treated for melanoma, pancreatic, esophageal, or non-small cell lung cancer. SNPs in the following DNA repair genes have been associated with the response of patients to cancer therapy: MGMT, XPA, XPC, XPD, XPE, XPG, ERCC1, ERCC3, XRCC1, XRCC2, and XRCC3 (25-32). Polymorphisms in some DNA repair genes, such as ERCC1 and XPD, have also been associated with increased cancer therapy toxicity (33), and MGMT polymorphisms are associated with increased risk of myelodysplastic syndromes following treatment with alkylating agents (34).

Major advantages of genomic profiling include the breadth of data that can be obtained for relatively small (and steadily decreasing) investment of resources using next generation sequencing (DNAseq), conceptual simplicity, and the universality of the approach; standardized sequencing procedures foster high inter-laboratory reproducibility (35). An important limitation of studies that aim to make predictions based on DNA sequence is that, with the possible exception of CpG methylation specific PCR (MSP), one cannot know *a priori* how well the gene with which the sequence variant has been associated is actually expressed; indeed differential allelic gene expression is common (36). In this regard genome sequence based assays may be regarded as the least direct means of measuring function (**Figure 1.3**).

Transcriptional profiling

Transcriptional profiling has revealed that DNA repair gene expression has important consequences for disease biology. For example, studies have identified prognostic gene expression signatures in cancer cells that correlate with breast cancer survival (37, 38), breast cancer recurrence (39), and lung cancer survival (40). Tumor gene expression profiles that include DNA repair genes and correlate with cancer therapy response have also been identified (41), and in some cases tumor expression of a single DNA repair gene correlates with treatment efficacy (42-44). Moreover, gene expression profiling has been used to identify bleomycin-induced changes in DNA repair gene expression that predict bleomycin sensitivity, and low level radiation induced changes in DNA repair gene expression in (non-cancerous) human lymphocytes that can be used as biomarkers for occupational exposure to ionizing radiation (45, 46). An additional study in human lymphocytes demonstrated an inverse correlation between radiosensitivity measured by a G2 challenge assay and expression of the *NFKB* gene; this study also showed an association between radiosensitivity and breast cancer risk (47).

One study with human lymphoblastoid cell lines revealed that expression of a set of just 48 genes was sufficient to predict sensitivity to MNNG (48), an alkylating agent that generates the same spectrum of DNA lesions as the chemotherapeutic drugs temozolomide, dacarbazine, procarbazine and streptozotocin (49). These findings were important in several respects; first it was possible to predict MNNG sensitivity in cells derived from apparently healthy individuals using a small set of basally expressed genes, supporting the notion that gene expression patterns could be used to help predict how tumor cells will respond to therapy. Second, there were large variations in sensitivity to two different DNA kinds of alkylating agents, and it was subsequently shown that sensitivity to one agent did not accurately predict sensitivity to a second DNA damaging agent, indicating a unique response to each agent (50). One of the predictive transcripts for MNNG sensitivity encodes the methylguanine DNA methyltransferase (MGMT) protein that repairs O^6-alkylguanine DNA lesions by DR. This might be expected, since O^6-methylguanine is one of the most toxic lesions generated by MNNG. However, *MGMT* expression alone was a much weaker predictor than expression of the combined set of 48 genes, indicating that sensitivity to DNA damaging agents reflects the integration of numerous biological pathways. The other 47 genes included at least one other known DNA repair gene (*MUTYH*), and it is possible that the other genes identified in this study affect DRC in ways that have yet to be established.

A major strength of gene expression profiling is the relatively new ability to complement or replace quantitative real time PCR (qPCR) and microarray gene expression assays with next generation RNA sequencing (RNAseq) to generate very accurate data assessing genome wide expression levels, along with splicing information, at steadily falling costs. However, as with genomic DNA sequencing analysis, gene expression profiling is limited because it remains a relatively indirect measure of function; the presence of a transcript does not guarantee that the translation product will be correctly folded, active, appropriately modified and localized to the correct cellular compartment.

Mutagen sensitivity assays

Mutagen sensitivity assays, using mutagen-induced chromosome and chromatid breaks, can also provide an indirect assessment of DRC phenotype (16, 20, 51). Epidemiological studies comparing the response of lymphocytes to mutagens including bleomycin, UV-light, and benzopyrene diol epoxide (BPDE) have revealed increased sensitivity to mutagen-induced chromosomal aberrations in lymphocytes from individuals with cancer versus lymphocytes from healthy individuals. This has been borne out in patients with cutaneous melanoma, basal cell carcinoma, squamous cell carcinoma of the head and neck, plus patients with lung, breast and bladder cancers (51). Of particular interest, a prospective study found that higher bleomycin sensitivity of lymphocytes (>0.5 chromatid breaks per cell) from patients with head and neck cancer was associated with an elevated risk (hazard ratio of 1.38) of developing second, unrelated primary tumors as well as recurrence of the original cancer (52); importantly the blood cells were drawn before development of second primary or recurrent tumors. A second prospective study found a significant association between bleomycin sensitivity in lymphoblastoid cell lines and combined risk of prostate, lung, colorectal and ovarian cancers in the patients from whom the cells were derived (53). However, this study found no significant associations between cancer risk and several other measures of DRC, including BPDE sensitivity, endogenous DNA damage levels measured by comet assays, or host cell reactivation of UV-irradiated plasmids (comet and host cell reactivation assays are discussed in detail below). These negative results may reflect the small sample size and the focus on a limited number of DNA repair pathways.

A major advantage of mutagen sensitivity assays is that, by measuring the response of whole cells to specific mutagens of interest, they integrate biological complexity, such as SNPs, gene expression, epigenetics, protein folding and cell cycle checkpoint activation pathways that may not be accounted for by other methods of measuring DRC. On the other hand, it should be noted that a

shortcoming of mutagen sensitivity assays is that they do not provide specific mechanistic information with regard to the identity of the genotoxic lesion or the pathways responsible for potentially defective repair, and they are relatively labor intensive.

3. Complications associated with indirect measurements of DRC exemplified by a simple repair pathway

The challenges associated with making accurate indirect DRC measurements can be illustrated by considering the performance of available methods for estimating MGMT activity; this protein essentially represents a one-protein DNA repair pathway. SNPs that may lead to MGMT defects have been associated with increased risk of some cancers, and better prognosis following chemotherapy with alkylating agents, as might be expected if the SNP leads to inefficient DNA repair (32). However, there are many examples of cancer cells in which *MGMT* is epigenetically silenced due to promoter hypermethylation (43, 54, 55); in these cases the sequence of the *MGMT* gene would be irrelevant to prognosis because the gene is not expressed. Thus, information about promoter methylation and/or gene expression may be needed to complement information obtained from DNA sequencing.

Epigenetic *MGMT* silencing due to promoter CpG hypermethylation in tumors has been detected by methylation-specific PCR methods and shown to correlate with the efficacy of cancer treatment with O^6-MeG generating chemotherapeutic agents such as temozolomide or dacarbazine (56). Nevertheless, even a combination of SNP data and promoter methylation status may fail to predict function (and therefore clinical outcome) for several reasons. MGMT methylation status is sometimes not predictive of transcript levels (57, 58), and transcript levels are not informative unless the transcripts are translated and the protein stably folded and localized to the nucleus. For example, a significant fraction of the human MGMT protein is inactive in some cells (59), possibly due to posttranslational modifications (60). Furthermore, environmental

exposures can alter the activity of DNA repair proteins, including MGMT (1, 2, 61). These phenomena may not be detected by the available indirect measurements of MGMT activity.

A defect in protein localization could also confound DRC assays. Mitochondria-associated OGG1 protein and activity levels are higher in the livers of old mice and in presenescent human fibroblasts compared to young mice and replicating human fibroblasts, respectively. However, a significant fraction of OGG1 remains inactive and sequestered in the mitochondrial outer membrane and intermembrane space, leading to accumulation of unrepaired oxidized bases in the mitochondrial DNA (62). An age-related localization defect was also observed for the mitochondrial uracil-DNA glycosylase (UNG) (62). A potential analogous localization defect for MGMT (or any other DNA repair protein) could be difficult to predict from the DNA sequence, and would not be detected by *in vitro* activity assays performed on whole cell lysates. In general, relatively laborious immunohistochemistry and subcellular fractionation techniques are required to detect protein localization defects.

In this section we have principally highlighted some of the pitfalls associated with making indirect measurements of DRC in the context of the simple one-protein MGMT pathway, and one might anticipate even more complex challenges for indirect measures of DRC in pathways that involve multi-protein complexes and multiple enzymatic steps (63-72). A major strength of *in vivo* functional DNA repair assays is the ability to integrate the complexity described above to reflect, as closely as possible, repair of genomic DNA damage. The next section discusses recent technological advances for making direct DRC measurements.

4. Evidence of inter-individual DRC differences from direct (functional) measurements

There are numerous methods for measuring DNA repair directly, and each has its strengths and weaknesses (**Table 1.2**). Some of the earliest protocols for

measuring DRC, such as unscheduled DNA synthesis, removal of radiolabeled alkylation damage from genomic DNA (73, 74), and methods using antibodies specific for DNA lesions including BPDE, 8-oxoG, O^6-MeG, pyrimidine dimers and cisplatin adducts (75), hold the advantage of measuring repair of genomic DNA in intact cells. While these assays have been used to detect a 3-5 fold range in inter-individual DRC (76, 77), they are relatively labor-intensive thus inhibiting their application to large-scale studies.

Activity assays with cell lysates

Pathway-specific DNA repair activity analyses in cell lysates have yielded considerable insight into inter-individual variation in DRC. An advantage of these assays is that because they measure levels of functional protein, they integrate much of the biological complexity that might confound indirect measures of DRC (**Figure 1.3**); indeed a low correlation between enzymatic activity and mRNA levels has been documented in some cases (78). Quantitative *in vitro* functional assays have been developed for various steps of BER (79-83), MMR (84), DR of alkylation damage by MGMT (85), NER (86), NHEJ (87), cross-link repair (88), and HR (89).

In vitro assays with cell-free extracts prepared from human lymphocytes have so far been used to measure inter-individual differences in MGMT activity and in the efficiency of several key steps in the BER pathway. These studies revealed an approximately 10-fold variation in MGMT activity (90), a 10-fold variation in activity of alkyladenine DNA glycosylase (AAG, a.k.a. MPG) (49, 78, 91, 92) that initiates BER of several types of alkylation damage, and a 3-fold interindividual variation in activity of 8-oxoguanine DNA glycosylase (OGG1) (78) that initiates repair of oxidative DNA damage. Measurements of the subsequent BER steps has revealed a 1.9-fold to 2.5-fold variation in AP endonuclease activity (79, 82), 1.3-fold variation in subsequent polymerase beta dependent gap filling (82), and 3.4-fold variation in DNA nick ligation (82).

Although cell-free assays are quantitative and permit specificity with regard to the type of lesion being repaired, they have some limitations. The disruption of cells necessary for cell-free analyses can lead to dissociation of protein complexes and protein unfolding, or can mask defects in protein localization that would be detected using an *in vivo* assay. Similarly, the assay buffers may fail to reflect the *in vivo* intracellular environment. In addition, the substrates, typically naked short oligonucleotides, may not fully represent the complexity of repair in chromatinized DNA.

Comet Assays

Comet assays provide a powerful means of measuring endogenous DNA damage, induced DNA damage, and repair of DNA damage in genomic DNA in live cells (93). The assay is named for the comet-like appearance of DNA after single cell gel electrophoresis, before and after treatment with DNA damaging agents; measuring the disappearance of DNA damage following DNA damage induction enables estimates of DNA repair kinetics. Double strand breaks are measured at neutral pH, whereas single strand breaks are measured under alkaline conditions that dissociate DNA strands. Cells may be treated with ionizing radiation or bleomycin, which directly induce strand breaks, or agents such as UV-light, BPDE, peroxides, and alkylating agents that induce DNA lesions that can be converted into strand breaks upon processing *in vivo* by DNA repair machinery (94, 95). Genomic DNA base damage levels can also be measured using the alkaline comet assay following treatment of permeabilized gel-embedded cells with purified lesion-specific enzymes such as Endonuclease III (thymine glycol), FPG and OGG1 (8-oxoG), T4 endonuclease V (pyrimidine dimers), AlkA (alkylation damage such as 3-methyladenine), and UNG (uracil) that convert their respective substrates to alkalai-labile abasic sites or to single strand breaks (93).

Comet assays have been used in several studies with human lymphocytes to measure inter-individual differences in DRC. One study demonstrated a 4-fold

inter-individual variation in BER of 8-oxoG and a 10-fold variation in NER of UV-induced damage (96). Differences in DRC provide biological insight; reduced DRC relative to healthy individuals has been associated with cancer risk in a number of studies (**Table 1.2**). An approximate 2-fold increase in endogenous DNA damage (suggestive of reduced DRC) has been observed in lymphocytes from brain cancer patients, relative to healthy individuals (97); up to 2-fold higher levels of bleomycin induced DNA damage have been observed in lymphocytes from breast cancer patients (98); and modestly reduced repair of bleomycin-induced DNA damage has been observed in lymphocytes from both breast cancer patients and non small cell lung cancer patients relative to healthy controls (98, 99). Furthermore, reduced repair of hydrogen peroxide-induced DNA damage was found in lymphocytes from lung cancer patients (100), and a 4-fold variation in the *rate* of repair for ionizing radiation induced DNA damage was found, wherein lymphocytes from head and neck cancer patients were more likely to exhibit slow repair (101). The aggregate data from these studies raise the possibility that individuals with lower DRC are more prone to cancer, and might be candidates for more aggressive cancer screening.

Finally, comet assays have also been used to show that lymphocytes from patients with extreme reactions to radiation treatment, defined as grade 4 (102), repair ionizing radiation induced DNA damage with slower kinetics than lymphocytes from normal responders (103). This work suggests that it may be possible to use DRC assays to predict radiation sensitivity, and to tailor treatment based on individual tolerance.

Until recently, the labor-intensive nature and large inter-laboratory variation in analysis of comet assay data constituted a significant barrier to the application of the comet assay in large studies. However, a chip-based comet assay with automated image analysis has opened the door to such studies (104); the comet chip assay provides a high throughput platform that significantly reduces the inherent noise in the conventional comet assay. The ability to measure repair of a wide variety of types of genomic DNA damage on a single comet chip represents a major step forward. A limitation of the assay is that while there are

strategies for extending the comet assay to measure many types of DNA damage, the methodology is limited to the subset of DNA lesions that either induce strand breaks or induce damage that can be converted to strand breaks.

Host Cell Reactivation Assays

Host cell reactivation (HCR) assays offer a powerful way to measure DRC in living cells. The foundation of the assay lies in the ability of transcription blocking DNA damage to impede expression of a transiently transfected reporter gene; repair restores transcription of the reporter gene, which may encode enzymes such as chloramphenicol acetyltransferase (CAT) and luciferase, or a fluorescent protein (105). A major strength of HCR assays, stemming from the *in vitro* generation of damaged reporter plasmid DNA, is the ability to measure the *in vivo* repair of specific DNA lesions in intact cells.

The same association between reduced DRC and cancer risk found using mutagen sensitivity, comet and cell free assays, was also demonstrated in several epidemiological studies using HCR assays. For most of these studies, reporter protein activity was measured in cell lysates prepared from transiently transfected human lymphocytes. Early HCR assays using UV-irradiated CAT reporter plasmids showed approximately a 10-fold range of inter-individual differences in NER capacity, with a significantly lower average DRC in lymphocytes from basal cell carcinoma (BCC) patients compared to those from controls (106). A second study also found lower average DRC in lymphocytes from BCC patients aged 40 years or less relative to cancer-free individuals in the same demographic, but found *higher* average DRC in older BCC patients (107). The unexpected higher DRC was associated with smokers, indicating that environmental exposures can affect DRC. HCR assays making use of UV-irradiated CAT or luciferase reporters have revealed 10-20% reduced DRC (relative to control) in lymphocytes from patients with either melanoma or non-melanoma skin cancer (108, 109). In other studies, repair of plasmids damaged with BPDE or the nitrosamine 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone

(NNK) was reduced between 10 and 60% in lymphocytes from patients with lung cancer (110-113), non-small cell lung cancer (114), breast cancer (115-117), squamous cell carcinoma of the head and neck (118, 119), and lung adenocarcinoma (113). Furthermore, DRC below the control median was associated with an increased risk of cancer, with odds ratios ranging from 1.5 to 5.7 (51). In further studies, lymphocytes from bladder cancer patients repaired plasmids damaged with 4-aminobiphenyl with ~10% reduced efficiency (120), and ~10% reduction in repair of plasmids alkylated with dimethyl sulfate was observed in lymphocytes from patients with lung adenocarcinoma (113). Assessment of NER capacity from apparently healthy individuals has also shown a 5.6 to 11-fold range of inter-individual variation and an inverse correlation with age and adiposity (121, 122). A plasmid end-joining assay found a statistically insignificant ~6% reduced average DRC in lymphocytes from breast cancer patients, but the lowest DRC quartile was positively associated with increased cancer risk (odds ratio 2.2) (123).

Additional HCR assays, including a few with fluorescent reporters that do not require cell lysates for analysis have been developed for measuring HR (124, 125), MMR (126), BER (127), NHEJ (128), inter-strand cross link repair (7, 129, 130) and repair of oxidative damage (131, 132). A multiplexed fluorescence-based flow cytometric HCR assay (FM-HCR) that uses different colored fluorescent reporter plasmids to measure repair of multiple doses or multiple types of DNA damage in a single assay was recently developed (133). FM-HCR is less labor intensive than HCR assays that require cell lysate preparation, and uses DNA lesion-induced transcriptional mutagenesis to measure repair of specific DNA lesions, such as O^6-methylguanine and 8-oxoguanine, that are bypassed by RNA polymerase and thus refractory to conventional HCR assays. An even higher throughput HCR assay that uses deep sequencing to measure and sequence reporter transcripts (HCR-seq) was also developed (133). While the epidemiological studies have so far been dominated by NER reporters, the availability of reporters for additional pathways and high throughput HCR assays should encourage future studies to examine multiple repair pathways.

HCR assays face some potential limitations. The repair of constitutively transcribed plasmid DNA measured by HCR may not accurately reflect repair of genomic DNA. However, the assays have been validated using a variety of cell lines and in primary human blood cells with DNA repair defects caused by mutated and inactivated DNA repair genes. Importantly, numerous epidemiological studies (discussed above) have confirmed that HCR assays can reproducibly measure small DRC differences in primary human tissues that are associated with disease. In further support of the notion that plasmid DNA transactions can be reflective of genomic DNA transactions, it appears that plasmids are readily complexed into a nucleosomal structure in human cells, i.e. they become chromatinized (134-137). Moreover, plasmid DNA damage induces histone modifications that affect expression of plasmid DNA (138), indicating a *functional* plasmid-chromatin structure.

5. The need for assays that measure DRC in more than one pathway

The majority of epidemiological studies that apply functional DRC assays have focused on a single DNA repair pathway, namely NER. However, data continue to emerge in support of the notion that DRC for more than one pathway will be required to gain maximal biological insight. Here we consider several contexts in which multiple DNA repair pathways, or the multiple steps within a single pathway, interact to influence disease risk or the sensitivity of cells and animals to DNA damaging agents.

Multiple repair defects and cancer

Treatment with S_N1 type alkylating agents such as temozolomide and decarbazine generates toxic O^6-MeG lesions that are repaired by MGMT. MGMT deficient cells are thus generally very sensitive to S_N1 alkylating agents. However because the toxicity of O^6-MeG lesions is mediated by MMR (139, 140), MGMT deficient cells can become resistant to alkylating agents if they acquire a

MMR deficiency (141). This chemoresistance or tolerance mechanism would confound efforts to predict treatment efficacy based on tumor MGMT status alone, and suggests a need to measure both pathways for improved prognosis. A second example where the status of two or more pathways determines the sensitivity of cells is seen in the context of poly(ADP-ribose) polymerase (PARP) inhibitors. PARP is involved in DNA single strand break repair, and PARP inhibition potentiates DNA damage-induced cell death (142). In the absence of DNA damaging agents, cells can generally tolerate either PARP inhibition or a defect in HR. However, because HR rescues the collapsed replication forks generated when the replication machinery encounters a single strand break, treatment of HR-deficient cells with PARP inhibitors leads to a synthetic lethality (142).

Polymorphisms in multiple DNA repair pathways or multiple steps within a pathway are associated with elevated cancer risk. A study of non-small cell lung cancer patients revealed small individual hazard ratios (up to ~1.4) for polymorphisms in genes involved in NER (XPA, XPD, XPG), BER (XRCC1), and HR (XRCC2, XRCC3), but a larger hazard ratio for patients with any combination of 4 or more polymorphisms in different genes within a pathway or in different pathways (hazard ratio 1.8) (28). For pathways that involve more than one repair protein, such as NER, there is potential for an additive or synergistic effect of combining modest functional defects in multiple steps along the pathway. A recent candidate gene association study revealed that SNPs in the NER genes XPG, XPD, XPA, and XPE are associated with worse prognosis following skin cancer diagnosis; a hazard ratio of 1.26 was calculated for individuals with a variant genotype in one of the genes, but the hazard ratio increased dramatically for individuals with variants in 2 or 3 of these genes to 3.90 and 34.3, respectively (30).

Functional DRC measurements in multiple DNA repair pathways have also revealed higher risk factors than measurements in any single pathway would indicate. A breast cancer study found that combined NER deficiency (measured by an immunohistochemical assay for BPDE repair) and NHEJ deficiency

(measured by a plasmid repair assay) represented a greater cancer risk factor (odds ratio 4.92) than deficiency in either pathway alone (odds ratio 1.16) (77, 123). A second case control study of both OGG1 and AAG activities showed that reduced OGG1 activity and *elevated* AAG activity were associated with a higher risk of lung cancer, and most important, that a combined score for the two enzyme activities was more strongly associated with cancer risk than either OGG1 or AAG activity alone (143). Recently, this study has been extended to incorporate APE1 into an integrated DNA repair score, termed "OMA" for OGG1, MPG (a.k.a AAG) and APE1; the OMA score varies over a 20-fold range and associates even more strongly with risk of lung cancer (odds ratio 5.6 comparing individuals with the lowest to highest tertile OMA scores) (144). These results emphasize that measuring repair capacity in more than one pathway has the potential to increase biological insight and reveal stronger correlations between DRC and disease risk. It should be noted that the lower OMA scores correspond to lower levels of OGG1 and APE1 activity, but *higher* AAG activity, underscoring the fact that for some pathways high levels of DRC are not always protective.

Imbalanced repair and toxic repair intermediates

Evidence for potentially harmful and tissue dependent effects from higher DRC levels has emerged from the characterization of Aag-dependent alkylation sensitivity in cells and animals. Aag deficiency has been associated with sensitivity to alkylating agents in mouse embryonic stem cells, consistent with a relationship wherein risk (defined for this specific example as the risk of cell death upon exposure to an alkylating agent) in one tissue decreases with increasing DRC (**Figure 1.2B**). However an unexpected phenotype was observed in mouse models, wherein Aag deficiency leads to extreme alkylation resistance in certain tissues (49), indicating, that for some tissues risk increases with *increasing* DRC (**Figure 1.2C**). Accordingly, overexpression of Aag leads to tissue-specific alkylation sensitivity in mice (92).

A lack of proper coordination among multiple DNA repair steps (repair imbalance) has been invoked to explain increased alkylation sensitivity in cells that overexpress Aag. Aag overexpression leads to the accumulation of DNA repair intermediates (49, 145), which include 5'-deoxyribose phosphate containing single strand breaks that can trigger hyperactivation of PARP. This enzyme modifies numerous other proteins, including several in DNA repair pathways (142, 146). PARP also facilitates both repair (SSBR and BER) as well as a cell death pathway involving NAD and ATP depletion, and an energetic crisis followed by cell death (147-149). Cells and whole animals that overexpress Aag but are genetically deficient for PARP show a complete rescue of wild type sensitivity to alkylating agents, confirming that SSB-stimulated PARP hyperactivation is responsible for hypersensitivity (92).

It should be noted that this situation contrasts with pharmacological PARP inhibition that generally leads to DNA alkylating agent sensitivity. The consequences pharmacological PARP inactivation using inhibitors may differ from the consequences of genetic depletion because inhibitors can induce formation of a stable 5'dRP:PARP:Inhibitor complex at SSBs that inhibits DNA repair and blocks replication (150-152), potentially leading to double strand breaks (153, 154). Thus PARP inhibition leads to alkylation hypersensitivity in cells that accumulate SSBs, including cells that overexpress AAG (155), and polymerase beta or ligase III deficient cells (156, 157).

SSBs also accumulate in cells from individuals with the neurodegenerative diseases spinocerebellar ataxia with axonal neuropathy-1 (SCAN1) and ataxia oculomotor apraxia-1 (AOA1) (158). For SCAN1, abortive topoisomerase-I reactions lead to SSBs covalently linked to the enzyme; SCAN1 patients are deficient for the TDP1 enzyme that hydrolyzes the 3'-phosphotyrosyl bond between stalled topoisomerase-I and a SSB to facilitate repair. A related enzyme, TDP2, hydrolyzes 5'-phosphotyrosyl bonds between topoisomerase-II and DNA at DSBs to facilitate NHEJ-dependent repair (159), and exhibits weak 3'-tyrosyl phosphodiesterase activity (160), however TDP2 has not as yet been associated with disease. AOA1 patients are deficient for APTX, an enzyme that

catalyzes reversal of premature 5'-adenylation at SSBs; although 5'adenylation is required for ligation of SSBs, if this modification occurs in the absence of a free 3' hydroxyl, ligation cannot be completed. It is of particular interest that SCAN1 and AOA1 manifest as neurodegenerative diseases, but do not predispose to cancer. To explain this disproportionate effect on terminally differentiated neurons, it has been proposed that TDP1 and APTX may be redundant in proliferating cells because alternative end processing factors and the HR pathway can resolve SSBs during replication (158).

An additional example of a phenotype caused by the accumulation of repair intermediates that depends on multiple DNA repair proteins, comes from a recent study implicating the DSB repair protein WRN in long patch BER of adenine opposite 8-oxoguanine (A:8-oxoG) (161). Cells deficient for WRN or polymerase λ are more sensitive than wild type to oxidizing agents such as hydrogen peroxide, due to inefficient BER of oxidative damage. However, WRN deficient cells and polymerase λ deficient cells that are also deficient for MUTYH exhibit wild type sensitivity to oxidizing agents, suggesting that MUTYH leads to toxic repair intermediates . Indeed, glycosylase mediated accumulation of toxic BER intermediates has also been invoked to explain sensitivity to a variety agents including alkylating agents (49, 145), ionizing radiation (162), and 5-fluorouracil (163).

Repair competition

Some DNA lesions are repaired by proteins from more than one of the canonical DNA repair pathways shown in **Table 1.1 and Table 1.2**. Multiple pathways may either complement or interfere with one another. For example, BER of 8-oxoG opposite cytosine is initiated by one of several DNA glycosylases (OGG1, NEIL1 and NIEL2) (164). Moreover, it was recently reported that proteins involved in transcription coupled NER (e.g. XPA, CSB and RNA polymerase II), also participate in an 8-oxoG repair in actively transcribed DNA (9). Another example of distinct repair proteins competing for the same lesions

arises for highly mutagenic etheno base lesions (165-167). Ethenocytosine (εC) can be bound by AAG but not excised by it, and this binding interaction prevents repair by ALKBH2 (167) and possibly TDG glycosylase. As a result, individual activity levels for AAG, TDG or ALKBH2 would not provide a complete picture for the repair of εC base lesions. Finally, although agents that form inter-strand cross-links (ICL) have been widely used as chemotherapeutics, the detailed mechanisms of ICL repair are only now beginning to be understood (168). Evidence exists for replication-dependent and replication-independent ICL repair involving proteins from several DNA repair pathways including the FA, HR, and NER pathways, as well as TLS polymerases (7, 8). Recent work shows that BER and MMR play an epistatic role in mediating cisplatin sensitivity, implicating proteins from these pathways in ICL repair as well (6, 10).

Immune dysfunction

Because immune function involves programmed induction of multiple types of DNA damage, DNA repair proteins from multiple pathways also play a critical role in the immune system, and some DRC defects are associated with immunodeficiency (Table 1). Numerous DSB repair proteins are required for V(D)J recombination, which takes place in both T and B lymphocytes and is essential for the development of specialized antigenic receptors known as T-cell receptors (TCR) and B-cell receptors (BCR), composed of an immunoglobulin molecule and a CD79 moiety. The process is initiated by the Rag1 and Rag2 recombinase enzymes that induce DSBs in specific recombination signal sequences flanking V, D, and J gene units. These DSBs are then repaired by NHEJ. Consequently, many human patients with NHEJ deficiencies also have V(D)J recombination defects and suffer from a particular group of diseases known as severe combined immunodeficiencies (SCID), in particular, radiosensitive SCID. SCID is characterized by impaired T and B lymphocyte differentiation that is sometimes accompanied by deficiencies in other lineages (169). Individuals with deficiencies in NHEJ proteins (DNA-PKcs (170), Artemis

(171, 172), LigIV (173) and NHEJ1/XLF/Cernunnos (174)) consistently present some degree of SCID.

Proteins from several DNA repair pathways are also involved in the terminal maturation of B lymphocytes during which two additional stages of DNA modification take place after VDJ recombination in order to increase the efficiency of the humoral response (175, 176). First, class switch recombination (CSR) exchanges the immunoglobulin (Ig) constant region to modify the Ig isotype (from IgM to IgG, IgA, etc.). In the second step, somatic hypermutation (SHM) introduces sequence diversity into the Ig variable domain to provide the potential for increased antigen affinity. Both CSR and SHM are initiated by the action of activation-induced cytidine deaminase (AID). Within hotspots, AID deaminates cytosine to uracil, creating U:G mismatches. CSR is induced when the BER protein UNG excises closely opposed uracils that are further processed by APE1 to generate a DSB that triggers processing by the HR machinery. U:G pairs escaping UNG recognition can go on to be processed by the MMR machinery and subsequently form DSB in a yet unidentified manner (177). The distinct process of SHM occurs by replication bypass of uracil (in the absence of repair by MMR or BER) inducing C to A transversions, or by translesion polymerase bypass of AP sites and gaps (generated by UNG and MMR proteins, respectively). As might be expected from their involvement in the immune response, deficiencies in both BER and MMR have been implicated in improper B-cell maturation; one of the autosomal forms of CSR deficiency known as hyper-IgM (HIGM) syndromes has been ascribed to mutations in the *UNG* gene (178). Similarly, three patients with deficiencies in the MMR protein PMS2 were shown to be deficient in CSR (179). Immunodeficiency syndromes are associated with increased risk of cancer (180), it has been recently hypothesized that germline mutations in genes involved in V(D)J recombination, SHM and CSR play a role in the lymphomagenesis of diffuse large B cell lymphomas (181).

Our understanding of the role that DNA repair plays during normal lymphocyte maturation continues to evolve as additional interactions between DNA repair pathways are discovered. For example, recent work suggests that

interactions between MMR proteins and the MBD4 DNA glycosylase may be important for efficient CSR (182). An intriguing possibility is that the pronounced DNA repair defects associated with some severe immune disorders may presage discovery of milder immunodeficiency that can be attributed to modest defects in multiple DNA repair pathways.

The diversity of disease states and sensitivity phenotypes associated with inefficient DNA repair in more than one pathway, or in more than one step within a pathway, calls for more studies that explore multiple repair activities. Recently accumulating data discussed above suggest subtle defects in multiple DNA repair pathways might promote disease, raising the prospect that multiplexed DRC assays could be of value in clinical diagnosis, prevention and treatment of disease.

6. Current status of DRC measurements for prevention and treatment of disease

A long-term goal that motivates many of the epidemiological studies discussed herein is to eventually apply direct or indirect estimates of DRC to the personalized treatment or prevention of disease. However current clinical practice is limited to diagnostics. Genetic testing is available for known mutations in many of the genes associated with DRC defects and disease, including *BRCA1/2*, *MLH1*, *MSH2/6*, *p53*, and *MUTYH* (14); individuals with these mutations are advised to undergo more aggressive screening, and in some cases prophylactic surgery. T-cell chromosome breakage or aberrations following treatment with DNA damaging agents such as mitomycin C have been used as a diagnostic for Fanconi Anemia (183), and UV-induced unscheduled DNA synthesis and sister chromatid exchange assays have been used for the molecular diagnosis of XP and Blooms syndrome, respectively (184). However, because they are labor intensive and/or expensive, both genetic testing and cell-based assays are typically used only in cases where the disease is already

suspected either because of pronounced symptoms or a family history of disease.

DRC Estimates may be close to finding application in cancer treatment. Relationships between DRC and improved tumor response to anticancer drugs have been reported for MMR proficiency in cisplatin, alkylating agents, and 5-fluorouracil cancer chemotherapy (141, 185-187). XRCC1 (BER) deficiency in tumors has been linked to cisplatin sensitivity (44, 188); MGMT deficiency has been linked to temozolomide and BCNU sensitivity (189, 190), and HR deficiency to PARP inhibitor sensitivity (142). A limited study suggests that functional assays could be useful for determining the maximum radiation dose that will be tolerated by a patient (103). Thus, if individuals could be pre-identified as therapy resistant, it may be possible to raise the treatment dose to improve efficacy. Despite these examples of potential clinical applications, enthusiasm for the use of DRC measurements to guide treatment decisions are dampened by concerns about assay standardization, assay reproducibility, and the lack of prospective, randomized studies. As a result, although DRC assays are used to retrospectively classify patients into good versus poor responders to cancer therapy (28, 43, 53, 188, 191, 192), these assays are not used currently to influence cancer patient management (20, 58, 193-198).

7. What is needed going forward

To speed the translation of functional DRC assays from a laboratory tool to biomedical applications, advances are needed in several areas. **(i)** Excepting a small number of prospective studies (52, 53, 199), virtually all studies associating DRC defects with cancer susceptibility have been retrospective, raising concerns that the observed DRC differences may reflect changes subsequent to cancer diagnosis, or to cancer treatment. Similar levels of DNA damage have been observed in lymphoblastoid cell lines derived before versus after cancer diagnosis (199), suggesting that cancer development may not alter DRC. Moreover, one line of evidence suggests that lymphocyte DRC may be

altered in cancer patients due to a systemic inflammatory response to the disease (200-204), and it is also possible that cancer treatment affects DRC. There is thus a need for additional large prospective studies to confirm that the DRC was a cause, rather than an effect, of the disease, and that subsequent treatment did not cause long-term DRC changes. *We advocate that as new studies are initiated, cells from patients and their tumors should be cryopreserved such that live cells can be recovered for the purpose of functional DRC assays.* Currently such samples are most often preserved for DNA, RNA and protein analyses, under conditions incompatible with *in vivo* functional assays. **(ii)** To carry out large studies, high throughput quantitative assays measuring DRC in multiple pathways are needed to maximize biological insight and prognostic potential. Standardized quantitative assays will help overcome concerns about reproducibility, an issue that is especially acute for assessments that are more subjective and less quantitative DRC indicators, such as microsatellite instability (MSI), where labs use different thresholds to distinguish MSI from microsatellite stability (205). **(iii)** The complexity of the relationships between DRC, other pathways, genetics, epigenetics and environmental exposure suggest that complementary approaches combining several of the techniques described above and in **Figure 1.3** may be needed to provide a comprehensive assessment of DRC. Collaborative projects using multiple approaches would be helpful for determining which approach or combination of approaches yields the most robust DRC-based disease prediction and diagnostic potential. **(iv)** The suitability of lymphocytes as a surrogate tissue for DRC in other tissues requires additional testing. The numerous epidemiological studies referenced above support the utility of measuring DRC in lymphocytes to predict disease susceptibility, and some investigators have found strong correlations between DRC in lymphocytes and other tissues (206), while others have not (90). **(v)** DRC variation among tissues is underexplored. In both humans and mice, large tissue-specific DRC variation has been measured for some pathways, indicating that it may be necessary to measure DRC directly in the tissue of interest (92, 207, 208). For example, DRC in liver cells might be most useful in assessing

liver cancer risk; however the need for an invasive biopsy represents a major barrier to tissue-specific DRC screening. A particularly promising solution to this problem would be to generate cells representative of various human tissues by differentiating induced pluripotent stem cells (iPS cells) generated from skin fibroblasts that can be obtained from a single, relatively less invasive biopsy. The commercial availability of multiple cell types from a single individual (including iPS cells) together with advancing methods of generating iPS cells from primary tissues (209), make an initial test this approach experimentally feasible. A long-term goal would be to develop methods of measuring tissue-specific DRC variation in a variety of human cell types derived from a skin biopsy from each individual. **(vi)** Finally, mitochondrial DNA repair represents a relatively understudied area, and given the severe health consequences of mitochondrial DNA depletion (210), as well as accumulating evidence of relationships among mitochondrial DNA damage, mitochondrial dysfunction and disease (211-216), it appears likely that defects in mitochondrial DRC can also be linked to disease susceptibility.

8. Conclusions

The aggregate data from several decades of molecular epidemiology indicate that DRC varies significantly among individuals, and that these variations associate with disease risk. No single DNA repair pathway is universally representative of DRC in general, and the effects of variation in repair efficiency at distinct steps or in separate pathways can combine to produce surprising and sometimes counterintuitive phenotypes. There are many ways to measure DRC. However each has its strengths and weaknesses. So far a lack of standardization and clinical validation, together with the relatively low throughput and labor-intensive nature of most methods of measuring DRC have precluded the application of functional DRC assays for personalized disease prevention and treatment. However a recent burst of technical advances, including the highly automated comet chip (104), a proof of concept for integrating DRC in multiple

pathways to calculate disease risk (144), and highly multiplexed HCR assays (133), support the notion that measuring DRC could become common clinical practice. To promote this transition, these emerging technologies should be further developed, standardized and validated across multiple laboratories in large (ideally prospective) epidemiological studies employing measurements of multiple DNA repair pathways.

Tables

Table 1.1. Human diseases associated with DNA repair deficiencies categorized by DNA repair pathway.

Repair pathway	Primary Lesions	Genes Associated with Disease	Diseases Associated	References
DR	O^6-meG	MGMT	- Esophageal, Lung **Cancer**	(217)
MMR	Mismatches, loops	MSH2, MLH1, MSH6, PMS2	- Hereditary Non-Polyposis Colon Cancer (**HNPCC**) - Class Switch Recombination (**CSR**) Defects - T-cell non-Hodgkin's lymphoma (**T-NHL**) - Diffuse large B-cell lymphoma (**DLBCL**)	(218) (219) (220) (221) (222) (179) (223) (224)
NER	Bulky adducts	XPA, XPB (ERCC3), XPC, XPD (ERCC2), XPE (DDB1 & DDB2), XPF (ERCC4), XPG (ERCC5), ERCC1, CSA, CSB, TTDA	- Xeroderma Pigmentosum (**XP**) - Cockayne Syndrome (**CS**) - Trichothiodystrophy (**TTD**)	(12) (225)
HR	DSB	BRCA1, BRCA2, NBS1	- Breast, Prostate **Cancer** - Nijmegen Breakage Syndrome (**NBS**)	(226) (227)
BER & SSBR	Damaged bases, SSB	MUTYH, UNG, OGG1, AAG, APE1, TDP1, APTX	- MUTYH-Associated, Polyposis (**MAP**) - Hyper-IgM syndrome (**HIGM**) type V - Lung **Cancer** - Spinocerebellar ataxia with axonal neuropathy 1 (**SCAN1**) - Ataxia-oculomotor apraxia 1 (**AOA1**)	(228) (178) (229) (158) (144)
NHEJ	DSB	DNA-PKcs, Artemis, LigIV, NHEJ1/XLF/Cernunnos	- Severe Combined Immunodeficiency (**SCID**)	(171) (173) (172) (174) (170)
FANC	Cross-links	FANCA, FANCB, FANCC, FANCD1/BRCA2, FANCD2, FANCE, FANCF, FANCG, FANCI, FANCJ/BRIP1, FANCL, FANCM, FANCN/PALB2, FANCO/RAD51C, FANCP/SLX4, NBS1	- Fanconi Anemia (**FA**) - Nijmegen Breakage Syndrome (**NBS**)	(183) (227) (8)

Table 1.2. Studies in which DNA repair assays have been used to evaluate human inter-individual and/or tumor-specific variability in DRC.

Publications are categorized by DNA repair pathway studied and the type of assay used. **Black** corresponds to studies in which a single repair pathway was assayed. **Blue** corresponds to studies in which two or more pathways were assayed as a consequence of a lesion being repaired by more than one pathway or a protein being involved in more than one pathway. **Red** corresponds to studies in which two or more pathways were assayed simultaneously either through sequencing of two or more genes or by the use of separate repair measurements.

	Evidence of inter-individual/tumor differences in DRC							
Repair pathway	***Indirect measurements***				**Mutagen Sensitivity**	***Direct measurements***		
	SNP/GWAS	**Gene Methylation**	**mRNA**	**Protein**		**Cell free extract**	**Comet**	**HCR**
DR	(32) (230) (20) (217)	(54) (55) (56) (57) (43) (58)	(48) (231) (232) (58)	(232)		(59) (90) (230)		
MMR	(233) (234) (20) (181)	(233) (63) (234)	(231) (57)	(63) (192)		(63) (235) (236)		
NER	(237) (238) (25) (26) (121) (28) (20) (29) (30) (31)	(70) (72)	(239) (240) (70) (231) (72)	(241) (70) (72)	(64) (66) (48) (42) (232) (132) (7) (52) (53)		(96) (242)	(106) (129) (239) (111) (243) (108) (132) (244) (109) (121) (130) (119) (122) (7)
HR	(238) (20)	(245) (66) (69) (246)	(66) (69) (231)	(69)		(89)	(98) (99) (101) (238) (97)	
BER & SSBR	(247) (238) (42) (82) (20)	(68) (71)	(78) (48) (68) (231) (42) (71) (127)	(71)		(79) (247) (78) (248) (82) (91) (143) (92) (144)	(238) (96) (100)	(132) (127)
NHEJ	(238) (249) (20) (181)	(69)	(69) (231)	(69)		(87)	(98) (99) (101) (238) (97)	(244) (123)
FANC	(67) (250)	(245) (64) (65) (66) (246)	(65) (66)	(64) (67)		(88)		(129) (130) (7)

Figures

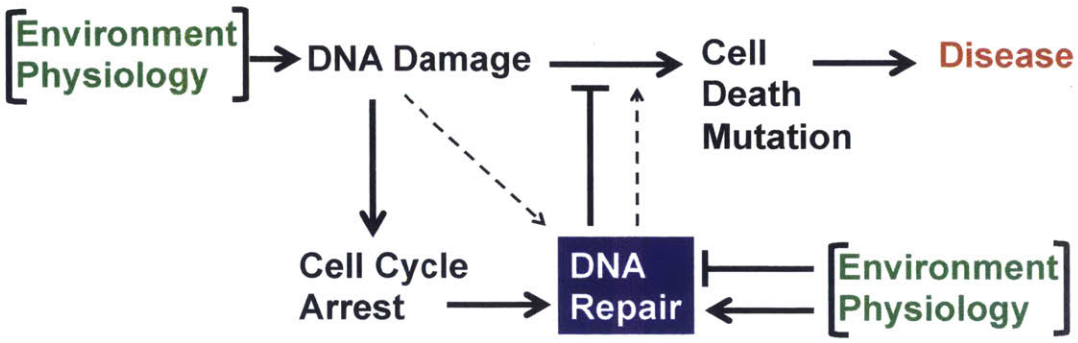


Figure 1.1. DNA damage, DNA repair, and disease.

The canonical role of DNA repair is to protect cells from death, mutation, and the inception of disease. As discussed in the main text, increasing DNA repair can also have the opposite effect, inducing cell death because of the potential accumulation of toxic repair intermediates. The environment and the physiology of the individual enter this diagram at two points; both factors may may increase DNA damage, or they may affect DRC either positively or negatively.

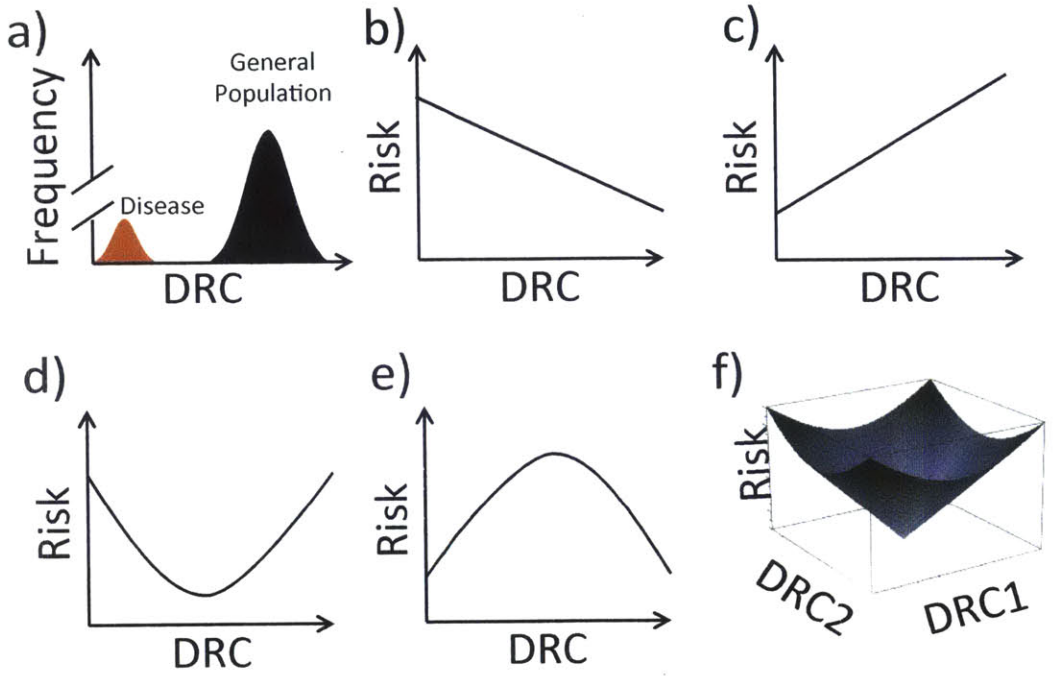


Figure 1.2. Potential consequences of DRC variability.

a) Both intuition and experimental data point to a distribution in DRC, shown here for a single pathway among healthy individuals (black curve), and a multimodal distribution when disease states (red curve) are included; this panel is inspired by a similar figure originally published by Grossman and Wei (15). **b)** The simplest assumption is that a generic representation of risk (cell death, disease diagnosis or mortality) will decrease as DRC increases. **c)** In some cases, such as when a level of glycosylase initiates BER leading to an accumulation of intermediates that are more toxic than the initial DNA damage, elevated DRC may be deleterious. **d and e)** In principle the combined influence of factors driving the relationships in panels b,c could lead to more complicated relationships between risk and DRC. **f)** The relationship between DRC and risk may be represented as a complex landscape that depends on DRC in more than one pathway. As discussed in section 5, multiple DRC defects can act synergistically, but can also produce surprising and counterintuitive phenotypes.

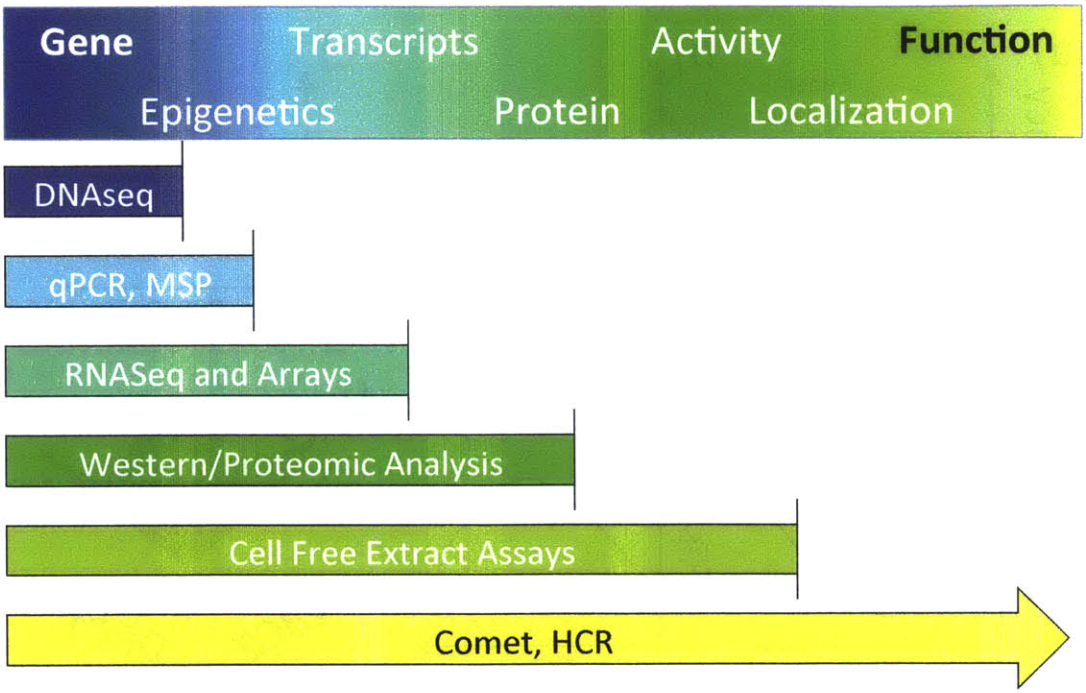


Figure 1.3. Methods of assessing DRC and their limitations.

The biological trajectory that runs from genes to function includes numerous intermediate steps at which a given assay may fail to predict the functional endpoint. The point along this trajectory at which a particular method may fail is indicated by color-coded bars running from left to right. Although any of the assays may accurately predict function, only functional (Comet and HCR) assays integrate the complexity of this entire trajectory into their readout.

References

1. Hartwig A, Blessing H, Schwerdtle T, Walter I. *Modulation of DNA repair processes by arsenic and selenium compounds.* Toxicology. 2003; **193** (1-2): 161-9.
2. Andrew AS, Burgess JL, Meza MM, Demidenko E, Waugh MG, Hamilton JW, Karagas MR. *Arsenic exposure is associated with decreased DNA repair in vitro and in individuals exposed to drinking water arsenic.* Environ Health Perspect. 2006; **114** (8): 1193-8.
3. Ahmed S, Khoda SME, Rekha RS, Gardner RM, Ameer SS, Moore S, Ekstrom EC, Vahter M, Raqib R. *Arsenic-Associated Oxidative Stress, Inflammation, and Immune Disruption in Human Placenta and Cord Blood.* Environ Health Perspect. 2011; **119** (2): 258-64.
4. Hengstler JG, Bolm-Audorff U, Faldum A, Janssen K, Reifenrath M, Gotte W, Jung DL, Mayer-Popken O, Fuchs J, Gebhard S, Bienfait HG, Schlink K, Dietrich C, Faust D, Epe B, Oesch F. *Occupational exposure to heavy metals: DNA damage induction and DNA repair inhibition prove co-exposures to cadmium, cobalt and lead as more dangerous than hitherto expected.* Carcinogenesis. 2003; **24** (1): 63-73.
5. Mattern J, Koomagi R, Volm M. *Smoking-related increase of O-6-methylguanine-DNA methyltransferase expression in human lung carcinomas.* Carcinogenesis. 1998; **19** (7): 1247-50.
6. Kothandapani A, Dangeti VSMN, Brown AR, Banze LA, Wang X-H, Sobol RW, Patrick SM. *Novel role of base excision repair in mediating cisplatin cytotoxicity.* J Biol Chem. 2011; **286** (16): 14564-74.
7. Enoiu M, Jiricny J, Schärer OD. *Repair of cisplatin-induced DNA interstrand crosslinks by a replication-independent pathway involving transcription-coupled repair and translesion synthesis.* Nucleic Acids Res. 2012; **40** (18): 8953-64.
8. Kim H, D'Andrea AD. *Regulation of DNA cross-link repair by the Fanconi anemia/BRCA pathway.* Genes Dev. 2012; **26** (13): 1393-408.
9. Guo J, Hanawalt PC, Spivak G. *Comet-FISH with strand-specific probes reveals transcription-coupled repair of 8-oxoGuanine in human cells.* Nucleic Acids Res. 2013.
10. Kothandapani A, Sawant A, Dangeti VSMN, Sobol RW, Patrick SM. *Epistatic role of base excision repair and mismatch repair pathways in mediating cisplatin cytotoxicity.* Nucleic Acids Res. 2013.
11. Lv L, Wang F, Ma X, Yang Y, Wang Z, Liu H, Li X, Liu Z, Zhang T, Huang M, Friedberg EC, Tang T-S, Guo C. *Mismatch repair protein MSH2 regulates translesion DNA synthesis following exposure of cells to UV radiation.* Nucleic Acids Res. 2013; **41** (22): 10312-22.
12. Kraemer KH, Lee MM, Scotto J. *DNA repair protects against cutaneous and internal neoplasia - evidence from xeroderma pigmentosum.* Carcinogenesis. 1984; **5** (4): 511-4.
13. O'Driscoll M. *Diseases Associated with Defective Responses to DNA Damage.* Cold Spring Harbor Perspectives in Biology. 2012; **4** (12).
14. Ellis NC. Obtaining and Using Genetic Information. Ellis NC, editor. New York: Springer; 2003.

15. Grossman L, Wei Q. *DNA repair and epidemiology of basal cell carcinoma.* Clinical chemistry. 1995; **41** (12): 1854-63.
16. Hsu TC. *Genetic instability in the human population - a working hypothesis* Hereditas. 1983; **98** (1): 1-9.
17. Cleaver JE. *Defective repair replication of DNA in xeroderma pigmentosum.* Nature. 1968; **218** (5142): 652-&.
18. Jackson SP, Bartek J. *The DNA-damage response in human biology and disease.* Nature. 2009; **461** (7267): 1071-8.
19. Waters LS, Minesinger BK, Wiltrout ME, D'Souza S, Woodruff RV, Walker GC. *Eukaryotic translesion polymerases and their roles and regulation in DNA damage tolerance.* Microbiology and Molecular Biology Reviews. 2009; **73** (1): 134-54.
20. Jalal S, Earley JN, Turchi JJ. *DNA Repair: From Genome Maintenance to Biomarker and Therapeutic Target.* Clinical Cancer Research. 2011; **17** (22): 6973-84.
21. Easton DF, Eeles RA. *Genome-wide association studies in cancer.* Human Molecular Genetics. 2008; **17**: R109-R15.
22. Varghese JS, Easton DF. *Genome-wide association studies in common cancers what have we learnt?* Curr Opin Genet Dev. 2010; **20** (3): 201-9.
23. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, McCarthy MI, Ramos EM, Cardon LR, Chakravarti A, Cho JH, Guttmacher AE, Kong A, Kruglyak L, Mardis E, Rotimi CN, Slatkin M, Valle D, Whittemore AS, Boehnke M, Clark AG, Eichler EE, Gibson G, Haines JL, Mackay TFC, McCarroll SA, Visscher PM. *Finding the missing heritability of complex diseases.* Nature. 2009; **461** (7265): 747-53.
24. Gossage L, Madhusudan S. *Cancer pharmacogenomics - Role of DNA repair genetic Polymorphisms in individualizing cancer therapy.* Mol Diagn Ther. 2007; **11** (6): 361-80.
25. Kamikozuru H, Kuramochi H, Hayashi K, Nakajima G, Yamamoto M. *ERCC1 codon 118 polymorphism is a useful prognostic marker in patients with pancreatic cancer treated with platinum-based chemotherapy.* Int J Oncol. 2008; **32** (5): 1091-6.
26. Kalikaki A, Kanaki M, Vassalou H, Souglakos J, Voutsina A, Georgoulias V, Mavroudis D. *DNA Repair Gene Polymorphisms Predict Favorable Clinical Outcome in Advanced Non-Small-Cell Lung Cancer.* Clin Lung Cancer. 2009; **10** (2): 118-23.
27. Gangawar R, Ahirwar D, Mandhani A, Mittal RD. *Impact of nucleotide excision repair ERCC2 and base excision repair APEX1 genes polymorphism and its association with recurrence after adjuvant BCG immunotherapy in bladder cancer patients of North India.* Med Oncol. 2010; **27** (2): 159-66.
28. Butkiewicz D, Rusin M, Sikora B, Lach A, Chorazy M. *An association between DNA repair gene polymorphisms and survival in patients with resected non-small cell lung cancer.* Mol Biol Rep. 2011; **38** (8): 5231-41.
29. He CY, Duan ZP, Li P, Xu Q, Yuan Y. *Role of ERCC5 promoter polymorphisms in response to platinum-based chemotherapy in patients with advanced non-small-cell lung cancer.* Anti-Cancer Drugs. 2013; **24** (3): 300-5.
30. Li CY, Yin M, Wang LE, Amos CI, Zhu DK, Lee JE, Gershenwald JE, Grimm EA, Wei QY. *Polymorphisms of Nucleotide Excision Repair Genes Predict Melanoma Survival.* Journal of Investigative Dermatology. 2013; **133** (7): 1813-21.

31. Rumiato E, Cavallin F, Boldrin E, Cagol M, Alfieri R, Basso D, Castoro C, Ancona E, Amadori A, Ruol A, Saggioro D. *ERCC1 C8092A (rs3212986) polymorphism as a predictive marker in esophageal cancer patients treated with cisplatin/5-FU-based neoadjuvant therapy*. Pharmacogenet Genomics. 2013; **23** (11): 597-604.
32. Pegg AE, Fang Q, Loktionova NA. *Human variants of O6-alkylguanine-DNA alkyltransferase*. DNA Repair. 2007; **6** (8): 1071-8.
33. Khrunin AV, Moisseev A, Gorbunova V, Limborska S. *Genetic polymorphisms and the efficacy and toxicity of cisplatin-based chemotherapy in ovarian cancer patients*. Pharmacogenomics J. 2010; **10** (1): 54-61.
34. Dubois J, Etienne G, Laroche-Clary A, Lascaux A, Bidet A, Lippert E, Ait-ouferoukh A, Saada V, Micol J, Bouabdallah K, Robert J. *Identification of methylguanine methyltransferase polymorphisms as genetic markers of individual susceptibility to therapy-related myeloid neoplasms*. European Journal of Cancer. 2014; **50**: 7.
35. Kohlmann A, Klein HU, Weissmann S, Bresolin S, Chaplin T, Cuppens H, Haschke-Becher E, Garicochea B, Grossmann V, Hanczaruk B, Hebestreit K, Gabriel C, Iacobucci I, Jansen JH, Kronnie GT, van de Locht L, Martinelli G, McGowan K, Schweiger MR, Timmermann B, Vandenberghe P, Young BD, Dugas M, Haferlach T. *The Interlaboratory RObustness of Next-generation sequencing (IRON) study: a deep sequencing investigation of TET2, CBL and KRAS mutations by an international consortium involving 10 laboratories*. Leukemia. 2011; **25** (12): 1840-8.
36. Lo HS, Wang ZN, Hu Y, Yang HH, Gere S, Buetow KH, Lee MP. *Allelic variation in gene expression is common in the human genome*. Genome Res. 2003; **13** (8): 1855-62.
37. van de Vijver MJ, He YD, van 't Veer LJ, Dai H, Hart AAM, Voskuil DW, Schreiber GJ, Peterse JL, Roberts C, Marton MJ, Parrish M, Atsma D, Witteveen A, Glas A, Delahaye L, van der Velde T, Bartelink H, Rodenhuis S, Rutgers ET, Friend SH, Bernards R. *A gene-expression signature as a predictor of survival in breast cancer*. New England Journal of Medicine. 2002; **347** (25): 1999-2009.
38. Wan YW, Qian Y, Rathnagiriswaran S, Castranova V, Guo NL. *A breast cancer prognostic signature predicts clinical outcomes in multiple tumor types*. Oncol Rep. 2010; **24** (2): 489-94.
39. Ma Y, Qian Y, Wei L, Abraham J, Shi XL, Castranova V, Harner EJ, Flynn DC, Guo L. *Population-based molecular prognosis of breast cancer by transcriptional profiling*. Clinical Cancer Research. 2007; **13** (7): 2014-22.
40. Shedden K, Taylor JMG, Enkemann SA, Tsao MS, Yeatman TJ, Gerald WL, Eschrich S, Jurisica I, Giordano TJ, Misek DE, Chang AC, Zhu CQ, Strumpf D, Hanash S, Shepherd FA, Ding K, Seymour L, Naoki K, Pennell N, Weir B, Verhaak R, Ladd-Acosta C, Golub T, Gruidl M, Sharma A, Szoke J, Zakowski M, Rusch V, Kris M, Viale A, Motoi N, Travis W, Conley B, Seshan VE, Meyerson M, Kuick R, Dobbin KK, Lively T, Jacobson JW, Beer DG, Director's Challenge Consortium M. *Gene expression-based survival prediction in lung adenocarcinoma: a multi-site, blinded validation study*. Nat Med. 2008; **14** (8): 822-7.
41. Bild AH, Yao G, Chang JT, Wang QL, Potti A, Chasse D, Joshi MB, Harpole D, Lancaster JM, Berchuck A, Olson JA, Marks JR, Dressman HK, West M, Nevins JR.

Oncogenic pathway signatures in human cancers as a guide to targeted therapies. Nature. 2006; **439** (7074): 353-7.
42. Dopeso H, Mateo-Lozano S, Elez E, Landolfi S, Pascual FJR, Hernández-Losa J, Mazzolini R, Rodrigues P, Bazzocco S, Carreras MJ. *Aprataxin tumor levels predict response of colorectal cancer patients to irinotecan-based treatment.* Clinical Cancer Research. 2010; **16** (8): 2375-82.
43. Amatu A, Sartore-Bianchi A, Moutinho C, Belotti A, Bencardino K, Chirico G, Cassingena A, Rusconi F, Esposito A, Nichelatti M. *Promoter CpG Island Hypermethylation of the DNA Repair Enzyme MGMT Predicts Clinical Response to Dacarbazine in a Phase II Study for Metastatic Colorectal Cancer.* Clinical Cancer Research. 2013; **19** (8): 2265-72.
44. Olaussen KA, Mountzios G, Soria JC. *ERCC1 as a risk stratifier in platinum-based chemotharapy for non-small-cell lung cancer.* Current Opinion in Pulmonary Medicine. 2007; **13** (4): 284-9.
45. Cloos J, de Boer WPH, Snel MHJ, van den Ijssel P, Ylstra B, Leemans CR, Brakenhoff RH, Braakhuis BJM. *Microarray analysis of bleomycin-exposed lymphoblastoid cells for identifying cancer susceptibility genes.* Molecular Cancer Research. 2006; **4** (2): 71-7.
46. Fachin AL, Mello SS, Sandrin-Garcia P, Junta CM, Ghilardi-Netto T, Donadi EA, Passos GAD, Sakamoto-Hojo ET. *Gene Expression Profiles in Radiation Workers Occupationally Exposed to Ionizing Radiation.* Journal of Radiation Research. 2009; **50** (1): 61-71.
47. Sims AH, Finnon P, Miller CJ, Bouffler SD, Howell A, Scott D, Clarke RB. *TPD52 and NFKB1 gene expression levels correlate with G2 chromosomal radiosensitivity in lymphocytes of women with and at risk of hereditary breast cancer.* Int J Radiat Biol. 2007; **83** (6): 409-20.
48. Fry RC, Svensson JP, Valiathan C, Wang E, Hogan BJ, Bhattacharya S, Bugni JM, Whittaker CA, Samson LD. *Genomic predictors of interindividual differences in response to DNA damaging agents.* Genes Dev. 2008; **22** (19): 2621-6.
49. Fu D, Calvo JA, Samson LD. *Balancing repair and tolerance of DNA damage caused by alkylating agents.* Nat Rev Cancer. 2012; **12** (2): 104-20.
50. Valiathan C, McFaline JL, Samson LD. *A rapid survival assay to measure drug-induced cytotoxicity and cell cycle effects.* DNA Repair. 2012; **11** (1): 92-8.
51. Li C, Wang L-E, Wei Q. *DNA repair phenotype and cancer susceptibility-A mini review.* Int J Cancer. 2009; **124** (5): 999-1007.
52. Wu XF, Gu J, Dong Q, Huang MS, Do KA, Hong WK, Spitz MR. *Joint effect of mutagen sensitivity and insulin-like growth factors in predicting the risk of developing secondary primary tumors and tumor recurrence in patients with head and neck cancer.* Clinical Cancer Research. 2006; **12** (23): 7194-201.
53. Sigurdson AJ, Jones IM, Wei QY, Wu XF, Spitz MR, Stram DA, Gross MD, Huang WY, Wang LE, Gu JA, Thomas CB, Reding DJ, Hayes RB, Caporaso NE. *Prospective analysis of DNA damage and repair markers of lung cancer risk from the Prostate, Lung, Colorectal and Ovarian (PLCO) Cancer Screening Trial.* Carcinogenesis. 2011; **32** (1): 69-73.
54. Watts GS, Pieper RO, Costello JF, Peng YM, Dalton WS, Futscher BW. *Methylation of discrete regions of the O-6-methylguanine DNA methyltransferase*

(MGMT) CpG island is associated with heterochromatinization of the MGMT transcription start site and silencing of the gene. Molecular and Cellular Biology. 1997; **17** (9): 5612-9.
55. Esteller M, Hamilton SR, Burger PC, Baylin SB, Herman JG. *Inactivation of the DNA repair gene O-6-methylguanine-DNA methyltransferase by promoter hypermethylation is a common event in primary human neoplasia.* Cancer Research. 1999; **59** (4): 793-7.
56. Hegi ME, Diserens A, Gorlia T, Hamou M, de Tribolet N, Weller M, Kros JM, Hainfellner JA, Mason W, Mariani L, Bromberg JEC, Hau P, Mirimanoff RO, Cairncross JG, Janzer RC, Stupp R. *MGMT gene silencing and benefit from temozolomide in glioblastoma.* New England Journal of Medicine. 2005; **352** (10): 997-1003.
57. Kreth S, Thon N, Eigenbrod S, Lutz J, Ledderose C, Egensperger R, Tonn JC, Kretzschmar HA, Hinske LC, Kreth FW. *O6-methylguanine-DNA methyltransferase (MGMT) mRNA expression predicts outcome in malignant glioma independent of MGMT promoter methylation.* PLoS One. 2011; **6** (2): e17156.
58. Cankovic M, Nikiforova MN, Snuderl M, Adesina AM, Lindeman N, Wen PY, Lee EQ. *The Role of MGMT Testing in Clinical Practice A Report of the Association for Molecular Pathology.* J Mol Diagn. 2013; **15** (5): 539-55.
59. Zhukovskaya N, Rydberg B, Karran P. *Inactive O6-methylguanine-DNA methyl transferase in human cells.* Nucleic Acids Res. 1992; **20** (22): 6081-90.
60. Ishiguro K, Shyam K, Penketh PG, Bauman RP, Sartorelli AC, Rutherford TJ, Ratner ES. *Expression of O6-Methylguanine-DNA Methyltransferase Examined by Alkyl-Transfer Assays, Methylation-Specific PCR, and Western Blots in Tumors, and Matched Normal Tissue.* Journal of Cancer Therapy. 2013; **4** (4): 14.
61. Iwitzki F, Schlepegrell R, Eichhorn U, Kaina B, Beyersmann D, Hartwig A. *Nickel(II) inhibits the repair of O-6-methylguanine in mammalian cells.* Archives of Toxicology. 1998; **72** (11): 681-9.
62. Szczesny B, Hazra TK, Papaconstantinou J, Mitra S, Boldogh I. *Age-dependent deficiency in import of mitochondrial DNA glycosylases required for repair of oxidatively damaged bases.* Proceedings of the National Academy of Sciences. 2003; **100** (19): 10670-5.
63. Herman JG, Umar A, Polyak K, Graff JR, Ahuja N, Issa J-PJ, Markowitz S, Willson JK, Hamilton SR, Kinzler KW. *Incidence and functional consequences of hMLH1 promoter hypermethylation in colorectal carcinoma.* Proceedings of the National Academy of Sciences. 1998; **95** (12): 6870-5.
64. Taniguchi T, Tischkowitz M, Ameziane N, Hodgson SV, Mathew CG, Joenje H, Mok SC, D'Andrea AD. *Disruption of the Fanconi anemia–BRCA pathway in cisplatin-sensitive ovarian tumors.* Nat Med. 2003; **9** (5): 568-74.
65. Gasco M, Sullivan A, Smith P, Farrell P, Numico G, Colantonio I, Merlano M, Crook T. *Transcriptional silencing of Fanconi anaemia genes and clinical outcome in head and neck cancer.* J Clin Oncol. 2004; **22**: 9546.
66. Narayan G, Arias-Pulido H, Nandula SV, Basso K, Sugirtharaj DD, Vargas H, Mansukhani M, Villella J, Meyer L, Schneider A. *Promoter hypermethylation of FANCF disruption of Fanconi Anemia-BRCA pathway in cervical cancer.* Cancer research. 2004; **64** (9): 2994-7.

67. Levran O, Attwooll C, Henry RT, Milton KL, Neveling K, Rio P, Batish SD, Kalb R, Velleuer E, Barral S. *The BRCA1-interacting helicase BRIP1 is deficient in Fanconi anemia*. Nature genetics. 2005; **37** (9): 931-3.
68. Peng B, Hurt EM, Hodge DR, Thomas SB, Farrar WL. *DNA hypermethylation and partial gene silencing of human thymine-DNA glycosylase in multiple myeloma cell lines*. Epigenetics. 2006; **1** (3): 138-45.
69. Lee M-N, Tseng R-C, Hsu H-S, Chen J-Y, Tzao C, Ho WL, Wang Y-C. *Epigenetic inactivation of the chromosomal stability control genes BRCA1, BRCA2, and XRCC5 in non–small cell lung cancer*. Clinical cancer research. 2007; **13** (3): 832-8.
70. Wu Y, Chang JT, Cheng Y, Wu T, Chen C, Lee H. *Xeroderma pigmentosum group C gene expression is predominantly regulated by promoter hypermethylation and contributes to p53 mutation in lung cancers*. Oncogene. 2007; **26** (33): 4761-73.
71. Wang P, Tang JT, Peng YS, Chen XY, Zhang YJ, Fang JY. *XRCC1 downregulated through promoter hypermethylation is involved in human gastric carcinogenesis*. Journal of Digestive Diseases. 2010; **11** (6): 343-51.
72. Yang J, Xu Z, Li J, Zhang R, Zhang G, Ji H, Song B, Chen Z. *XPC epigenetic silence coupled with p53 alteration has a significant impact on bladder cancer outcome*. The Journal of urology. 2010; **184** (1): 336-43.
73. Rasmussen RE, Painter RB. *Evidence for repair of ultra-violet damaged deoxyribonucleic acid in cultured mammalian cells* Nature. 1964; **203** (495): 1360-&.
74. Pegg AE, Hui G. *Formation and subsequent removal of O6-methylguanine from deoxyribonucleic acid in rat liver and kidney after small doses of dimethylnitrosamine*. Biochem J. 1978; **173** (3): 739-48.
75. Santella RM. *Immunological methods for detection of carcinogen-DNA damage in humans*. Cancer Epidemiology Biomarkers & Prevention. 1999; **8** (9): 733-9.
76. Pero RW, Ostlund C. *Direct comparison, in human resting lymphocytes, of the inter-individual variations in unscheduled DNA synthesis induced by N-acetoxy-2-acetylaminofluorene and ultraviolet radiation*. Mutation Research. 1980; **73** (2): 349-61.
77. Kennedy DO, Agrawal M, Shen J, Terry MB, Zhang FF, Senie RT, Motykiewicz G, Santella RM. *DNA repair capacity of lymphoblastoid cell lines from sisters discordant for breast cancer*. J Natl Cancer Inst. 2005; **97** (2): 127-32.
78. Paz-Elizur T, Elinger D, Leitner-Dagan Y, Blumenstein S, Krupsky M, Berrebi A, Schechtman E, Livneh Z. *Development of an enzymatic DNA repair assay for molecular epidemiology studies: Distribution of OGG activity in healthy individuals*. DNA Repair. 2007; **6** (1): 45-60.
79. Redaelli A, Magrassi R, Bonassi S, Abbondandolo A, Frosina G. *AP endonuclease activity in humans: Development of a simple assay and analysis of ten normal individuals*. Teratogenesis Carcinogenesis and Mutagenesis. 1998; **18** (1): 17-26.
80. Shen GP, Galick H, Inoue M, Wallace SS. *Decline of nuclear and mitochondrial oxidative base excision repair activity in late passage human diploid fibroblasts*. DNA Repair. 2003; **2** (6): 673-93.
81. Parsons JL, Dianova, II, Dianov GL. *APE1-dependent repair of DNA single-strand breaks containing 3 '-end 8-oxoguanine*. Nucleic Acids Res. 2005; **33** (7): 2204-9.

82. Wilson DM, Kim D, Berquist BR, Sigurdson AJ. *Variation in base excision repair capacity*. Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis. 2011; **711** (1): 100-12.
83. Georgiadis P, Polychronaki N, Kyrtopoulos SA. *Progress in high-throughput assays of MGMT and APE1 activities in cell extracts*. Mutat Res-Fundam Mol Mech Mutagen. 2012; **736** (1-2): 25-32.
84. Lu AL, Clark S, Modrich P. *Methyl-directed repair of DNA base-pair mismatches in vitro* Proceedings of the National Academy of Sciences of the United States of America-Biological Sciences. 1983; **80** (15): 4639-43.
85. Wu RS, Hurstcalderone S, Kohn KW. *Measurement of O-6-alkylguanine DNA alkyltransferase activity in human cells and tumor tissues by restriction endonuclease inhibition*. Cancer Research. 1987; **47** (23): 6229-35.
86. Mu D, Hsu DS, Sancar A. *Reaction mechanism of human DNA repair excision nuclease*. J Biol Chem. 1996; **271** (14): 8285-94.
87. Zhong Q, Boyer TG, Chen PL, Lee WH. *Deficient nonhomologous end-joining activity in cell-free extracts from Brca1-null fibroblasts*. Cancer Research. 2002; **62** (14): 3966-70.
88. Guainazzi A, Scharer OD. *Using synthetic DNA interstrand crosslinks to elucidate repair pathways and identify new therapeutic targets for cancer chemotherapy*. Cell Mol Life Sci. 2010; **67** (21): 3683-97.
89. Kucherlapati RS, Spencer J, Moore PD. *Homologous recombination catalyzed by human cell extracts*. Molecular and Cellular Biology. 1985; **5** (4): 714-20.
90. O'Donnell PNS, Barber PV, Margison GP, Povey AC. *Association between O-6-alkylguanine-DNA- alkyltransferase activity in peripheral blood lymphocytes and bronchial epithelial cells*. Cancer Epidemiology Biomarkers & Prevention. 1999; **8** (7): 641-5.
91. Crosbie PAJ, Watson AJ, Agius R, Barber PV, Margison GP, Povey AC. *Elevated N3-methylpurine-DNA glycosylase DNA repair activity is associated with lung cancer*. Mutat Res-Fundam Mol Mech Mutagen. 2012; **732** (1-2): 43-6.
92. Calvo JA, Moroski-Erkul CA, Lake A, Eichinger LW, Shah D, Jhun I, Limsirichai P, Bronson RT, Christiani DC, Meira LB, Samson LD. *Aag DNA Glycosylase Promotes Alkylation-Induced Tissue Damage Mediated by Parp1*. Plos Genetics. 2013; **9** (4).
93. Azqueta A, Collins AR. *The essential comet assay: a comprehensive guide to measuring DNA damage and repair*. Archives of Toxicology. 2013; **87** (6): 949-68.
94. McKenna DJ, McKeown SR, McKelvey-Martin VJ. *Potential use of the comet assay in the clinical management of cancer*. Mutagenesis. 2008; **23** (3): 183-90.
95. Decordier I, Loock KV, Kirsch-Volders M. *Phenotyping for DNA repair capacity*. Mutation Research-Reviews in Mutation Research. 2010; **705** (2): 107-29.
96. Gaivão I, Piasek A, Brevik A, Shaposhnikov S, Collins AR. *Comet assay-based methods for measuring DNA repair in vitro; estimates of inter-and intra-individual variation*. Cell biology and toxicology. 2009; **25** (1): 45-52.
97. Kalthur G, Kumar P, Devi U, Ali S, Upadhya R, Pillai S, Rao A. *Susceptibility of peripheral lymphocytes of brain tumour patients to in vitro radiation-induced DNA damage, a preliminary study*. Clinical and Experimental Medicine. 2008; **8** (3): 147-50.

98. Jaloszynski P, Kujawski M, Czub-Swierczek M, Markowska J, Szyfter K. *Bleomycin-induced DNA damage and its removal in lymphocytes of breast cancer patients studied by comet assay*. Mutation Research-DNA Repair. 1997; **385** (3): 223-33.
99. Rajaee-Behbahani N, Schmezer P, Risch A, Rittgen W, Kayser KW, Dienemann H, Schulz V, Drings P, Thiel S, Bartsch H. *Altered DNA repair capacity and bleomycin sensitivity as risk markers for non-small cell lung cancer*. Int J Cancer. 2001; **95** (2): 86-91.
100. El-Zein RA, Monroy CM, Cortes A, Spitz MR, Greisinger A, Etzel CJ. *Rapid method for determination of DNA repair capacity in human peripheral blood lymphocytes amongst smokers*. BMC Cancer. 2010; **10**.
101. Palyvoda O, Polanska J, Wygoda A, Rzeszowska-Wolny J. *DNA damage and repair in lymphocytes of normal individuals and cancer patients: Studies by the comet assay and micronucleus tests*. Acta Biochim Pol. 2003; **50** (1): 181-90.
102. Cox JD, Stetz J, Pajak TF. *Toxicity criteria of the radiation therapy oncology group (RTOG) and the European organization for research and treatment of cancer (EROTC)* International Journal of Radiation Oncology Biology Physics. 1995; **31** (5): 1341-6.
103. Alapetite C, Thirion P, de la Rochefordiere A, Cosset JM, Moustacchi E. *Analysis by alkaline comet assay of cancer patients with severe reactions to radiotherapy: Defective rejoining of radioinduced dna strand breaks in lymphocytes of breast cancer patients*. Int J Cancer. 1999; **83** (1): 83-90.
104. Wood DK, Weingeist DM, Bhatia SN, Engelward BP. *Single cell trapping and DNA damage analysis using microwell arrays*. Proceedings of the National Academy of Sciences of the United States of America. 2010; **107** (22): 10008-13.
105. Johnson JM, Latimer JJ. Analysis of DNA Repair Using Transfection-Based Host Cell Reactivation. In: Keohavong P, Grant GG, editors. Molecular Toxicology Protocols. Totowa, NJ: Humana Press; 2005.
106. Athas WF, Hedayati MA, Matanoski GM, Farmer ER, Grossman L. *Development and field-test validation of an assay for DNA-repair in circulating human lymphocytes*. Cancer Research. 1991; **51** (21): 5786-93.
107. D'Errico M, Calcagnile A, Iavarone I, Sera F, Baliva G, Chinni LM, Corona R, Pasquini P, Dogliotti E. *Factors that influence the DNA repair capacity of normal and skin cancer-affected individuals*. Cancer Epidemiology Biomarkers & Prevention. 1999; **8** (6): 553-9.
108. Wei QY, Lee JE, Gershenwald JE, Ross MI, Mansfield PF, Strom SS, Wang LE, Guo ZZ, Qiao YW, Amos CI, Spitz MR, Duvic M. *Repair of UV light-induced DNA damage and risk of cutaneous malignant melanoma*. J Natl Cancer Inst. 2003; **95** (4): 308-15.
109. Wang LE, Li CY, Strom SS, Goldberg LH, Brewster A, Guo ZZ, Qiao YW, Clayman GL, Lee JJ, El-Naggar AK, Prieto VG, Duvic M, Lippman SM, Weber RS, Kripke ML, Wei QY. *Repair capacity for UV light-induced DNA damage associated with risk of nonmelanoma skin cancer and tumor progression*. Clinical Cancer Research. 2007; **13** (21): 6532-9.

110. Wei QY, Gu J, Cheng L, Bondy ML, Jiang H, Hong WK, Spitz MR. *Benzo(a)pyrene diol epoxide-induced chromosomal aberrations and risk of lung cancer.* Cancer Research. 1996; **56** (17): 3975-9.
111. Wei Q, Cheng L, Amos CI, Wang LE, Guo Z, Hong WK, Spitz MR. *Repair of tobacco carcinogen-induced DNA adducts and lung cancer risk: a molecular epidemiologic study.* J Natl Cancer Inst. 2000; **92** (21): 1764-72.
112. Spitz MR, Wei QY, Dong Q, Amos CI, Wu XF. *Genetic susceptibility to lung cancer: The role of DNA damage and repair.* Cancer Epidemiology Biomarkers & Prevention. 2003; **12** (8): 689-98.
113. Wang L, Wei QY, Shi QL, Guo ZS, Qiao YW, Spitz MR. *A modified host-cell reactivation assay to measure repair of alkylating DNA damage for assessing risk of lung adenocarcinoma.* Carcinogenesis. 2007; **28** (7): 1430-6.
114. Shen HB, Spitz MR, Qiao YW, Guo ZZ, Wang LE, Bosken CH, Amos CI, Wei QY. *Smoking, DNA repair capacity and risk of nonsmall cell lung cancer.* Int J Cancer. 2003; **107** (1): 84-8.
115. Ramos JM, Ruiz A, Colen R, Lopez ID, Grossman L, Matta JL. *DNA repair and breast carcinoma susceptibility in women.* Cancer. 2004; **100** (7): 1352-7.
116. Shi QL, Wang LE, Bondy ML, Brewster A, Singletary SE, Wei QY. *Reduced DNA repair of benzo a pyrene diol epoxide-induced adducts and common XPD polymorphisms in breast cancer patients.* Carcinogenesis. 2004; **25** (9): 1695-700.
117. Matta J, Echenique M, Negron E, Morales L, Vargas W, Gaetan FS, Lizardi ER, Torres A, Rosado JO, Bolanos G, Cruz JG, Laboy J, Barnes R, Medina SS, Romero A, Martinez R, Dutil J, Suarez E, Alvarez-Garriga C, Bayona M. *The association of DNA Repair with breast cancer risk in women. A comparative observational study.* BMC Cancer. 2012; **12.**
118. Cheng L, Eicher SA, Guo ZZ, Hong WK, Spitz MR, Wei QY. *Reduced DNA repair capacity in head and neck cancer patients.* Cancer Epidemiology Biomarkers & Prevention. 1998; **7** (6): 465-8.
119. Wang LE, Hu ZB, Sturgis EM, Spitz MR, Strom SS, Amos CI, Guo ZZ, Qiao YW, Gillenwater AM, Myers JN, Clayman GL, Weber RS, El-Naggar AK, Mao L, Lippman SM, Hong WK, Wei QY. *Reduced DNA Repair Capacity for Removing Tobacco Carcinogen-Induced DNA Adducts Contributes to Risk of Head and Neck Cancer but not Tumor Characteristics.* Clinical Cancer Research. 2010; **16** (2): 764-74.
120. Lin J, Kadlubar FF, Spitz MR, Zhao H, Wu XF. *A modified host cell reactivation assay to measure DNA repair capacity for removing 4-aminobiphenyl adducts: A pilot study of bladder cancer.* Cancer Epidemiology Biomarkers & Prevention. 2005; **14** (7): 1832-6.
121. Tyson J, Caple F, Spiers A, Burtle B, Daly AK, Williams EA, Hesketh JE, Mathers JC. *Inter-individual variation in nucleotide excision repair in young adults: effects of age, adiposity, micronutrient supplementation and genotype.* British Journal of Nutrition. 2009; **101** (9): 1316.
122. Mendez P, Taron M, Moran T, Fernandez MA, Requena G, Rosell R. *A modified host-cell reactivation assay to quantify DNA repair capacity in cryopreserved peripheral lymphocytes.* DNA Repair. 2011; **10** (6): 603-10.
123. Machella N, Terry MB, Zipprich J, Gurvich I, Liao YY, Senie RT, Kennedy DO, Santella RM. *Double-strand breaks repair in lymphoblastoid cell lines from sisters*

discordant for breast cancer from the New York site of the BCFR. Carcinogenesis. 2008; **29** (7): 1367-72.
124. Slebos RJC, Taylor JA. *A novel host cell reactivation assay to assess homologous recombination capacity in human cancer cell lines.* Biochemical and Biophysical Research Communications. 2001; **281** (1): 212-9.
125. Kiziltepe T, Yan A, Dong M, Jonnalagadda VS, Dedon PC, Engelward BP. *Delineation of the chemical pathways underlying nitric oxide-induced homologous recombination in mammalian cells.* Chemistry & Biology. 2005; **12** (3): 357-69.
126. Lei XF, Zhu Y, Tomkinson A, Sun LZ. *Measurement of DNA mismatch repair activity in live cells.* Nucleic Acids Res. 2004; **32** (12).
127. Raetz AG, Xie Y, Kundu S, Brinkmeyer MK, Chang C, David SS. *Cancer-associated variants and a common polymorphism of MUTYH exhibit reduced repair of oxidative DNA damage using a GFP-based assay in mammalian cells.* Carcinogenesis. 2012.
128. Rahal EA, Henricksen LA, Li YL, Williams RS, Tainer JA, Dixon K. *ATM regulates Mre11-dependent DNA end-degradation and microhomology-mediated end joining.* Cell Cycle. 2010; **9** (14): 2866-77.
129. Sun Y, Moses R. *Reactivation of psoralen-reacted plasmid DNA in Fanconi anemia, xeroderma pigmentosum, and normal human fibroblast cells.* Somatic cell and molecular genetics. 1991; **17** (3): 229-38.
130. Hlavin EM, Smeaton MB, Noronha AM, Wilds CJ, Miller PS. *Cross-link structure affects replication-independent DNA interstrand cross-link repair in mammalian cells.* Biochemistry. 2010; **49** (18): 3977-88.
131. Brooks PJ, Wise DS, Berry DA, Kosmoski JV, Smerdon MJ, Somers RL, Mackie H, Spoonde AY, Ackerman EJ, Coleman K, Tarone RE, Robbins JH. *The oxidative DNA lesion 8,5 '-(S)-cyclo-2 '-deoxyadenosine is repaired by the nucleotide excision repair pathway and blocks gene expression in mammalian cells.* J Biol Chem. 2000; **275** (29): 22355-62.
132. Spivak G, Hanawalt PC. *Host cell reactivation of plasmids containing oxidative DNA lesions is defective in Cockayne syndrome but normal in UV-sensitive syndrome fibroblasts.* DNA Repair. 2006; **5** (1): 13-22.
133. Nagel ZD, Thompson CM, Chaim IA, McRee SK, Mazzucato P, Ahmad A, Abo RP, Butty VL, Forget AL, Samson LD. *Multiplexed DNA repair assays for multiple lesions and multiple doses via transcription inhibition and transcriptional mutagenesis.* Proceedings of the National Academy of Sciences. 2014; ***In Press***.
134. Reeves R, Gorman CM, Howard B. *Minichromosome assembly of non-integrated plasmid DNA transfected into mammalian cells* Nucleic Acids Res. 1985; **13** (10): 3599-615.
135. Jeong S, Stein A. *DNA sequence affects nucleosome ordering on replicating plasmids in transfected COS-1 cells and in vitro.* J Biol Chem. 1994; **269** (3): 2197-205.
136. Jeong SW, Stein A. *Micrococcal nuclease digestion of nuclei reveals extended nucleosome ladders having anomalous lengths for chromatin assembled on nonreplicating plasmids in transfected cells* Nucleic Acids Res. 1994; **22** (3): 370-5.

137. Mladenova V, Mladenov E, Russev G. *Organization of plasmid DNA into nucleosome-like structures after transfection in eukaryotic cells.* Biotechnology & Biotechnological Equipment. 2009; **23** (1): 1044-7.
138. Khobta A, Anderhub S, Kitsera N, Epe B. *Gene silencing induced by oxidative DNA base damage: association with local decrease of histone H4 acetylation in the promoter region.* Nucleic Acids Res. 2010; **38** (13): 4285-95.
139. Fedier A, Fink D. *Mutations in DNA mismatch repair genes: Implications for DNA damage signaling and drug sensitivity (review).* Int J Oncol. 2004; **24** (4): 1039-47.
140. Stojic L, Brun R, Jiricny J. *Mismatch repair and DNA damage signalling.* DNA Repair. 2004; **3** (8-9): 1091-101.
141. Kat A, Thilly WG, Fang WH, Longley MJ, Li GM, Modrich P. *An alkylation-tolerant, mutator human cell line is deficient in strand-specific mismatch repair* Proceedings of the National Academy of Sciences of the United States of America. 1993; **90** (14): 6424-8.
142. Rouleau M, Patel A, Hendzel MJ, Kaufmann SH, Poirier GG. *PARP inhibition: PARP1 and beyond.* Nat Rev Cancer. 2010; **10** (4): 293-301.
143. Leitner-Dagan Y, Sevilya Z, Pinchev M, Kramer R, Elinger D, Roisman LC, Rennert HS, Schechtman E, Freedman L, Rennert G, Livneh Z, Paz-Elizur T. *N-Methylpurine DNA Glycosylase and OGG1 DNA Repair Activities: Opposite Associations With Lung Cancer Risk.* J Natl Cancer Inst. 2012; **104** (22): 1765-9.
144. Sevilya Z, Leitner-Dagan Y, Pinchev M, Kremer R, Elinger D, Rennert HS, Schechtman E, Freedman LS, Rennert G, Paz-Elizur T. *Low integrated DNA repair score and lung cancer risk.* Cancer Prevention Research. 2013: canprevres. 0318.2013.
145. Posnick LM, Samson LD. *Imbalanced base excision repair increases spontaneous mutation and alkylation sensitivity in Escherichia coli.* Journal of Bacteriology. 1999; **181** (21): 6763-71.
146. Jungmichel S, Rosenthal F, Altmeyer M, Lukas J, Hottiger MO, Nielsen ML. *Proteome-wide Identification of Poly(ADP-Ribosyl)ation Targets in Different Genotoxic Stress Responses.* Molecular cell. 2013; **52** (2): 272-85.
147. Sobol RW, Kartalou M, Almeida KH, Joyce DF, Engelward BP, Horton JK, Prasad R, Samson LD, Wilson SH. *Base excision repair intermediates induce p53-independent cytotoxic and genotoxic responses.* J Biol Chem. 2003; **278** (41): 39951-9.
148. Berger NA. *Poly(ADP-ribose) polymerase in the cellular response to DNA damage.* Radiat Res. 1985; **101** (1): 4-15.
149. Schreiber V, Dantzer F, Ame JC, de Murcia G. *Poly(ADP-ribose): novel functions for an old molecule.* Nature Reviews Molecular Cell Biology. 2006; **7** (7): 517-28.
150. Horton JK, Wilson SH. *Predicting Enhanced Cell Killing through PARP Inhibition.* Molecular Cancer Research. 2013; **11** (1): 13-8.
151. Strom CE, Johansson F, Uhlen M, Szigyarto CA, Erixon K, Helleday T. *Poly (ADP-ribose) polymerase (PARP) is not involved in base excision repair but PARP inhibition traps a single-strand intermediate.* Nucleic Acids Res. 2011; **39** (8): 3166-75.

152. Murai J, Huang SN, Renaud A, Zhang YZ, Ji J, Takeda S, Morris J, Teicher B, Doroshow JH, Pommier Y. *Stereospecific PARP trapping by BMN 673 and comparison with olaparib and rucaparib.* Molecular Cancer Therapeutics. 2014; **13** (2): 12.
153. Heacock ML, Stefanick DF, Horton JK, Wilson SH. *Alkylation DNA damage in combination with PARP inhibition results in formation of S-phase-dependent double-strand breaks.* DNA Repair. 2010; **9** (8): 929-36.
154. Kedar PS, Stefanick DF, Horton JK, Wilson SH. *Increased PARP-1 Association with DNA in Alkylation Damaged, PARP-Inhibited Mouse Fibroblasts.* Molecular Cancer Research. 2012; **10** (3): 360-8.
155. Tang JB, Svilar D, Trivedi RN, Wang XH, Goellner EM, Moore B, Hamilton RL, Banze LA, Brown AR, Sobol RW. *N-methylpurine DNA glycosylase and DNA polymerase beta modulate BER inhibitor potentiation of glioma cells to temozolomide.* Neuro-oncology. 2011; **13** (5): 471-86.
156. Horton JK, Watson M, Stefanick DF, Shaughnessy DT, Taylor JA, Wilson SH. *XRCC1 and DNA polymerase beta in cellular protection against cytotoxic DNA single-strand breaks.* Cell Res. 2008; **18** (1): 48-63.
157. Tobin LA, Robert C, Rapoport AP, Gojo I, Baer MR, Tomkinson AE, Rassool FV. *Targeting abnormal DNA double-strand break repair in tyrosine kinase inhibitor-resistant chronic myeloid leukemias.* Oncogene. 2013; **32** (14): 1784-93.
158. Caldecott KW. *Single-strand break repair and genetic disease.* Nature Reviews Genetics. 2008; **9** (8): 619-31.
159. Gomez-Herreros F, Romero-Granados R, Zeng ZH, Alvarez-Quilon A, Quintero C, Ju LM, Umans L, Vermeire L, Huylebroeck D, Caldecott KW, Cortes-Ledesma F. *TDP2-Dependent Non-Homologous End-Joining Protects against Topoisomerase II-Induced DNA Breaks and Genome Instability in Cells and In Vivo.* Plos Genetics. 2013; **9** (3).
160. Zeng ZH, Sharma A, Ju LM, Murai J, Umans L, Vermeire L, Pommier Y, Takeda S, Huylebroeck D, Caldecott KW, El-Khamisy SF. *TDP2 promotes repair of topoisomerase I-mediated DNA damage in the absence of TDP1.* Nucleic Acids Res. 2012; **40** (17): 8371-80.
161. van Loon B, Hubscher U. *An 8-oxo-guanine repair pathway coordinated by MUTYH glycosylase and DNA polymerase lambda.* Proceedings of the National Academy of Sciences of the United States of America. 2009; **106** (43): 18201-6.
162. Yang N, Galick H, Wallace SS. *Attempted base excision repair of ionizing radiation damage in human lymphoblastoid cells produces lethal and mutagenic double strand breaks.* DNA Repair. 2004; **3** (10): 1323-34.
163. Berger SH, Pittman DL, Wyatt MD. *Uracil in DNA: consequences for carcinogenesis and chemotherapy.* Biochemical Pharmacology. 2008; **76** (6): 10.
164. Dou H, Mitra S, Hazra TK. *Repair of oxidized bases in DNA bubble structures by human DNA glycosylases NEIL1 and NEIL2.* J Biol Chem. 2003; **278** (50): 49679-84.
165. Hang B, Medina M, Fraenkel-Conrat H, Singer B. *A 55-kDa protein isolated from human cells shows DNA glycosylase activity toward 3,N-4-ethenocytosine and the G/T mismatch.* Proceedings of the National Academy of Sciences of the United States of America. 1998; **95** (23): 13561-6.
166. Ringvoll J, Moen MN, Nordstrand LM, Meira LB, Pang B, Bekkelund A, Dedon PC, Bjelland S, Samson LD, Falnes PO, Klungland A. *AlkB homologue 2-mediated*

repair of ethenoadenine lesions in mammalian DNA. Cancer Research. 2008; **68** (11): 4142-9.
167. Fu D, Samson LD. *Direct repair of 3,N4-ethenocytosine by the human ALKBH2 dioxygenase is blocked by the AAG/MPG glycosylase*. DNA repair. 2011.
168. Deans AJ, West SC. *DNA interstrand crosslink repair and cancer*. Nat Rev Cancer. 2011; **11** (7): 467-80.
169. Rivera-Munoz P, Malivert L, Derdouch S, Azerrad C, Abramowski V, Revy P, Villartay J-Pd. *DNA repair and the immune system: From V (D) J recombination to aging lymphocytes*. European journal of immunology. 2007; **37** (S1): S71-S82.
170. van der Burg M, Ijspeert H, Verkaik NS, Turul T, Wiegant WW, Morotomi-Yano K, Mari P-O, Tezcan I, Chen DJ, Zdzienicka MZ, van Dongen JJ, van Gent DC. *A DNA-PKcs mutation in a radiosensitive T-B- SCID patient inhibits Artemis activation and nonhomologous end-joining*. The Journal of clinical investigation. 2009; **119** (1): 91.
171. Moshous D, Callebaut I, de Chasseval R, Corneo B, Cavazzana-Calvo M, Le Deist F, Tezcan I, Sanal O, Bertrand Y, Philippe N, Fischer A, de Villartay J-P. *Artemis, a novel DNA double-strand break repair/V (D) J recombination protein, is mutated in human severe combined immune deficiency*. Cell. 2001; **105** (2): 177-86.
172. Moshous D, Pannetier C, de Chasseval Rg, le Deist Fo, Cavazzana-Calvo M, Romana S, Macintyre E, Canioni D, Brousse N, Fischer A, Casanova J-L, de Villartay J-P. *Partial T and B lymphocyte immunodeficiency and predisposition to lymphoma in patients with hypomorphic mutations in Artemis*. Journal of Clinical Investigation. 2003; **111** (3): 381-7.
173. O'Driscoll M, Cerosaletti KM, Girard P-M, Dai Y, Stumm M, Kysela B, Hirsch B, Gennery A, Palmer SE, Seidel J. *DNA ligase IV mutations identified in patients exhibiting developmental delay and immunodeficiency*. Molecular cell. 2001; **8** (6): 1175-85.
174. Buck D, Malivert L, de Chasseval R, Barraud A, Fondaneche M-C, Sanal O, Plebani A, Stephan J-L, Hufnagel M, le Deist F. *Cernunnos, a novel nonhomologous end-joining factor, is mutated in human immunodeficiency with microcephaly*. Cell. 2006; **124** (2): 287-99.
175. Petersen S, Casellas R, Reina-San-Martin B, Chen HT, Difilippantonio MJ, Wilson PC, Hanitsch L, Celeste A, Muramatsu M, Pilch DR. *AID is required to initiate Nbs1/ γ -H2AX focus formation and mutations at sites of class switching*. Nature. 2001; **414** (6864): 660-5.
176. Petersen-Mahrt SK, Harris RS, Neuberger MS. *AID mutates E. coli suggesting a DNA deamination mechanism for antibody diversification*. Nature. 2002; **418** (6893): 99-104.
177. Rada C, Di Noia JM, Neuberger MS. *Mismatch recognition and uracil excision provide complementary paths to both Ig switching and the A/T-focused phase of somatic mutation*. Molecular cell. 2004; **16** (2): 163-71.
178. Imai K, Slupphaug G, Lee W-I, Revy P, Nonoyama S, Catalan N, Yel L, Forveille M, Kavli B, Krokan HE, Ochs HD, Fischer A, Durandy A. *Human uracil-DNA glycosylase deficiency associated with profoundly impaired immunoglobulin class-switch recombination*. Nature immunology. 2003; **4** (10): 1023-8.

179. Péron S, Metin A, Gardes P, Alyanakian M-A, Sheridan E, Kratz CP, Fischer A, Durandy A. *Human PMS2 deficiency is associated with impaired immunoglobulin class switch recombination*. The Journal of experimental medicine. 2008; **205** (11): 2465-72.
180. de Miranda N, Bjorkman A, Pan-Hammarstrom Q, Annals NYAS. *DNA repair: the link between primary immunodeficiency and cancer*. AnnNY AcadSci. 2011; **1246**: 50-63.
181. de Miranda NF, Peng R, Georgiou K, Wu C, Sörqvist EF, Berglund M, Chen L, Gao Z, Lagerstedt K, Lisboa S. *DNA repair genes are selectively mutated in diffuse large B cell lymphomas*. The Journal of experimental medicine. 2013; **210** (9): 1729-42.
182. Grigera F, Bellacosa A, Kenter AL. *Complex Relationship between Mismatch Repair Proteins and MBD4 during Immunoglobulin Class Switch Recombination*. PloS one. 2013; **8** (10): e78370.
183. Alter BP. Diagnostic Evaluation of FA. In: Eiler ME, Frohnmayer D, Frohnmayer L, Larsen K, Owen J, editors. Fanconi Anemia: Guidelines for Diagnosis and Management. 3 ed: Fanconi Anemia Research Fund, Inc.; 2008.
184. Bernstam VA. CRC Handbook of Gene Level Diagnostics in Clinical Practice: CRC Press; 1992.
185. Aebi S, KurdiHaidar B, Gordon R, Cenni B, Zheng H, Fink D, Christen RD, Boland CR, Koi M, Fishel R, Howell SB. *Loss of DNA mismatch repair in acquired resistance to cisplatin*. Cancer Research. 1996; **56** (13): 3087-90.
186. Carethers JM, Hawn MT, Chauhan DP, Luce MC, Marra G, Koi M, Boland CR. *Competency in mismatch repair prohibits clonal expansion of cancer cells treated with N-methyl-N'-nitro-N-nitrosoguanidine*. Journal of Clinical Investigation. 1996; **98** (1): 199-206.
187. Wyatt MD, Wilson DM. *Participation of DNA repair in the response to 5-fluorouracil*. Cell Mol Life Sci. 2009; **66** (5): 788-99.
188. Olaussen KA, Dunant A, Fouret P, Brambilla E, Andre F, Haddad V, Taranchon E, Filipits M, Pirker R, Popper HH, Stahel R, Sabatier L, Pignon J, Tursz T, Le Chevalier T, Soria JC, Investigators IB. *DNA repair by ERCC1 in non-small-cell lung cancer and cisplatin-based adjuvant chemotherapy*. New England Journal of Medicine. 2006; **355** (10): 983-91.
189. Esteller M, Garcia-Foncillas J, Andion E, Goodman SN, Hidalgo OF, Vanaclocha V, Baylin SB, Herman JG. *Inactivation of the DNA-repair gene MGMT and the clinical response of gliomas to alkylating agents*. New England Journal of Medicine. 2000; **343** (19): 1350-4.
190. Sarkaria JN, Kitange GJ, James CD, Plummer R, Calvert H, Weller M, Wick W. *Mechanisms of chemoresistance to alkylating agents in malignant glioma*. Clinical Cancer Research. 2008; **14** (10): 2900-8.
191. Boland CR. *Clinical uses of microsatellite instability testing in colorectal cancer: an ongoing challenge*. Journal of clinical oncology. 2007; **25** (7): 754-6.
192. Öhrling K, Edler D, Hallström M, Ragnhammar P. *Mismatch repair protein expression is an independent prognostic factor in sporadic colorectal cancer*. Acta Oncologica. 2010; **49** (6): 797-804.

193. Stupp R, Tonn JC, Brada M, Pentheroudakis G, Grp EGW. *High-grade malignant glioma: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up*. Annals of Oncology. 2010; **21**: v190-v3.
194. Fruh M, De Ruysscher D, Popat S, Crino L, Peters S, Felip E, Grp EGW. *Small-cell lung cancer (SCLC): ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up*. Annals of Oncology. 2013; **24**: 99-105.
195. Horwich A, Parker C, de Reijke T, Kataja V, Grp EGW. *Prostate cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up*. Annals of Oncology. 2013; **24**: 106-14.
196. Labianca R, Nordlinger B, Beretta GD, Mosconi S, Mandala M, Cervantes A, Arnold D, Grp EGW. *Early colon cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up*. Annals of Oncology. 2013; **24**: 64-72.
197. Senkus E, Kyriakides S, Penault-Llorca F, Poortmans P, Thompson A, Zackrisson S, Cardoso F, Grp EGW. *Primary breast cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up*. Annals of Oncology. 2013; **24**: 7-23.
198. Vansteenkiste J, De Ruysscher D, Eberhardt WEE, Lim E, Senan S, Felip E, Peters S, Grp EGW. *Early and locally advanced non-small-cell lung cancer (NSCLC): ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up*. Annals of Oncology. 2013; **24**: 89-98.
199. Bhatti P, Sigurdson AJ, Thomas CB, Iwan A, Alexander BH, Kampa D, Bowen L, Doody MM, Jones IM. *No evidence for differences in DNA damage assessed before and after a cancer diagnosis*. Cancer Epidemiology Biomarkers & Prevention. 2008; **17** (4): 990-4.
200. Jaiswal M, LaRusso NF, Burgart LJ, Gores GJ. *Inflammatory cytokines induce DNA damage and inhibit DNA repair in cholangiocarcinoma cells by a nitric oxide-dependent mechanism*. Cancer Research. 2000; **60** (1): 184-90.
201. Tsoncheva VL, Todorova KA, Ivanov IG, Maximova VA. *Influence of interferons on the repair of UV-damaged DNA*. ZNaturforsch(C). 2008; **63** (3-4): 303-7.
202. Chechlinska M, Kowalewska M, Nowak R. *Systemic inflammation as a confounding factor in cancer biomarker discovery and validation*. Nat Rev Cancer. 2010; **10** (1): 2-U13.
203. Machida K, Tsukamoto H, Liu JC, Han YP, Govindarajan S, Lai MMC, Akira S, Ou JHJ. *c-Jun Mediates Hepatitis C Virus Hepatocarcinogenesis Through Signal Transducer and Activator of Transcription 3 and Nitric Oxide-Dependent Impairment of Oxidative DNA Repair*. Hepatology. 2010; **52** (2): 480-92.
204. Yakovlev VA. *Nitric Oxide-Dependent Downregulation of BRCA1 Expression Promotes Genetic Instability*. Cancer Research. 2013; **73** (2): 706-15.
205. Soreide K. *Molecular testing for microsatellite instability and DNA mismatch repair defects in hereditary and sporadic colorectal cancers - Ready for prime time?* Tumor Biol. 2007; **28** (5): 290-300.
206. Paz-Elizur T, Krupsky M, Blumenstein S, Elinger D, Schechtman E, Livneh Z. *DNA repair activity for oxidative damage and risk of lung cancer*. J Natl Cancer Inst. 2003; **95** (17): 1312-9.

207. Myrnes B, Giercksky KE, Krokan H. *Interindividual variation in the activity of O6-methyl guanine DNA methyltransferase and uracil DNA glycosylase in human organs* Carcinogenesis. 1983; **4** (12): 1565-8.
208. Gerson SL. *Clinical relevance of MGMT in the treatment of cancer.* Journal of Clinical Oncology. 2002; **20** (9): 2388-99.
209. Park IH, Zhao R, West JA, Yabuuchi A, Huo HG, Ince TA, Lerou PH, Lensch MW, Daley GQ. *Reprogramming of human somatic cells to pluripotency with defined factors.* Nature. 2008; **451** (7175): 141-U1.
210. Chan SSL, Longley MJ, Copeland WC. *The common A467T mutation in the human mitochondrial DNA polymerase (POLG) compromises catalytic efficiency and interaction with the accessory subunit.* J Biol Chem. 2005; **280** (36): 31341-6.
211. Lin MT, Beal MF. *Mitochondrial dysfunction and oxidative stress in neurodegenerative diseases.* Nature. 2006; **443** (7113): 787-95.
212. Lauritzen KH, Moldestad O, Eide L, Carlsen H, Nesse G, Storm JF, Mansuy IM, Bergersen LH, Klungland A. *Mitochondrial DNA Toxicity in Forebrain Neurons Causes Apoptosis, Neurodegeneration, and Impaired Behavior.* Molecular and Cellular Biology. 2010; **30** (6): 1357-67.
213. Liu PF, Demple B. *DNA Repair in Mammalian Mitochondria: Much More Than We Thought?* Environ Mol Mutagen. 2010; **51** (5): 417-26.
214. Mao PZ, Reddy PH. *Aging and amyloid beta-induced oxidative DNA damage and mitochondrial dysfunction in Alzheimer's disease: Implications for early intervention and therapeutics.* Biochimica Et Biophysica Acta-Molecular Basis of Disease. 2011; **1812** (11): 1359-70.
215. Meyer JN, Bess AS. *Involvement of autophagy and mitochondrial dynamics in determining the fate and effects of irreparable mitochondrial DNA damage.* Autophagy. 2012; **8** (12): 1822-3.
216. Furda AM, Marrangoni AM, Lokshin A, Van Houten B. *Oxidants and not alkylating agents induce rapid mtDNA loss and mitochondrial dysfunction.* DNA Repair. 2012; **11** (8): 684-92.
217. Du L, Wang H, Xiong T, Ma Y, Yang J, Huang J, Zeng D, Wang X, Huang H, Huang J. *The polymorphisms in the MGMT gene and the risk of cancer: a meta-analysis.* Tumor Biol. 2013: 1-11.
218. Fishel R, Lescoe MK, Rao M, Copeland NG, Jenkins NA, Garber J, Kane M, Kolodner R. *The human mutator gene homolog MSH2 and its association with hereditary nonpolyposis colon cancer.* cell. 1993; **75** (5): 1027-38.
219. Leach FS, Nicolaides NC, Papadopoulos N, Liu B, Jen J, Parsons R, Peltomäki P, Sistonen P, Aaltonen LA, Nyström-Lahti M. *Mutations of a mutS homolog in hereditary nonpolyposis colorectal cancer.* Cell. 1993; **75** (6): 1215-25.
220. Bronner C, Baker S, Morrison P, Warren G, Smith L, Lescoe M, Kane M, Earabino C, Lipford J, Lindblom A. *Mutation in the DNA mismatch repair gene homologue hMLH1 is.* Nature. 1994; **368** (6468): 258-61.
221. Papadopoulos N, Nicolaides NC, Wei Y-F, Ruben SM, Carter KC, Rosen CA, Haseltine WA, Fleischmann RD, Fraser CM, Adams MD. *Mutation of a mutL homolog in hereditary colon cancer.* Science. 1994; **263** (5153): 1625-9.
222. Wu Y, Berends MJ, Mensink RG, Kempinga C, Sijmons RH, van der Zee AG, Hollema H, Kleibeuker JH, Buys CH, Hofstra RM. *Association of Hereditary*

Nonpolyposis Colorectal Cancer–Related Tumors Displaying Low Microsatellite Instability with MSH6 Germline Mutations. The American Journal of Human Genetics. 1999; **65** (5): 1291-8.
223. Ripperger T, Beger C, Rahner N, Sykora KW, Bockmeyer CL, Lehmann U, Kreipe HH, Schlegelberger B. *Constitutional mismatch repair deficiency and childhood leukemia/lymphoma–report on a novel biallelic MSH6 mutation.* Haematologica. 2010; **95** (5): 841-4.
224. Couronné L, Ruminy P, Waultier-Rascalou A, Rainville V, Cornic M, Picquenot J-M, Figeac M, Bastard C, Tilly H, Jardin F. *Mutation mismatch repair gene deletions in diffuse large B-cell lymphoma.* Leukemia & lymphoma. 2013; **54** (5): 1079-86.
225. Cleaver JE, Lam ET, Revet I. *Disorders of nucleotide excision repair: the genetic and molecular basis of heterogeneity.* Nature Reviews Genetics. 2009; **10** (11): 756-68.
226. Roy R, Chun J, Powell SN. *BRCA1 and BRCA2: different roles in a common pathway of genome protection.* Nat Rev Cancer. 2011; **12** (1): 68-78.
227. Chrzanowska KH, Gregorek H, Dembowska-Bagińska B, Kalina MA, Digweed M. *Nijmegen breakage syndrome (NBS).* Orphanet J Rare Dis. 2012; **7** (1): 13.
228. Jones S, Emmerson P, Maynard J, Best JM, Jordan S, Williams GT, Sampson JR, Cheadle JP. *Biallelic germline mutations in MYH predispose to multiple colorectal adenoma and somatic G: C→ T: A mutations.* Human Molecular Genetics. 2002; **11** (23): 2961-7.
229. Baglioni S, Melean G, Gensini F, Santucci M, Scatizzi M, Papi L, Genuardi M. *A kindred with MYH - associated polyposis and pilomatricomas.* American Journal of Medical Genetics Part A. 2005; **134** (2): 212-4.
230. Povey AC, Margison GP, Santibanez-Koref MF. *Lung cancer risk and variation in MGMT activity and sequence.* DNA Repair. 2007; **6** (8): 1134-44.
231. Saviozzi S, Ceppi P, Novello S, Ghio P, Iacono ML, Borasio P, Cambieri A, Volante M, Papotti M, Calogero RA. *Non–Small Cell Lung Cancer Exhibits Transcript Overexpression of Genes Associated with Homologous Recombination and DNA Replication Pathways.* Cancer research. 2009; **69** (8): 3390-6.
232. Johannessen T-CA, Prestegarden L, Grudic A, Hegi ME, Tysnes BB, Bjerkvig R. *The DNA repair protein ALKBH2 mediates temozolomide resistance in human glioblastoma cells.* Neuro-oncology. 2013; **15** (3): 269-78.
233. Kane MF, Loda M, Gaida GM, Lipman J, Mishra R, Goldman H, Jessup JM, Kolodner R. *Methylation of the hMLH1 promoter correlates with lack of expression of hMLH1 in sporadic colon tumors and mismatch repair-defective human tumor cell lines.* Cancer Research. 1997; **57** (5): 808-11.
234. Mueller J, Gazzoli I, Bandipalliam P, Garber JE, Syngal S, Kolodner RD. *Comprehensive molecular analysis of mismatch repair gene defects in suspected Lynch syndrome (hereditary nonpolyposis colorectal cancer) cases.* Cancer Research. 2009; **69** (17): 7053-61.
235. Drost M, van Dijk L, Morreau H, Tops CM, Vasen HF, Wijnen JT, de Wind N. *A cell-free assay for the functional analysis of variants of the mismatch repair protein MLH1.* Human mutation. 2010; **31** (3): 247-53.

236. Drost M, Zonneveld J, van Hees S, Rasmussen LJ, Hofstra RM, de Wind N. *A rapid and cell-free assay to test the activity of lynch syndrome-associated MSH2 and MSH6 missense variants.* Human mutation. 2012; **33** (3): 488-94.
237. Cheng L, Sturgis EM, Eicher SA, Spitz MR, Wei Q. *Expression of nucleotide excision repair genes and the risk for squamous cell carcinoma of the head and neck.* Cancer. 2002; **94** (2): 393-7.
238. Vodicka P, Stetina R, Polakova V, Tulupova E, Naccarati A, Vodickova L, Kumar R, Hanova M, Pardini B, Slyskova J. *Association of DNA repair polymorphisms with DNA repair functional outcomes in healthy human subjects.* Carcinogenesis. 2006; **28** (3): 657-64.
239. Vogel U, Dybdahl M, Frentz G, Nexo BA. *DNA repair capacity: inconsistency between effect of over-expression of five NER genes and the correlation to mRNA levels in primary lymphocytes.* Mutation Research/DNA Repair. 2000; **461** (3): 197-210.
240. Cobo M, Isla D, Massuti B, Montes A, Sanchez JM, Provencio M, Viñolas N, Paz-Ares L, Lopez-Vivanco G, Muñoz MA, Felip E, Alberola V, Camps C, Domine M, Sanchez JJ, Sanchez-Ronco M, Danenberg K, Taron M, Gandara D, Rosell R. *Customizing cisplatin based on quantitative excision repair cross-complementing 1 mRNA expression: a phase III trial in non-small-cell lung cancer.* Journal of Clinical Oncology. 2007; **25** (19): 2747-54.
241. Wei Q, Wang L-E, Sturgis EM, Mao L. *Expression of nucleotide excision repair proteins in lymphocytes as a marker of susceptibility to squamous cell carcinomas of the head and neck.* Cancer Epidemiology Biomarkers & Prevention. 2005; **14** (8): 1961-6.
242. Allione A, Russo A, Ricceri F, Loock KV, Guarrera S, Voglino F, Kirsch-Volders M, Matullo G. *Validation of the nucleotide excision repair comet assay on cryopreserved PBMCs to measure inter-individual variation in DNA repair capacity.* Mutagenesis. 2013; **28** (1): 65-70.
243. Qiao YW, Spitz MR, Guo ZZ, Hadeyati M, Grossman L, Kraemer KH, Wei QY. *Rapid assessment of repair of ultraviolet DNA damage with a modified host-cell reactivation assay using a luciferase reporter gene and correlation with polymorphisms of DNA repair genes in normal human lymphocytes.* Mutat Res-Fundam Mol Mech Mutagen. 2002; **509** (1-2): 165-74.
244. Thoms KM, Baesecke J, Emmert B, Hermann J, Roedling T, Laspe P, Leibeling D, Truemper L, Emmert S. *Functional DNA repair system analysis in haematopoietic progenitor cells using host cell reactivation.* Scand J Clin Lab Invest. 2007; **67** (6): 580-8.
245. Marsit CJ, Liu M, Nelson HH, Posner M, Suzuki M, Kelsey KT. *Inactivation of the Fanconi anemia/BRCA pathway in lung and oral cancers: implications for treatment and survival.* Oncogene. 2003; **23** (4): 1000-4.
246. Szaumkessel M, Richter J, Giefing M, Jarmuz M, Kiwerska K, Tönnies H, Grenman R, Heidemann S, Szyfter K, Siebert R. *Pyrosequencing-based DNA methylation profiling of Fanconi anemia/BRCA pathway genes in laryngeal squamous cell carcinoma.* Int J Oncol. 2011; **39** (2): 505-14.
247. Starcevic D, Dalal S, Sweasy JB. *Is there a link between DNA polymerase beta and cancer?* Cell Cycle. 2004; **3** (8): 996-9.

248. Pons B, Belmont A-S, Masson-Genteuil G, Chapuis V, Oddos T, Sauvaigo S. *Age-associated modifications of Base Excision Repair activities in human skin fibroblast extracts*. Mechanisms of ageing and development. 2010; **131** (11): 661-5.
249. Hsu C-F, Tseng H-C, Chiu C-F, Liang S-Y, Tsai C-W, Tsai M-H, Bau D-T. *Association between DNA double strand break gene Ku80 polymorphisms and oral cancer susceptibility*. Oral oncology. 2009; **45** (9): 789-93.
250. Berwick M, Satagopan JM, Ben-Porat L, Carlson A, Mah K, Henry R, Diotti R, Milton K, Pujara K, Landers T. *Genetic heterogeneity among Fanconi anemia heterozygotes and risk of cancer*. Cancer research. 2007; **67** (19): 9591-6.

Chapter II: Multiplexed DNA repair assays for multiple lesions and multiple doses via transcription inhibition and transcriptional mutagenesis.

Zachary D. Nagel, Carrie M. Thompson, Isaac. A. Chaim, Siobhan K. McRee, Patrizia Mazzucato, Anwaar, Ahmad, Ryan. P. Abo, Vincent. L. Butty, Anthony. L. Forget, Leona. D. Samson.

Published as:

Nagel, Z.D., C.M. Margulies, I.A. Chaim, S.K. McRee, P. Mazzucato, A. Ahmad, R.P. Abo, V.L. Butty, A.L. Forget, and L.D. Samson, *Multiplexed DNA repair assays for multiple lesions and multiple doses via transcription inhibition and transcriptional mutagenesis.* Proceedings of the National Academy of Sciences, 2014. **111**(18): p. E1823-E1832.

Table of Contents

Multiplexed DNA repair assays for multiple lesions and multiple doses via transcription inhibition and transcriptional mutagenesis.

Z. D. Nagel[1,2], C. M. Thompson[1,2], I. A. Chaim[1,2], S. K. McRee[1,2], P. Mazzucato[1,2], A. Ahmad[1,2], R. P. Abo[1,2,3,4], V. L. Butty[1,2,3,4], A. L. Forget[1,2], L. D. Samson[1,2,3,4,*]

[1]Department of Biological Engineering, [2]Center for Environmental Health Sciences, [3]Department of Biology, [4]The David H. Koch Institute for Integrative Cancer Research, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

*To whom correspondence should be addressed: Address: 77 Massachusetts Avenue, MIT Building 56-235, Cambridge, MA 02139, USA; Phone: 617-253-6220; email: lsamson@mit.edu

High-Throughput DNA Repair Assays

Key Words: Host cell reactivation; DNA repair capacity, flow cytometry, high throughput sequencing, transcriptional mutagenesis, multiplex assays

Abstract

The capacity to repair different types of DNA damage varies among individuals making them more or less susceptible to the detrimental health consequences of such exposures. Current methods for measuring DNA repair capacity (DRC) are relatively labor intensive, often indirect and usually limited to a single repair pathway. Here we describe a fluorescence-based multiplex flow-cytometric host cell reactivation assay (FM-HCR) that measures the ability of human cells to repair plasmid reporters each bearing a different type of DNA damage or different doses of the same type of DNA damage. FM-HCR simultaneously measures repair capacity in any four of the following pathways, NER, MMR, BER, NHEJ, HR and MGMT. We show that FM-HCR can measure interindividual DRC differences a panel of 24 cell lines derived from genetically diverse apparently healthy individuals, and we show that FM-HCR can be used to identify inhibitors or enhancers of DRC. We further develop a next generation sequencing-based HCR assay (HCR-Seq) that detects rare transcriptional mutagenesis events due to lesion bypass by RNA polymerase, providing an added dimension to DRC measurements. FM-HCR and HCR-Seq provide powerful tools for exploring relationships among global DRC, disease susceptibility, and optimal treatment.

Significance Statement

DNA, the blueprint of the cell, is constantly damaged by chemicals and radiation. Because DNA damage can cause cell death or mutations that can lead to diseases like cancer, cells are armed with an arsenal of several distinct mechanisms for repairing the many types of DNA damage that occur. DNA repair capacity (DRC) varies among individuals, and reduced DRC is associated with disease risk, however the available DRC assays are labor intensive and only measure one pathway at a time. Herein, we present powerful new assays that measure DRC in multiple pathways in a single assay. We use the assays to

measure inter-individual DRC differences, inhibition of DNA repair, and to uncover unexpected error-prone transcriptional bypass of a thymine dimer.

Introduction

DNA is under constant assault from damaging agents that produce a vast array of lesions. Left unrepaired, these lesions have the potential to alter cellular function and compromise the health of the organism, leading to degenerative diseases, cancer, and premature aging [1-5]. Consequently, inter-individual variations in DNA repair capacity (DRC) are thought to contribute to the fact that people have different susceptibilities to these diseases [6-10]. Furthermore, the efficacy of cancer chemotherapy with DNA damaging agents is dependent on the DRC of the targeted cells [11]. Thus, DRC measurements could potentially be used to personalize both treatment and prevention of disease.

We define DRC as the basal ability of cells to eliminate DNA damage from the genome, however it should be noted that some DRC assays, such as mutagen sensitivity assays, may also reflect changes in gene expression and the activation of non-DNA repair pathways upon treatment of cells with DNA damaging agents. A wide variety of methods are used to estimate DRC. Many studies focus on indirect assessments of DRC through transcriptional profiling, proteomics, and single nucleotide polymorphism (SNP) screens [12-16]. However, SNPs in DNA repair genes are not informative if the relevant gene is not expressed. Likewise, gene expression data are not informative for cases in which the gene product is inactive or cannot be assembled into a functional complex. More direct measurements of DRC *in vitro* using cell lysates overcome some of this complexity by integrating these factors into a single readout [17-20]; however these methods require separate assays for measurements in more than one repair pathway, and may not be representative of DRC in intact cells. Measuring the consequences of DNA repair in intact cells by monitoring sister chromatid exchanges, chromosome aberrations, or DNA strand breaks by comet assays also integrate complexity into a single readout, but require labor-intensive

analyses that make them refractory to implementation in a clinical setting [21-23]. Although recently developed high throughput comet assays provide an excellent alternative, they are limited to DNA damage that leads to, or can be converted to DNA strand breaks [24].

Host cell reactivation (HCR) assays report the ability of cells to repair DNA damage that blocks transcription of a transiently transfected reporter gene [8, 9, 25-28]. Repair of the transcription-blocking lesion reactivates reporter gene expression, thus providing a quantitative readout for DRC. However, HCR assays cannot generally report repair of DNA lesions that do not block the progression of the RNA polymerase, and current methods for measuring DRC are further limited by the need for separate experiments to measure repair capacity in more than one pathway, or at more than one dose of DNA damage. Herein we introduce a new high-throughput, fluorescence-based multiplex HCR assay (FM-HCR) for measuring DRC in living cells that overcomes these limitations. We first present a multicolor fluorescence assay that simultaneously measures DNA repair at multiple doses of DNA damage. We then demonstrate simultaneous DRC measurements for up to four repair pathways in human and rodent cells, using reporters for the repair of both transcription-blocking lesions and lesions that are bypassed by RNA polymerase. To demonstrate potential applications of FM-HCR to population studies, we measure interindividual DRC differences in five pathways in a panel of 24 lymphoblastoid cell lines derived from apparently healthy individuals. We also show that FM-HCR can be used to identify agents that inhibit or enhance DRC. Finally, to further increase throughput and to establish a single generalized detection method for the repair of any lesion that alters transcription of a reporter gene, we added to the HCR reporter protein paradigm a reporter transcript assay that leverages the extraordinary power of next generation sequencing (HCR-Seq). We use HCR-Seq to measure 20 independent reporter signals in a single assay, and detect error-prone transcriptional bypass at a bulky DNA lesion in human cells. FM-HCR and HCR-Seq provide rapid, high throughput methods of assessing DRC in multiple pathways and represent a major improvement over standard methods

currently used in basic, clinical and epidemiological research addressing the relationship among DNA damage, DNA repair and disease susceptibility.

Results

Validation of FM-HCR

FM-HCR was used to assay DRC in 55 cell lines (**Table 2.1**). Expression levels of five fluorescent reporter proteins were quantitated simultaneously using flow cytometry (**Figure S2.1**). Use of 96-well electroporation plates reduced the time required for transfection to less than an hour per plate, and a BD High Throughput Sampler permitted data acquisition in less than 10 minutes active time.

In vitro treatment of plasmids with UVC light resulted in a dose-dependent reduction in reporter expression. When each of the five fluorescent reporter plasmids was treated with a unique dose of UVC (Plasmid combination #1 in **Table 2.2**), and subsequently co-transfected into cells, a dose-response curve was generated from a single experiment that required only two transfections (**Figure 2.1a**). Dose-response curves spanning up to 3 decades of percent reporter expression (%R.E.) were obtained for 7 lymphoblastoid cell lines (**Figure 2.1b,c**), chosen because they were previously characterized over 20 years ago for their capacity to repair UV-irradiated plasmid DNA by another much more laborious method [8]. Differences in DRC were most pronounced at the highest dose to plasmid (800 J/m^2), with % reporter expression (%R.E.) values varying over a range of about 100-fold among the cell lines. As expected, the highest DRC was observed for lymphoblastoid cell lines derived from apparently healthy individuals, referred to as wild type (WT) (**Table 2.1**). Moderately reduced DRC was observed for two XPC cell lines, and severe defects were evident for XPA and XPD cell lines. Between 18 and 40 hours, %R.E. increased for most cell lines (**Figure 2.1b,c**), consistent with time-dependent repair of transcription blocking lesions.

The FM-HCR data presented in **Figure 2.1c** reproduce those from the previous study that also monitored transcription inhibition on UV-damaged plasmids 40 hours after transfection [8]. In that study, chloramphenicol acetyl transferase (CAT) levels in cell-free extracts were used as the reporter. Two complementary methods were used to compare our data to the historical data. First, the percent CAT expression (%CAT) reported at a single dose of UV irradiation (300 J/m^2 in ref. [8]) was highly correlated ($R^2 = 0.92$, $p = 0.0006$) with %R.E. at a single dose (400 J/m^2) in the present study (**Figure 2.1d)**. The relative repair capacity of multiple cell lines can also be compared by calculating the parameter D_o, corresponding to the dose at which HCR falls below 37% R.E. [29]. The D_o values calculated from our experimental data were also highly correlated with the historical D_o values ($R^2 = 0.92$, $p < 0.0001$) (**Figure 2.1e)**.

To confirm that the dose response curves in **Figs. 1b** and **1c** could be obtained independently of the choice of fluorescent reporters, the experiment was repeated with the plasmids shuffled with regard to which plasmid received a particular UV dose (Plasmid combination #2 in **Table 2.2**). The resulting dose response curves obtained at 18 and 40 hours are presented in **Figs. 1f** and **1g**, respectively. Once again, the FM-HCR data collected at 40 hours reproduce the historical data [8] (**Figure 2.1h**). Likewise, the FM-HCR data collected with plasmid combination #2 reproduce those obtained using the plasmid combination #1 (**Figure 2.1i**).

FM-HCR assays were also carried out on 7 primary untransformed skin fibroblast cell lines and compared to Epstein-Barr virus transformed lymphoblastoid cell lines derived from the same individuals (represented as individuals i-vii in **Table 2.1**); cells were from 4 apparently healthy individuals and 3 XP patients. A similar pattern of dose response curves was obtained for both fibroblasts and lymphoblastoid cells (**Figs. 1j** and **1k, respectively**). Overall, absolute NER capacity measured in fibroblasts appeared to be somewhat higher than that in the lymphoblastoid cell lines; however, a comparison of DRC measured at 800 J/m^2 indicated that NER phenotype is strongly correlated ($R^2 = 0.94$, $p = 0.0003$) between the two cell types (**Figure 2.1l**).

Development of FM-HCR assays for DNA mismatch repair and direct reversal of O^6-MeG

Fluorescent plasmid reporters for direct reversal of O^6-MeG and mismatch repair capacity were developed (**Figure S2.2**). For mismatch repair assays, a G:G mismatch containing plasmid was constructed; it has previously been shown that G:G mismatches are repaired inefficiently in extracts from MMR deficient cells [30], and we confirmed this observation in MMR-deficient HCT116 cells and MMR-proficient HCT116+3 cells that have been complemented with human chromosome 3 (**Figure S2.2d**), as well as MMR-deficient MT1 cells that lack MSH6, and MMR-proficient TK6 cells (**Figure 2.2**). This reporter expresses a non-fluorescent protein unless a repair event restores a cytosine in the transcribed strand, leading to wild type orange fluorescent protein. Because the plasmid lacks a strand discrimination signal, the theoretical upper limit of reporter expression (relative to a similarly constructed wild type homoduplex control) is 50%. For direct reversal of O^6-MeG, a plasmid that encodes a non-fluorescent protein in the absence of the lesion was prepared. Introduction of a site-specific O^6-MeG lesion into the transcribed strand causes transcription errors [31], producing transcripts encoding the wild type mPlum fluorescent protein. Thus, cells deficient for methylguanine methyltransferase (MGMT) mediated O^6-MeG repair express relatively high levels of the mPlum reporter, whereas cells expressing MGMT remove the source of transcription errors, thus reducing reporter expression.

Simultaneous measurement of DRC in 3 pathways with FM-HCR

DRC for three pathways, namely NER, mismatch repair (MMR), and the direct reversal of O^6-MeG (MGMT) was measured in five lymphoblastoid cell lines (**Figure 2.2**). DRC for each pathway was first measured in separate experiments. The severe NER defect for the XPA-deficient GM02344 cell line was reproduced, whereas relatively high NER capacity was confirmed for the

four other cell lines with no known NER defect (GM01953, MT1, TK6, and TK6+MGMT). The lymphoblastoid cell line MT1, known to be deficient for mismatch repair [32], expressed the MMR reporter at a level up to 10-fold lower than that of the other four cell lines that have no known MMR defects. As expected, the O^6-MeG direct reversal reporter was highly expressed in MT1 and TK6, owing to the absence of MGMT. Expression of the same reporter was reduced nearly 1000-fold in TK6+MGMT cells expressing a high level of MGMT, and was reduced ~ 8-fold and 250-fold in GM01953 and GM02344, respectively, both of which express active MGMT, but at different levels [33].

One of our goals is to increase the throughput of DRC assays by measuring the activity of multiple DNA repair pathways in a single assay. To test whether our DRC reporters could be combined in a single experiment without affecting the accuracy of the assay, we co-transfected three reporter plasmids, each targeting a different pathway, along with an internal transfection control (plasmid combination #3 in **Table 2.2**). This yielded nearly identical DRC profiles for NER, MMR and MGMT as those obtained when the reporters were transfected in separate experiments (**Figure 2.2b)**.

Simultaneous measurement of four DNA repair pathways including BER and NHEJ.

We next sought to measure DRC in four pathways including BER or DSB repair (**Figure 2.3**). To add a fourth pathway to the FM-HCR in **Figure 2.2**, a BFP reporter for repair of a double strand break by the NHEJ pathway was developed (**Figure 2.3a and Figure S2.3**). This assay was validated using the M059J and M059K cell lines, which were derived from a single glioblastoma [34]. The M059J cell line is deficient for DNA PKcs that is required for NHEJ [35]. As expected, M059J cells expressed ~40-fold lower levels of the NHEJ reporter relative to the wild type M059K cells when the reporter was transfected separately from other reporters (**Figure 2.3b**). To test whether DRC could be measured in four pathways simultaneously, we co-transfected the NHEJ reporter

with the reporters described above for NER, MMR, and MGMT (plasmid combination #4 in **Table 2.2**). As with the three-pathway measurements described above, the co-transfected reporters yielded nearly identical DNA repair profiles as the separately transfected reporters (**Figs. 3b** and **3c**). MMR and MGMT capacity was similar in the two cell lines, however NER capacity was reduced in M059J by approximately 7-fold relative to M059K. It has been observed previously that the XPC and ERCC2 genes are overexpressed in M059J vs M059K cells, and that the M059J cells are slightly more sensitive than M059K cells to UV irradiation [36]. Inefficient NER in the presence of excess XPC protein has also been noted in vitro [37].

To further demonstrate the versatility of FM-HCR, we performed an assay that includes BER, NER, MMR and MGMT (**Figure 2.3d**). An mOrange fluorescent reporter for base excision repair of 8-oxoguanine (8-oxoG) was developed that produces wild type mOrange transcripts when RNA polymerase incorporates adenine opposite a site-specific 8-oxoG lesion. Deficient 8-oxoG repair is expected to result in a higher reporter expression. In keeping with this expectation, mouse embryonic fibroblasts (MEFs) deficient for 8-oxoG DNA glycosylase (Ogg1) expressed an approximately 20-fold higher level of mOrange than wild type MEFs when the reporter was transfected separately from the other reporters (**Figure 2.3e**). When the 8-oxoG reporter was co-transfected with three other reporters for NER, MMR and MGMT (plasmid combination #5 in **Table 2.2**), we once again observed DRC profiles that were nearly identical to those obtained when the reporters were transfected separately (**Figs. 3e** and **3f**). MMR and NER capacity were similar in the two cell lines, but MGMT reporter expression was approximately 5-fold higher in the wild type MEFs. Since the wild type MEFS were derived from C57BL/6J mice, the differences in MGMT repair capacity could be due to the mixed (C57BL/6J and 129SV) background of the $Ogg1^{-/-}$ mouse from which the MEFs were derived [38].

Analysis of DRC for five pathways in a panel of 24 cell lines derived from apparently healthy individuals

A previously described assay for HR was incorporated into FM-HCR experiments and was validated using cell lines with known defects in double strand break repair (**Figure S2.4**) [39]. Assays for NER, MMR, MGMT, HR and NHEJ capacity (see plasmid combinations #3 and #6) were then carried out on a panel of 3 control cell lines (TK6, MT1, and TK6+MGMT) and 24 human lymphoblastoid cell lines derived from apparently healthy individuals of diverse ancestry [40]. Each of the 24 cell lines exhibited a unique DNA repair profile (**Figure 4a**), and a range of DRC was observed across the 24 cell lines for each pathway (MGMT, 285-fold; MMR, 4.4-fold; HR, 3.7-fold; NER, 3.2-fold, and NHEJ, 2.1-fold). To further validate the FM-HCR assay for MGMT, transcript levels measured by TaqMan qPCR for the cell lines in **Figure 2.4a** were compared with fluorescent reporter expression. A non-linear relationship was observed between MGMT FM-HCR % Reporter Expression and transcript levels (not shown), however log-transformed FM-HCR data correlated extremely well with MGMT transcript levels (**Figure 2.4b**).

Application of FM-HCR to assays for DRC inhibition

To further demonstrate the potential applications of FM-HCR, the assay was used to detect inhibition of DRC by metals and a small molecule. Cadmium and arsenic have been shown previously to inhibit NER [41, 42]. FM-HCR confirmed NER inhibition in the presence of low micromolar concentrations of the two metals (**Figure 2.4c**), whereas no effect on NER was detected in the presence of Compound 401, known to inhibit a critical NHEJ factor, namely the DNA dependent protein kinase catalytic subunit (DNA PK_{cs}) [43]. Moreover, Compound 401 was shown to inhibit NHEJ in a dose dependent manner (**Figure 2.4d**), while MMR, NER, and HR were unaffected.

Taken together, the FM-HCR data demonstrate a versatile method for measuring in a single assay the repair of multiple DNA damage substrates, with either different doses or with different types of damage. While measuring four DNA repair pathways simultaneously represents a significant advance, the degree of multiplexing possible for the FM-HCR assay is dependent upon the number of fluorescent reporters that can be measured simultaneously. We therefore developed an additional assay that does not require the detection of fluorescent proteins.

Deep sequencing analysis of cells transfected with reporter plasmids

To increase the potential number of reporters that can be detected in a single assay we developed HCR-seq, a method of distinguishing and quantifying multiple full-length reporter transcripts using next generation sequencing. Two cell lines exhibiting a large difference in their NER capacity (GM02344 and GM01953) were selected for a direct comparison of DRC measured by HCR-seq versus FM-HCR (**Figure 2.5**). Each cell line was transfected with plasmid combination #7 (**Table 2.2)**, and at 18 hours cells were analyzed by both HCR-seq and FM-HCR. Plasmid combination #7 included a modified GFP reporter containing a single site-specific cyclobutane pyrimidine dimer (CPD). This reporter was included to allow a focused analysis of possible transcriptional errors induced by a bulky DNA lesion.

For HCR-seq analysis, total RNA was isolated and subjected to standard Illumina mRNAseq sample preparation and analysis [44]. A total of 358,281,302 reads were generated for two replicates of 4 multiplexed samples, with each replicate analyzed in a separate HiSeq lane (**Table S2.1**). Between 30 and 50 million reads were assigned to each original sample. 315,574,792 reads (88%) mapped properly to genes annotated for the human genome plus the five reporter sequences. In each sequencing lane, all 5 reporter transcripts were detected for each of the 4 samples; each sequencing lane simultaneously measured expression levels for 20 reporters. Alignment statistics, the criteria

used to define proper alignment and reasons for excluding the remaining 12% of reads from subsequent analysis are detailed in **Tables S2-S5**.

DNA repair and transcriptional mutagenesis detected by RNA sequencing

Relative transcript levels in WT (GM01953) versus XPA (GM02344) cell lines were determined for both host genes and plasmid reporter genes. Plasmid reporters were found to be among the most highly expressed genes (**Figure S2.5a**). As expected, reporter expression from UV-treated plasmids was reduced in a dose-dependent manner (**Figure 2.6a**), and the reduced expression for the XPA cell line (GM02344) was far greater than that for WT cells (GM01953). More importantly, reporter transcript expression mirrored closely the dose response curves obtained from the same transfected cells using FM-HCR (**Figure 2.6b**). Reporter expression from plasmids containing a single site-specific CPD in the transcribed strand was likewise reduced relative to that from undamaged plasmids (**Figure S2.5b**).

With respect to global gene expression in the transfected cells, fewer than 10 host transcripts showed a > 2-fold change in expression in the presence of UV treated plasmids versus undamaged plasmids (**Table S2.5**). Among these, only three (*SMN1*, *RPL21*, and *RN5-8S1*) were observed in both cell lines, but in no case was a change in the same direction observed for both replicates. Thus, no significant transcriptional response to the presence of DNA damage in plasmids was evident under our experimental conditions. However, consistent with the FM-HCR data (**Figure 2.2)** that suggested higher MGMT activity in GM02344 cells compared with GM01953 cells, the mRNAseq data indicated an approximately 3-fold higher expression of the *MGMT* transcript in GM02344 versus GM01953. Furthermore, XPA transcripts were expressed at lower levels in GM02344 versus GM01953, and they were only rarely spliced correctly in GM02344 (**Figure S2.5c**). These data reproduce previously reported splicing errors in GM02344 due to a homozygous 555G>C mutation in the XPA gene [45]. Finally, to assess the potential for DNA contamination in RNAseq samples,

the density of reads aligning to intergenic regions (which are not expected to be represented in transcripts) was compared to the density of reads aligning to exons, and the ratio of exonic/intergenic reads was found to be greater than 1000, indicating an RNA purity >99.9%.

Sequence-level analysis of mRNAseq data revealed base substitutions in reporter transcripts at the position corresponding to the site-specific CPD; this was true for both cell lines (**Figure 2.6c**). The most frequently observed base change, an A→G mutation at the 5' Adenine in the ApA sequence opposite the CPD (hereafter AA→GA), was detected at a frequency of 1.3% in cells with no known repair defect (GM01953), and 5.8% in NER-deficient GM02344 cells. In transcripts expressed from the undamaged plasmid, the frequency of the AA→GA mutations at this position was less than 0.2%. A potential experimental concern is that trace contaminating plasmid DNA might be amplified during Illumina sample preparation, thus giving rise to DNA fragments with base substitutions due to error-prone CPD bypass by DNA polymerase. However, nearly identical frequencies for AA→GA mutations were found using a second sample preparation method that excludes the possibility of contaminating lesion-containing plasmid (**Figure S2.5d**). It therefore appears that human RNA polymerase can bypass a thymine dimer *in vivo*, albeit in an error prone manner.

AA→GA mutations were also induced in a dose-dependent manner in transcripts expressed from UV-irradiated reporter plasmids containing thymine dimers that were not site-specific (**Figure 2.6d**). As expected for randomly induced DNA damage, the absolute frequency of the base substitution was much lower than that observed for transcripts expressed from the reporter with the site-specific thymine dimer. Once again, base changes occurred at a higher frequency in NER-deficient versus wild type cells. These data provide additional evidence for error prone transcriptional bypass of thymine dimers.

Discussion

We have established new methods that enable rapid, high throughput measurements of DRC for DNA lesions that either block RNA polymerase II-mediated transcription, or alter the sequence of reporter transcripts. These methods do not require the generation of cell lysates or the use of *in vitro* assays, and can simultaneously measure in *vivo* DRC at multiple doses or in multiple repair pathways. As a result, these assays outperform current methods of measuring DRC [20, 27], and have the potential to be used to personalize the prevention and treatment of cancer and other diseases caused by inefficient repair of DNA damage.

We have demonstrated an application of the FM-HCR to the question of whether NER capacity in human lymphoblastoid cells is representative of repair capacity in other tissues. Lymphoblastoid cells provide a convenient source of cells for use in human variability studies, however the extent to which they represent a faithful surrogate for other cells in primary tissues has been called into question [46-48]. The present data indicate a strong correlation between NER capacity in primary human skin fibroblasts and transformed B-lymphoblastoid cells from the same individuals (**Figure 2.1I**). The strong correlation further illustrates that the assay can be carried out reproducibly in primary or transformed cells from multiple tissues.

To our knowledge, our use of HCR to simultaneously measure combinations of NER, MMR, BER, NHEJ, HR or MGMT capacity is the first example of a quantitative assay capable of measuring repair of DNA damage by multiple distinct pathways in parallel. One of the strengths of FM-HCR is that it yields a single readout (fluorescence) in place of multiple unique outputs from very different experimental procedures that have been used previously to characterize the same repair pathways in the cell lines for which data are presented (**Figure 2.2** and **Figure 2.3**) [8, 32, 33, 35, 38, 49]. The fluorescent reporters for direct reversal of O^6-MeG and BER of 8-oxoG illustrate the use of transcriptional mutagenesis to measure the repair of DNA lesions that are

bypassed in an error-prone manner by RNA polymerase. This paradigm, where the presence of a DNA lesion changes the expressed reporter sequence to one that encodes a functional protein, holds promise as a general method of measuring DRC, because a wide variety of toxic and mutagenic DNA lesions are known to induce transcriptional errors [50, 51].

We have presented several applications of FM-HCR to demonstrate the broad utility of the assay. The flow cytometric fluorescence-based FM-HCR method accurately reproduces data collected previously for a set of cell lines known to differ in NER capacity [8]. A screen of a much larger set of cell lines derived from apparently healthy individuals illustrates the potential to efficiently measure DRC in multiple pathways in large sets of samples (**Figure 2.4a**), and to identify agents that inhibit or enhance DRC (**Figure 4c** and **4d**). Screens for DRC inhibitors or enhancers are expected to identify some agents for which the mechanism of action is unknown; indeed, uncertainty remains as to the precise mechanisms by which arsenic and cadmium exposure lead to reduced DRC [41, 52]; the strength of FM-HCR lies in the ability to measure changes in DRC as an important functional endpoint.

By using multiple fluorescent reporters, a 96-well format, and automated flow cytometric sample processing, the method is rapid and less labor intensive than the standard CAT-based HCR assay. For example, the total active laboratory time required to perform the analysis to generate the triplicate data in **Figure 2.1b** and **1c** is approximately 12 hours, or 1-2 hours per cell line, using flow cytometers equipped with a high throughput sampler to enable automated data acquisition. In addition, experimental error is reduced by co-transfection of reporters, allowing normalization of expression from a damaged plasmid to that of an undamaged control plasmid included in every transfection. Through these technical improvements, FM-HCR removes a major barrier to epidemiological studies of DRC that include large populations and multiple DNA repair pathways. Furthermore, because standard oncology labs are equipped with flow cytometers, the assay also has the potential to be of use in a clinical setting.

The use of next generation sequencing to essentially count reporter transcripts (HCR-seq), rather than measuring their fluorescent translation products, presents an opportunity to vastly increase throughput, and overcomes important limitations on assay throughput and versatility that are otherwise imposed by the need to detect fluorescent reporter proteins. We have validated the HCR-seq approach by showing that HCR of UV-irradiated plasmids analyzed by mRNAseq yields a pattern of dose-response curves similar to those obtained previously using a CAT-based HCR assay [8], as well as those obtained in the current study by FM-HCR analysis (**Figure 2.1**). Because next generation sequencing can be used to quantitate the expression levels of thousands of transcripts simultaneously, our assay has the potential to measure expression of dozens of reporters for multiple individuals in a single experiment; this would make characterization of global DRC in large populations both efficient and affordable (See **supplementary note**).

HCR-seq constitutes a paradigm shift in the quantitation of DRC because of the ability to measure the repair of any lesion that either inhibits transcription or induces transcriptional mutagenesis. Base misincorporation opposite DNA lesions that are bypassed by DNA polymerase during replication has been extensively studied for many lesions [53]. Misincorporation during transcription by RNA polymerase has been documented for a growing number lesions, and often mirrors that of DNA polymerase during replication [54]. As a result, most mutagenic lesions can be expected to have a transcriptional mutagenic signature. The HCR-seq strategy should therefore be useful in DRC measurements for nearly any pathway. The data in **Figure 2.6** also illustrate the power of this unbiased approach to detect rare events that are specific to transcription of damaged DNA.

The two major applications to human health that we foresee for these assays relate to personalized prevention and treatment of cancer. The available published data indicate that DRC is an important factor both in cancer susceptibility and in the efficacy of cancer treatment with DNA damaging agents, and that plasmid-based HCR assays can readily be applied to primary human

tissue samples, including stimulated peripheral blood mononuclear cells [8-11, 20, 27]. FM-HCR and HCR-Seq now open the door to a comprehensive analysis of DRC as a biomarker for disease susceptibility. For personalized disease prevention, FM-HCR could be applied to human blood cells to identify individuals who may have a higher risk of disease. In terms of personalized treatment, the assays could be used to measure DRC in blood cells to predict patient tolerance for a particular cancer therapy [55], or to measure DRC in cancer cells to predict the efficacy of treatment with DNA damaging chemotherapeutic agents in a manner analogous to using MGMT promoter methylation to predict the response of cancers to alkylating chemotherapy agents such as temozolomide [56]. Indeed, the data in **Figure 2.4** show that FM-HCR data reproduce the results of a standard TaqMan qPCR assay for MGMT gene expression in lymphoblastoid cell lines. The functional FM-HCR and HCR-Seq assays might be expected to outperform promoter methylation assays because (i) they provide a direct, quantitative readout of repair activity rather than an indirect estimate of DNA repair gene expression, and (ii) they provide data for repair capacity in additional pathways such as MMR, which also contributes importantly to alkylation sensitivity [32]. Finally, the ability of FM-HCR to identify agents that either inhibit or enhance DRC in human cells (**Figure 2.4c** and **4d**) opens the door to screens for novel compounds that could be used either to potentiate the effects of DNA damage-based anticancer agents or to mitigate the deleterious effects of environmental exposure to DNA damaging agents [57-59].

In addition to the possible clinical applications described above, HCRseq has the potential to reveal new biological phenomena in the basic research setting. The mRNAseq data presented here provide evidence that transcriptional errors result when human RNA polymerase II bypasses a CPD. Because the plasmids are not replicated in the cell, and transcript sequence changes were observed at an elevated rate in repair deficient cells, these changes are likely to reflect transcriptional mutagenesis events due to unrepaired DNA lesions in the transcribed DNA strand. While it has been reported previously that *in vivo* bypass of a CPD by RNA polymerase may result in rare deletions, and bypass of

a bulky 8,5'-cyclo-2'-deoxyadenosine lesion may result in both deletions and base substitutions [60], our observation of frequent base misincorporation opposite a CPD by RNA polymerase II appears to be without precedent. A recent *in vitro* analysis indicated that transcriptional CPD bypass followed a so-called A-rule, resulting in error-free bypass [61]; athough base misincorporation was observed, subsequent extension of transcripts beyond the misincorporated base was strongly inhibited. The present data provide *in vivo* evidence of error-prone transcriptional bypass of bulky DNA lesions in human cells followed by completion and polyadenylation of the transcript. A lower limit (about 6%) for the frequency of bypass events resulting in an AA→GA mutation can be estimated from the data in **Figure 2.6c**. Since it is expected that reporter plasmids that have already been repaired will be transcribed at a higher rate, and because error-free bypass (according to an A-rule) cannot be distinguished from transcripts arising from repaired plasmid, the rate of bypass is likely higher than 6%.

Conclusions

FM-HCR and HCRseq represent powerful new tools for high throughput measurements of human DRC and provide a rapid functional characterization that complements existing, indirect measures of DRC. FM-HCR permits the simultaneous measurement of repair for up to four different doses of DNA damage, or types of DNA damage, in a single assay. HCR-seq has the potential to measure thousands of reporter sequences in a single assay, with barcodes providing unique identifiers for the type or dose of DNA damage as well as for the individual whose cells are being analyzed. Both methods expand the scope of lesions whose repair can be measured to include those that do not block transcription, and as additional substrates are developed, we anticipate that our assays will permit measurements of DRC in all of the major DNA repair pathways in a single assay. Our assays hold an advantage over *in vitro* assays because the transcription-based reporters limit the readout to DNA that has been repaired

in vivo in chromatinized DNA, thus increasing the likelihood of recapitulating physiological DNA repair phenotypes. The assays have the potential to reduce the cost and labor required for DRC measurements to a level compatible with large-scale epidemiological studies and clinical diagnostic/prognostic applications. The data presented herein also illustrate the utility of the assays as a research tool that can reveal mechanisms of DNA repair and damage tolerance and that can provide a new means of screening chemical libraries for inhibitors or enhancers of DRC.

Materials and Methods

DNA Repair Reporter Plasmids

Detailed methodology for the construction of reporter plasmids and the methods used to transfect plasmids into cells can be found in the supplemental information (**Figs S2-S4**). Briefly, Plasmids for expression of the fluorescent proteins AmCyan, EGFP, mOrange, and mPlum were purchased from Clontech, and that for tagBFP was purchased from Axxora. Reporter genes were subcloned into the pmax cloning vector (Lonza). NER reporters were prepared by irradiating plasmids with UVC light. The resulting DNA damage prevents fluorescent reporter expression by blocking transcription; repair by NER restores reporter expression. The NHEJ reporter comprised a linearized fluorescent reporter; because double strand breaks constitute an absolute block to transcription, NHEJ-dependent recircularization of the plasmid is required to restore reporter expression. MMR reporters consisted of heteroduplex DNA engineered such that the transcribed strand encoded a non-fluorescent protein. Repair of a single, site-specific mismatch restores the wild type sequence to the transcribed strand, and results in fluorescent reporter expression. Reporters for repair of 8-oxoG or O^6-MeG were engineered such that transcriptional mutagenesis in the presence of the DNA lesion lead to expression of wild type fluorescent reporter protein. Because repair removes the source of

transcriptional mutagenesis, repair of these plasmids is inversely proportional to the measured fluorescence.

Flow Cytometry

Cells suspended in culture media were analyzed for fluorescence on a BD LSRII cytometer, running FACSDIVA software. Cell debris, doublets and aggregates were excluded based on their side scatter and forward scatter properties. TOPRO-3 was added to cells 5-10 minutes prior to analysis, and used to exclude dead cells from the analysis. The following fluorophores and their corresponding detectors (in parentheses) were used: tagBFP (Pacific Blue), AmCyan (AmCyan), EGFP (FITC), mOrange (PE), mPlum (PE-Cy5-5), and TOPRO-3 (APC). The linear range for the corresponding photomultiplier tubes was determined using BD Rainbow fluorescent beads and unlabeled polystyrene beads based on the signal-to-noise ratio, %CV, and M1/M2 parameters as previously described [62]. Compensation was set using single color controls. Regions corresponding to cells positive for each of the 5 fluorescent proteins were established using single color dropout controls. For reporters that required compensation in more than one detector channel, fluorescence in the reporter channel was plotted separately against each of the channels requiring compensation. Using these plots, both single controls and the dropout control (in which the reporter of interest was excluded from the transfection) were used to establish regions corresponding to positive cells (**Figure S2.1a**). Equations used to calculate fluorescent reporter expression are detailed in the supplemental information.

mRNAseq

Total RNA was isolated using standard procedures detailed in the supplemental information. Total RNA samples were submitted to the MIT BioMicroCenter for preparation and sequencing. Briefly, total RNA was poly-A

purified, fragmented, and converted to cDNA using the Illumina Tru-Seq™ protocol. Library construction from cDNA was performed using the Beckman Coulter SPRI-works system. During library amplification, a unique bar-code was introduced for each of 8 samples corresponding to the four transfections performed in duplicate (**Table S2.1**), and from which total RNA was generated. Four samples from each replicate were clustered on a separate sequencing lane and run on an Illumina HiSeq 2000 instrument. Image analysis, base calling and sequence alignment to a synthetic genome consisting of the human genome and the five fluorescent reporter genes were performed using the Illumina Pipeline. Aberrant expression of the XPA gene in GM02344 cells provided an internal confirmation of the identity of the cell lines; reduced expression and an expected lack of regular splicing junction reads spanning intron 4 of the XPA gene from GM02344 was observed (**Figure S2.5c**), confirming a previously reported missplice in XPA transcripts due to the homozygous 555G>C mutation [45].

To ensure that trace DNA contamination of the RNAseq samples did not contribute significantly to the observed frequency of base substitutions in transcripts expressed from reporter plasmids (**Figure 2.6**), a second complementary sample preparation was performed and analyzed by Illumina sequencing. Details of the experimental procedures are available in the supplemental information; briefly, mRNA isolated from cells transfected with reporter plasmids was treated with DNAse and reverse transcribed to generate a cDNA library. PCR amplification of reporter cDNA was not detected in when mRNA that was not reverse transcribed was used as a template (**Figure S2.5d**), confirming cDNA as the template for PCR amplification, and hence ruling out significant plasmid contamination. Amplicons were fragmented and submitted for standard Illumina sample preparation.

Next Generation Sequencing Data Analysis

Illumina sequencing data were analyzed using the Tuxedo software suite. Mapped reads were aligned to the hg19 human genome assembly and the

five reporter gene sequences using Tophat v 2.0.6, and junction reads determined. Additional details of all analyses including input parameters are available in **tables S2-S6**. Cufflinks v 2.0.2 was run to quantify reads in terms of reads per kilobase of transcript per million mapped reads (RPKM) [63]. Samtools mpileup (v 0.1.16 r963:234) was used to aggregate reads at all positions in the alignment file. Using the pileup file as input, single nucleotide variants, as well as insertions and deletions (indels) present in the mRNAseq data were identified using the software package VarScan v2.3.4 [64]. All positions meeting a minimum read depth of 8 were considered further, however no minimum variant frequency threshold was set in order to detect rare variants and to establish the sequencing error rate. Custom Python scripts were used to generate a list of all deletions spanning an ApA sequence. The frequencies for base substitutions at each ApA sequence in the reporter transcripts were also determined.

Statistics

Statistics were performed with the GraphPad Prism 5.0 software package. The correlations between data sets in **Figs. 1 and 4** were assessed using a linear regression model that reports R^2 for the goodness of fit and a p value for the slope of the line being significantly different from zero. The p values in **Figure 2.6** were calculated from a two-tailed unpaired t-test. Error bars in **figs. 1,2,3,4** and **5** report the standard deviation of at least three biological replicates

Acknowledgements

We acknowledge funding support from NIH grant DP1-ES022576. We thank Dr. Jennifer Calvo for establishing the immortalized MEF cell lines, Prof. Penny Jeggo for the V79 and xrs6 cell lines, and Prof. Ryan Jensen for the VC8 cell line. Authors declare no competing interests.

References

1. Lindahl T & Wood RD (1999) Quality control by DNA repair. (Translated from English) *Science* 286(5446):1897-1905 (in English).
2. Hoeijmakers JHJ (2001) Genome maintenance mechanisms for preventing cancer. (Translated from English) *Nature* 411(6835):366-374 (in English).
3. Jackson SP & Bartek J (2009) The DNA-damage response in human biology and disease. (Translated from English) *Nature* 461(7267):1071-1078 (in English).
4. Ciccia A & Elledge SJ (2010) The DNA Damage Response: Making It Safe to Play with Knives. (Translated from English) *Mol. Cell* 40(2):179-204 (in English).
5. Fu D, Calvo JA, & Samson LD (2012) Balancing repair and tolerance of DNA damage caused by alkylating agents. *Nat. Rev. Cancer* 12(2):104-120.
6. Ralhan R, Kaur J, Kreienberg R, & Wiesmuller L (2007) Links between DNA double strand break repair and breast cancer: Accumulating evidence from both familial and nonfamilial cases. (Translated from English) *Cancer Letters* 248(1):1-17 (in English).
7. Wilson DM, Kim D, Berquist BR, & Sigurdson AJ (2011) Variation in base excision repair capacity. *Mutat. Res.-Fundam. Mol. Mech. Mutagen.* 711(1-2):100-112.
8. Athas WF, Hedayati MA, Matanoski GM, Farmer ER, & Grossman L (1991) Development and field-test validation of an assay for DNA-repair in circulating human lymphocytes. (Translated from English) *Cancer Research* 51(21):5786-5793 (in English).
9. Decordier I, Loock KV, & Kirsch-Volders M (2010) Phenotyping for DNA repair capacity. (Translated from English) *Mutat. Res.-Rev. Mutat. Res.* 705(2):107-129 (in English).
10. Jalal S, Earley JN, & Turchi JJ (2011) DNA Repair: From Genome Maintenance to Biomarker and Therapeutic Target. *Clinical Cancer Research* 17(22):6973-6984.
11. Sarkaria JN, *et al.* (2008) Mechanisms of chemoresistance to alkylating agents in malignant glioma. *Clinical Cancer Research* 14(10):2900-2908.
12. Chin L & Gray JW (2008) Translating insights from the cancer genome into clinical practice. *Nature* 452(7187):553-563.
13. van 't Veer LJ & Bernards R (2008) Enabling personalized cancer medicine through analysis of gene-expression patterns. *Nature* 452(7187):564-570.
14. Ellis NC (2003) *Obtaining and Using Genetic Information* (Springer, New York).
15. Li CY, *et al.* (2013) Polymorphisms of Nucleotide Excision Repair Genes Predict Melanoma Survival. *J. Invest. Dermatol.* 133(7):1813-1821.
16. Warmoes M, *et al.* (2012) Proteomics of Mouse BRCA1-deficient Mammary Tumors Identifies DNA Repair Proteins with Potential

Diagnostic and Prognostic Value in Human Breast Cancer. *Molecular & Cellular Proteomics* 11(7).
17. Redaelli A, Magrassi R, Bonassi S, Abbondandolo A, & Frosina G (1998) AP endonuclease activity in humans: Development of a simple assay and analysis of ten normal individuals. *Teratogenesis Carcinogenesis and Mutagenesis* 18(1):17-26.
18. Geng H, *et al.* (2011) In vitro studies of DNA mismatch repair proteins. (Translated from English) *Anal. Biochem.* 413(2):179-184 (in English).
19. Georgiadis P, Polychronaki N, & Kyrtopoulos SA (2012) Progress in high-throughput assays of MGMT and APE1 activities in cell extracts. (Translated from English) *Mutat. Res.-Fundam. Mol. Mech. Mutagen.* 736(1-2):25-32 (in English).
20. Leitner-Dagan Y, *et al.* (2012) N-Methylpurine DNA Glycosylase and OGG1 DNA Repair Activities: Opposite Associations With Lung Cancer Risk. (Translated from English) *J. Natl. Cancer Inst.* 104(22):1765-1769 (in English).
21. Evans RG & Norman A (1968) Radiation stimulated incorporation of thymidine into DNA of human lymphocytes. (Translated from English) *Nature* 217(5127):455-& (in English).
22. Perry P & Evans HJ (1975) Cytological detection fo mutagen carcinogen exposure by sister chromatid exchange (Translated from English) *Nature* 258(5531):121-125 (in English).
23. Parshad R, Sanford KK, & Jones GM (1983) Chromatid damage after G2 phase X-irradiation of cells from cancer-prone individuals implicates deficiency in DNA repair (Translated from English) *Proceedings of the National Academy of Sciences of the United States of America-Biological Sciences* 80(18):5612-5616 (in English).
24. Wood DK, Weingeist DM, Bhatia SN, & Engelward BP (2010) Single cell trapping and DNA damage analysis using microwell arrays. *Proceedings of the National Academy of Sciences of the United States of America* 107(22):10008-10013.
25. Qiao YW, *et al.* (2002) Rapid assessment of repair of ultraviolet DNA damage with a modified host-cell reactivation assay using a luciferase reporter gene and correlation with polymorphisms of DNA repair genes in normal human lymphocytes. *Mutat. Res.-Fundam. Mol. Mech. Mutagen.* 509(1-2):165-174.
26. Ramos JM, *et al.* (2004) DNA repair and breast carcinoma susceptibility in women. (Translated from English) *Cancer* 100(7):1352-1357 (in English).
27. Li C, Wang L-E, & Wei Q (2009) DNA repair phenotype and cancer susceptibility-A mini review. *International Journal of Cancer* 124(5):999-1007.
28. Mendez P, *et al.* (2011) A modified host-cell reactivation assay to quantify DNA repair capacity in cryopreserved peripheral lymphocytes. *DNA Repair* 10(6):603-610.

29. Jagger J (1976) Ultraviolet inactivation of biological systems. *Photochemistry and Photobiology of Nucleic Acids*, ed Wang SY (Academic Press, New York), Vol 2, pp 147-186.
30. Parsons R, *et al.* (1993) Hypermutability and mismatch repair deficiency in RER+ tumor cells. (Translated from English) *Cell* 75(6):1227-1236 (in English).
31. Burns JA, Dreij K, Cartularo L, & Scicchitano DA (2010) O-6-Methylguanine induces altered proteins at the level of transcription in human cells. *Nucleic Acids Research* 38(22):8178-8187.
32. Kat A, *et al.* (1993) An alkylation-tolerant, mutator human cell line is deficient in strand-specific mismatch repair (Translated from English) *Proceedings of the National Academy of Sciences of the United States of America* 90(14):6424-6428 (in English).
33. Zhukovskaya N, Rydberg B, & Karran P (1992) Inactive O6-methylguanine-DNA methyl transferase in human cells. (Translated from English) *Nucleic Acids Research* 20(22):6081-6090 (in English).
34. Allalunis-Turner MJ, Barron GM, Day RS, Dobler KD, & Mirzayans R (1993) Isolation of two cell lines from a human malignant glioma specimen differing in sensitivity to radiation and chemotherapeutic drugs. *Radiation Research* 134(3):349-354.
35. Leesmiller SP, *et al.* (1995) Absence of p350 subunit of DNA activated protein kinase from a radiosensitive human cell line. *Science* 267(5201):1183-1185.
36. Galloway AM & Allalunis-Turner J (2000) cDNA expression array analysis of DNA repair genes in human glioma cells that lack or express DNA-PK. *Radiation Research* 154(6):609-615.
37. Mu D, Hsu DS, & Sancar A (1996) Reaction mechanism of human DNA repair excision nuclease. *J. Biol. Chem.* 271(14):8285-8294.
38. Klungland A, *et al.* (1999) Accumulation of premutagenic DNA lesions in mice defective in removal of oxidative base damage. *Proceedings of the National Academy of Sciences of the United States of America* 96(23):13300-13305.
39. Kiziltepe T, *et al.* (2005) Delineation of the chemical pathways underlying nitric oxide-induced homologous recombination in mammalian cells. *Chemistry & Biology* 12(3):357-369.
40. Collins FS, Brooks LD, & Chakravarti A (1998) A DNA polymorphism discovery resource for research on human genetic variation. *Genome Res.* 8(12):1229-1231.
41. Hartwig A (2010) Mechanisms in cadmium-induced carcinogenicity: recent insights. (Translated from English) *Biometals* 23(5):951-960 (in English).
42. Hartwig A, *et al.* (1997) Interaction of arsenic(III) with nucleotide excision repair in UV-irradiated human fibroblasts. *Carcinogenesis* 18(2):399-405.
43. Griffin RJ, *et al.* (2005) Selective benzopyranone and pyrimido 2,1-alpha isoquinolin-4-one inhibitors of DNA-dependent protein kinase: Synthesis, structure-activity studies, and radiosensitization of a human tumor cell line

in vitro. (Translated from English) *J. Med. Chem.* 48(2):569-585 (in English).
44. Mortazavi A, Williams BA, McCue K, Schaeffer L, & Wold B (2008) Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature Methods* 5(7):621-628.
45. Satokata I, Tanaka K, Yuba S, & Okada Y (1992) Identification of splicing mutations of the last nucleotides of exons, a nonsense mutation, and a missense mutation of the XPAC gene as causes of group A xeroderma pigmentosum. (Translated from English) *Mutation Research* 273(2):203-212 (in English).
46. Choy E, *et al.* (2008) Genetic Analysis of Human Traits In Vitro: Drug Response and Gene Expression in Lymphoblastoid Cell Lines. *Plos Genetics* 4(11).
47. Davis AR & Kohane IS (2009) Expression differences by continent of origin point to the immortalization process. *Human Molecular Genetics* 18(20):3864-3875.
48. Stark AL, *et al.* (2010) Heritable and non-genetic factors as variables of pharmacologic phenotypes in lymphoblastoid cell lines. *Pharmacogenomics Journal* 10(6):505-512.
49. Hickman MJ & Samson LD (1999) Role of DNA mismatch repair and p53 in signaling induction of apoptosis by alkylating agents. *Proceedings of the National Academy of Sciences of the United States of America* 96(19):10764-10769.
50. Bregeon D, Peignon PA, & Sarasin A (2009) Transcriptional Mutagenesis Induced by 8-Oxoguanine in Mammalian Cells. *Plos Genetics* 5(7).
51. Bregeon D & Doetsch PW (2006) Assays for transcriptional mutagenesis in active genes. *DNA Repair, Pt B,* Methods in Enzymology, (Elsevier Academic Press Inc, San Diego), Vol 409, pp 345-357.
52. Bhattacharjee P, Banerjee M, & Giri AK (2013) Role of genomic instability in arsenic-induced carcinogenicity. A review. (Translated from English) *Environ. Int.* 53:29-40 (in English).
53. Shrivastav N, Li DY, & Essigmann JM (2010) Chemical biology of mutagenesis and DNA repair: cellular responses to DNA alkylation. *Carcinogenesis* 31(1):59-70.
54. Bregeon D & Doetsch PW (2011) Transcriptional mutagenesis: causes and involvement in tumour development. (Translated from English) *Nat. Rev. Cancer* 11(3):218-U288 (in English).
55. Alapetite C, Thirion P, de la Rochefordiere A, Cosset JM, & Moustacchi E (1999) Analysis by alkaline comet assay of cancer patients with severe reactions to radiotherapy: Defective rejoining of radioinduced dna strand breaks in lymphocytes of breast cancer patients. (Translated from English) *International Journal of Cancer* 83(1):83-90 (in English).
56. Hegi ME, *et al.* (2005) MGMT gene silencing and benefit from temozolomide in glioblastoma. *New England Journal of Medicine* 352(10):997-1003.

57. Srinivasan A & Gold B (2012) Small-molecule inhibitors of DNA damage-repair pathways: an approach to overcome tumor resistance to alkylating anticancer drugs. *Future Medicinal Chemistry* 4(9):1093-1111.
58. Kim K, *et al.* (2011) High throughput screening of small molecule libraries for modifiers of radiation responses. (Translated from English) *Int. J. Radiat. Biol.* 87(8):839-845 (in English).
59. Zellefrow CD, *et al.* (2012) Identification of Druggable Targets for Radiation Mitigation Using a Small Interfering RNA Screening Assay. (Translated from English) *Radiation Research* 178(3):150-159 (in English).
60. Marietta C & Brooks PJ (2007) Transcriptional bypass of bulky DNA lesions causes new mutant RNA transcripts in human cells. (Translated from English) *EMBO Rep.* 8(4):388-393 (in English).
61. Walmacq C, *et al.* (2012) Mechanism of Translesion Transcription by RNA Polymerase II and Its Role in Cellular Resistance to DNA Damage. *Mol. Cell* 46(1):18-29.
62. Perfetto SP, Ambrozak D, Nguyen R, Chattopadhyay P, & Roederer M (2006) Quality assurance for polychromatic flow cytometry. *Nat. Protoc.* 1(3):1522-1530.
63. Trapnell C, *et al.* (2012) Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. (Translated from English) *Nat. Protoc.* 7(3):562-578 (in English).
64. Koboldt DC, *et al.* (2012) VarScan 2: Somatic mutation and copy number alteration discovery in cancer by exome sequencing. (Translated from English) *Genome Res.* 22(3):568-576 (in English).

Figure Legends

Figure 2.1. Measurements of DRC by FM-HCR.

(**a**) DNA lesions are introduced into fluorescent reporter plasmids *in vitro*. Numbers labeling the plasmids represent the dose (in J/m^2) of UV radiation. Following treatment, plasmids were combined and co-transfected into cells. After 18 or 40 hours incubation, cells were assayed for fluorescence by flow cytometry. Comparison of fluorescence signals to those from cells transfected with undamaged plasmids yields a dose-response curve (experimental data for GM02344 with plasmid combination #1 in **Table 2.2**) (**b**) Dose-response curves for seven cell lines 18 hours after transfection with plasmid combination #1 (**Table 2.2**) (**c**) Dose response curves for the cells in (b) at 40 hours. (**d**) Comparison of % reporter expression as measured by FM-HCR at 400 J/m^2 is plotted against %CAT as measured by conventional HCR for the same cell lines at 300 J/m^2. (**e**) D_0 values calculated from FM-HCR data plotted against those reported in the literature. Error bars represent the standard deviation calculated from biological triplicates. (**f**) Dose-response curves for seven cell lines 18 hours after transfection with plasmid combination #2 (**Table 2.2**). (**g**) Dose response curves at 40 hours. (**h**) Comparison of % reporter expression as measured by FM-HCR at 400 J/m^2 with plasmid combination #2 is plotted against %CAT as measured by conventional HCR for the same cell lines at 300 J/m^2. (**i**) Comparison of FM-HCR data for plasmids treated at 400 J/m^2 in experiments #1 and #2. (**j**) Dose-response curves generated by FM-HCR for lymphoblastoid cell lines 40 hours after transfection with plasmid combination #2 (**Table 2.2**). (**k**) Corresponding dose response curves for primary skin fibroblasts from the same seven individuals. (**l**) Correlation between % reporter expression from plasmids irradiated at 800 J/m^2 in the lymphoblastoid and fibroblast cells isolated from the same individuals. Each color in panels **j**, **k** and **l** corresponds to one of the individuals (i-vii) in **Table 2.1**. Error bars represent the standard deviation calculated from biological triplicates. See also **Fig S1**.

Figure 2.2 Simultaneous measurements of DRC in three pathways.
(**a**) Plasmids used in the multi-pathway FM-HCR (also see plasmid combination #3 in **Table 2.2**). (**b**) DRC for several cell lines obtained by assaying each pathway in a separate transfection experiment (left) simultaneously following co-transfection of the reporter plasmids (right). Error bars represent the standard deviation calculated from biological triplicates. See also **Fig S2**.

Figure 2.3. Simultaneous measurements of DRC in four pathways.
(**a**) Plasmids used in the FM-HCR for NHEJ, NER, MMR and MGMT (also see plasmid combinations #4 **Table 2.2**). Note that the undamaged plasmid (AmCyan) included to control for transfection efficiency is not shown. (**b**) DRC for several cell lines obtained by assaying each pathway in a separate transfection experiment. (**c**) DRC measured simultaneously following co-transfection of the reporter plasmids. (**d**) Plasmids used in the FM-HCR for NER, MMR, BER, and MGMT (also see plasmid combinations #5 **Table 2.2**). (**e**) DRC for several cell lines obtained by assaying each pathway in a separate transfection experiment. (**f**) DRC measured simultaneously following co-transfection of the reporter plasmids. Error bars represent the standard deviation calculated from biological triplicates. See also **Fig S3**.

Figure 2.4. Applications of FM-HCR to interindividual DRC differences and identifying DNA repair inhibitors and enhancers.
(a) FM-HCR analysis of repair capacity in 5 pathways for 27 cell lines. Cells were transfected with plasmid combination #3 or plasmid combination #6. (b) Correlation between MGMT transcript levels measured by taqMan qPCR and % Reporter Expression (log transformed) from the MGMT HCR. (c) FM-HCR measurements of NER inhibition. (d) FM-HCR measurements of Compound 401 inhibition of DNA repair capacity in four pathways. For all experiments, cells were assayed by flow cytometry 18 hours after transfection. See also **Figure S2.4**.

Figure 2.5. Workflow for experiments comparing two methods of analyzing reporter expression.

Following transfection, an aliquot of cells is analyzed by flow cytometry. From the remaining cells, RNA is isolated and an aliquot is subjected to Illumina sample preparation and sequencing. FM-HCR analysis of fluorescent reporter expression, is compared to HCR-Seq analysis of reporter transcript expression, measured as RPKM. See also **Tables S1-5**.

Figure 2.6. mRNAseq analysis of reporter expression.

(a) Dose response curves for reporter expression from randomly damaged plasmids generated from mRNAseq analysis. (b) Dose response curves for the same cells generated from flow cytometric (FM-HCR) analysis. (c) Sequence variants detected in transcripts at the position corresponding to the site-specific thymine dimer in the absence (top) or presence of the lesion (bottom). Frequencies are reported for the expected sequence (AA) as well as all variants that were observed in at least one sample. (d) Frequencies of AA→GA mutations in transcripts expressed from randomly damaged plasmids as a function of dose (combination #7 in **Table 2.2**). The undamaged case (0 J/m^2) refers to the frequency of mutations as measured in transcripts expressed from the BFP transfection control. "U" refers to HCR-Seq data from cells in which all of the reporter plasmids were undamaged, and "D" refers data for cells transfected with reporters irradiated as indicated in Table 2.2. Symbols (*) represent differences that were deemed to be statistically significant ($p<0.05$) by a *t* test. See also **Figure S2.5**.

Figures

Figure 2.1

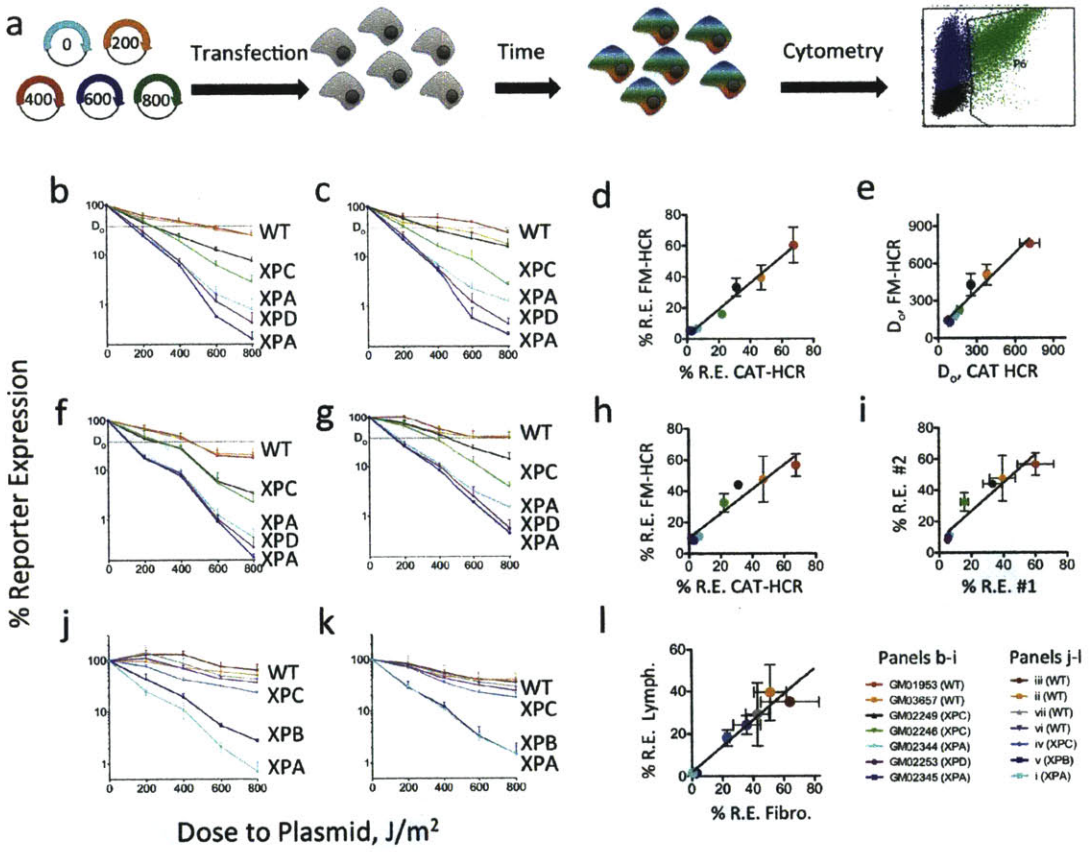


Figure 2.2

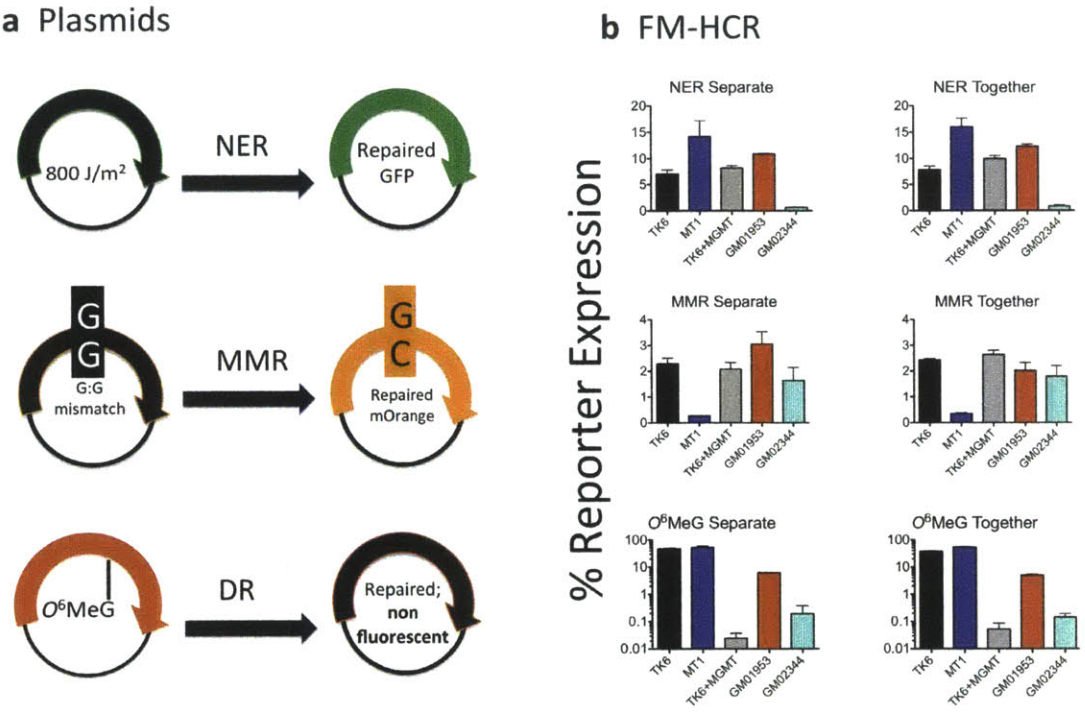

a Plasmids
800 J/m²
NER
Repaired GFP
G
G
G:G mismatch
MMR
G
C
Repaired mOrange
O⁶MeG
DR
Repaired; non fluorescent
b FM-HCR
% Reporter Expression
NER Separate
NER Together
MMR Separate
MMR Together
O⁶MeG Separate
O⁶MeG Together
TK6
MT1
TK6+MGMT
GM01953
GM02344

Figure 2.3

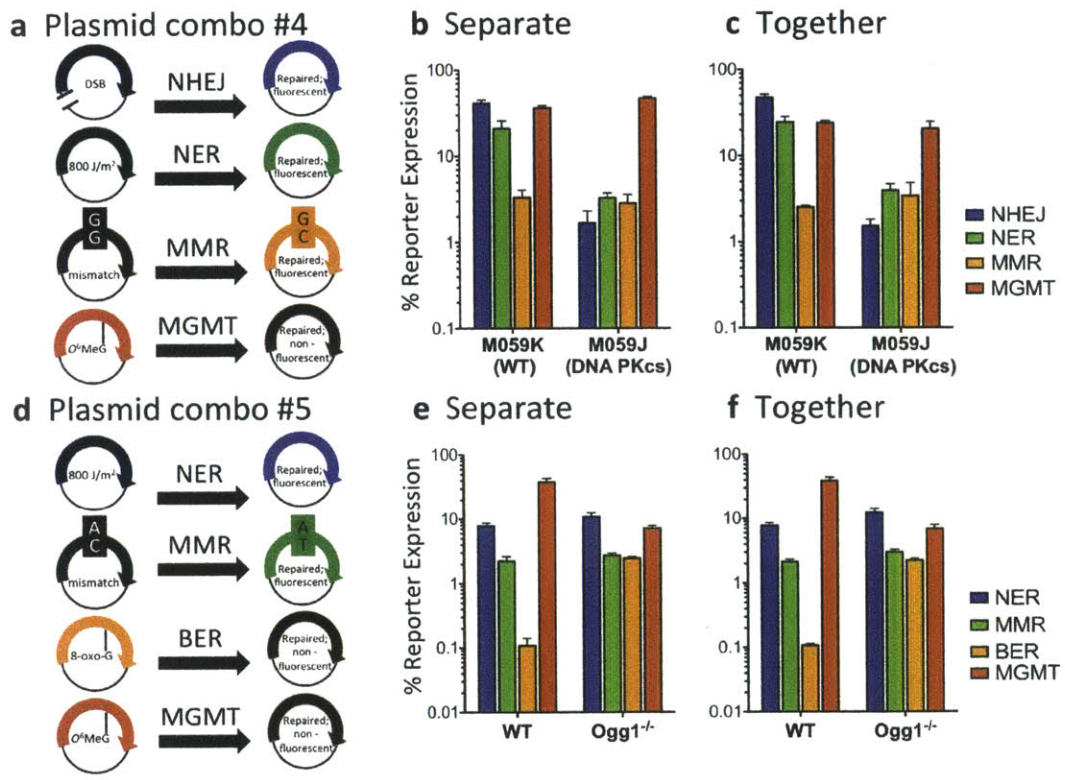

a Plasmid combo #4
DSB
NHEJ
Repaired; fluorescent
800 J/m²
NER
Repaired; fluorescent
G
G
mismatch
MMR
G
C
Repaired; fluorescent
O⁶MeG
MGMT
Repaired; non - fluorescent
b Separate
% Reporter Expression
100
10
1
0.1
M059K (WT)
M059J (DNA PKcs)
c Together
100
10
1
0.1
M059K (WT)
M059J (DNA PKcs)
NHEJ
NER
MMR
MGMT
d Plasmid combo #5
800 J/m²
NER
Repaired; fluorescent
A
C
mismatch
MMR
A
T
Repaired; fluorescent
8-oxo-G
BER
Repaired; non - fluorescent
O⁶MeG
MGMT
Repaired; non - fluorescent
e Separate
% Reporter Expression
100
10
1
0.1
0.01
WT
Ogg1-/-
f Together
100
10
1
0.1
0.01
WT
Ogg1-/-
NER
MMR
BER
MGMT

Figure 2.4

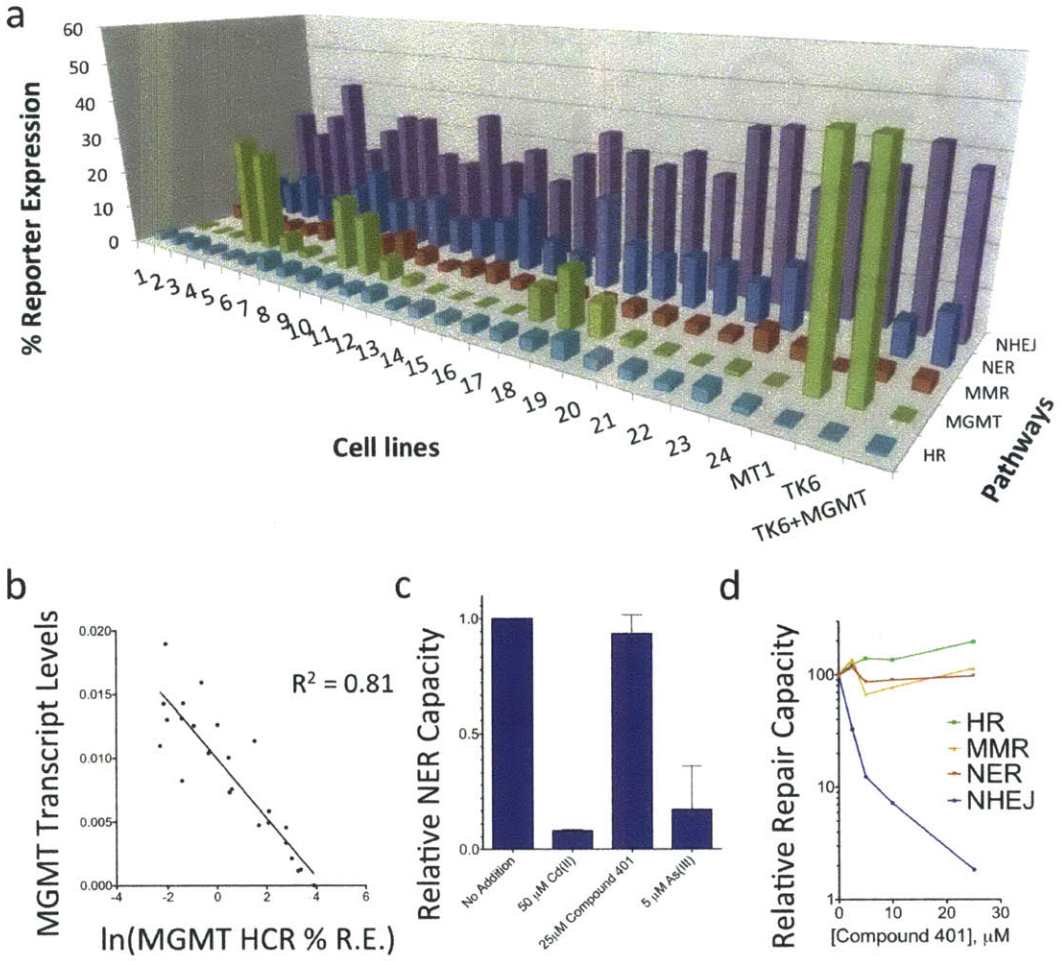

a
% Reporter Expression
60
50
40
30
20
10
0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 MT1 TK6 TK6+MGMT
Cell lines
NHEJ
NER
MMR
MGMT
HR
Pathways
b
MGMT Transcript Levels
0.020
0.015
0.010
0.005
0.000
-4 -2 0 2 4 6
R² = 0.81
ln(MGMT HCR % R.E.)
c
Relative NER Capacity
1.0
0.5
0.0
No Addition
50 μM Cd(II)
25μM Compound 401
5 μM As(III)
d
Relative Repair Capacity
100
10
1
0 10 20 30
[Compound 401], μM
HR
MMR
NER
NHEJ

Figure 2.5

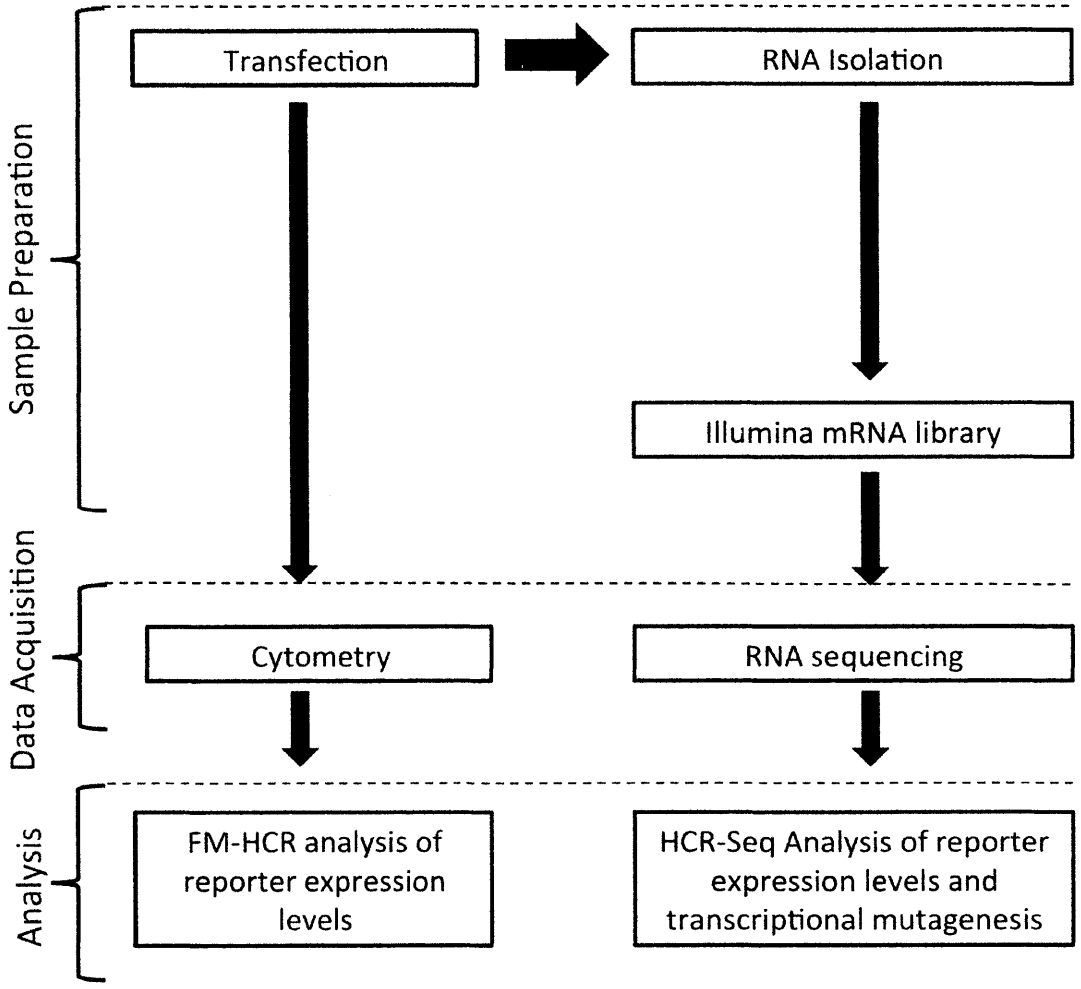
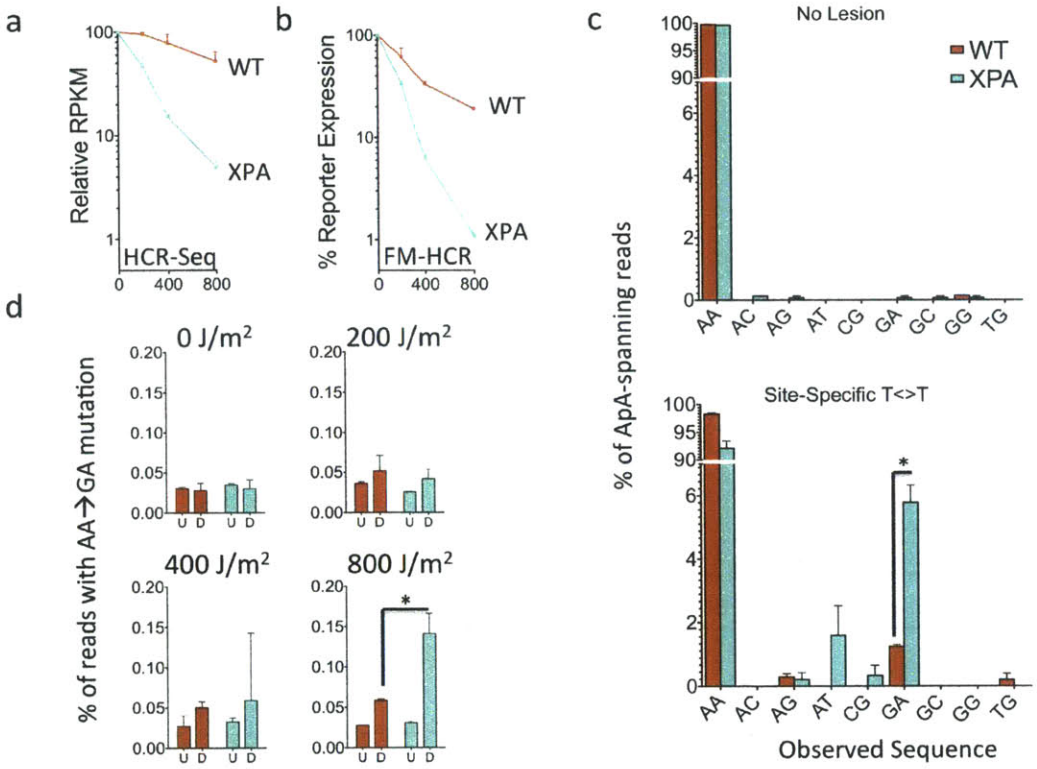

Sample Preparation
Transfection
RNA Isolation
Illumina mRNA library
Data Acquisition
Cytometry
RNA sequencing
Analysis
FM-HCR analysis of reporter expression levels
HCR-Seq Analysis of reporter expression levels and transcriptional mutagenesis

Figure 2.6


a
Relative RPKM
100
10
1
WT
XPA
HCR-Seq
0
400
800
b
% Reporter Expression
100
10
1
WT
XPA
FM-HCR
0
400
800
c
No Lesion
WT
XPA
% of ApA-spanning reads
100
95
90
6
4
2
0
AA
AC
AG
AT
CG
GA
GC
GG
TG
Site-Specific T<>T
*
Observed Sequence
d
% of reads with AA→GA mutation
0 J/m²
200 J/m²
400 J/m²
800 J/m²
0.20
0.15
0.10
0.05
0.00
U
D
*

Tables:

Table 2.1. 55 cell lines used for this study.

To facilitate comparison of data, the seven individuals from whom both lymphoblastoid and fibroblast cultures were derived have been assigned indexes i through vii.

Cell Line	Cell Type	Genotype	Repair Defect
GM01630 (i)	Fibroblast	XPA	NER, severe
GM01953	Lymphoblastoid	WT	None
GM02246	Lymphoblastoid	XPC	NER, moderate
GM02249	Lymphoblastoid	XPC	NER, mild
GM02253	Lymphoblastoid	XPD	NER, severe
GM02344 (i)	Lymphoblastoid	XPA	NER, severe
GM02345	Lymphoblastoid	XPA	NER, severe
GM03657 (ii)	Lymphoblastoid	WT	None
GM03658 (ii)	Fibroblast	WT	None
GM07752 (iii)	Lymphoblastoid	WT	None
GM07753 (iii)	Fibroblast	WT	None
GM14878 (iv)	Lymphoblastoid	XPC	NER, very mild
GM14879 (iv)	Fibroblast	XPC	NER, very mild
GM21071 (v)	Fibroblast	XPB	NER, severe
GM21148 (v)	Lymphoblastoid	XPB	NER, severe
GM21677 (vi)	Lymphoblastoid	WT	None
GM21833 (vii)	Lymphoblastoid	WT	None
GM23249 (vi)	Fibroblast	WT	None
GM23251 (vii)	Fibroblast	WT	None
TK6	Lymphoblastoid	MGMT	DR of O^6MeG
MT1	Lymphoblastoid	MGMT, MSH6	MMR and DR of O^6MeG
TK6+MGMT	Lymphoblastoid	WT	None

HCT116	Colorectal Carcinoma	MLH1	MMR
HCT116+3	Colorectal Carcinoma	WT	None
M059J	Glioblastoma	DNA PKcs	NHEJ
M059K	Glioblastoma	WT	None
WT MEFS	MEFs	WT	None
$Ogg1^{-/-}$ MEFS	MEFs	Ogg1	BER of 8-oxoG
V79 (Hamster)	Fibroblasts	WT	None
VC8 (Hamster)	Fibroblasts	BRCA2	HR
xrs6 (Hamster)	CHO	Ku80	NHEJ
GM15029 (#1)	Lymphoblastoid	WT	None
GM15036 (#2)	Lymphoblastoid	WT	None
GM15215 (#3)	Lymphoblastoid	WT	None
GM15223 (#4)	Lymphoblastoid	WT	None
GM15245 (#5)	Lymphoblastoid	WT	None
GM15224 (#6)	Lymphoblastoid	WT	None
GM15236 (#7)	Lymphoblastoid	WT	None
GM15510 (#8)	Lymphoblastoid	WT	None
GM15213 (#9)	Lymphoblastoid	WT	None
GM15221 (#10)	Lymphoblastoid	WT	None
GM15227 (#11)	Lymphoblastoid	WT	None
GM15385 (#12)	Lymphoblastoid	WT	None
GM15590 (#13)	Lymphoblastoid	WT	None
GM15038 (#14)	Lymphoblastoid	WT	None
GM15056 (#15)	Lymphoblastoid	WT	None
GM15072 (#16)	Lymphoblastoid	WT	None
GM15144 (#17)	Lymphoblastoid	WT	None
GM15216 (#18)	Lymphoblastoid	WT	None
GM15226 (#19)	Lymphoblastoid	WT	None
GM15242 (#20)	Lymphoblastoid	WT	None

GM15268 (#21)	Lymphoblastoid	WT	None
GM15324 (#22)	Lymphoblastoid	WT	None
GM15386 (#23)	Lymphoblastoid	WT	None
GM15061 (#24)	Lymphoblastoid	WT	None

Table 2.2. Combinations of reporter plasmids and types of DNA damage used in each experiment.

Combination	tagBFP	AmCyan	EGFP	mOrange	mPlum
#1	600 J/m^2	No lesion	800 J/m^2	200 J/m^2	400 J/m^2
#2	No lesion	200 J/m^2	400 J/m^2	600 J/m^2	800 J/m^2
#3	No lesion	-	800 J/m^2	G:G[C]	O^6-MeG[D]
#4	DSB	No lesion	800 J/m^2	G:G[C]	O^6-MeG[D]
#5	800 J/m^2	No lesion	A:C[A]	8-oxoG	O^6-MeG[D]
#6	DSB	-	DSB	-	No Lesion
#7	No lesion	200 J/m^2	T<>T[B]	400 J/m^2	800 J/m^2

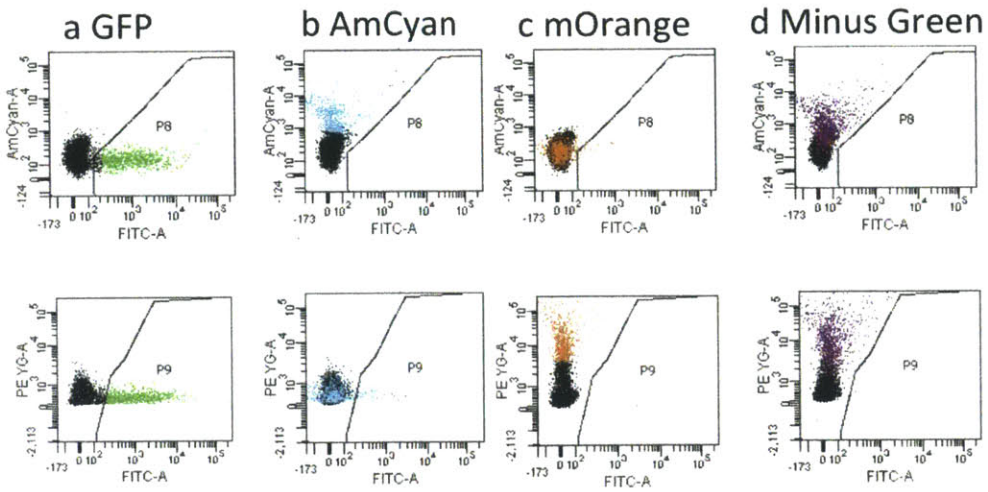
[A]A:C mismatch

[B]Site specific thymine dimer

[C]G:G mismatch

[D] O^6-MeG

Supplemental Figures



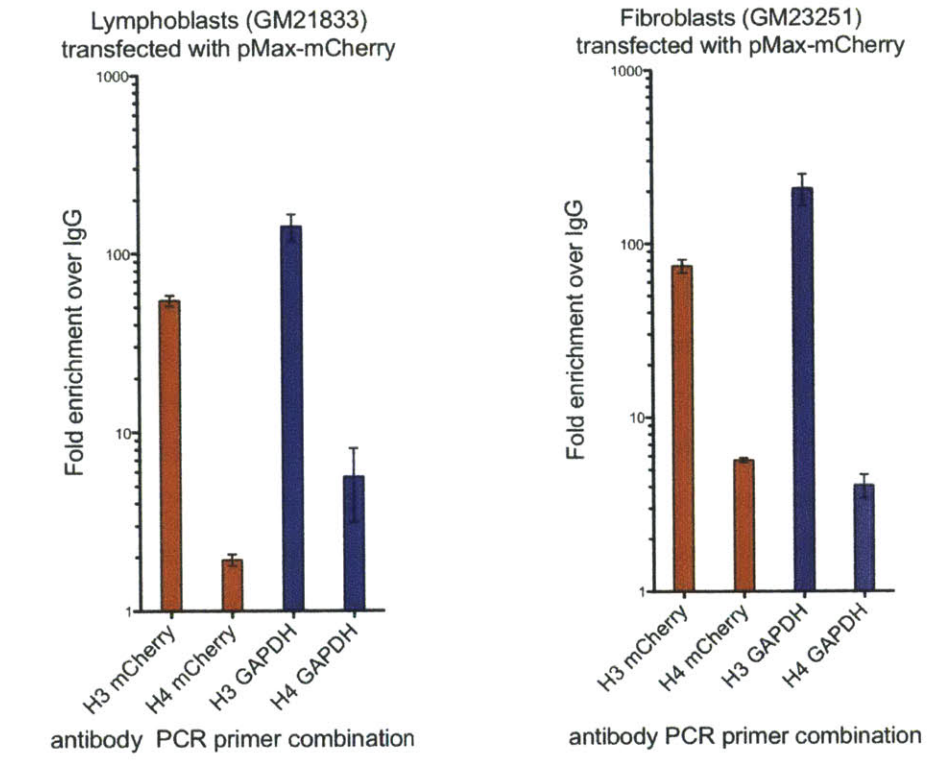


Figure S2.1; related to **Figure 2.1**. Transfection controls and method of establishing the positive region for a fluorophore requiring compensation in two or more channels.

Both the mOrange and AmCyan reporters overlap significantly with the GFP reporter; they result in significant signal in the GFP (FITC) channel even when compensation is applied, creating the potential for false positives. In this example, we illustrate why no single gate is sufficient to identify true positive GFP cells in the presence of mOrange and AmCyan. The problem is overcome by establishing a region that represents the union of two or more gates; cells must appear in *both* gates to be counted as positive for GFP. Each panel in the figure indicates the fluorescent reporter or mixture of reporters being expressed in the cells analyzed. The detector channels are as follows: AmCyan = Cyan, FITC = GFP, PE YG = mOrange. **a)** Flow cytometry plots for cells transfected with the GFP reporter. Bright fluorescence is detected in the FITC channel **b)** Flow cytometry plots for cells tranfected with the AmCyan reporter plasmid. Application of compensation leads to a typical "funnel" shape (the brightest cells in the AmCyan channel fan out into the FITC channel). Gate P8 is drawn in such a way as to exclude them as false positives. However, on a plot of fluorescence in the FITC channel against that in the PE YG channel, these cells are indistinguishable from GFP positive cells, and would be counted as false positives if gate P9 were used alone as the region corresponding to GFP positive cells. **c)** Plots for cells transfected with the mOrange reporter. Some cells appearing in gate P8 are indistinguishable from the GFP positive cells in panel (a). When fluorescence in the same cells is plotted as FITC vs. PE YG, it is seen that the false positives arise from the same "funnel" phenomenon occurring for AmCyan in in the upper plot in panel (b). As a result, using P8 alone would also be insufficient for the identification of true GFP positive cells in the presence of the other two fluorophores (mOrange and AmCyan). However, these false positives can be excluded using gate P9. Taken together, the observations in (b) and (c) indicate that a GFP positive region defined as the *union* of regions P8 and P9 represents true positives. **d)** Plots for cells transfected with all reporters

except for the GFP reporter. If needed, gates are further adjusted to minimize the number of cells appearing in the region P8+P9. **e)** Chromatin immunoprecipitation of plasmid DNA. The lymphoblastoid cell line GM21833 and the primary human fibroblast culture GM23251 were electroporated with the pmax:mCherry reporter plasmid and analyzed for plasmid chromatinization. Antibodies raised against the human histone H3 and H4 were used in all experiments. H3 precipitation yielded an enrichment of at least 40-fold for the reporter plasmid DNA in both cell types compared with chromatin precipitated with the nonspecific antibody IgG. Error bars represent the standard deviation of triplicate measurements. Similar enrichment was observed for the host gene glyceraldehyde 3-phosphate dehydrogenase (GAPDH). When DNA was immunoprecipitated with the antibody to histone H4, an enrichment of at least 2-fold was observed for the plasmid DNA and similar enrichment was found for GAPDH. Overall, the results are consistent with incorporation of plasmid DNA into nucleosomes.

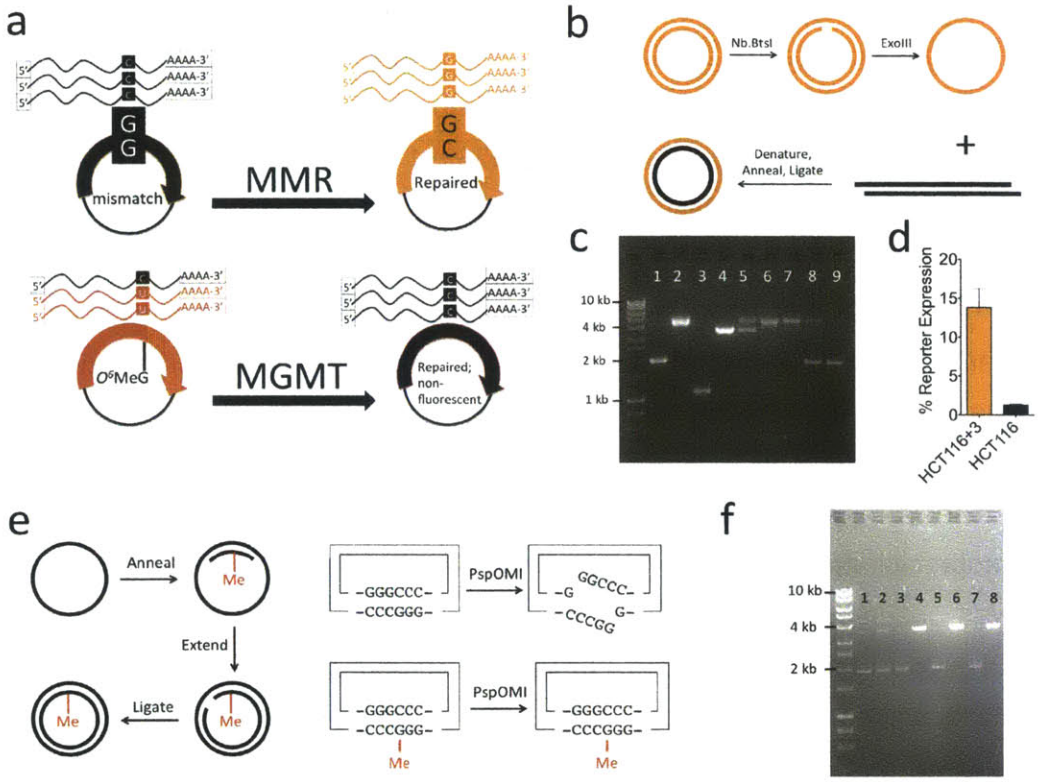


Figure S2.2; related to **Figure 2.2** Construction and validation of plasmid reporters for MMR and MGMT.

a) Fluorescent reporters for mismatch repair and MGMT. Top, repair of a G:G mismatch restores the wild type sequence to the transcribed strand, and results in expression of orange fluorescent protein. Bottom, use of transcriptional mutagenesis to measure repair of a site-specific O^6-methylguanine lesion. The lesion induces misincorporation of uracil into transcripts. The uracil containing transcripts are translated into wild type protein. Following repair of the O^6-methylguanine lesion, transcripts contain cytosine at the relevant position, and encode a non-fluorescent protein. **b**) Synthetic method for generation of heteroduplex plasmids. The transcribed strand of the reporter plasmid is nicked with a strand specific nicking endonuclease. The nicked strand is then digested with exonuclease III. The resulting closed circular single stranded DNA (ssDNA) is then combined with linearized double stranded DNA prepared from a second plasmid that differs by a single nucleotide; this sequence change prevents

expression of fluorescent protein. The two molecules are denatured with sodium hydroxide, and then brought to neutral pH to facilitate annealing between the circular ssDNA and the complementary strand from the linearized DNA. Unwanted side products (ssDNA and linear DNA) are selectively digested by plasmid safe ATP dependent nuclease (Epicentre). The desired heteroduplex DNA is finally ligated to produce a closed circular double stranded heteroduplex. **c**) Gel analysis of starting materials, intermediates and products of MMR substrate preparation. The material in lane 9 was used for HCR assays. Lane 1, Uncut; lane 2, *Nb.BtsI* (Nicked DNA); lane 3, *Nb.BtsI* + *ExoIII* (ssDNA); lane 4, *NheI* (linear DNA); lane 5, Annealing product of #3 and #4; lane 6, #5 + PSAD; lane 7, #6 following gel extraction; lane 8, #7 following ligation; lane 9, #8 following gel extraction. **d**) G:G mismatch repair in MMR-proficient HCT116+3 cells and MMR-deficient HCT116 cells. **e**) Synthetic method for O^6-methylguanine reporter, and molecular basis for resistance to reporter cleavage by *PspOMI*. An oligo containing a site-specific O^6-methylguanine lesion is annealed to single stranded DNA (non-transcribed strand), followed by primer extension and ligation to yield closed circular DNA. In the absence of O^6-methylguanine, *PspOMI* cleaves the mPlum C207G:T208C reporter plasmid at the recognition site *GGGCCC*. The presence of O^6-methylguanine prevents recognition by the enzyme, thus the reporter plasmid containing the site specific lesion in the transcribed strand is resistant to cleavage. **f**) Assay for site specific incorporation of O^6-methylguanine into the C207G:T208C mutant of the mPlum reporter plasmid. This assay was performed as an independent confirmation of the presence of the lesion, and to minimize the possibility that fluorescent signal might arise due to the presence of unmodified bases or other lesions. Primer extension reactions were performed with single stranded DNA from the C207G:T208C mutant of the mPlum reporter plasmid, which contains the GGGCCC recognition site for the restriction enzyme *PspOMI*. It has been shown previously that O^6-methylguanine can abolish recognition of restriction sites [1]. In the analytical digest shown below, we find that O^6-methylguanine blocks cleavage of the plasmid to at least 90%, and that removal of the lesion by MGMT

restores *PspOMI* cleavage at the restriction site to near 100%. The data are consistent with greater than 90% incorporation of the desired lesion at the intended position. Lane 1, O^6-MeG plasmid, no treatment; lane 2, O^6-MeG plasmid + *PspOMI*; lane 3, O^6-MeG plasmid + MGMT; lane 4, O^6-MeG plasmid + MGMT + *PspOMI*; lane 5, Homoduplex; lane 6, Homoduplex + *PspOMI*; lane 7, Homoduplex + MGMT; lane 8, Homoduplex + MGMT + *PspOMI*.

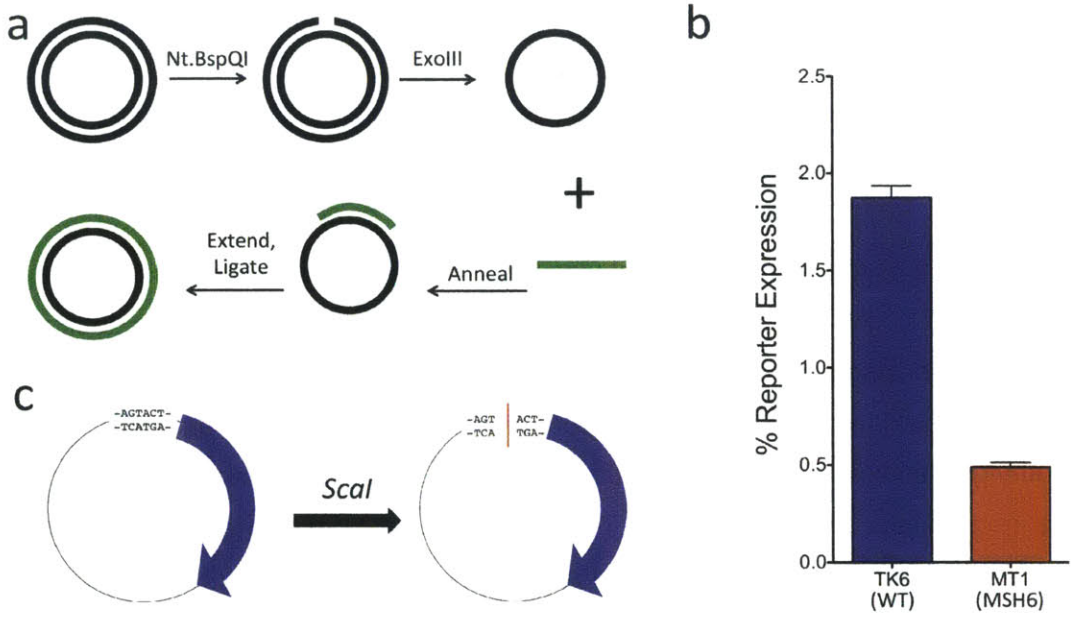


Figure S2.3; related to **Figure 2.3**. Construction and validation of reporters for MMR and NHEJ.

a) Synthetic method for A:C mismatch in a GFP reporter. The non-transcribed strand of the C289T mutant GFP reporter plasmid is nicked with *Nt.BspQI*, and digested from the plasmid with *ExoIII*. An oligo with the wild type sequence (G at the position opposite C289) and complementary to the region of the plasmid that has been mutated is annealed and extended to form a heteroduplex with an A:C mismatch. **b**) Validation of GFP reporter for MMR repair of an A:C mismatch. As

expected, reduced reporter expression was observed in the MSH6-deficient MT1 cell line relative to the WT TK6 cell line. **C**) Structure of the BFP reporter for NHEJ. An *ScaI* recognition site (AGTACT) was inserted upstream of the reporter gene (BFP, represented by a blue arrow). Plasmids were linearized with the *ScaI* restriction enzyme, which leaves a blunt-end double strand break in the 5' untranslated region of the reporter plasmid.

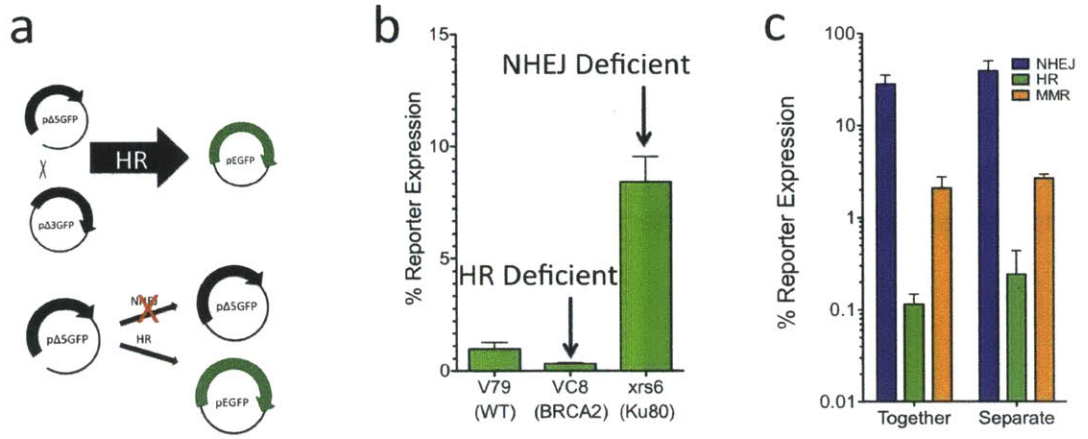


Figure S2.4; related to **Figure 2.4**. Homologous recombination plasmid reporters and assay validation.

(a) Recombination between two plasmids (one circular and one linearized) that express truncated non-fluorescent proteins results in a plasmid that expresses full-length green fluorescent protein. NHEJ-mediated repair of the linearized reporter plasmid does not give rise to fluorescent protein expression. (b) BRCA2-deficient VC8 hamster cells exhibit deficient homologous recombination relative to wild type V79 hamster cells. Ku80-deficient xrs6 hamster cells are NHEJ deficient, and elevated HR is observed, consistent with the competition between NHEJ and HR for repair of DSBs. (c) Cells were transfected with a combination of reporter plasmids for NHEJ, HR and MMR either in separate transfections or together in a single transfection. % Reporter expression under the two conditions was indistinguishable within the experimental error of the measurements (error bars represent the standard deviation of 3 experiments).

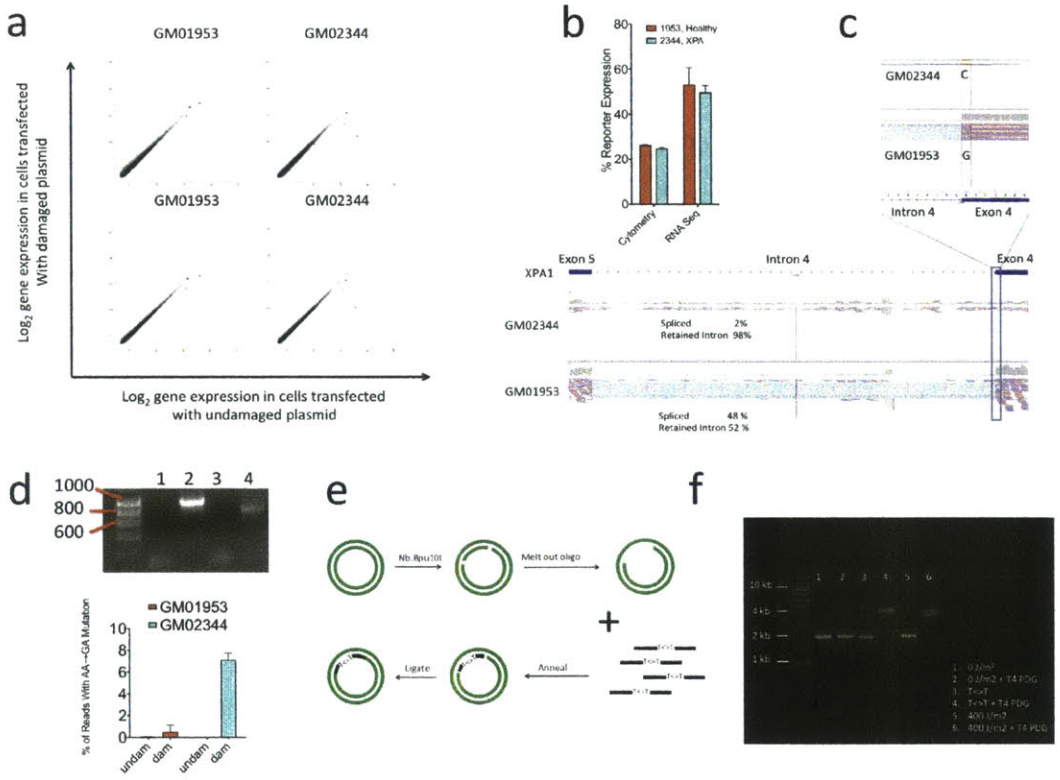


Figure S2.5; related to **Figure 2.6**. RNAseq analysis of transfected cells and construction of a plasmid with a site-specific thymine dimer.

a) Gene expression profile of cells transfected with damaged or undamaged reporter plasmids. Panels a and b represent the two replicate measurements, for the cell lines indicated above each plot. Levels of expression in cells that were transfected with damaged plasmids are plotted on the vertical axis, and expression levels in cells transfected with the undamaged (control) plasmids are plotted on the horizontal. Genes expressed at the same level under both conditions appear on the diagonal, and this is overwhelmingly the case for endogenous transcripts (black and gray circles), indicating no major changes in transcription in cells in response to the presence of damaged plasmid DNA. Reporter transcripts are colored in blue, cyan, orange, green, and magenta. These reporters are seen to be among the most highly expressed in all samples. Reduced expression in the presence of DNA damage (due to transcription blocking lesions) is reflected in these points falling below the diagonal. **b**)

Expression of GFP mRNAfrom reporter plasmid containing a site-specific thymine dimer in the transcribed strand. Expression was assayed by flow cytometry or RNAseq. Error bars represent the standard deviation of two biological replicates. **c**) XPA read coverage and junction reads for GM02344 and GM01953. Exons 4 and 5 are shown. A 555G>C mutation in exon 4 of the XPA gene has previously been reported to induce transcript splicing errors in GM02344 [2]. Reads (indicated as small red and blue bars) are aligned to the region of the genome that encodes the XPA gene. The majority of reads in transcripts from the wild type cell line GM01953 align, as expected, to the exons. Read coverage for GM01953 is overall higher than that for GM02344; XPA expression was ~ 2.5-fold higher in in GM01953. The expected intron-spanning reads (indicated as light gray lines that run between exons) are abundant for GM01953, and ~ 50% of the transcripts were correctly spliced. By contrast, read coverage is lower, intron-spanning reads are nearly absent, and many reads fall within the introns for GM02344. Only ~2% of the XPA transcripts in the mutant cell line were correctly spliced. **d**) Agarose gel (top) and sequencing (bottom) analysis of PCR amplicons generated from reporter cDNAs. In the gel, PCR products were analyzed from reactions where the template was excluded (lane 1), plasmid DNA (lane 2), purified mRNA that was not reverse transcribed (lane 3), or cDNA generated by reverse transcription of mRNA (lane 4). No product was observed in the absence of reverse transcription, and amplicons generated from cDNA templates migrated 100-200 bp below amplicons generated from plasmid DNA templates; this size difference corresponds to a 136 bp intron expected to be absent from cDNA, but retained in plasmid DNA. The frequency of AA→GA base substitutions in transcripts expressed in GM01953 and GM02344 from plasmids containing a site-specific thymine dimer are within experimental error of the frequencies measured for AA→GA substitutions using RNAseq (**Figure 2.6c**). **e**) Preparation of a reporter with a site-specific thymine dimer. The GFP reporter plasmid contains two recognition sites, 18 bp apart, for the strand specific nicking endonuclease *Nb.Bpu10I*. The plasmid is nicked with this enzyme, heated to melt the oligonucleotide away from the plasmid, and then

rapidly cooled to prevent the oligo from re-annealing. The oligo is then removed from the mixture using a Qiagen PCR cleanup kit. Plasmid is then combined with an excess of a 5'-phosphorylated oligonucleotide containing a site specific thymine dimer, heated to 80 C, and cooled slowly to facilitate annealing of the oligonucleotide. Finally, the resulting nicked plasmid is ligated and the desired closed circular DNA is purified from the mixture by gel electrophoresis. **f**) Assay for site-specific incorporation of a cyclobutane pyrimidine dimer into the pmax:GFP plasmid. 500 ng of plasmid was incubated in a 50 µL volume for 16 hours at 37°C with 40 units of the bifunctional thymine dimer specific glycosylase / AP lyase (T4 PDG, New England Biolabs), which nicks DNA that contains thymine dimers. Untreated plasmid (lane 2) is resistant to cleavage, whereas plasmid irradiated at 800 J/m2 is completely digest (lane 6). A second band in lane 6 migrates as linearized DNA (~3.7 kb) and likely reflects nicking at closely opposed DNA lesions. Nearly complete digest of the plasmid in Lane 4 to nicked DNA is consistent with at least 95% incorporation of the lesion.

Supplemental Tables

Table S2.1; related to **Figure 2.2.5**. Samples submitted for next generation sequencing.

8 total samples were submitted for complete RNA-seq using 40bp paired end libraries. These included damaged and undamaged samples for the two cell lines and two replicates for each cell line.

Sample ID	Sample name	Cell line	Replicate
D12-4969	WT undam	GM01953	1
D12-4970	WT dam	GM01953	
D12-4971	XPA mut undam	GM02344	
D12-4972	XPA mut dam	GM02344	
D12-4973	WT undam	GM01953	2
D12-4974	WT dam	GM01953	
D12-4975	XPA mut undam	GM02344	
D12-4976	XPA mut dam	GM02344	

Table S2.2; related to **Figure 2.2.5**. Cufflinks and Tophat parameters.

Tophat parameter	Value	Description
--min-anchor-length	6	TopHat will report junctions spanned by reads with at least this many bases on each side of the junction. Note that individual spliced alignments may span a junction with fewer than this many bases on one side. However, every junction involved in spliced alignments is supported by at least one read with this many bases on each side. This must be at least 3 and the default is 8.
--splice-mismatches	0	The maximum number of mismatches that may appear in the "anchor" region of a spliced alignment. The default is 0.
--min-intron-length	10	minimum intron size allowed in genome
--max-intron-length	1000000	maximum intron size allowed in genome
--min-isoform-fraction	0.0	The minimum frequency of any isoform to consider. The default is 0.15
--max-multihits	20	Instructs TopHat to allow up to this many alignments to the reference for a given read, and suppresses all alignments for reads with more than this many alignments. The default is 20 for read mapping.
--no-novel-juncs	True	Only look for reads across junctions indicated in the supplied GFF or junctions file.
--segment-length	20	Each read is cut up into segments, each at least this long. These segments are mapped independently. The default is 25.
--library-type	fr-unstranded	library prep used for input reads
--solexa1.3-quals	True	As of the Illumina GA pipeline version 1.3, quality scores are encoded in Phred-scaled base-64. Use this option for FASTQ files from pipeline 1.3 or later.
--mate-inner-dist	200	This is the expected (mean) inner distance between mate pairs. For, example, for paired end runs with fragments selected at 300bp, where each end is 50bp, you should set -r to be 200. The default is 50bp.
--mate-std-dev	100	The standard deviation for the distribution on inner distances between mate pairs. The default is 20bp.
Cufflinks parameters	**Value**	**Description**
--min-intron-length	10	minimum intron size allowed in genome
--max-intron-length	1000000	maximum intron size allowed in genome
--min-isoform-fraction	0.0	The minimum frequency of any isoform to consider. The default is 0.15

Table S2.3; related to **Figure 2.2.5**. RPKM values for the five reporter genes across samples.

		RPKM vaues			
Replicate	**Reporter gene**	**XPA mut undam**	**XPA mut dam**	**Norm undam**	**Norm dam**
1	BFP	1737.01	1213.61	1607.05	2076.39
	AmCyan	14975.3	5993.3	13792	16456.5
	GFP_615	1434.24	463.966	1251.22	984.1
	mOrange	2192.32	239.698	1853.39	2164.26
	mPlum	4257.75	162.343	3723.82	2928.89
2	BFP	1652.01	1096.74	1467.33	1608.85
	AmCyan	18999.2	4596.23	15349	16521.4
	GFP_615	1283.83	447.682	1314.83	652.47
	mOrange	1998.8	193.741	2119.65	1532.83
	mPlum	3706.3	108.021	4299.3	2027.84

Table S2.4; related to **Figure 2.2.5**. Read counts for RNA-seq samples and numbers of aligned reads using TopHat.

Sample names			Norm undam	Norm dam	XPA mut undam	XPA mut dam	Norm undam	Norm dam	XPA mut undam	XPA mut dam
Replicate			1				2			
Sample ID			D12-4969	D12-4970	D12-4971	D12-4972	D12-4973	D12-4974	D12-4975	D12-4976
Total sequences			46727196	42643848	42485978	41062392	49343810	45472418	49352500	41193160
Total mapped reads			55847402	50879779	52191219	47970477	57600436	54420536	58422772	48971710
Mapped properly	Total		41264548	37791752	38687218	35630500	42814870	40359416	43174990	35851498
	BFP		23230	26733	22671	11979	21405	19408	23312	13147
	AmCyan		245632	265119	232280	78184	277105	275550	357715	71185
	GFP_615		16261	12068	17476	3728	17887	7859	16816	5011
	mOrange		21258	22148	24496	1596	26262	13090	23339	1978
	mPlum		35921	29101	41173	1194	44851	17103	38084	1531
	Plasmids		342302	355169	338096	96681	387510	333010	459266	92852
	Other genes		40922246	37436583	38349122	35533819	42427360	40026406	42715724	35758646
Unmapped reads			245774	243730	228043	192470	337249	241637	215238	164415
Pair not mapped properly	Total		14582854	13088027	13504001	12339977	14785566	14061120	15247782	13120212
	Read mapped, mate unmapped	Other genes	1449291	1385025	1359536	1159168	1606305	1427621	1463587	1257407
		Plasmids	98603	75626	79587	57255	94179	71455	100645	20241
	Read and mate mapped, insert size too large	Other genes	7914228	6810016	7155202	7004002	7688172	7239820	8356254	7280056
		Plasmids	718	736	818	676	720	564	802	276
	Read and mate mapped to different chrom	Other genes	5118728	4815446	4907660	4117100	5395452	5321062	5325798	4561628
		Plasmids	1286	1178	1198	1776	738	598	696	604

Table S2.5; related to **Figure 2.2.5**. Genes with log2 fold change >= 1 when comparing cells transfected with undamaged plasmid to those transfected with damaged plasmids.

Sample	Replicate	Gene Name	Chr	Bp	Log2 Fold Change
GM02344	1	GFP_615	GFP_615	0-951	-1.62
		AmCyan	AmCyan	0-921	-1.32
		MALAT1	chr11	65265232-65273939	-1.29
		SCARNA9	chr11	93454679-93455032	-1.06
		RPL21	chr13	27825691-27830702	-1.38
		NDUFA3	chr19	54606159-54610281	-1.13
		INO80B-WBP1	chr2	74682149-74688018	-1.00
		HIST1H4H	chr6	26285353-26285727	-1.76
		RN5-8S1	chrUn_gl000220	112024-112180	-4.07
		RN5-8S1	chrUn_gl000220	155996-156152	-4.07
		mOrange	mOrange	0-942	-3.19
		mPlum	mPlum	0-912	-4.70
	2	GFP_615	GFP_615	0-951	-1.52
		AmCyan	AmCyan	0-921	-2.04
		mOrange	mOrange	0-942	-3.37
GM01953	1	mPlum	mPlum	0-912	-5.10
		C1orf86	chr1	2115898-2139172	1.02
		NUBP2	chr16	1832932-1839192	1.05
		ATP5D	chr19	1241748-1244824	1.00
		FAM108A1	chr19	1876974-1885518	1.05
		TMEM160	chr19	47549166-47551882	1.05
		SCAND1	chr20	34541538-34543281	1.34
		C4orf48	chr4	2043719-2045697	1.42
		TMUB1	chr7	150778171-150780620	1.00
		C9orf16	chr9	130922538-130926207	1.02
		FBXW5	chr9	139834884-139839206	1.01
		BCYRN1	chrX	70430034-70948962	1.41
	2	RPL21	chr13	27825691-27830702	2.53
		SMN1	chr5	69345349-69373418	1.50
		SMN1	chr5	70220767-70248838	1.43
		RN5-8S1	chrUn_gl000220	112024-112180	1.92
		RN5-8S1	chrUn_gl000220	155996-156152	1.92

			20		
		GFP_615	GFP_615	0-951	-1.01
		C16orf13	chr16	684428-686347	-1.01
		ATP5D	chr19	1241748-1244824	-1.07
		LOC100129250	chr9	32551141-32553015	-1.00
		mPlum	mPlum	0-912	-1.08

Supplemental Experimental Procedures

Plasmids

The AmCyan, EGFP, mOrange, and mPlum and tagBFP reporter genes were subcloned into the pmax cloning vector (Lonza) between the *KpnI* and *SacI* restriction sites in the multiple cloning site. The Kozak translation initiation consensus sequence and an additional *NheI* restriction site were introduced at the 5' end of each reporter, and a *HindIII* restriction site was added to the 3' end. The pmax cloning vector places reporter genes under the CMV Intermediate-Early promoter. Plasmids were amplified using *E. coli* DH5α (Invitrogen), and purified using Qiagen endotoxin-free maxi and giga kits. Constructs were confirmed by DNA sequencing and restriction enzyme digestion.

UV-Irradiated Substrates

Plasmids were irradiated in TE buffer (10 mM Tris-HCl, 1 mM EDTA, pH 7.0) at a DNA concentration of 50 ng/μL in a volume of 1.5 mL in 10 cm polystyrene petri dishes (without lids) with UVC light generated by a Stratalinker 2000 box. Following treatment, reporter plasmids were combined in the following ratio: 1 part tagBFP, 10 parts AmCyan, 1 part GFP, 2 parts mOrange, and 4 parts mPlum; these proportions were used to compensate for weaker fluorescence intensities observed for some of the reporters. The same plasmid mixture without UV irradiation was prepared as described above, except without UV-treatment. Data obtained following transfections with the plasmid mixture containing irradiated plasmids have been labeled "damaged" and those from untreated plasmids have been labeled "undamaged"; however every transfection included an undamaged reporter. The undamaged transfection reporter was used to normalize transfection efficiency. Further details regarding the UVC dose delivered to each plasmid are available in **Table 2.2**. Plasmid mixtures were ethanol precipitated, washed with 70% ethanol, and redissolved in TE

buffer at ~ 1.5 µg/µL; damaged and undamaged plasmid mixtures were adjusted to the same final concentration, confirmed using a Nanodrop spectrophotometer.

Substrates containing a G:G mismatch

Substrates were prepared using a method based on a previously published protocol [3]. The pmax:mOrange plasmid was nicked with *Nb.BstI* (New England Biolabs) to generate a single strand break in the transcribed strand (**Figure S2.2b**). The nicked strand was then digested with exonuclease III, and the remaining single stranded circular DNA (ssDNA) purified using a 1% agarose gel. 20 µg of the ssDNA was combined with 40 µg of G299C mutant pmax:mOrange plasmid linearized with *NheI* (New England Biolabs); the mixture was denatured by addition of 0.3N sodium hydroxide, and then returned to neutral pH to facilitate annealing between wild type ssDNA and the complementary strand of the linearized mutant sequence to yield a heteroduplex containing a G:G mismatch at position 299 of the mOrange gene. Subsequently, reactions were cleaned up using a Qiagen PCR cleanup kit, and unwanted linear and single stranded DNA side products were digested with Plasmid Safe ATP dependent DNAse (Epicentre). Nicked plasmid was purified using a 1% agarose gel, and ligated using 800 units T4 DNA ligase (New England Biolabs). Finally, a second gel purification was performed to isolate closed circular products (**Figure S2.2c**). Homoduplex DNA was prepared using the same procedure, except that linearized DNA was prepared from the wild type pmax:mOrange.

Substrates containing a site-specific O^6-MeG

A nonfluorescent variant of the pmax:mPlum reporter (T208C) was identified. This construct was further modified with a mutation that does not change the encoded protein sequence (C207G) to generate a unique recognition site (GGGCCC) for the restriction enzyme *PspOMI* (New England Biolabs). Substrates were prepared based on a previously described method [3], with

minor modifications (**Figure S2.2e**). Single stranded DNA was prepared as described above. 5 picomoles of a phosphorylated O^6-MeG containing oligonucleotide (5'-CACGTAGGCCTTGGXCCCGTACATGATCTG-3', where X = O^6-MeG) was combined with 2.5 µg of single stranded pmax:mPlum:C207G:T208C plasmid DNA in 1X *Pfu* polymerase buffer (Agilent Biotechnologies) in a total volume of 50 µL. The mixture was heated to 85°C in a thermal cycler for 6', and then allowed to anneal by cooling to 40°C at 1°C per minute. To extend the primer, 2.5 units *Pfu* polymerase (Agilent) and 0.2 µM dNTPs were added and the reaction, and then incubated for 1 hour at 68°C. The reaction was then cooled to 37°C, and supplemented with an additional 0.5 µM dNTPs, 1 mM ATP, 1.5 units T4 DNA polymerase (New England Biolabs), and 40 units T4 DNA ligase, and incubated for an additional hour at 37°C to yield closed circular plasmid. Finally, the product was purified from a 1% agarose gel using a Qiagen gel extraction kit. A homoduplex control plasmid that expresses the wild type mPlum fluorescent reporter protein was prepared using identical conditions with single stranded pmax:mPlum:C207G and the following oligonucleotide: 5'-CACGTAGGCCTTGGACCCGTACATGATCTG-3'. The plasmid containing O^6-MeG was resistant to cleaveage by the restriction enzyme *PspOMI*, whereas the lesion free homoduplex generated with the same ssDNA was readily digested under the same conditions. Treatment of the O^6-MeG containing plasmid with human methylguanine methyltransferase (hMGMT) (Alexis Biochemicals) resulted in >95% *PspOMI* cleavable material (**Figure S2.2f**).

Substrates containing a site-specific 8-oxoG

A non-fluorescent variant of mOrange (A215C) that lacks a critical tyrosine that forms the chromophore was identified. The plasmid encoding this non-fluorescent protein was used to prepare ssDNA using the same protocol described above for preparation of the O^6-MeG containing plasmid. The following 8-oxoG containing oligonucleotide was annealed and extended using the same

conditions that were used to prepare the O^6-MeG containing plasmid: 5'-GTAGGCCTTGGAGCCGXAGGTGAACTGAGG-3', where X represents the 8-oxoG lesion. An mOrange homoduplex was prepared using the following oligo with ssDNA prepared from the wild type mOrange plasmid: 5'-GTAGGCCTTGGAGCCGTAGGTGAACTGAGG-3'.

Substrates containing an A:C mismatch

Substrates were prepared using a method similar that described above for preparation of the O^6-MeG and 8-oxoG containing plasmids (**Figure S2.3a**). A non-fluorescent GFP variant (C289T) was identified. The protein expressed from this construct lacks a conserved arginine that is required for chromophore maturation. Single stranded DNA was prepared from this plasmid as described above, except the nicking enzyme was *Nt.BspQI*, which nicks the non-transcribed strand. Thus, after *ExoIII* digest, the remaining ssDNA comes from the transcribed strand. The following 5'-phosphrylated oligonucleotide was annealed to the ssDNA: 5'-P-GGCTACGTCCAGGAGCGCACCATCTTCTTC-3'. Primer extension was carried out as described above, except the extension temperature was lowered to 61°C to increase yield. The resulting substrate is a heteroduplex in which the transcribed strand has the mutant sequence (A) and the non-transcribed strand has the wild type sequence (C) at position 289. Mismatch repair activity restores the wild type sequence to the transcribed strand and leads to GFP expression. A wild type homoduplex was prepared identically using wild type plasmid DNA as the starting material, and the substrates were validated using the MT1 and TK6 cell lines (**Figure S2.3b**).

Substrates containing a blunt-end double strand break

A unique *ScaI* restriction site was inserted into the 5' untranslated region of the pmax:BFP reporter plasmid, immediately 5' of the reporter gene (**Figure**

S2.3c). Plasmids were linearized with *ScaI* restriction enzyme, purified by phenol/chloroform extraction and ethanol precipitation, and digest completeness was confirmed by gel electrophoresis. The uncut pmax:BFP reporter with the *ScaI* restriction site was used as the undamaged control in experiments measuring repair of the linearized reporter.

Substrates and methods for measuring homologous recombination

GFP-based HR reporter plasmids have been described previously [4], and were a generous gift from Prof. Bevin Engelward. The D5G plasmid expresses a non-fluorescent GFP reporter that is truncated at the 5' end of the gene (**Fig S4a**). This plasmid was modified to include a StuI restriction site so that a blunt-end double strand break could be introduced into the plasmid with the goal of reducing the likelihood of unwanted re-ligation of the plasmid, because this process does not yield fluorescent signal. Cells were transfected with 0.5 micrograms of StuI-linearized D5G, 5 micrograms of D3G, plus 0.5 micrograms of an undamaged plasmid that was used as a control for transfection efficiency. Homology directed repair of the DSB in the D5G reporter plasmid that uses the D3G plasmid as a donor sequence leads to a full length GFP-encoding gene, and results GFP expression.

Substrates containing a site-specific thymine dimer

A site-specific thymine dimer spanning positions 614-615 of the GFP sequence was successfully introduced into the transcribed strand of the pmax GFP reporter plasmid using previously described methods [5] (**Figure S2.5e**). Briefly, two nicking sites for the enzyme *Nb.Bpu10I* (Thermo Scientific) near the 3' end of the GFP reporter gene were used to excise a single stranded oligonucleotide of 18 bp in length: 5'-TCAGGGCGGATTGGGTGC-3'. The nicking sites flank a silent mutation that was introduced to generate a TpT sequence in the transcribed strand of the plasmid. A synthetic oligonucleotide 5'-

TCAGGGCGGAT<>TGGGTGC-3' containing a thymine-thymine cis-syn cyclobutane dimer indicated by T<>T (synthesized by TriLink BioTechnologies using a cis-syn thymine dimer phosphoramidite (Glenn Research)) was annealed and ligated into the gapped plasmid. Incorporation of the site-specific thymine dimer in the plasmid was confirmed by endonucleolytic digestion with thymine dimer specific glycosylase / AP lyase (T4 PDG, New England Biolabs). Greater than 95% of the resulting product migrated as nicked plasmid DNA (**Figure S2.5f**), indicating at least 95% of the plasmids contained the lesion.

Isolation of total RNA for mRNAseq

At 18 hours, transfected cells were harvested by centrifugation, washed three times with PBS, and resuspended in 1 mL Trizol reagent. The suspension was extracted with 200 µL chloroform. The aqueous phase was removed, combined with one volume of absolute ethanol, and applied to a Qiagen RNeasy mini-prep spin column. The column was then washed two times with 500 µL buffer RPE (Qiagen), and finally eluted in 40 µL diethylpyrocarbonate (DEPC) treated water. From this point forward, RNA was handled in Eppendorf DNA LoBind tubes to minimize loss of material. The quality of the RNA preparation was determined using a bioanalyzer to confirm a RIN of at least 9.0. 1 µg of total RNA was stored in TE Buffer at -80°C until submission for mRNAseq.

Isolation of mRNA and synthesis of cDNA

mRNA was isolated using a Qiagen Oligotex kit, using the manufacturer's protocol, but substituting Eppendorf DNA LoBind tubes for those provided with the kit. In the final step, mRNA was eluted in 20 uL buffer OEB preheated to 70 °C. 5 µL of the eluate was transferred to a LoBind tube, combined with 1 µL of DNAse buffer and 1 unit of DNAseI (Invitrogen). The mixture was brought up to a 10 µL volume with DEPC treated water, and incubated for 15 minutes at room temperature. DNAse was inactivated by addition of 1 µL of 25 mM EDTA,

followed by incubation at 65 °C for 10 minutes. A cocktail comprised of 2X RT buffer (Qiagen), oligo-dT(12-18) (125 ng/uL; invitrogen), 4 units of RNAse inhibitor (Qiagen), 5 mM dNTPs, and 4 units of reverse transcriptase (Omniscript; Qiagen) was prepared, and 8 µL added to the DNAse digest. The reaction was incubated for 1 hour at 37 °C. No-RT controls were performed identically, except for the exclusion of the reverse transcriptase.

Specific amplification of reporter cDNA by PCR

cDNA samples were amplified with primers specific to the 3' and 5' UTR regions of the pMax vector. The following primers were synthesized for specific amplification of reporter cDNA:

5UTR: 5'- TTG CTA ACG CAG TCA GTG CT -3'
3UTR: 5'- GCA TTC TAG TTG TGG TTT GTC C -3'

1.5 µL of cDNA was PCR amplified in a 25 µL reaction volume with 1X PCR buffer (Denville), 0.5 µM primers, 0.2 mM dNTPs, and 1 unit Taq polymerase (Denville). Specific amplification was confirmed by gel electrophoresis and analysis on a bioanalyzer chip. Water and EGFP encoded plasmids were used as negative and positive controls, respectively. Finally, reactions were cleaned up using a Qiagen PCR cleanup kit according to the manufacturer's protocol, and eluted in 50 uL of TE.

Fragmentation of DNA

250 ng of PCR product was diluted to a total volume of 130 uL in TE buffer. The DNA was fragmented in a Covaris microTUBE using a Covaris S2 sonicator (Duty Cycle 10%, Intensity 5, 200 cycles per burst, 180 seconds. Fragmentation to a target base pair peak of 150 bp was checked using a Agilent BioAnalyzer.

Validation of MGMT FM-HCR

Total RNA was isolated as described immediately above for mRNAseq from the panel of cell lines for which data are presented in **Figure 2.4a**. Poly-dT oligonucleotides were used to generate cDNA, and MGMT was quantified by TaqMan qPCR as described previously [6]. MGMT transcript levels were quantitated using the DDCT method, and GAPDH was used as the internal control. Transcript levels are reported in Figure 2.4b relative to transcript levels in the cell line TK6 +MGMT, which overexpresses MGMT and was used a positive control.

Supplementary Note

In this manuscript we have presented a proof of concept for high throughput DNA repair capacity assays. Because our experiments required only a small fraction of the theoretical maximum throughput achievable by HCR-seq, they were relatively expensive to perform. In what follows, we describe how the cost of performing HCR-seq on a per sample basis decreases as more samples are added to a given experiment.

Whereas FM-HCR allowed for the simultaneous detection of 5 independent repair reporters, the HCR-seq permitted the measurement of 20 reporters (5 reporter genes x 4 bar-codes) in a single experiment. The 20 RNAseq reporters were detected at sufficient coverage to obtain highly reproducible dose response curves (**Figure 2.5**). Because these transcripts represented less than 1% of the total mapped reads, it can be estimated that at least 2000 reporters (or 200 dose-response curves) could be independently assayed on a single lane if host transcripts were excluded from the assay.

The four dose response curves derived from sequencing data and presented in **Figure 5a** were acquired at a cost of approximately $800 per curve. However, several considerations would reduce the cost of sequencing-based assays if deployed in large-scale population studies. As cost of sequencing

continues to fall, and particularly if a large number of samples is multiplexed on single lane, sample preparation can be expected to dominate the cost of the assay, with sequencing accounting for a small fraction of the overall cost. In the present work bar-codes were introduced as part of the Illumina sample preparation pipeline, however an equivalent means of distinguishing among samples would be to introduce bar-codes into the library of reporter plasmids. This configuration would permit sample pooling prior to sequencing library preparation, leaving the cost of cell culture and transfection reagents as the major contribution to the remaining cost of the assay. Such a workflow would also essentially limit the active time in the research laboratory to that required to handle and transfect the cells of interest. Furthermore, selective amplification of reporter transcripts could be used to exclude host transcripts, which represented the overwhelming majority of mapped sequences in the present work.

Supplemental References

1. Wu, R.S., S. Hurstcalderone, and K.W. Kohn, *Measurement of O-6-alkylguanine DNA alkyltransferase activity in human cells and tumor tissues by restriction endonuclease inhibition.* Cancer Research, 1987. **47**(23): p. 6229-6235.
2. Satokata, I., K. Tanaka, S. Yuba, and Y. Okada, *Identification of splicing mutations of the last nucleotides of exons, a nonsense mutation, and a missense mutation of the XPAC gene as causes of group A xeroderma pigmentosum.* Mutation Research, 1992. **273**(2): p. 203-212.
3. Baerenfaller, K., F. Fischer, and J. Jiricny, *Characterization of the "mismatch repairosome" and its role in the processing of modified nucleosides in vitro*, in *DNA Repair, Pt A*, J. Campbell and P. Modrich, Editors. 2006. p. 285-303.
4. Kiziltepe, T., A. Yan, M. Dong, V.S. Jonnalagadda, P.C. Dedon, and B.P. Engelward, *Delineation of the chemical pathways underlying nitric oxide-induced homologous recombination in mammalian cells.* Chemistry & Biology, 2005. **12**(3): p. 357-369.
5. Kitsera, N., D. Stathis, B. Luhnsdorf, H. Muller, T. Carell, B. Epe, and A. Khobta, *8-Oxo-7,8-dihydroguanine in DNA does not constitute a barrier to transcription, but is converted into transcription-blocking damage by OGG1.* Nucleic Acids Research, 2011. **39**(14): p. 5926-5934.
6. Kitange, G.J., B.L. Carlson, M.A. Schroeder, P.T. Grogan, J.D. Lamont, P.A. Decker, W.T. Wu, C.D. James, and J.N. Sarkaria, *Induction of MGMT expression is associated with temozolomide resistance in glioblastoma xenografts.* Neuro-Oncology, 2009. **11**(3): p. 281-291.

Chapter III: Assessing BER Capacity in the human population

Isaac A. Chaim and Zachary D. Nagel

An edited version of this chapter will be published as:

Nagel, Z. D. and Chaim, I. A. "Assesing BER Capacity in the Human Population" In: Wilson III, D.M. The Base Excision Repair Pathway, Molecular Mechanisms and Role in Disease Development and Therapeutic Strategies. (To be published in 2016, *World Scientific Publishing Co*)

Table of Contents

Decades of biochemical, structural and mechanistic investigations of human DNA repair pathways and advances in technologies for measuring their function in human samples have recently set the stage for studies focused on inter-individual variation in DNA repair capacity, and its role in disease etiology, reviewed in [1]. Extreme deficiencies in the activity of glycosylases that initiate base excision repair (BER) are associated with cancer and immunodeficiency in humans [2, 3], and inefficient repair intermediate processing has been associated with neurological disorders in humans and cancer and immune disorders in animal models [4]. This Chapter reviews studies that have explored small variations in BER capacity across human populations and the relationships between these variations and altered disease susceptibility. We review the methods that are available for measuring BER capacity in human populations (Section 1), and the data that have emerged from studies conducted so far (Section 2). In Section 3, we discuss both the methods and the results of studies in the broader context of the field of DNA repair at the population level. Finally, in Section 4, we highlight research avenues that can be pursued in the near term to advance the goal of using functional assessments of BER capacity in human populations to promote personalized disease treatment and prevention.

1. Methods for Measuring BER Capacity in Populations

An ideal quantitative method for measuring BER capacity in human populations would assess repair under physiological conditions and satisfy two interrelated goals: a) Sufficiently high throughput and precision to measure BER in a large number of samples and identify differences between individuals with aberrant BER. b) Sufficiently high resolution to understand the mechanistic basis for inter-individual differences in BER capacity. These goals are motivated by the anticipated downstream translational applications of BER assays, which include potential preventive and therapeutic interventions for diseases associated with aberrant BER.

In practice, population studies have so far made use of two broadly defined approaches for measuring BER. According to the first approach, assays measure the efficiency of a specific BER step(s) or the levels of repair intermediates (e.g. [5]). In the second approach, assays measure the integrated efficiency of all pathway steps by detecting repair products (e.g. [6]). As discussed below, some methods encompass both approaches. For a basic overview of the BER process, and the enzymatic steps involved, please see the book Introduction.

Current methods represent an approximation of physiological repair because of the need to work in isolated cells or cell lysates, or the need to use exogenous DNA-damaging agents to produce DNA damage levels in excess of what is normally encountered by cells. As such, no single assay developed so far approaches the ideal described above. However, in the last 10-15 years, several novel high-throughput methodologies for measuring BER capacity have emerged, creating an opportunity to carry out functional characterizations of BER in populations that were previously unfeasible. Here, we highlight the major features of 5 types of assay that are well-suited for measuring BER capacity in human populations.

1.1 Oligonucleotide-Based Assays

The key principle underlying oligonucleotide-based assays is that BER processing of a damage-containing DNA fragment changes the fragment's electrophoretic mobility in a denaturing polyacrylamide gel (**Figure 3.1**). These assays hold the advantage of measuring the activity of a specific step or steps of BER using chemically defined substrates. A single-stranded oligonucleotide containing a BER substrate lesion or BER intermediate is synthesized, end-labeled with ^{32}P or a fluorescent dye, and annealed to a complementary unlabeled strand. The resulting double-stranded DNA is incubated *in vitro* with a protein extract. Bifunctional glycosylases or apurinic/apyrimidinic endonuclease 1 (APE1) directly incise at the lesion (damaged bases or AP sites, respectively),

yielding a DNA fragment that is shorter than the starting material and migrates more rapidly in a gel. For monofunctional glycosylases, which lack the AP lyase activity found in bifunctional enzymes, a subsequent alkali treatment is needed to cleave AP sites formed by the repair protein. Similarly, substrates that measure the final ligation step of BER yield products that are longer than the starting material and migrate more slowly in a gel. Repair efficiency is calculated from the relative amounts of substrate versus product.

Activity for any DNA glycosylase can be measured directly by this method, provided a site-specific damaged base that represents a substrate for the appropriate glycosylase can be incorporated into an oligonucleotide. Assays for the 11 known-to-date human DNA glycosylases have been developed, specifically: for alkyladenine glycosylase (AAG, aka MPG) and 8-oxoguanine glycosylase (OGG1) ([7] and [8], respectively); MutY homolog (MUTYH) ([9] and [10]); endonuclease VIII-like 1 and 2 (NEIL1, NEIL2) and endonuclease III homolog-like 1 (NTHL1) [11]; NEIL3 [12]; uracil N-glycosylase (UNG) and single-strand-selective monofunctional uracil DNA glycosylase (SMUG1) [13]; methyl-CpG binding domain protein 4 (MBD4) and thymine DNA glycosylase (TDG) [14]. As mentioned above, by the use of an AP site (or the more chemically stable AP site analog, tetrahydrofuran, THF), the activity of APE1 can be quantitated as well [15]. Other downstream steps such as gap-filling, flap endonuclease and ligase activity have been measured using a similar approach [16]. Although repair activities determined using oligonucleotide-based assays are usually attributed to a single enzyme (i.e. a specific DNA glycosylase or APE1), each oligonucleotide actually reports on the overall cellular extract capacity to repair a particular *lesion*. This repair can involve either several enzymes from the BER pathway (as is the case for glycosylases with overlapping substrate specificities) or even other repair pathways. For example, while OGG1 has been shown to account for 90% of 8-oxoguanine (8-oxoG) repair, both *in vitro* and in human cell lines, this lesion can also be a substrate for the NEIL1 and NEIL2 DNA glycosylases and to a lesser degree AAG in murine systems [11, 17-19]. Uracil in DNA can be excised by UNG, SMUG1, MBD4 and TDG, and is recognized by

the mismatch repair machinery [13, 20]. A major practical advantage of oligonucleotide-based BER assays for use in population studies is that they do not require live cells, which are often not available due to the time and cost requirements associated with their storage and maintenance.

1.2 Molecular Beacons

The key principle underlying molecular beacon assays is that enzymatic cleavage of a damage-containing DNA fragment leads to increased fluorescent signal by separating a fluorophore from a quencher (**Figure 3.2**). Like the oligonucleotide-based assays, molecular beacons report the activity of a specific step(s) (glycosylase and AP site incision) in the BER pathway, but they hold the advantage of being compatible with both extracts and in cells, and do not require radioactive labels. Molecular beacons are constructed from single-stranded oligonucleotides designed to form a stable hairpin and to hold a fluorophore and a quencher in close proximity. Excision of a damaged base positioned near the 3'- or 5'-end, followed by incision at the AP site breaks the covalent linkage that keeps together the fluorophore and quencher, allowing them to diffuse apart and leading to an increase in fluorescence. Most beacons make use of synthetic fluorophores and quenchers; however, in one case, the DNA lesion itself has been used as a quencher to generate a direct reporter of 8-oxoG excision [21]. A similar probe, but for uracil repair, has also been developed and tested both in cell lysates and in living cells by the same group [22]. Although considerable effort can be required to optimize molecular beacons for use in BER assays, once designed they are easily prepared; a comprehensive protocol for designing and building *in vitro* molecular beacons for BER is described by [23].

Molecular beacons have already found diverse applications. They have been used in cell lysates to calculate accurate kinetic parameters (k_{cat}/K_M) for UNG and APE1 that have been validated against oligonucleotide-based assays [24]. They have also been used in HeLa and NIH3T3 cells to measure AAG (hypoxanthine, Hx), MUTYH (A:8-oxoG), UNG (Uracil, U) and APE1 (THF)

activities, and to monitor BER in real time [24]. This work has been extended to include a molecular beacon for OGG1 that measures activity in both the nuclear and mitochondrial compartments [25]. The molecular versatility and potential for high-throughput applications makes molecular beacons especially well-suited for use in future population studies.

1.3 Microchip-Immobilized Fluorescent Reporter Systems

The key principle underlying microchip-based BER assays is that cleavage of a fluorescent end-labeled oligonucleotide probe that is covalently attached to a solid support releases a fluorophore that is then washed away (**Figure 3.3**) [26]. Probes that contain site-specific oxidized, deaminated, or alkylated bases are chemically immobilized in a spatially addressed manner on the surface of each well of a 96-well-plate. The immobilized probes are incubated with purified proteins or cell extracts, treated with Endo V or Endo VIII to incise glycosylase-generated AP sites (required when measuring monofunctional glycosylase activity), and finally washed to remove cleaved fluorophore. The resulting decrease in fluorescence is used as a quantitative measure of enzymatic activity. Importantly, the high-throughput microchip assay has been validated against traditional ^{32}P-end-labeled oligonucleotide assays, and yielded similar estimates of glycosylase activity in HeLa extracts for substrates containing 5-(hydroxymethyl)uracil, 5-formyluracil, 5-carboxyuracil, formylamine and uracil [27]. The 96-well format and plate reader detection, as well as the ability to measure repair of specific repair steps are the major strengths of this *in vitro* approach for use in population studies.

1.4 Plasmid-Based Reporter Systems for use in Cell Lysates

The key principle underlying *in vitro* plasmid-based assays for BER capacity is that processing of damage-containing plasmids by BER enzymes changes the physical properties of the plasmid; these may include a change in

electrophoretic mobility in a gel (**Figure 3.4**), a change in radioactivity, or a change in fluorescence properties imparted to the DNA by incorporation of fluorescent nucleotides. Depending on how they are configured, plasmid-based assays can measure the activity of specific BER steps, or complete repair, and they may include either chemically defined site-specific DNA lesions, or randomly induced mixtures of DNA lesions produced by exposure of the plasmids to a DNA-damaging agent.

Both electrophoretic- and fluorescence-based assays for repair of plasmid DNA in cell lysates have been used in population studies. An electrophoretic assay that involves quantifying the fraction of plasmid in which supercoiling is relaxed by the nicking action of APE1 has been used in peripheral blood mononuclear cell (PBMC) extracts [28]. An alternative approach, which has been generalized to several different types of DNA damage, measures the incorporation of fluorescently labeled nucleotides into plasmid DNA during repair synthesis [29]. Plasmids are treated *in vitro* with one of several DNA-damaging agents that induce a spectrum of DNA lesions including 8-oxoG, thymine glycol (Tg), ethenoguanine (εG) and AP sites. Treated plasmids containing different types of damage are then immobilized on a hydrogel surface to create spatially addressed microchips. The resulting microarrays are incubated with cell extracts in the presence of fluorescently labeled nucleotides, and the rate of BER synthesis is determined for each type of DNA damage by quantitating total spot fluorescence intensity (**Figure 3.5**). The ability to generalize the fluorescent nucleotide incorporation approach to different DNA-damaging agents combined with fluorescence detection in a microarray format makes this a versatile, high-throughput approach for measuring repair in large numbers of samples.

1.5 Plasmid-Based Reporter Systems for use in Cells (Host Cell Reactivation)

The key principle underlying plasmid-based BER assays for use in cells is that the introduced DNA lesion or the action of repair enzymes and proteins on a

transiently transfected damaged plasmid alters the efficiency and/or the fidelity of transcription of a reporter gene, which encodes a reporter protein that is not normally expressed in the cells of interest. Chloramphenicol acetyltransferase, luciferase and fluorescent proteins have all been used as reporters. Traditionally, these assays have measured the ability of transiently transfected "host cells" to repair randomly induced transcription blocking lesions from plasmids, thereby "reactivating" expression of the reporter gene, hence the name Host Cell Reactivation (HCR). Damaging agents including hydrogen peroxide (H_2O_2) and methylene blue plus visible light (VL) have been used to generate BER substrates that include transcription blocking DNA lesions [30-32]. A fluorescent reporter for APE1-dependent repair of the transcription blocking abasic site analog tetrahydrofuran has also recently been developed (Chaim *et al.,* in preparation). Since both the initial damage (a mixture of DNA lesions) and the intermediates of repair (abasic sites and nicks) block transcription, these assays measure completion of BER, including AP site incision, gap-filling and ligation. More recently the HCR assay has been adapted to include chemically defined site-specific lesions that report on the activity of specific steps of the BER pathway.

Many DNA lesions that are BER substrates do not block transcription, presenting a challenge to measuring their repair by the canonical HCR methodology. One strategy for overcoming this limitation of HCR assays has made use of the coding properties of a florescent reporter plasmid to study a panel of human MUTYH variants expressed in *Mutyh*$^{-/-}$ mouse embryonic fibroblasts [33]. The reporter consisted of both an unmodified dsRed reporter gene and a modified green fluorescent protein (*GFP*) gene in one plasmid. A stop codon that prevents expression of full-length GFP was engineered into the *GFP* gene. Furthermore, a site-specific 8-oxoG was incorporated opposite an adenine in the stop codon. MUTYH-dependent excision of adenine opposite 8-oxoG, followed by incorporation of cytosine opposite 8-oxoG during repair synthesis results in a change of sequence that abrogates the stop codon, and restores expression of full-length GFP.

A modified fluorescence-based multiplex HCR assay (FM-HCR) was recently developed [34], and presents a general strategy for measuring repair of both transcription blocking and transcriptionally bypassed DNA lesions (**Figure 3.6**). Transcriptional mutagenesis is the process by which RNA polymerase incorporates an "incorrect" nucleotide during transcriptional bypass of a DNA lesion. Fluorescent reporter genes can be site-specifically mutated in a manner that results in expression of a non-fluorescent protein. When a DNA lesion is positioned at the point of the mutation in the transcribed strand of the reporter, transcriptional mutagenesis can lead to restored fluorescent protein expression [35]. Based on this principle, a fluorescent reporter for misincorporation of adenine opposite a site-specific 8-oxoG lesion has recently been developed and validated as a reporter for repair of 8-oxoG in mammalian cells [34]. Excision of 8-oxoG eliminates the source of transcriptional mutagenesis, leading to expression of a non-fluorescent protein. Unlike canonical HCR assays, for which higher levels of reporter expression represent higher repair capacity, for the fluorescent 8-oxoG HCR assay, high levels of fluorescence represent low levels of DNA glycosylase activity. The same group has developed additional reporters that use transcriptional mutagenesis to report excision of uracil, hypoxanthine, and undamaged adenine opposite 8-oxoG (Chaim *et al.,* in preparation). Importantly, each site-specific lesion has been positioned in a reporter with a different fluorescence emission spectrum, allowing for simultaneous, flow cytometric quantitation of all the repair activities. Importantly, and in contrast to canonical HCR assays that measure completion of BER, these assays specifically measure the activity of the DNA glycosylase-catalyzed excision step. HCR assays hold the major advantage of reporting on repair in living cells, and can be used either to measure repair completion or the activity of specific steps in the BER pathway. The recent development of FM-HCR assays to measure repair capacity for multiple BER steps or multiple DNA repair pathways in a single experiment using a fluorescent readout can be adapted to 96-well format for high-throughput analyses.

1.6 Comet Assays

The key principle underlying the comet assay with respect to BER is that enzymatic processing of damaged genomic DNA alters its electrophoretic mobility in a gel (**Figure 3.7**). Comet assays make use of single cell electrophoresis, which consists of embedding single cells in low melting point agarose on a microscope slide, and lysing cells to permit migration of nuclear DNA during electrophoresis. An important limitation of the comet assay is that DNA damage can only be detected directly if it comes in the form of a single- or double-strand break (SSB and DSB, respectively); any other kind of DNA damage has to be converted to a strand break for its detection. As such, if the DNA presents no damage, if damage is present but hasn't been converted to a SSB or DSB, or if repair has already been completed by the time of analysis, little or no DNA migration is detected beyond the margins of the gel-embedded nucleus (the comet head). By contrast, damaged DNA in the form of SSBs and DSBs - including incomplete repair intermediates - migrates faster through the gel due to relaxation of supercoiling in the DNA, and appears as a tail in the gel. A ratio of the amount of DNA in the tail and the head, the length of the tail itself, or a combination of both (known as the Olive moment) is used to quantify DNA damage in comet assays. Intact cells can be treated with a DNA-damaging agent and analyzed at different time-points to study the kinetics of lesion processing, although it is not possible to measure repair activity in real time as the comet assay requires cell lysis.

There are two flavors of the comet assay that depend on the pH of the buffer used before/while carrying out electrophoresis. The neutral comet assay detects DSBs and SSBs, which may be induced directly by the agent used to damage the nuclear DNA, or may be introduced by processing of damaged bases or AP sites by BER enzymes (and potentially other pathways). By contrast, the alkaline comet assay additionally detects alkaline labile sites, such as AP sites, by converting them to a SSB. The complexity arising out of potential differences in initial DNA damage levels, differences in the rate of initiating BER

and differences in the rate of BER intermediate processing necessitates time-dependent measurements for the clearest interpretation of comet assays used in population studies.

Two important basic variations on the comet assay (in addition to neutral versus alkaline) have been used in population studies: one measures the DNA repair capacity of living cells challenged with a DNA-damaging agent; or alternatively, one measures the repair activity of protein extracts from the cells of interest using a damaged genomic DNA substrate derived from the nucleus of a different cell that is embedded in agarose. Agents commonly used to produce damaged genomic DNA substrates for these *in vitro* comet assays to measure BER capacity include potassium bromate [36] or the photosensitizer Ro19-8022 and VL to generate primarily 8-oxoG [37].

Comet assays are highly versatile and have been adapted in numerous ways to improve their reproducibility as well as to increase their specificity for particular repair steps and repair pathways. In yet another variation of the comet assay, damaged bases are detected by their conversion to SSBs if nuclei are incubated with a glycosylase "specific" for the damaged base to be studied; for example 8-oxoG levels have been evaluated by comparing comet assays performed with and without incubation of lysed cells with exogenously added Fpg (the bacterial functional counterpart to OGG1) or OGG1 enzyme [38]. Whereas the use of randomly induced DNA damage has usually limited the use of comet assays to estimating overall BER efficiency, such enzyme-based approaches make it possible to study the efficiency of particular steps of the BER pathway. Historically, the comet assay has faced challenges arising out of inter-assay and inter-lab variability. However, a technology called CometChip was recently developed to overcome many of these limitations by standardizing the method [39].

In many cases the comet assay is used to determine steady-state levels of oxidative damage or SSBs in cells that have not been exposed *in vitro* to exogenous DNA-damaging agents. Although these damage levels depend on cellular BER activity, they also depend on endogenous production of reactive

oxygen species (ROS), and antioxidant levels in the cells. As such, this application of the comet assay is not within the scope of this Chapter, and we refer the reader to an extensive review on this topic [40].

Alkaline comet assay readouts (tail lengths or the percentage of DNA in the comet tail) are time-dependent and information-rich because they are influenced by numerous processes that take place simultaneously in cells (**Figure 3.8**). The strand breaks that produce the comet represent any directly induced strand breaks (particularly for ionizing radiation) and a mixture of repair intermediates that may accumulate between the moment of initiation of BER via a glycosylase to the moment of BER completion by DNA ligase. Steady-state levels of DNA damage, represented in region 1 of **Figure 3.8**, may vary among individuals due to differences in: environmental exposure; the production of endogenous DNA-damaging agents (namely ROS); the capacity of cells to repair DNA damage; and the ability of cells to prevent DNA damage by neutralizing DNA-damaging agents before they are able to target DNA. The initial level of damage induced by an agent in region 2 and the additional damage that accumulates during BER processing of damaged bases in region 3 in **Figure 3.8**, may likewise vary among individuals due to differences in the ability of cells to prevent reactive species from attacking DNA (e.g. different glutathione levels that can neutralize ROS). Because *in vitro* comet assays are carried out with cell lysates and uniformly prepared isolated nuclei, they are not subject to the sources of variation represented in regions 1 and 2. Variation in the efficiency of any step of the BER pathway can influence the observed kinetics of repair completion.

Comet assays are uniquely suited to measuring BER capacity in genomic DNA in large numbers of individuals. Although they are limited to repair of DNA strand breaks and types of DNA damage that can be converted to strand breaks, comet assays provide information about both repair completion and the efficiency of some specific steps in the BER pathway.

1.7 General Considerations for the Application of BER Capacity Assays in Human Populations

No single assay meets both of the goals established at the beginning of this Section. Every assay requires disruption of cells from their physiological conditions or cell lysis, which can affect their ability to represent BER capacity *in vivo*. Furthermore, inter-individual variation in biological parameters beyond DNA repair can potentially confound efforts to measure inter-individual differences in BER capacity. For example, differences in repair protein stability in dilute solutions could affect assays in cell lysates, differences in cytoplasmic processing of plasmids could affect HCR assays, and differences in the cellular response to culture at ambient oxygen tension could affect *ex vivo* comet assays in PBMCs and other primary cells. A powerful approach for overcoming these potential pitfalls, as well as the technical limitations of each type of assay, will be to combine more than one approach for measuring BER capacity in the same set of samples in future studies.

2. Studies on BER Capacity Variation in Human Populations

In this Section we discuss population-based studies that have been carried out to test for inter-individual differences in BER capacity, and to test whether subpopulations defined according to age, disease status, or environmental exposure exhibit different levels of BER capacity. Studies fall under two major categories. The first category includes studies that test the hypothesis that there is significant inter-individual variation in BER capacity among individuals (**Table 3.1**). These studies are necessary for establishing a foundation for a second category of studies that test for relationships between BER capacity and other variables including age, diet and lifestyle, cancer, neurologic disorders, and immune disorders (**Tables 3.2, 3.3** and 3.**4**). Studies in the second category are an essential step toward establishing the potential for

translation of BER capacity assays to personalized prevention and treatment of disease (see Chapters 17 and 18).

Because **Tables 3.1-3.4** are formatted similarly, a general legend explaining the contents under the headings that are common to all four tables is provided here. The BER steps for which repair capacity is reported are defined as follows: initiation of BER by specific DNA glycosylases (S for Start); processing of specific BER intermediates by APE1, DNA polymerase, or DNA ligase (M for Middle); combinations of steps that contribute collectively to the readout of the assay but where repair has not necessarily been completed (T for Total); and combinations of steps from beginning (glycosylase) to end (ligase) where readout is a consequence of complete repair (C for Complete). Tissue refers to the tissue of origin for the cells used in the study, format refers to the type of assay used (see Section 1). The column titled "treatment/enzyme/lesion" describes the particular enzyme activity that was measured and the type of DNA damage that was used. N refers to the number of subjects in the study. For case-control studies, two numbers are given; the first is the number of cases, and the second is the number of control subjects.

2.1 Variation in BER Capacity Among Apparently Healthy Individuals

Studies measuring BER capacity in healthy individuals (**Table 3.1**) have focused on variation in the activities primarily executed by APE1 and three different BER-initiating glycosylases (AAG, OGG1 and UNG). A 3.3- to a 10-fold range of AAG activity has been measured in cell lysates with either a Hx - or an ethenoadenine (εA)-containing oligo (N = 80 to 139) [7, 41-43]. A 2 to 3-fold range in OGG1 activity has been measured using labeled oligonucleotides containing an 8-oxoG:C base-pair (N = 34 to 120) [8, 44, 45]. UNG activity measured by a method that utilized calf thymus DNA with radioactive uracil revealed substantial tissue-dependent variation among individuals: 65-fold in small intestine (N=12), 5.5-fold in colon (N=10), 3.2-fold in stomach (N=5) and 3.1-fold in liver (N=9) [46]. This study also included activity measurements for a

second protein, O^6-methylguanine DNA methyltransferase (MGMT), which operates in a different pathway (Direct Reversal). Similarly, tissue-dependent variation in MGMT activity was observed among the same individuals: 42-fold in small intestine, 9.6-fold in colon, 1.7-fold in stomach and 7.8-fold in liver. Notably, MGMT activity was not correlated with UNG activity, indicating that BER capacity cannot be inferred from the activity of other pathways. Two studies focused on the processing of abasic sites that form as BER intermediates, or spontaneously via hydrolysis. In a relatively small study (N = 10), AP endonuclease activity in PBMC lysates was found to vary 2.5-fold using depurinated plasmids; nicking at AP sites was used to calculate APE1 activity [28]. When APE1 activity was measured in a larger study (N = 100) using labeled oligonucleotides containing an AP site, a 4.9-fold variation was observed [47]. Quite different ranges in BER capacity have been reported when using the *in vitro* comet assay, 4- and 41-fold in a pair of methodologically similar studies ([48] and [49], respectively). In both reports, PBMC lysates were assayed using HeLa nuclei pretreated with a photosensitizing agent, Ro 19-8022, and VL, which together induce 8-oxoG. The disparity in reported fold-range of repair capacity might be explained by differences in the incubation time following exposure to DNA damage (10 and 20 min, respectively) and the parameter used to report repair capacity ("tail length" versus a calculated parameter called "incision activity").

Genetic variants of unknown significance present a major barrier to using genetic testing to identify individuals with DNA repair defects. Functional assays have been used to test whether an *OGG1* polymorphism that leads to an amino acid substitution (S326C) significantly affects OGG1 function. One report using phytohemagglutinin (PHA)-stimulated PBMCs (N = 34) found significantly decreased OGG1 activity in cell lysates from individuals with the Cys/Cys genotype versus Ser/Ser [50], whereas a second report using an identical radiolabeled oligonucleotide assay with unstimulated PBMCs (N = 78) found no effect [44]. Using a comet assay following exposure of intact, unstimulated PBMCs to ionizing radiation, which generates a mixture of DNA lesions including 8-oxoG, one report found an increase in residual DNA damage for both Ser/Cys

and Cys/Cys *OGG1* genotypes versus Ser/Ser, suggesting a defect in repair of γ-irradiation induced oxidative lesions [51]. A second comparable study reported a decrease in repair of oxidative DNA damage in the variant genotype, but did not find a difference in repair of γ-irradiation-induced DNA damage between the genotypes [52]. The apparently conflicting results may be due to the different types of DNA damage and different assays used in the studies. It is also possible that relatively small differences in OGG1 activity associated with the S326C substitution are overwhelmed by inter-individual variation in OGG1 activity due to potential variation in other genes, and differences in lifestyle and environmental exposures that may affect total OGG1 activity (see below).

Overall, the available data from studies of BER capacity in apparently healthy individuals mirror a larger body of accumulating data indicating that DNA repair capacity varies significantly across human populations [1, 53]. Both assays in cell lysates and comet assays with intact cells provide evidence of inter-individual differences in BER capacity for at least three DNA glycosylases and APE1. The following Sections discuss associations among inter-individual variation in BER capacity, aging, and human health.

2.2 Associations Between BER Capacity and Age or Lifestyle

Some studies have explored the possibility that BER capacity may be a dynamic property influenced by age and environmental exposures (**Table 3.2**). The mechanisms of aging are not fully understood, and a longstanding hypothesis is that age-dependent changes in DNA repair capacity may be responsible for a constellation of age-associated health problems including cancer and neurodegeneration [54]. Changes in BER capacity with increasing age in mice, and with increasing passage number in human cells have provided indirect evidence of age-dependent BER capacity changes in humans [55]. Recently, several studies have directly explored variation in BER capacity as a function of age in human populations (see also Chapter 20).

Most investigations assessing the potential for age-dependent changes in BER capacity involve measurements in donors of different ages at the time of the study. Measurements in PBMC extracts and labeled oligonucleotides yield consistent results among similar studies even when conducted in different laboratories. In particular, AAG activity has been found to be invariant with age when comparing individuals under and over 65 (or 66), in three relatively large studies (N > 100) [7, 42, 43]. By contrast, OGG1 activity has been reported to decrease with age in four different studies. Specifically, 9.2 % higher mean OGG1 activity was observed among individuals aged 60 or younger versus those over 60 (N = 68) [45]. A larger study (N = 93) showed the same trend, with a 9.3 % difference in mean OGG1 activity, but between individuals under or over 65 [56]. In a subsequent study with 120 subjects, the trend was again observed but only for males over 55 versus under 55 (11.7 % lower OGG1 activity among older individuals) [8]. Importantly *OGG1* mRNA and OGG1 activity were not strongly correlated in this study, indicating that posttranscriptional regulation of BER may confound efforts to predict function from transcript levels. In contrast to the studies carried out with oligonucleotides and extracts from PBMCs, one study that employed an oligonucleotide array (see Section 1) reported an *increase* in 8-oxoG excision activity with age in skin fibroblasts [57]. This study compared several BER activities in lysates from primary skin fibroblasts from 29 individuals in three different age groups: 20-33, 40-50 and 61-68. The efficiency of processing for three other DNA lesions, including an abasic site analog (THF), thymine glycol, and uracil opposite either guanine, or adenine were found to decrease with age. Although a relatively small study measuring activity using depurinated plasmids and PBMC extracts from 23 women aged 27-57 showed that AP endonuclease activity did not to correlate with age [58], a larger (N = 100), more recent study did report a decrease in APE1 activity with age [47], .

One study used a GFP-based HCR assay to measure repair of oxidative damage generated in plasmids with methylene blue plus VL, followed by transfection into foreskin fibroblasts isolated from 19 donors aged 20-64. This assay is expected to report primarily on completion of OGG1-initiated BER of

transcription blocking oxidative damage. BER capacity was found to decline significantly with age. This study also explored mechanisms governing the apparent decline of BER capacity. It was found that fibroblasts from older individuals transfected with a vector expressing Sirt6 (but neither DNA polymerase β (POLβ) nor X-ray cross-complementing protein 1 (XRCC1)) exhibit higher BER activity in a poly(ADP-ribose) polymerase (PARP)-dependent manner; treating with a PARP inhibitor (PJ34) yields lower activity, and Sirt6 does not increase BER activity in the presence of PJ34 [59].

Multiple studies using comet assays have yielded conflicting results with respect to possible age-dependent changes in BER capacity. Some of these conflicting results may be due to the different DNA-damaging agents and time-points employed in the studies (**Table 3.2**). Comet assays often exhibit complex kinetics that have been attributed to multiple rate constants for repair of different types of DNA damage and repair intermediates, particularly when ionizing radiation is used to generate DNA damage [60]. As a result, differences in the time at which comet assays are performed after exposure to DNA-damaging agents affects which DNA repair pathway(s) and which steps within the BER pathway (e.g. initiation versus intermediate processing) most strongly influence comet tail length and intensity (**Figure 3.8**). In 3 out of 4 studies in which comet assays were used to measure DNA repair at time-points of 40 minutes or longer following exposure to DNA damage, repair capacity was found to *decrease* significantly with age. By contrast, among 5 studies in which DNA repair was assayed at 20 minutes or less following exposure to DNA damage, 3 found no relationship between age and repair capacity, and 2 found that repair capacity *increased* with age. Because BER initiation initially leads to an increase in comet tail length and intensity, the shorter or less intense comets that have been interpreted as an increase in BER capacity with age when measured at short times could reflect a decrease in glycosylase activity with age (fewer intermediates formed), rather than more rapid completion of repair. However, even a single research group using ionizing radiation and a 40 minute time-point reached opposite conclusions with regard to the question of whether the rate of

repair for DNA damage induced by ionizing radiation increases or decreases with age [61, 62].

There have been relatively few longitudinal studies of BER capacity. In one study, BER capacity was measured using comet assays for repair of oxidative damage induced by the photosensitizer Ro 19-8022 and VL in lymphocytes isolated from 35 subjects at 6 time-points over a period of 5 months. When data were compared for time-points taken 2 months apart, a correlation between BER capacity within individuals was relatively weak (R^2=0.30). The mean coefficient of variation (CV) reported for intra-individual variation in BER capacity was 27 %, suggesting that there may be significant variation in BER capacity on the timescale of months [48]. By contrast, a second study, discussed in more detail below, found that OGG1 activity from PBMC extracts was stable when measured in PBMCs from cancer patients up to 4 months before and over 1 year after treatment, surgery and remission [45]. Some studies have found evidence that environmental exposure and lifestyle can influence BER capacity. Some examples are: increased SSB levels in genomic DNA following coenzyme Q10 supplementation [63]; more efficient repair of oxidative DNA damage following consumption of 1, 2 or 3 kiwi fruits a day [64], dietary antioxidant supplementation [49], or consumption of antioxidant-rich plant products [65]; and more efficient SSB repair among individuals who engage in medium and high levels of physical activity versus low levels of physical activity [66].

In summary, studies that have directly assessed the relationship between age and BER capacity have led to apparently conflicting conclusions. The apparent conflicts may be explained in part by important methodological differences among the studies, including the use of assays at different time-points, cells of different origin, and different substrates or sources of DNA damage. As discussed in Section 4, a more direct comparison among multiple methods and more detailed kinetic analyses to provide mechanistic insights that can be gleaned from comet assays may be helpful for future studies.

2.3 Associations Between BER Capacity and Cancer

There is increasing evidence of links between BER capacity and cancer etiology, and several studies have functionally assessed BER capacity in cancer patients versus healthy controls (**Table 3.3**) (see also Chapter 13). These relatively limited studies have found evidence of an association between cancer and differences in the activity of several BER enzymes, including APE1, OGG1, and AAG. All of these studies have been carried out with lysates from PBMCs.

Four studies have yielded apparently conflicting results with regard to whether APE1 activity is significantly different in individuals with cancer versus healthy controls. Two early case-control studies compared APE1 activities of breast and lung cancer patients to healthy volunteers by quantitating nicking of depurinated plasmids in PBMC lysates (see Section 1) [67, 68]. Neither study found a significant difference in activities in cells from cancer patients versus healthy controls. However, the small sample sizes (8/8 and 10/10 case/control, for breast and lung cancer, respectively) likely limit the power of these studies to measure small differences. More recently, larger studies have found significantly lower APE1 activity in PBMC lysates from cancer patients versus healthy controls using oligonucleotide-based assays [47].

OGG1 activity has been measured in studies comparing BER capacity in PBMCs from healthy controls versus lung cancer, colon cancer, or head and neck cancer. Lower mean OGG1 activity has been measured in non-small-cell lung cancer (NSCLC) patients versus healthy controls ($P < 0.001$) [45]. Three different logistic regression models, where OGG1 activity was either considered (a) a continuous variable, (b) a dichotomized variable (by the median) or (c) a tertile, yielded a significantly increased odds ratio for lung cancer among individuals with lower OGG1 activity. Importantly, a direct comparison between OGG1 activity in PBMCs and in lung cells from a subset of 7 individuals revealed a strong correlation ($R^2 = 0.86$, $p = 0.003$). Although the sample size is relatively small, this strong correlation supports the notion that PBMCs are a reasonable surrogate for DNA repair capacity in other tissues (in this particular case, OGG1

in the lung). Furthermore, as part of the same study, OGG1 activity was measured in PBMC lysates from patients over time, both before and after surgery, and found to be stable. A subset of 8 individuals was tested over a period of 3 years and CV ranged from 2% to 12%, with an average of 7% (95% CI = 4% to 10%). An association between low OGG1 activity and lung cancer risk was independently observed by a second group in the same year [69], and has been confirmed in larger, more recent studies [7, 47]. Analogous studies have found a statistically significant association between reduced OGG1 activity and squamous cell carcinoma of the head and neck (SCCHN) [56]. Interestingly, both in the case of NSCLC and SCCHN, no statistical interaction between smoking and reduced OGG1 activity was found, implying that the two are independent risk factors for these two cancers, and highlighting the importance of both genes and environment in risk assessment. Strikingly, in contrast to what was observed for NSCLC and SCCHN, *higher* median OGG1 activity was measured in PBMCs from colorectal cancer patients versus healthy controls [70]. If one assumes that inter-individual differences in OGG1 activity in PBMCs are representative of inter-individual differences in OGG1 activity in other tissues, there would appear to be opposite associations between OGG1 activity and colon cancer, versus OGG1 activity and lung or head and neck cancer. This observation suggests that relative differences in DNA repair capacity can have opposite consequences for different tissues and cell-types. Furthermore, contrary to the intuitive notion that higher BER capacity is protective, this particular example (and more below) indicates that increased BER capacity can also be deleterious.

Several studies have tested for associations between cancer and the activity of other DNA glycosylases, or combinations of activities for multiple BER enzymes. Significantly, higher AAG-catalyzed εA incision activity was measured in lysates obtained from PBMCs isolated from lung cancer patients versus healthy controls [42]. Individuals in the highest εA incision activity tertile had a 3-fold increased risk of lung cancer when compared to individuals in the lowest tertile. This result was replicated in PBMC extracts from lung cancer patients and healthy controls in a more recent study that measured activity for both OGG1 and

AAG [7]. In addition to high AAG activity, low OGG1 activity was associated with lung cancer in this study. Conditional logistic regression analysis treating OGG1 activity and AAG activity as a continuous variable found that the odds ratios for lung cancer associated with a 1 standard deviation increase in AAG activity (2.0) or a 1 standard deviation decrease in OGG1 activity (1.6) were both significantly greater than 1.0. The odds ratio for lung cancer associated with a 1 standard deviation decrease in a combined score calculated for the two activities was even larger (2.3), suggesting that the two enzymatic activities could provide a more robust biomarker for lung cancer risk than either activity alone [7]. A study in which activity was measured for three BER enzymes in the same set of NSCLC cases and healthy controls found associations between lung cancer and lower OGG1, higher MPG (AAG), and lower APE1 activity. The authors of this study have developed a combined "OMA" score composed of three enzymatic activities for each subject. The score was defined in such a way that a lower OMA score corresponds to lower OGG1 activity, lower APE1 activity, and higher AAG activity. The odds ratio for lung cancer for individuals in the lowest tertile versus the highest tertile of the DNA repair OMA score was 9.7 (95% CI, 3.1–29.8; $P < 0.001$), larger than the odds ratio for lung cancer for any of the individual enzymatic activities [47]. This work highlights the advantages of measuring repair with multiple DNA repair substrates to obtain a more comprehensive assessment of BER capacity.

Seemingly at odds with the studies discussed above, a different study found *lower* levels εA incision activity in PBMCs from lung cancer patients versus healthy volunteers. Furthermore, this study found significantly lower εA incision activity in adenocarcinoma patients versus squamous cell carcinoma patients [71]. This difference in εA incision activity was also observed in lysates derived from non-dysplastic lung tissue from the two lung cancer types patients. However, unlike the assays used in studies that found the opposite association between AAG activity and cancer [42], the assays used in these studies actually measure a combination of AAG and APE1 activity, because the oligonucleotide substrates are not alkali treated following lysate incubation. As a result, even in

the presence of high AAG activity, a low APE1 activity would result in a decreased combined (AAG/APE1) εA incision.

Because the studies discussed immediately above were all conducted retrospectively, important questions remain about causality of the observed differences in BER capacity in cancer cases versus healthy controls. Is altered BER capacity predictive of higher cancer risk? Or on the contrary, does the development of cancer itself modify BER? One study did explore whether BER capacity changes over the course of cancer treatment, and found that OGG1 activity in PBMCs did not change after surgery and chemotherapy when compared to activity measured at time of diagnosis [45]. Although OGG1 activity was not measured before cancer development, these observations are consistent with BER capacity (at least for OGG1) being independent of cancer status and stable within the detection limits of the assay for over periods of up to three years.

Taken together, the available data suggest there may be complex relationships between cancer and BER capacity. Either higher or lower BER capacity may be associated with cancer depending on the enzymatic activity in question and the tissue of cancer origin. The existence of significant associations between BER capacity in PBMCs and at least three types of cancer suggest that BER capacity in blood cells could be a useful biomarker for cancer risk, and may be a surrogate for BER capacity in other tissues. Future studies will be needed to determine whether variation in the activity of additional DNA glycosylases and enzymes involved in BER intermediate processing are associated with specific types of cancer.

2.4 BER Capacity Changes Associated with Neurodegenerative Disorders

Several relatively small studies (N < 30) have found relationships between aberrant BER capacity and neurodegenerative disorders including Alzheimer's Disease (AD) and mild cognitive impairment (MCI) (**Table 3.4**). The available data point strongly to an association between neurodegenerative disorders and

lower BER capacity in brain tissue, and these results are partially corroborated by studies carried out in blood cells.

The most direct tests for associations between BER capacity and neurological disease have come from reports of BER capacity in brain tissue from case control studies (see also Chapter 15). A study that used oligonucleotide-based glycosylase assays found lower OGG1 activity in nuclear extracts from three different brain tissues in AD patients versus healthy controls [72]. An additional study found reduced OGG1 activity in both nuclear and mitochondrial extracts derived from multiple brain tissues from AD patients and MCI patients versus healthy controls [73]. A more recent study reported lower BER capacity for several different substrates in cell lysates obtained from cerebellar tissue from 10 Alzheimer's patients versus 10 controls [74]. This study used oligonucleotides with 8-oxoG, THF, Uracil, or a single nucleotide gap into which a radiolabeled nucleotide (^{32}P-dCTP) can be incorporated during repair synthesis by POLβ. "Total BER" was measured by ^{32}P-dCTP incorporation into a double-stranded substrate with a G:U base pair (achieving incorporation only if glycosylase, APE1, and POLβ act on the lesion). Glycosylase activity, gap-filling, and "total BER" were all lower in AD patients; however, AP site incision was not significantly different versus healthy controls. Furthermore, in a cohort of 9 patients with MCI, who are at high risk for developing AD, there was a statistically significant inverse correlation between BER capacity and the severity of MCI clinical diagnosis (neurological impairment measured by Braak stage).

Two studies have focused exclusively on BER capacity in the mitochondrial compartment of cells isolated from brain tissues in healthy individuals versus AD patients. In one study, immunohistochemical analysis revealed reduced expression of the mitochondrial isoform of OGG1 in brain tissue isolated from AD patients; however, the methods used in this study are indirect and may not reflect functional BER capacity [75]. In a more recent study, glycosylase activity, AP site incision, gap-filling, and ligation have been measured in mitochondrial extracts from postmortem tissue isolated from the inferior parietal region of the brain of 6 AD patients and 6 control subjects [76].

This study showed lower 5-hydroxyuracil incision and DNA ligase activity in lysates from patients versus lysates from controls, but similar levels of uracil incision, AP site (THF) incision, and gap-filling. Ligation was measured using a nicked, radiolabeled substrate, and a gapped duplex pre-treated with POLβ; both substrates yielded similar results.

The small numbers of subjects in BER capacity studies using brain tissue in part reflect the technical challenges of obtaining biopsies or postmortem samples, and have motivated efforts to test for DNA repair biomarkers in more accessible blood cells. One study has tested whether AD is associated with differences in BER capacity in PBMCs. Isolated lymphocytes from 13 Alzheimer's patients showed slower repair kinetics versus 14 elderly controls when treated with H_2O_2 and analyzed by alkaline comet assay [77]. Future studies will be needed to further test whether differences in BER capacity in PBMCs are associated with neurological diseases.

3. Methodology and Study Design: Practical Considerations for Studies of BER Capacity in Human Populations

Mounting experimental data detailed in Section 2 support the notion that variation in BER capacity is associated with disease, and to some extent, with aging (see Chapter 20). The aggregate data provide a solid foundation and strong motivation for larger and more comprehensive studies of BER capacity in human populations. However, the data also highlight areas where future work may be able to overcome limitations of existing methodologies, and to maximize the insights that can be gleaned from future population studies. As noted in Section 2, studies exploring links among BER capacity and environmental exposure, age and healthspan have in some cases yielded conflicting results. Furthermore, whereas most studies have focused on repair of a single lesion by a single pathway, BER is a multidimensional and time-dependent process (**Figure 3.9**). For a given individual, BER capacity may vary from tissue to tissue, from cell to cell, and from one chromatin environment to another within the cell.

At the molecular level, BER capacity comprises the efficiency of multiple steps within the pathway, and its relative contribution to the repair of a specific DNA lesion depends on the extent to which other pathways may compete for repair of the same lesion. In this Section, we discuss possible origins of conflicting results in some studies, and we propose some best practices for future studies, with particular emphasis on the need for collaborative efforts employing multiple methods for measuring BER capacity in human populations.

3.1 Tissue-Dependent Variation in BER Capacity (and the Need for Surrogate Tissues)

The availability of tissues for analysis represents a major practical consideration for population-based studies of BER capacity that is balanced against a goal of approximating, as closely as possible, the repair capacity of cells in a tissue of interest *in vivo*. There is evidence that BER capacity and the consequences of changes in BER capacity can vary greatly among different tissues in both humans and in animal models ([46], and **Table 3.1**)). The consequences of unrepaired DNA damage are also expected to vary among tissues; for example, whereas unrepaired lesions can trigger apoptosis in some cells without leading to disease in proliferating tissues such as skin, apoptosis in postmitotic cells such as neurons could lead to catastrophic illness. Some tissues sustain much greater DNA damage insults than others; as a consequence, a lower repair capacity in a given tissue may not represent a deficiency, so long as it is sufficient for processing the level of DNA damage normally sustained in that tissue. These tissue-dependent differences in BER capacity may be due to relatively stable epigenetic programing; however, it should also be borne in mind that human cells can undergo transcriptional reprograming as a function of tissue microenvironment that can dramatically alter phenotype, and may not be preserved during functional characterization *ex vivo* [78]. Taken together, these observations raise questions about the appropriateness of measuring BER capacity in cells from a single tissue *ex vivo* under standard tissue culture

conditions that are in some cases very different from the physiological context in which cells are found *in vivo*. Because at present it is unfeasible to measure BER repair capacity *in vivo*, and further, because it is not feasible to sample most tissues in a non-invasive manner, blood cells and skin cells have dominated investigations of BER capacity in populations.

Although there is limited evidence that BER capacity measurements in one tissue can represent BER capacity in other tissues [45, 71], it will be important to test this idea in future studies. It is possible that repair capacity varies in such a way across tissues so as to preserve the rank order for repair capacity among different individuals; viz. if one were to rank several individuals from lowest to highest repair capacity in blood cells, and then rank them from lowest to highest repair capacity in neurons, the order might be the same despite potentially large differences in the absolute repair capacity between the tissues. This ideal scenario is represented in the top panel of **Figure 3.10**, and could explain why BER capacity in blood cells can serve as a biomarker for risk of cancers originating in lung, breast, and colon tissues, or neurodegenerative disorders that affect brain tissues. At the opposite extreme, represented in the bottom panel of **Figure 3.10**, it is possible that BER capacity varies independently across tissues. This latter scenario could explain instances where studies in different tissues have yielded conflicting results, such as with regard to the relationship between OGG1 activity and lung cancer versus colon cancer [56, 70], or the relationship between OGG1 activity and age in blood cells versus skin cells [8, 45, 56, 57]. Given their unique physiological property of sampling nearly all other tissues, it is conceivable that blood lymphocytes represent an average of repair capacity over multiple tissues for a given individual.

To address the fundamental question of whether either of the scenarios in **Figure 3.10** applies to human tissues, future studies should include, wherever possible, measuring BER capacity in at least two tissues or cell types for the same set of subjects. Collaborative studies making use of emerging "organ on a chip" technologies that allow a closer approximation to the *in vivo* cell environment in laboratory experiments will be also useful for testing whether *ex*

vivo BER capacity measurements can be refined to yield a more accurate representation of inter-individual differences [79]. Furthermore, emerging induced pluripotent stem cell (iPS) technology could prove invaluable in such an effort, by providing access to and enabling to repair measurements in virtually every tissue and cell type.

3.2 Subcellular variation in BER / Mitochondrial BER

In addition to the multidimensionality and potential temporal variation of BER capacity discussed above, it should be borne in mind that 1% of the DNA in human cells resides in the mitochondrial compartment. The fate of damaged mitochondrial DNA and the activity of DNA repair pathways in the mitochondrial compartment remains much less well understood than DNA damage processing in the nuclear compartment [80, 81]. Although it is well established that the BER pathway is active in mitochondria (see Chapter 19), at this time only a very limited subset of studies have tested whether inter-individual differences in mitochondrial BER capacity are associated with aging or human disease [76, 82]. However, as evidence continues to accumulate in support of a role for mitochondrial dysfunction and accompanying oxidative stress in a variety of diseases [83], further studies testing whether inter-individual differences in mitochondrial BER capacity are associated with disease are warranted. It will be of particular interest to determine whether nuclear BER capacity can be used a surrogate for mitochondrial BER capacity.

3.3 Variation in BER Efficiency at the Molecular Level

Associations continue to emerge between disease and imbalance among the rates at which the multiple steps in the BER pathway proceed [7, 41, 47], and mechanistic studies indicate extensive crosstalk between the BER pathway and other DNA repair pathways [84-90]. This complexity can lead to counterintuitive relationships, as in the context of lung and colon cancers, where *higher*

glycosylase activity is associated with cancer; these associations may be due to the accumulation of BER intermediates, including strand breaks and abasic sites, which are more toxic to cells than the original DNA lesions [7, 70]. The health consequences and mechanistic underpinnings of variation in BER capacity among individuals may therefore only be fully understood in the broader context of DNA repair capacity measured at multiple repair steps and in multiple pathways. Thus, future studies would be strengthened by including measurements of the efficiency of multiple BER steps, and repair capacity measurements in additional pathways.

3.4 Temporal Changes in BER Capacity

A key assumption in many population studies of BER capacity is that measurements taken on a single day can be regarded as a snapshot representative of stable differences among individuals. There is some evidence in support of this assumption; OGG1 activity in PBMCs from 8 individuals was very stable over the course of 3 years with a %CV between 2% and 10% [45]. However, the aging studies in **Table 3.2** provide indirect evidence that BER capacity changes over an individual's lifetime. Furthermore, there is evidence that environmental exposures, diet, and lifestyle can influence repair capacity [49, 63-66, 91, 92], and animal models provide evidence of diurnal variation in repair capacity [93]. These data suggest that intra-individual differences in BER capacity need to be taken into consideration when studying inter-individual variation in BER capacity. Future studies should include, wherever possible, longitudinal measurements of BER capacity in cells isolated at different times from the same individual, and additional studies testing the effects of nutrition, environmental exposures, and disease on BER capacity are warranted.

3.5 A need for Collaborative Studies Using Multiple Methods

DNA repair capacity is a highly multidimensional process (**Figure 3.9**), and the available methods for measuring BER capacity in populations vary in their ability to address these dimensions. A strength of *in vitro* assays with cell lysates is that they can be performed with chemically defined DNA lesions for any cell type or tissue that can be isolated and processed to generate protein extracts. However, each extraction and incubation step can create potential sources of inter-sample and inter-laboratory variability, and information about subcellular localization of repair proteins and cell type(s) within a tissue are lost. Assays in cell extracts also represent a snapshot of repair capacity at a single time-point, and therefore place limits on the ability to measure the kinetics of BER in cells. By contrast, HCR assays and some molecular beacons provide a platform for generating chemically defined DNA lesions *in vitro*, and then measuring their repair in cells. However, DNA repair assays using exogenously generated DNA, including plasmids, beacons and microarray chips lack at least some of the chromatin structures that are present in genomic DNA, and so may not be amenable to assessing variation in BER capacity as a function of chromatin environment. Comet assays provide a means for measuring BER capacity in genomic DNA, can be carried out in living cells, and allow for measurements of repair kinetics. However, it is not possible to generate chemically defined lesions in comet assays. Instead, the use of DNA-damaging agents results in a mixture of DNA lesions, and can also damage other cellular components when carried out in living cells. Furthermore, interpretation of the results can be complicated as each time-point represents a mixture of BER initiation, processing and ligation (**Figure 3.8**). Because of the complexity of the task and the potential need to measure BER capacity in a large number of individuals, a new generation of tools for measuring BER capacity in human populations with still higher-throughput and precision may need to be developed.

Some of the apparently conflicting results among multiple studies exploring the same relationship between BER capacity and variables, including

age and healthspan, may be a consequence of using different assays to measure activity. Best practices for leveraging the strengths of different assays, while overcoming the challenges associated with inter-assay and inter-study differences in BER capacity assessments, should include performing collaborative population studies that apply multiple complementary methods on a single set of samples.

3.6 Key Questions to Address in Future Studies

Decades of mechanistic work and new technologies for measuring BER capacity leave us poised to make major advances toward translating what is known about BER at the molecular level into strategies for treating or preventing disease. At the frontier of these efforts lie a number of exciting questions that can potentially be answered in the coming years using methodologies now at hand:

1. What is the magnitude of variation in BER capacity among individuals? What is the shape of the distribution?
2. To what extent does BER capacity vary in time, and how is it influenced by environmental factors?
3. To what extent can BER capacity be predictive of healthspan or lifespan?
4. Do PBMCs represent an appropriate surrogate for other tissues?
5. How does BER capacity vary across tissues, and to what extent does it change during organismal development and cell differentiation?
6. Are the relationships between BER capacity and disease monotonic, or are the relationships more complex?
7. What is the relationship between BER capacity and the hazards associated with environmental exposures?
8. For how many specific BER steps does repair capacity need to be measured to yield sufficient information for assessments of disease susceptibility or predictions about the ability of individuals to tolerate environmental and medical exposures to DNA-damaging agents?

9. What enabling technologies are needed before functional measurements of BER capacity and repair capacity in other pathways can improve prevention and enhance the standard of care for patients with diseases that are associated with DNA repair defects?

4. Conclusions

A perhaps unanticipated follow-on of the genomic revolution has been the renaissance of functional assays and high-throughput methods that do not involve sequencing DNA. Although sequencing and other 'omics' approaches have yielded tremendous insights into the origins of inter-individual differences, efforts to use them to predict function have been challenged by the complexity arising out of multiple layers of regulation and the influence of environmental exposures. A variety of innovative methods suitable for functionally measuring BER capacity in human populations have been developed within the past decade. These methods have allowed us to begin to elucidate the level of inter-individual variability with respect to BER capacity and its individual enzymatic steps, as well as correlations between small variations in BER with healthspan. Nevertheless, no single method provides a complete and mechanistic picture of the BER pathway. Large, collaborative, ideally prospective DNA repair studies using multiple methods, including structural and biochemical analyses, will refine our understanding of the relationship between variation in BER capacity and health- and life-span, elucidate the mechanistic underpinnings of these relationships, and identify targets for preventive or therapeutic intervention.

Tables

Table 3.1 BER capacity variation among apparently healthy individuals.

The reported fold variation represents the ratio of the highest measured repair capacity to the lowest measured repair capacity. The BER steps for which repair capacity is reported are defined as follows: initiation of BER by specific DNA glycosylases (S for Start); processing of specific BER intermediates by APE1, DNA polymerase, or DNA ligase (M for Middle); combinations of steps that contribute collectively to the readout of the assay but where repair has not necessarily been completed (T for Total). Tissue refers to the tissue of origin for the cells used in the study, and format refers to the type of assay used (see Section 1). The column titled "treatment/enzyme/lesion" gives the enzyme activity that was measured and the type of DNA damage that was used. N refers to the number of subjects in the study.

Variation	Step(s)	Tissue	Format	treatment/enzyme/ lesion	N	Reference
≈ 3.5-fold	**S**	PBMC	Lysate, oligo	AAG-Hx	100 [A]	[7]
10-fold	**S**	PBMC	Lysate, oligo	AAG-εA	80	[41]
≈ 3.3-fold	**S**	PBMC	Lysate, oligo		100 [A]	[43]
2-fold	**S**	PBMC	Lysate, oligo	OGG1-8-oxoG	34	[44]
≈ 2.9-fold	**S**	PBMC	Lysate, oligo		68	[45]
2.8-fold	**S**	PBMC	Lysate, oligo		120	[8]
65-fold	**S**	Intestine	Lysate	UNG	12	[46]
5.5-fold	**S**	Colon			10	
3.2-fold	**S**	Stomach			5	
3.1-fold	**S**	Liver			9	
6.2-fold	**S+M**	PBMC	Lysate, oligo	AAG/APE1-εA	139	[42]
2.5-fold	**M**	PBMC	Lysate, plasmid	APE1 depurinated-plasmid	10	[28]
4.9-fold	**M**	PBMC	Lysate, oligo	APE1-THF	100 [A]	[47]
4-fold	**T**	PBMC	Comet, alkaline (10 min repair)	Ro 19-8022 + VL treated substrate nuclei	35	[48]
41-fold	**T**	PBMC	Comet, alkaline (20 min repair)		48	[49]

[A] These studies were carried out on the same set of subjects

Table 3.2. Age-associated changes in BER capacity.

The range of ages included in the study is provided in the leftmost column. Arrows indicate whether age is associated with higher (↑), lower (↓), or no significant difference (=) in BER capacity. The BER steps for which repair capacity is reported are defined as follows: initiation of BER by specific DNA glycosylases (S for Start); processing of specific BER intermediates by APE1, DNA polymerase, or DNA ligase (M for Middle); combinations of steps that contribute collectively to the readout of the assay but where repair has not necessarily been completed (T for Total); and combinations of steps from beginning (glycosylase) to end (ligase) where readout is a consequence of complete repair (C for Complete). Tissue refers to the tissue of origin for the cells used in the study, and format refers to the type of assay used (see Section 1). The column titled "treatment/enzyme/lesion" gives the enzyme activity that was measured and the type of DNA damage that was used. N refers to the number of subjects in the study.

Age range	Age-effect	Step(s)	Tissue	Format	treatment/enzyme/ lesion	N	Reference
≤65 vs >65	=	**S**	PBMC	Lysate, oligo	AAG-Hx, AAG-εA	100 [A]	[7, 43]
Range 0-91	↓	**S**	PBMC + PHA	Lysate, oligo	OGG1-8-oxoG	78	[50]
≤60 vs >60	↓	**S**	PBMC	Lysate, oligo		68	[45]
≤65 vs >65	↓	**S**	PBMC	Lysate, oligo		93	[56]
≤55 vs >55, males	↓	**S**	PBMC	Lysate, oligo		120	[8]
Range 20-68	↑	**S**	Skin Fibroblasts	Lysate, oligo-chip	8-oxoG	29	[57]
Range 20-68	↓	**S**	Skin Fibroblasts	Lysate, oligo-chip	Uracil	29	
Range 20-68	↓	**S**	Skin Fibroblasts	Lysate, oligo-chip	Tg	29	
≤66 vs >66	=	**S+M**	PBMC	Lysate, oligo	AAG/APE1-εA	139	[42]
Range 27-57, females	=	**M**	PBMC	Lysate, plasmid	APE1 depurinated-plasmid	23	[58]
Range 20-68	↓	**M**	Skin Fibroblasts	Lysate, oligo-chip	THF	29	[57]
≤65 vs >65	↓	**M**	PBMC	Lysate, oligo	APE1-THF	100 [A]	[47]
Range 20-64	↓	**C**	Skin Fibroblasts	HCR	Methylene Blue	19	[59]
Age not reported, males	↓	**T**	Whole Blood + PHA	Comet, alkaline (5, 60 and 120 min repair)	**Y-rays**	60	[51]
≈ Range 20s-50s	↑	**T**	PBMC	Comet, alkaline (40 min repair)		61	[62]
≤41.5 vs >41.5	↓	**T**	PBMC	Comet, alkaline (40 min repair)		96	[61]
Range 21-88	↑	**T**	PBMC	Comet, alkaline (10 min repair)	Ro 19-8022 + VL treated substrate nuclei	388	[94]
Range 20-82	↑	**T**	PBMC	Comet, alkaline (10 min repair)		97	[37]
Range 21-58	=	**T**	PBMC	Comet, alkaline (10 min repair)		141	[92]
Range 18-30	=	**T**	PBMC	Comet, alkaline (20 min repair)		48	[49]
Range 18-80	=	**T**	PBMC	Comet, alkaline (20 min repair)		309	[95]
Range 18-83	↓	**T**	PBMC	Comet, alkaline (45 min repair)	$KBrO_3$ treated substrate nuclei	80	[36]

[A] These studies were carried out on the same set of subjects

Table 3.3. Comparisons of BER capacity in cancer patients versus healthy controls.

Arrows indicate whether the relevant cancer is associated with higher (↑), lower (↓), or no significant difference (=) in BER capacity. The BER steps for which repair capacity is reported are defined as follows: initiation of BER by specific DNA glycosylases (S for Start); processing of specific BER intermediates by APE1, DNA polymerase, or DNA ligase (M for Middle); combinations of steps that contribute collectively to the readout of the assay but where repair has not necessarily been completed (T for Total). Tissue refers to the tissue of origin for the cells used in the study, and format refers to the type of assay used (see Section 1). The column titled "treatment/enzyme/lesion" describes the enzyme activity that was measured and the type of DNA damage that was used. N refers to the number of subjects in the study. For case-control studies, two numbers are given; the first is the number of cases, and the second is the number of control subjects.

Cancer-type	Cancer-effect	Step(s)	Tissue	Format	treatment/enzyme/ lesion	N[A]	Reference
Lung	↓	S	PBMC	Lysate, oligo	AAG-εA/Hx	100/100 [B]	[7, 43]
Lung	↑	S	PBMC	Lysate, oligo	AAG-Hx	100/100 [B]	[47]
Lung	↓	S	PBMC	Lysate, oligo	OGG1-8-oxoG	64/51	[69]
Lung	↓	S	PBMC	Lysate, oligo		68/68	[45]
Head and Neck	↓	S	PBMC	Lysate, oligo		93/37	[56]
Colon	↑	S	PBMC	Lysate, oligo		90/68	[70]
Lung	↑	S	PBMC	Lysate, oligo		100/100 [B]	[7]
Lung	↓	S	PBMC	Lysate, oligo		100/100 [B]	[47]
Lung	↓	S+M	PBMC	Lysate, oligo	AAG/APE1-εA	25/56	[71]
Lung	↑	S+M	PBMC	Lysate, oligo		88/51	[42]
Breast	=	M	PBMC	Lysate, plasmid	APE1 depurinated-plasmid	8/8	[67]
Lung	=	M	PBMC	Lysate, plasmid		10/10	[68]
Lung	↓	M	PBMC	Lysate, oligo	APE1-THF	100/100 [B]	[47]
Lung	↓	M	PBMC	Lysate, oligo		100/100 [B]	[15]

[A] The number of cases is given first; the number of controls is given second

[B] These studies were carried out on the same set of subjects

Table 3.4. Comparisons of BER capacity in individuals with neurological diseases versus healthy controls.

Arrows indicate whether Alzheimer Disease (AD) or Mild Cognitive Impairment (MCI) is associated with higher (↑), lower (↓), or no significant difference (=) in BER capacity. The BER steps for which repair capacity is reported are defined as follows: initiation of BER by specific DNA glycosylases (S for Start); processing of specific BER intermediates by APE1, DNA polymerase, or DNA ligase (M for Middle); combinations of steps that contribute collectively to the readout of the assay but where repair has not necessarily been completed (T for Total); and combinations of steps from beginning (glycosylase) to end (ligase) where readout is a consequence of complete repair (C for Complete). Tissue refers to the tissue of origin for the cells used in the study, and format refers to the type of assay used (see Section 1). The column titled "treatment/enzyme/lesion" describes the enzyme activity that was measured and the type of DNA damage that was used. N refers to the number of subjects in the study. For case-control studies, two or three numbers are given; the first (and second when there are three) is the number of cases, and the last is the number of control subjects.

Disease-effect	Step(s)	Tissue	Format	treatment/enzyme/ lesion	N[A]	Reference
↓	S	Brain Nuclear Extracts	Lysate, oligo	OGG1-8-oxoG	10/8	[72]
↓	S	Brain Cell	Lysate, oligo		10/10/9	[74]
↓	S	Brain Cell Nuclear and Mitochondrial Extracts	Lysate, oligo		15/6	[73]
↓	S	Brain Cell Mitochondria	Lysate, oligo	5-OHU	6/6	[76]
↓	S	Brain Cell	Lysate, oligo	Uracil	10/10/9	[74]
=	S	Brain Cell Mitochondria	Lysate, oligo		6/6	[76]
=	M	Brain Cell Mitochondria	Lysate, oligo	APE1-THF	6/6	[76]
=	M	Brain Cell	Lysate, oligo		10/10/9	[74]
↓	M	Brain Cell	Lysate, oligo	Single Nucleotide Gap, ^{32}P incorporation	10/10/9	[74]
=	M	Brain Cell Mitochondria	Lysate, oligo		6/6	[76]
↓	M	Brain Cell Mitochondria	Lysate, oligo	Ligase	6/6	[76]
↓	C	Brain Cell	Lysate, oligo	U:G bp and ^{32}P-dCTP incorporation	10/10/9	[74]
↓	T	PBMC + PHA	Comet, Alkaline, 0.5, 2, and 6 hours	H_2O_2	14/13	[77]

[A] The number of cases is given first; the number of controls is given second. Where three numbers are given, the first number is for AD cases, the second number is for MCI cases, and the third number is for controls.

Figures

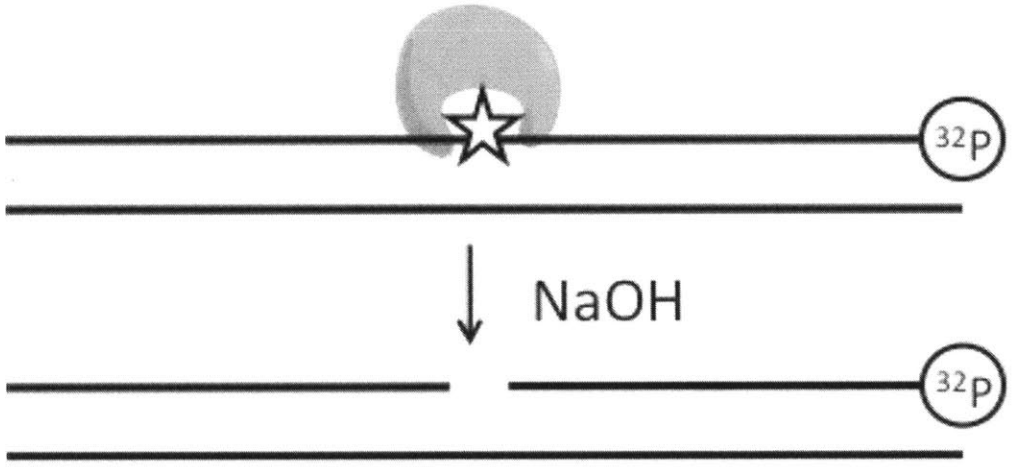


Figure 3.1: Oligonucleotide-nicking-based assay.
A glycosylase (gray crescent) recognizes and excises a damaged base (star). Alkaline conditions (e.g. NaOH) are used to convert all abasic sites to strand breaks. Repair products are then separated from unmodified substrates by gel electrophoresis, and quantified using a radiolabel (sphere marked ^{32}P) or fluorescent dye that is appended to the damage-containing strand of the oligonucleotide duplex.

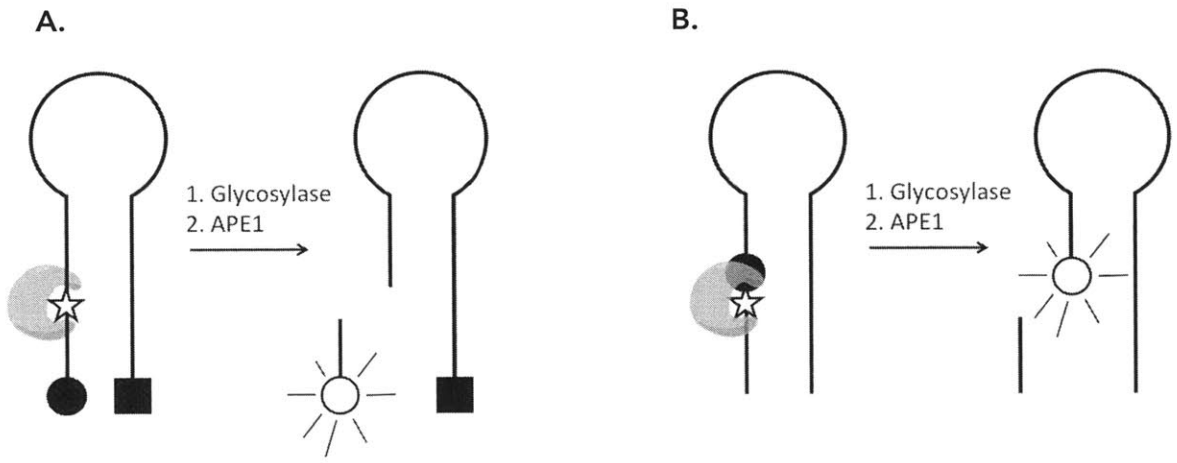


Figure 3.2: Molecular beacons assays for BER capacity.

Molecular beacons are composed of a single strand of DNA with a fluorophore (black sphere) appended to one end and a quencher (black square) appended to the opposite end. The beacon is designed to form a stem loop structure that brings the DNA ends into proximity and quenches fluorescence. A glycosylase (gray crescent) binds and excises a lesion (star) that is positioned near the end of the molecular beacon that bears the fluorophore. Strand cleavage of the resulting abasic site due to bifunctional glycosylase activity or APE1 activity releases a short strand of DNA bound to the fluorophore from the beacon, and leads to an increase in fluorescent signal (open sphere). B) A variation on the beacon assay makes use of the lesion itself (i.e. 8-oxoG, again represented by a star) as the quencher for an adjacent fluorophore (black sphere). Excision of the damaged base, with or without subsequent strand cleavage, is sufficient to separate the quencher from the beacon, leading to an increase in fluorescence.

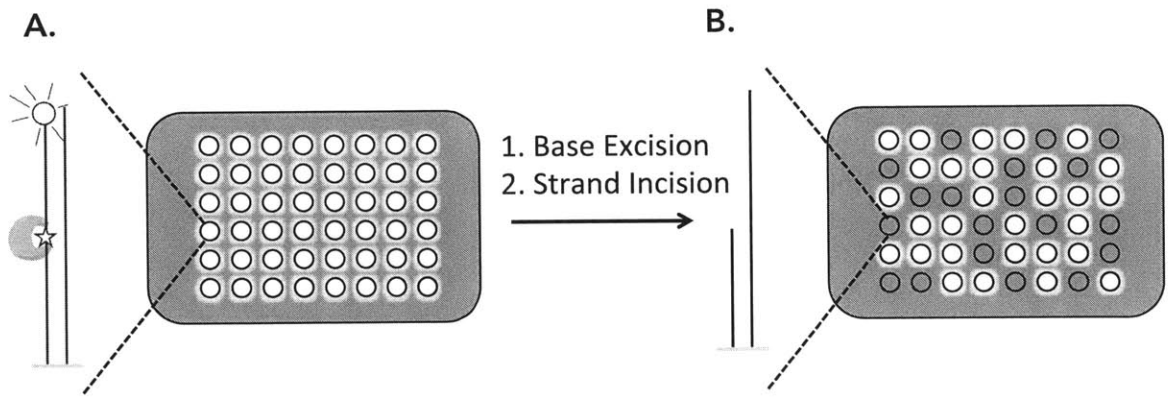


Figure 3.3: Microchip oligonucleotide assay.

A) Each well in a 96-well plate is covalently modified with an oligonucleotide duplex that includes a fluorophore (open sphere) and a site-specific DNA lesion (star) on the same strand. A DNA glycosylase (gray crescent) excises the DNA lesion, followed by DNA strand cleavage by an exogenously added endonuclease at the resulting abasic site. B) After washing, the cleaved fluorescently labeled fragment is removed from oligonucleotides on which glycosylase activity has initiated BER, and a decrease in fluorescence is detected.

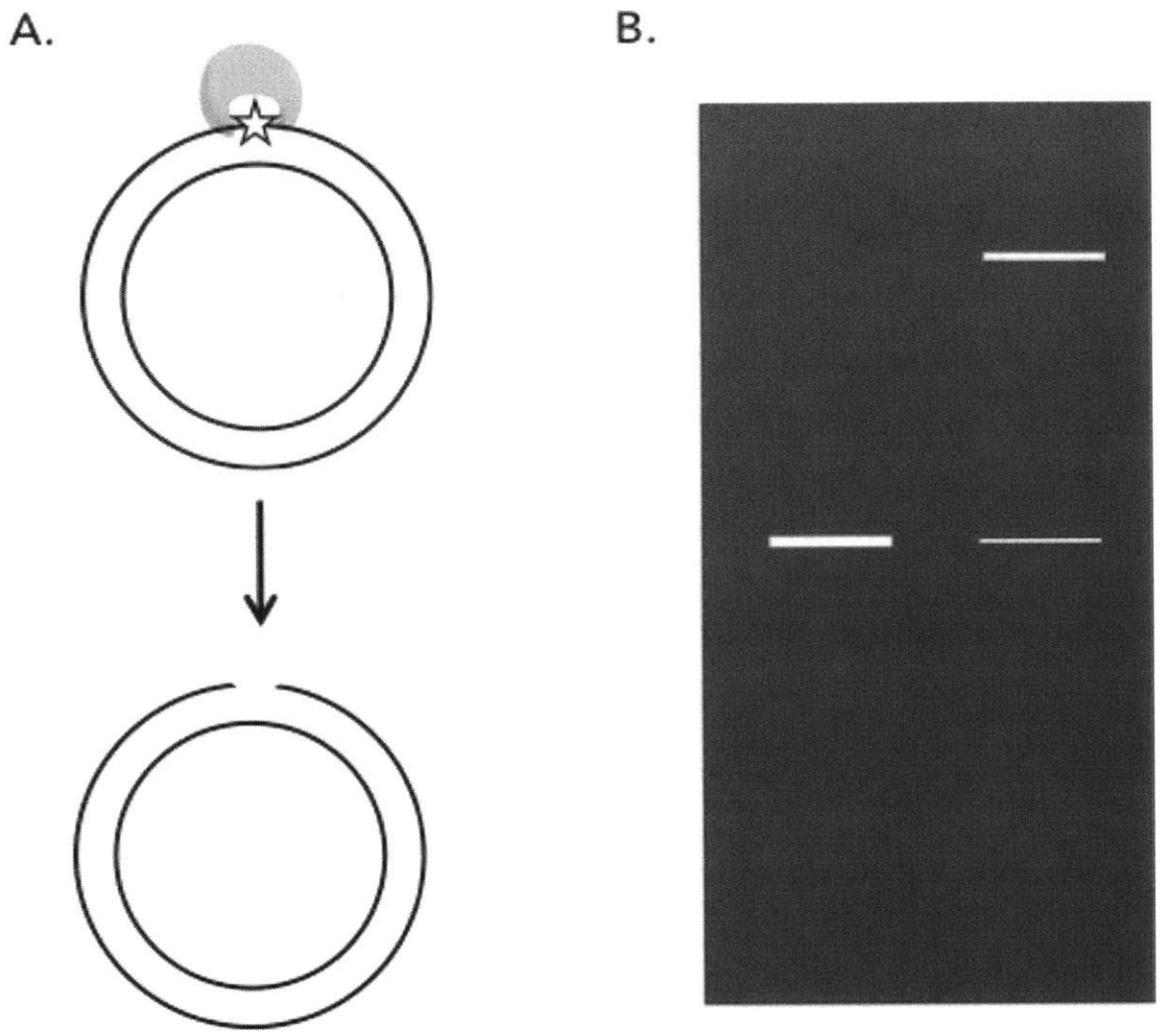


Figure 3.4: Plasmid-nicking gel-shift assay.

A) APE1 (gray crescent) generates a SSB at the site of an abasic site (star) in a plasmid (each DNA strand is represented by a black circle). Abasic sites are randomly distributed in the plasmid, and there may be more than one abasic site present. B) APE1 nicked plasmid is separated from closed circular plasmid and quantified using an agarose gel and densitometry.

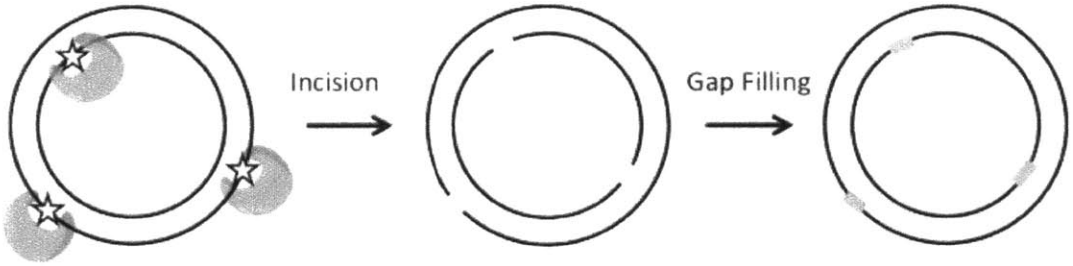


Figure 3.5: Microchip-plasmid-based assays.

Plasmids (each DNA strand is represented by a black circle) containing specific types of DNA damage (stars) are immobilized in a hydrogel. Glycosylase excision of damaged bases followed by enzymatic incision at resulting abasic sites results in a nicked plasmid intermediate. Gap-filling by DNA polymerase leads to incorporation of fluorescent nucleotides (represented by gray bars). Fluorescence intensity is used to quantitate repair.

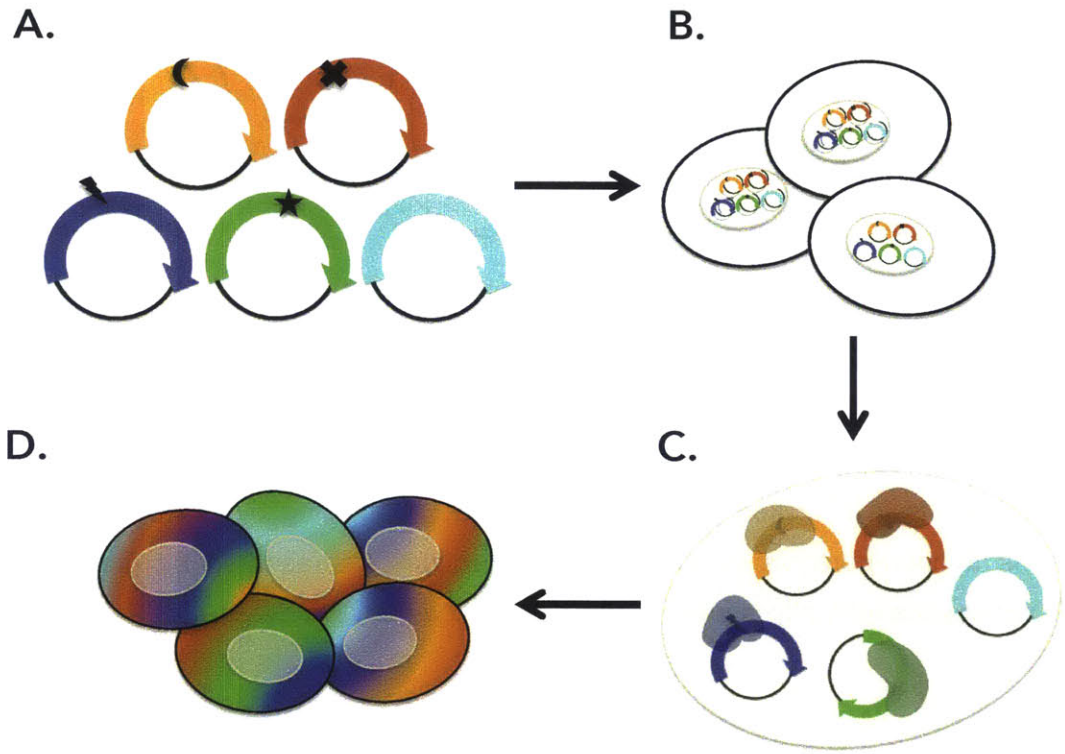


Figure 3.6: Fluorescence Multiplexed Host Cell Reactivation Assay (FM-HCR). A) A unique type of DNA damage (represented by various black shapes) is introduced *in vitro* into each of several reporter plasmids encoding different fluorescent reporter genes (represented by colored arrows). One reporter plasmid (here, cyan) is left undamaged and is used as a control for transfection efficiency. The presence of DNA damage affects the ability of the reporter plasmid to express a fluorescent reporter protein. In some cases, the DNA lesion blocks transcription, leading to *low* levels of expression of the fluorescent reporter protein in the absence of repair. In other reporter plasmids, lesion-induced transcriptional mutagenesis leads to *high* levels of fluorescent reporter expression in the absence of repair. B) A mixture of reporter plasmids is co-transfected into cells. A second transfection (not shown) is performed with the same mixture of plasmids without DNA damage. C) Once inside the cells, damaged plasmids are repaired by the host cell DNA repair machinery. D) Fluorescence intensity in each reporter channel is measured by flow cytometry.

The level of fluorescent protein expression is normalized to the undamaged (cyan) control. BER capacity is calculated from the percentage of fluorescent reporter expression in cells transfected with damaged plasmids relative to fluorescent reporter expression in cells transfected with undamaged plasmids.

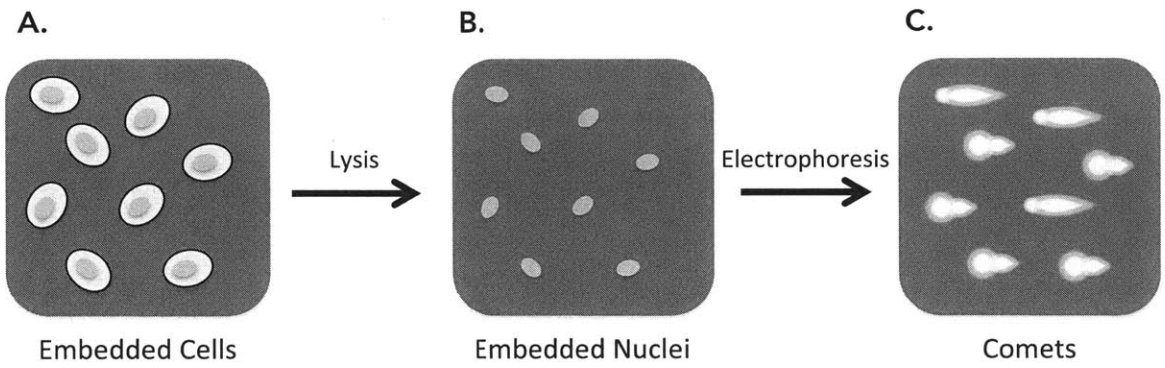


Figure 3.7: Comet Assay.

A) Live cells are embedded in agarose and, in some cases, treated with a DNA-damaging agent. When DNA-damaging agents are used, cells are allowed time to repair after exposure to DNA damage. B) After the desired incubation time for DNA repair, cells are lysed using a detergent and high salt solution, either under neutral pH or alkaline conditions. Alternately, isolated nuclei that have been treated with a DNA damaging agent in vitro can be embedded in agarose, and then incubated with lysates from the cells of interest. C) Electrophoresis is carried out either under neutral pH conditions or alkaline conditions. Strand breaks result in a loss of supercoiling that allows some DNA to migrate into the gel in a pattern resembling a comet (hence the name of the assay). DNA is visualized by fluorescent staining, and the percentage of DNA in the comet tail relative to the comet head is used to quantify the level of DNA damage in cells.

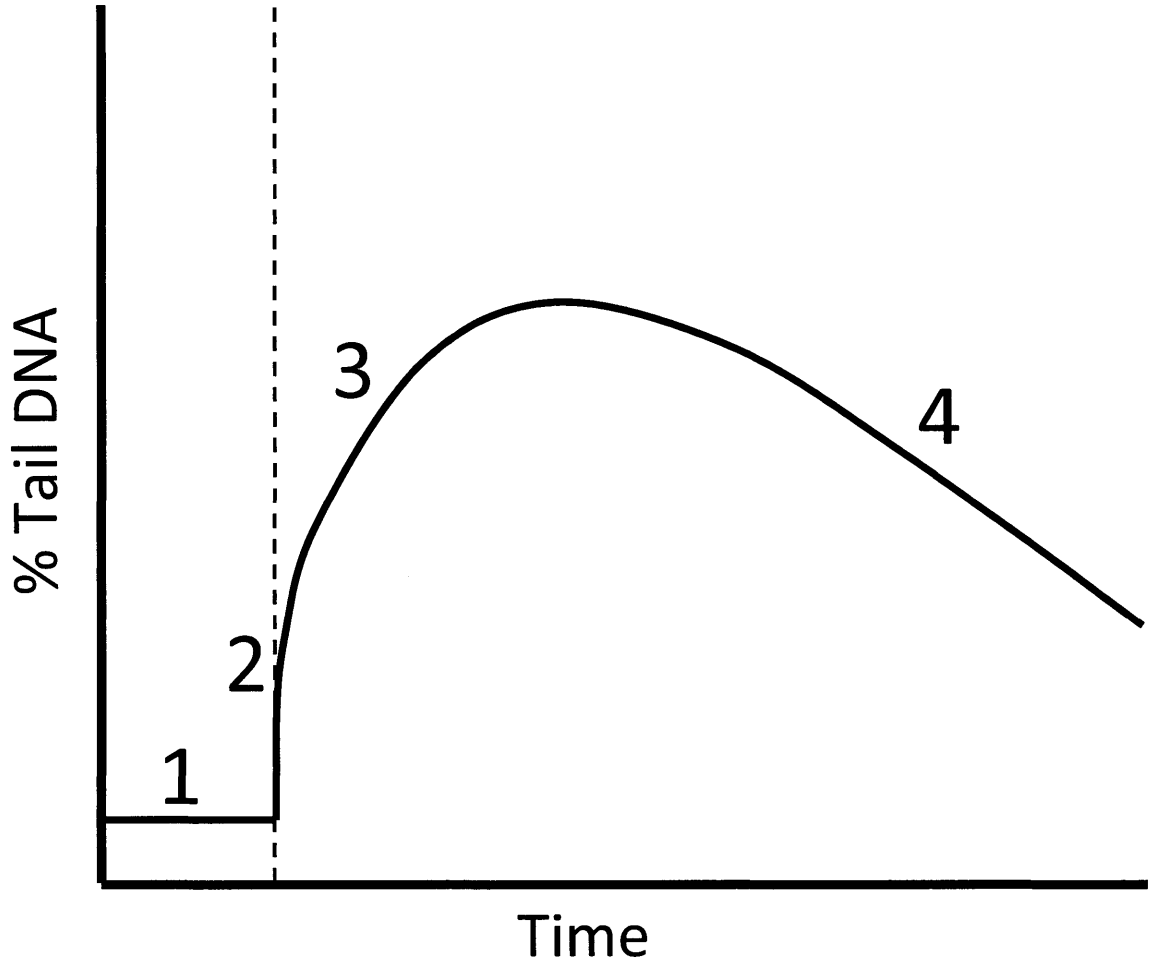


Figure 3.8: Comet assays yield a time-dependent signal that is influenced by many cellular processes.

Numbers represent features in a theoretical time course of % comet tail DNA. Cells from different individuals may vary in: the steady-state levels of DNA damage present in their DNA (1), the initial level of damage directly induced by an agent such as ionizing radiation (2), the rate and extent of repair intermediate accumulation that can include strand breaks and, for alkaline assays, abasic sites (3), and the rate at which repair is completed (4). Note that the initial level of DNA damage in region 2 is generally not generated instantaneously, as represented in the figure for illustrative purposes. Instead the initial damage occurs over a period of time that coincides with the increase in % tail DNA in region 3 due to repair initiation during exposure to the DNA-damaging agent.

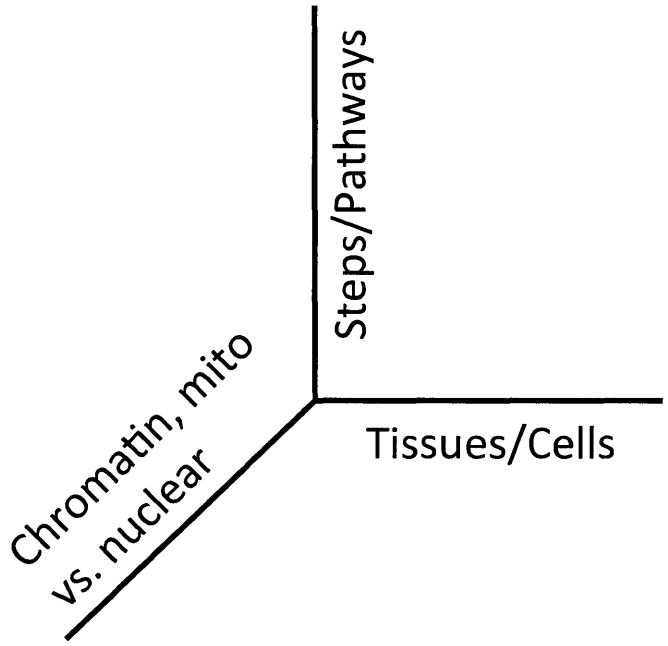


Figure 3.9. A coordinate system is used to represent the multidimensional nature of BER capacity.

Rather than a single number, BER capacity for a given individual can be conceptualized as a series of points in coordinate space defined by the particular pathway step, cell, chromatin environment, and cell compartment in which the repair is taking place. How many data points are required to define the DNA repair landscape for the purpose of comparing individuals remains to be determined experimentally.

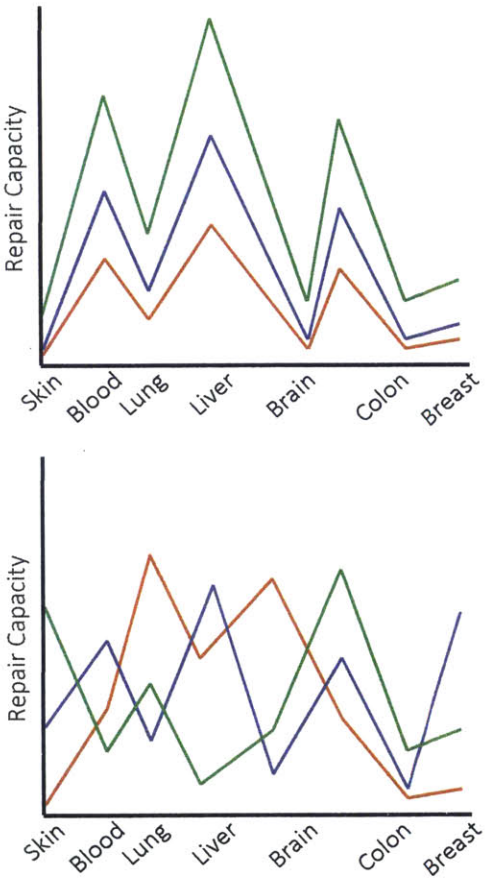


Figure 3.10. BER capacity variation in tissues

BER capacity might vary across tissues in a manner that preserves the relative repair capacity of different individuals (top), or it might vary in a manner that scrambles the relative repair capacity of the individuals (bottom). Green, blue and red lines connect points corresponding to hypothetical BER capacity in different tissues for three different individuals.

References

1. Nagel, Z.D., I.A. Chaim, and L.D. Samson, *Inter-individual variation in DNA repair capacity: A need for multi-pathway functional assays to promote translational DNA repair research.* DNA repair, 2014. **19**: p. 199-213.
2. Al-Tassan, N., N.H. Chmiel, J. Maynard, N. Fleming, A.L. Livingston, G.T. Williams, A.K. Hodges, D.R. Davies, S.S. David, J.R. Sampson, and J.P. Cheadle, *Inherited variants of MYH associated with somatic G:C-->T:A mutations in colorectal tumors.* Nat Genet, 2002. **30**(2): p. 227-32.
3. Imai, K., G. Slupphaug, W.I. Lee, P. Revy, S. Nonoyama, N. Catalan, L. Yel, M. Forveille, B. Kavli, H.E. Krokan, H.D. Ochs, A. Fischer, and A. Durandy, *Human uracil-DNA glycosylase deficiency associated with profoundly impaired immunoglobulin class-switch recombination.* Nat Immunol, 2003. **4**(10): p. 1023-8.
4. Senejani, A.G., Y. Liu, D. Kidane, S.E. Maher, C.J. Zeiss, H.J. Park, M. Kashgarian, J.M. McNiff, D. Zelterman, A.L. Bothwell, and J.B. Sweasy, *Mutation of POLB causes lupus in mice.* Cell Rep, 2014. **6**(1): p. 1-8.
5. Myrnes, B., K.E. Giercksky, and H. Krokan, *Interindividual variation in the activity of O6-methyl guanine DNA methyltransferase and uracil DNA glycosylase in human organs* Carcinogenesis, 1983. **4**(12): p. 1565-1568.
6. Dianov, G., A. Price, and T. Lindahl, *Generation of single-nucleotide repair patches following excision of uracil residues from DNA.* Molecular and Cellular Biology, 1992. **12**(4): p. 1605.
7. Leitner-Dagan, Y., Z. Sevilya, M. Pinchev, R. Kramer, D. Elinger, L.C. Roisman, H.S. Rennert, E. Schechtman, L. Freedman, G. Rennert, Z. Livneh, and T. Paz-Elizur, *N-Methylpurine DNA Glycosylase and OGG1 DNA Repair Activities: Opposite Associations With Lung Cancer Risk.* Journal of the National Cancer Institute, 2012. **104**(22): p. 1765-1769.
8. Paz-Elizur, T., D. Elinger, Y. Leitner-Dagan, S. Blumenstein, M. Krupsky, A. Berrebi, E. Schechtman, and Z. Livneh, *Development of an enzymatic DNA repair assay for molecular epidemiology studies: distribution of OGG activity in healthy individuals.* DNA Repair, 2007. **6**(1): p. 45-60.
9. Goto, M., K. Shinmura, Y. Nakabeppu, H. Tao, H. Yamada, T. Tsuneyoshi, and H. Sugimura, *Adenine DNA glycosylase activity of 14 human MutY homolog (MUTYH) variant proteins found in patients with colorectal polyposis and cancer.* Human mutation, 2010. **31**(11): p. E1861.
10. Molatore, S., M.T. Russo, V.G. D'Agostino, F. Barone, Y. Matsumoto, A.M. Albertini, A. Minoprio, P. Degan, F. Mazzei, and M. Bignami, *MUTYH mutations associated with familial adenomatous polyposis: functional characterization by a mammalian cell - based assay.* Human mutation, 2010. **31**(2): p. 159-166.
11. Dou, H., S. Mitra, and T.K. Hazra, *Repair of oxidized bases in DNA bubble structures by human DNA glycosylases NEIL1 and NEIL2.* Journal of Biological Chemistry, 2003. **278**(50): p. 49679-49684.

12. Liu, M., V. Bandaru, A. Holmes, A.M. Averill, W. Cannan, and S.S. Wallace, *Expression and purification of active mouse and human NEIL3 proteins.* Protein expression and purification, 2012. **84**(1): p. 130-139.
13. Kavli, B., O. Sundheim, M. Akbari, M. Otterlei, H. Nilsen, F. Skorpen, P.A. Aas, L. Hagen, H.E. Krokan, and G. Slupphaug, *hUNG2 is the major repair enzyme for removal of uracil from U: A matches, U: G mismatches, and U in single-stranded DNA, with hSMUG1 as a broad specificity backup.* Journal of Biological Chemistry, 2002. **277**(42): p. 39926-39936.
14. Yoon, J.H., S. Iwai, T.R. O' Connor, and G.P. Pfeifer, *Human thymine DNA glycosylase (TDG) and methyl - CpG - binding protein 4 (MBD4) excise thymine glycol (Tg) from a Tg: G mispair.* Nucleic acids research, 2003. **31**(18): p. 5399-5404.
15. Sevilya, Z., Y. Leitner-Dagan, M. Pinchev, R. Kremer, D. Elinger, F. Lejbkowicz, H.S. Rennert, L.S. Freedman, G. Rennert, T. Paz-Elizur, and Z. Livneh, *Development of APE1 enzymatic DNA repair assays: low APE1 activity is associated with increase lung cancer risk.* Carcinogenesis, 2015: p. bgv082.
16. McNeill, D.R., A. Narayana, H.-K. Wong, and D.M. Wilson 3rd, *Inhibition of Ape1 nuclease activity by lead, iron, and cadmium.* Environmental Health Perspectives, 2004. **112**(7): p. 799.
17. Lee, C.Y.I., J.C. Delaney, M. Kartalou, G.M. Lingaraju, A. Maor-Shoshani, J.M. Essigmann, and L.D. Samson, *Recognition and Processing of a New Repertoire of DNA Substrates by Human 3-Methyladenine DNA Glycosylase (AAG).* Biochemistry, 2009. **48**(9): p. 1850-1861.
18. Monden, Y., T. Arai, M. Asano, E. Ohtsuka, H. Aburatani, and S. Nishimura, *Human MMH (OGG1) type 1a protein is a major enzyme for repair of 8-hydroxyguanine lesions in human cells.* Biochemical and biophysical research communications, 1999. **258**(3): p. 605-610.
19. Bessho, T., R. Roy, K. Yamamoto, H. Kasai, S. Nishimura, K. Tano, and S. Mitra, *Repair of 8-hydroxyguanine in DNA by mammalian N-methylpurine-DNA glycosylase.* Proceedings of the National Academy of Sciences, 1993. **90**(19): p. 8901-8904.
20. Wilson, T.M., A. Vaisman, S.A. Martomo, P. Sullivan, L. Lan, F. Hanaoka, A. Yasui, R. Woodgate, and P.J. Gearhart, *MSH2–MSH6 stimulates DNA polymerase η, suggesting a role for A: T mutations in antibody genes.* The Journal of experimental medicine, 2005. **201**(4): p. 637-645.
21. Edwards, S.K., T. Ono, S. Wang, W. Jiang, R.M. Franzini, J.W. Jung, K.M. Chan, and E.T. Kool, *In Vitro Fluorogenic Real - Time Assay of the Repair of Oxidative DNA Damage.* ChemBioChem, 2015. **16**(11): p. 1637-1646.
22. Ono, T., S.K. Edwards, S. Wang, W. Jiang, and E.T. Kool, *Monitoring eukaryotic and bacterial UDG repair activity with DNA-multifluorophore sensors.* Nucleic acids research, 2013: p. gkt309.
23. Svilar, D., C. Vens, and R.W. Sobol, *Quantitative, real-time analysis of base excision repair activity in cell lysates utilizing lesion-specific molecular beacons.* Journal of visualized experiments: JoVE, 2012(66).

24. Maksimenko, A., A.A. Ishchenko, G. Sanz, J. Laval, R.H. Elder, and M.K. Saparbaev, *A molecular beacon assay for measuring base excision repair activities.* Biochemical and biophysical research communications, 2004. **319**(1): p. 240-246.
25. Mirbahai, L., R.M. Kershaw, R.M. Green, R.E. Hayden, R.A. Meldrum, and N.J. Hodges, *Use of a molecular beacon to track the activity of base excision repair protein OGG1 in live cells.* DNA repair, 2010. **9**(2): p. 144-152.
26. Sauvaigo, S., V. Guerniou, D. Rapin, D. Gasparutto, S. Caillat, and A. Favier, *An oligonucleotide microarray for the monitoring of repair enzyme activity toward different DNA base damage.* Analytical biochemistry, 2004. **333**(1): p. 182-192.
27. Guerniou, V., D. Rapin, J.-F. Millau, E. Bufflier, A. Favier, J. Cadet, and S. Sauvaigo, *Repair of oxidative damage of thymine by HeLa whole-cell extracts: simultaneous analysis using a microsupport and comparison with traditional PAGE analysis.* Biochimie, 2005. **87**(2): p. 151-159.
28. Redaelli, A., R. Magrassi, S. Bonassi, A. Abbondandolo, and G. Frosina, *AP endonuclease activity in humans: development of a simple assay and analysis of ten normal individuals.* Teratogenesis, carcinogenesis, and mutagenesis, 1998. **18**(1): p. 17-26.
29. Millau, J.-F., A.-L. Raffin, S. Caillat, C. Claudet, G. Arras, N. Ugolin, T. Douki, J.-L. Ravanat, J. Breton, T. Oddos, C. Dumontet, A. Sarasin, S. Chevillard, A. Faviera, and S. Sauvaigoa, *A microarray to measure repair of damaged plasmids by cell lysates.* Lab on a Chip, 2008. **8**(10): p. 1713-1722.
30. Philpott, S.M. and G.C. Buehring, *Defective DNA repair in cells with human T-cell leukemia/bovine leukemia viruses: role of tax gene.* Journal of the National Cancer Institute, 1999. **91**(11): p. 933-942.
31. Kassam, S.N. and A.J. Rainbow, *Deficient base excision repair of oxidative DNA damage induced by methylene blue plus visible light in xeroderma pigmentosum group C fibroblasts.* Biochemical and biophysical research communications, 2007. **359**(4): p. 1004-1009.
32. Wang, S., Z. Gong, R. Chen, Y. Liu, A. Li, G. Li, and J. Zhou, *JWA regulates XRCC1 and functions as a novel base excision repair protein in oxidative-stress-induced DNA single-strand breaks.* Nucleic acids research, 2009. **37**(6): p. 1936-1950.
33. Raetz, A.G., Y. Xie, S. Kundu, M.K. Brinkmeyer, C. Chang, and S.S. David, *Cancer-associated variants and a common polymorphism of MUTYH exhibit reduced repair of oxidative DNA damage using a GFP-based assay in mammalian cells.* Carcinogenesis, 2012. **33**(11): p. 2301-2309.
34. Nagel, Z.D., C.M. Margulies, I.A. Chaim, S.K. McRee, P. Mazzucato, A. Ahmad, R.P. Abo, V.L. Butty, A.L. Forget, and L.D. Samson, *Multiplexed DNA repair assays for multiple lesions and multiple doses via transcription inhibition and transcriptional mutagenesis.* Proceedings of the National Academy of Sciences, 2014. **111**(18): p. E1823-E1832.
35. Burns, J.A., K. Dreij, L. Cartularo, and D.A. Scicchitano, *O-6-Methylguanine induces altered proteins at the level of transcription in human cells.* Nucleic Acids Research, 2010. **38**(22): p. 8178-8187.

36. Løhr, M., A. Jensen, L. Eriksen, M. Grønbæk, S. Loft, and P. Møller, *Association between age and repair of oxidatively damaged DNA in human peripheral blood mononuclear cells.* Mutagenesis, 2015: p. gev031.
37. Humphreys, V., R.M. Martin, B. Ratcliffe, S. Duthie, S. Wood, D. Gunnell, and A.R. Collins, *Age-related increases in DNA repair and antioxidant protection: a comparison of the Boyd Orr Cohort of elderly subjects with a younger population sample.* Age and ageing, 2007. **36**(5): p. 521-526.
38. Azqueta, A. and A.R. Collins, *The essential comet assay: a comprehensive guide to measuring DNA damage and repair.* Arch Toxicol, 2013. **87**(6): p. 949-68.
39. Wood, D.K., D.M. Weingeist, S.N. Bhatia, and B.P. Engelward, *Single cell trapping and DNA damage analysis using microwell arrays.* Proceedings of the National Academy of Sciences, 2010. **107**(22): p. 10008-10013.
40. Collins, A., G. Koppen, V. Valdiglesias, M. Dusinska, M. Kruszewski, P. Møller, E. Rojas, A. Dhawan, I. Benzie, and E. Coskun, *The comet assay as a tool for human biomonitoring studies: the ComNet project.* Mutation Research/Reviews in Mutation Research, 2014. **759**: p. 27-39.
41. Calvo, J.A., C.A. Moroski-Erkul, A. Lake, L.W. Eichinger, D. Shah, I. Jhun, P. Limsirichai, R.T. Bronson, D.C. Christiani, L.B. Meira, and L.D. Samson, *Aag DNA Glycosylase Promotes Alkylation-Induced Tissue Damage Mediated by Parp1.* Plos Genetics, 2013. **9**(4).
42. Crosbie, P.A.J., A.J. Watson, R. Agius, P.V. Barber, G.P. Margison, and A.C. Povey, *Elevated N3-methylpurine-DNA glycosylase DNA repair activity is associated with lung cancer.* Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis, 2012.
43. Leitner-Dagan, Y., Z. Sevilya, M. Pinchev, R. Kremer, D. Elinger, H.S. Rennert, E. Schechtman, L. Freedman, G. Rennert, and Z. Livneh, *Enzymatic MPG DNA repair assays for two different oxidative DNA lesions reveal associations with increased lung cancer risk.* Carcinogenesis, 2014. **35**(12): p. 2763-2770.
44. Janssen, K., K. Schlink, W. Götte, B. Hippler, B. Kaina, and F. Oesch, *DNA repair activity of 8-oxoguanine DNA glycosylase 1 (OGG1) in human lymphocytes is not dependent on genetic polymorphism Ser 326/Cys 326.* Mutation Research/DNA Repair, 2001. **486**(3): p. 207-216.
45. Paz-Elizur, T., M. Krupsky, S. Blumenstein, D. Elinger, E. Schechtman, and Z. Livneh, *DNA repair activity for oxidative damage and risk of lung cancer.* Journal of the National Cancer Institute, 2003. **95**(17): p. 1312-1319.
46. Myrnes, B., K.-E. Giercksky, and H. Krokan, *Interindividual variation in the activity of O6-methyl guanine-DNA methyltransferase and uracil-DNA glycosylase in human organs.* Carcinogenesis, 1983. **4**(12): p. 1565-1568.
47. Sevilya, Z., Y. Leitner-Dagan, M. Pinchev, R. Kremer, D. Elinger, H.S. Rennert, E. Schechtman, L.S. Freedman, G. Rennert, and T. Paz-Elizur, *Low integrated DNA repair score and lung cancer risk.* Cancer Prevention Research, 2013: p. canprevres. 0318.2013.
48. Gaivão, I., A. Piasek, A. Brevik, S. Shaposhnikov, and A.R. Collins, *Comet assay-based methods for measuring DNA repair in vitro; estimates of inter-and intra-individual variation.* Cell biology and toxicology, 2009. **25**(1): p. 45-52.

49. Caple, F., E.A. Williams, A. Spiers, J. Tyson, B. Burtle, A.K. Daly, J.C. Mathers, and J.E. Hesketh, *Inter-individual variation in DNA damage and base excision repair in young, healthy non-smokers: effects of dietary supplementation and genotype.* British Journal of Nutrition, 2010. **103**(11): p. 1585-1593.
50. Chen, S.-K., W.A. Hsieh, M.-H. Tsai, C.-C. Chen, A.I. Hong, Y.-H. Wei, and W.P. Chang, *Age-associated decrease of oxidative repair enzymes, human 8-oxoguanine DNA glycosylases (hOgg1), in human aging.* Journal of radiation research, 2003. **44**(1): p. 31-35.
51. Aka, P., R. Mateuca, J.-P. Buchet, H. Thierens, and M. Kirsch-Volders, *Are genetic polymorphisms in OGG1, XRCC1 and XRCC3 genes predictive for the DNA strand break repair phenotype and genotoxicity in workers exposed to low dose ionising radiations?* Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis, 2004. **556**(1): p. 169-181.
52. Vodicka, P., R. Stetina, V. Polakova, E. Tulupova, A. Naccarati, L. Vodickova, R. Kumar, M. Hanova, B. Pardini, and J. Slyskova, *Association of DNA repair polymorphisms with DNA repair functional outcomes in healthy human subjects.* Carcinogenesis, 2006. **28**(3): p. 657-664.
53. Decordier, I., K.V. Loock, and M. Kirsch-Volders, *Phenotyping for DNA repair capacity.* Mutat Res, 2010. **705**(2): p. 107-29.
54. Soares, J.P., A.M. Silva, S. Fonseca, M.M. Oliveira, F. Peixoto, I. Gaivao, and M.P. Mota, *How can age and lifestyle variables affect DNA damage, repair capacity and endogenous biomarkers of oxidative stress?* Experimental Gerontology, 2015. **62**: p. 45-52.
55. Gorbunova, V., A. Seluanov, Z. Mao, and C. Hine, *Changes in DNA repair during aging.* Nucleic Acids Res, 2007. **35**(22): p. 7466-74.
56. Paz-Elizur, T., R. Ben-Yosef, D. Elinger, A. Vexler, M. Krupsky, A. Berrebi, A. Shani, E. Schechtman, L. Freedman, and Z. Livneh, *Reduced repair of the oxidative 8-oxoguanine DNA damage and risk of head and neck cancer.* Cancer research, 2006. **66**(24): p. 11683-11689.
57. Pons, B., A. Belmont, G. Masson-Genteuil, V. Chapuis, T. Oddos, and S. Sauvaigo, *Age-associated modifications of Base Excision Repair activities in human skin fibroblast extracts.* Mechanisms of ageing and development, 2010. **131**(11-12): p. 661-665.
58. Rossi, O., F. Carrozzino, and G. Frosina, *Age - independency of AP site incision capacity in women.* Environmental and molecular mutagenesis, 1999. **34**(4): p. 256-259.
59. Xu, Z., L. Zhang, W.J. Zhang, D. Meng, H.X. Zhang, Y. Jiang, X.J. Xu, M. Van Meter, A. Seluanov, V. Gorbunova, and Z.Y. Mao, *SIRT6 rescues the age related decline in base excision repair in a PARP1-dependent manner.* Cell Cycle, 2015. **14**(2): p. 269-276.
60. Olive, P.L., *The role of DNA single- and double-strand breaks in cell killing by ionizing radiation.* Radiat Res, 1998. **150**(5 Suppl): p. S42-51.
61. Vodicka, P., R. Kumar, R. Stetina, L. Musak, P. Soucek, V. Haufroid, M. Sasiadek, L. Vodickova, A. Naccarati, and J. Sedikova, *Markers of individual susceptibility*

and DNA repair rate in workers exposed to xenobiotics in a tire plant. Environmental and molecular mutagenesis, 2004. **44**(4): p. 283.
62. Vodicka, P., R. Kumar, R. Stetina, S. Sanyal, P. Soucek, V. Haufroid, M. Dusinska, M. Kuricova, M. Zamecnikova, and L. Musak, *Genetic polymorphisms in DNA repair genes and possible links with DNA repair rates, chromosomal aberrations and single-strand breaks in DNA.* Carcinogenesis, 2004. **25**(5): p. 757-763.
63. Tomasetti, M., R. Alleva, B. Borgui, and A.R. Collins, *In vivo supplementation with coenzyme Q10 enhances the recovery of human lymphocytes from oxidative DNA damage.* The FASEB Journal, 2001. **15**(8): p. 1425-1427.
64. Collins, A.R., V. Harrington, J. Drew, and R. Melvin, *Nutritional modulation of DNA repair in a human intervention study.* Carcinogenesis, 2003. **24**(3): p. 511-515.
65. Brevik, A., A. Karlsen, A. Azqueta, A.T. Estaban, R. Blomhoff, and A. Collins, *Both base excision repair and nucleotide excision repair in humans are influenced by nutritional factors.* Cell biochemistry and function, 2011. **29**(1): p. 36-42.
66. Cash, S.W., S.A. Beresford, T.L. Vaughan, P.J. Heagerty, L. Bernstein, E. White, and M.L. Neuhouser, *Recent physical activity in relation to DNA damage and repair using the comet assay.* Journal of physical activity & health, 2014. **11**(4): p. 770.
67. Rossi, O., F. Carrozzino, E. Cappelli, F. Carli, and G. Frosina, *Analysis of repair of abasic sites in early onset breast cancer patients.* Int J Cancer, 2000. **85**(1): p. 21-6.
68. Rossi, O., R. Filiberti, M. Neri, S. Biggi, L. Satragno, R. Puntoni, and G. Frosina, *Incision of AP sites in lung cancer patients: a pilot study.* Teratogenesis Carcinogenesis and Mutagenesis, 2002. **22**(4): p. 251-256.
69. Gackowski, D., E. Speina, M. Zielinska, J. Kowalewski, R. Rozalski, A. Siomek, T. Paciorek, B. Tudek, and R. Olinski, *Products of oxidative DNA damage and repair as possible biomarkers of susceptibility to lung cancer.* Cancer research, 2003. **63**(16): p. 4899-4902.
70. Obtułowicz, T., M. Swoboda, E. Speina, D. Gackowski, R. Rozalski, A. Siomek, J. Janik, B. Janowska, J.M. Cieśla, and A. Jawien, *Oxidative stress and 8-oxoguanine repair are enhanced in colon adenoma and carcinoma patients.* Mutagenesis, 2010: p. geq028.
71. Speina, E., M. Zielińska, A. Barbin, D. Gackowski, J. Kowalewski, M.A. Graziewicz, J.A. Siedlecki, R. Oliński, and B. Tudek, *Decreased repair activities of 1, N6-ethenoadenine and 3, N4-ethenocytosine in lung adenocarcinoma patients.* Cancer research, 2003. **63**(15): p. 4351-4357.
72. Lovell, M.A., C.S. Xie, and W.R. Markesbery, *Decreased base excision repair and increased helicase activity in Alzheimer's disease brain.* Brain Research, 2000. **855**(1): p. 116-123.
73. Shao, C., S. Xiong, G.-M. Li, L. Gu, G. Mao, W.R. Markesbery, and M.A. Lovell, *Altered 8-oxoguanine glycosylase in mild cognitive impairment and late-stage Alzheimer's disease brain.* Free Radical Biology and Medicine, 2008. **45**(6): p. 813-819.

74. Weissman, L., D.G. Jo, M.M. Sorensen, N.C. de Souza-Pinto, W.R. Markesbery, M.P. Mattson, and V.A. Bohr, *Defective DNA base excision repair in brain from individuals with Alzheimer's disease and amnestic mild cognitive impairment.* Nucleic Acids Res, 2007. **35**(16): p. 5545-55.
75. Marcon, G., G. Tell, L. Perrone, R. Garbelli, F. Quadrifoglio, F. Tagliavini, and G. Giaccone, *APE1/Ref-1 in Alzheimer's disease: An immunohistochemical study.* Neuroscience Letters, 2009. **466**(3): p. 124-127.
76. Canugovi, C., R.A. Shamanna, D.L. Croteau, and V.A. Bohr, *Base excision DNA repair levels in mitochondrial lysates of Alzheimer's disease.* Neurobiol Aging, 2014. **35**(6): p. 1293-300.
77. Leandro, G.S., R.R. Lobo, D.V. Oliveira, J.C. Moriguti, and E.T. Sakamoto-Hojo, *Lymphocytes of patients with Alzheimer's disease display different DNA damage repair kinetics and expression profiles of DNA repair and stress response genes.* Int J Mol Sci, 2013. **14**(6): p. 12380-400.
78. Mueller, M.M. and N.E. Fusenig, *Friends or foes - bipolar effects of the tumour stroma in cancer.* Nat Rev Cancer, 2004. **4**(11): p. 839-49.
79. van der Meer, A.D. and A. van den Berg, *Organs-on-chips: breaking the in vitro impasse.* Integrative Biology, 2012. **4**(5): p. 461-470.
80. Kazak, L., A. Reyes, and I.J. Holt, *Minimizing the damage: repair pathways keep mitochondrial DNA intact.* Nature Reviews Molecular Cell Biology, 2012. **13**(10): p. 659-671.
81. Alexeyev, M., I. Shokolenko, G. Wilson, and S. LeDoux, *The maintenance of mitochondrial DNA integrity--critical analysis and update.* Cold Spring Harb Perspect Biol, 2013. **5**(5): p. a012641.
82. Iida, T., A. Furuta, K. Nishioka, Y. Nakabeppu, and T. Iwaki, *Expression of 8-oxoguanine DNA glycosylase is reduced and associated with neurofibrillary tangles in Alzheimer's disease brain.* Acta Neuropathologica, 2002. **103**(1): p. 20-25.
83. Nunnari, J. and A. Suomalainen, *Mitochondria: In Sickness and in Health.* Cell, 2012. **148**(6): p. 1145-1159.
84. Allan, J.M., B.P. Engelward, A.J. Dreslin, M.D. Wyatt, M. Tomasz, and L.D. Samson, *Mammalian 3-methyladenine DNA glycosylase protects against the toxicity and clastogenicity of certain chemotherapeutic DNA cross-linking agents.* Cancer Res, 1998. **58**(17): p. 3965-73.
85. Kovtun, I.V. and C.T. McMurray, *Crosstalk of DNA glycosylases with pathways other than base excision repair.* DNA Repair (Amst), 2007. **6**(4): p. 517-29.
86. Bernardes de Jesus, B.M., M. Bjoras, F. Coin, and J.M. Egly, *Dissection of the molecular defects caused by pathogenic mutations in the DNA repair factor XPC.* Mol Cell Biol, 2008. **28**(23): p. 7225-35.
87. Parlanti, E., M. D'Errico, P. Degan, A. Calcagnile, A. Zijno, I. van der Pluijm, G.T. van der Horst, D.S. Biard, and E. Dogliotti, *The cross talk between pathways in the repair of 8-oxo-7,8-dihydroguanine in mouse and human cells.* Free Radic Biol Med, 2012. **53**(11): p. 2171-7.
88. Fu, D. and L.D. Samson, *Direct repair of 3,N(4)-ethenocytosine by the human ALKBH2 dioxygenase is blocked by the AAG/MPG glycosylase.* DNA Repair (Amst), 2012. **11**(1): p. 46-52.

89. Guo, J., P.C. Hanawalt, and G. Spivak, *Comet-FISH with strand-specific probes reveals transcription-coupled repair of 8-oxoGuanine in human cells.* Nucleic Acids Res, 2013. **41**(16): p. 7700-12.
90. Chou, W.C., L.Y. Hu, C.N. Hsiung, and C.Y. Shen, *Initiation of the ATM-Chk2 DNA damage response through the base excision repair pathway.* Carcinogenesis, 2015. **36**(8): p. 832-40.
91. Andrew, A.S., J.L. Burgess, M.M. Meza, E. Demidenko, M.G. Waugh, J.W. Hamilton, and M.R. Karagas, *Arsenic exposure is associated with decreased DNA repair in vitro and in individuals exposed to drinking water arsenic.* Environmental Health Perspectives, 2006. **114**(8): p. 1193-1198.
92. Staruchova, M., A.R. Collins, K. Volkovova, C. Mislanová, Z. Kovacikova, J. Tulinska, A. Kocan, L. Staruch, L. Wsolova, and M. Dusinska, *Occupational exposure to mineral fibres. Biomarkers of oxidative damage and antioxidant defence and associations with DNA damage and repair.* Mutagenesis, 2008. **23**(4): p. 249-260.
93. Palombo, P., M. Moreno-Villanueva, and A. Mangerich, *Day and night variations in the repair of ionizing-radiation-induced DNA damage in mouse splenocytes.* DNA Repair, 2015. **28**: p. 37-47.
94. Dušinská, M., Z. Dzupinková, L. Wsólová, V. Harrington, and A.R. Collins, *Possible involvement of XPA in repair of oxidative DNA damage deduced from analysis of damage, repair and genotype in a human population study.* 2006.
95. Slyskova, J., Y. Lorenzo, A. Karlsen, M.H. Carlsen, V. Novosadova, R. Blomhoff, P. Vodicka, and A.R. Collins, *Both genetic and dietary factors underlie individual differences in DNA damage levels and DNA repair capacity.* DNA repair, 2014. **16**: p. 66-73.

Chapter IV: Quantitating Inter-individual Differences in Cellular DNA Base Excision Repair Capacity – *in vivo* measurements of DNA glycosylases and APE1 activities

Isaac A. Chaim, Zachary D. Nagel, Jennifer J. Jordan
Leona D. Samson

Contributions: I.A.C. and L.D.S. designed experiments. I.A.C. conducted all experiments unless otherwise stated. I.A.C and Z.D.N. conceived the OGG1 reporter. J.J.J knocked out and overexpressed MUTYH and performed corresponding *in vitro* glycosylase assays I.A.C. analyzed data.

Table of Contents

Introduction

The integrity of our DNA is challenged by roughly 100,000 lesions per cell per day [1, 2]. These challenges arise endogenously, from products of our own cellular metabolic processes, as well as exogenously, from environmental damaging agents. Failure to cope with DNA damage can lead to cancer, degenerative disease and premature aging [3-6]. Evolution has provided cells with a variety of mechanisms that have allowed the repair DNA of damage and amelioration of its consequences.

Amongst these repair mechanisms is Base Excision Repair (BER). BER is known for the recognition and repair of small, non-bulky DNA lesions that are products of base alkylation, deamination or oxidation [6]. In its most simple form BER involves five mains steps: 1) recognition and removal of a damaged base by a DNA glycosylase, forming an abasic site, 2) incision of the phosphodiester DNA backbone at the abasic site, 3) processing and removal of remaining sugar moieties, 4) gap-filling by a DNA polymerase and 5) nick-sealing by a DNA ligase. BER can be further divided into short- and long-patch; the main difference between short- and long-patch repair are the enzymes that trim the single-strand break, the number of nucleotides incorporated into the patch and, the polymerases that perform this step (**Figure 4.1**) [2].

Only a few deficiencies in the BER pathway have been causally linked with human genetic disorders, with MUTYH and UNG being the only glycosylase examples. Deficiencies in MUTYH and UNG lead to MUTYH-associated polyposis [7] and Hyper IgM-Syndrome [8], respectively. Increasing evidence suggests that small variations in the activities of these and other BER enzymes, such as AAG, OGG1 and APE1, can be correlated with disease risk [9-13]. This being the case, there is a clear need in the field for assays that can reliably measure small inter-individual differences in BER capacity, preferably in a fast and high-throughout manner that would enable epidemiological studies.

Importantly, BER is involved in the repair of different lesions caused cancer chemotherapeutic agents [3, 14]. As such, being able to determine the "BER-status" of both tumor and healthy tissue, could potentially allow personalized disease treatment by targeting more accurately deficiencies of tumor cells, while minimizing damage in healthy tissues.

A variety of methods have been used to measure BER activity. Labeled oligonucleotides containing lesions repaired by specific BER enzymes have been the workhorse of functional assays in recent years. Specifically, methods have been developed for measuring the *in vitro* activity of all human DNA glycosylases (AAG [15]; OGG1 [16]; MUTYH [17, 18]; NEIL1, NEIL2 and NTHL1 [19] ; NEIL3 [19]; UNG and SMUG1 [20]; MBD4 and TDG [21]), incision by APE1 [22] as well as downstream steps such as gap filling, flap endonuclease and DNA ligase activity [23]. Molecular Beacons for assessment of 8oxoG, uracil and abasic site repair *in vitro*, and in some instances *in vivo*, have also gained some recent interest [24-26]. Microchip-immobilized fluorescent reporter systems have emerged as an *in vitro* method to probe the activity of cell lysates on a variety of BER substrates simultaneously [27, 28].

Even though these methods achieve the ultimate goal of measuring BER, each one of them has its own drawbacks. *In vitro* assays are based on cell lysates, which might not recapitulate physiological DNA repair conditions. Moreover single-cell resolution and intracellular compartmentalization are obviously lost upon cell lysis. Furthermore, with the exception of the aforementioned microchip-assay, multiplexing for different glycosylases and downstream BER steps is not straightforward. In an effort to overcome these problems we recently described a modified host cell reactivation (HCR) assay, specifically a fluorescence-based multiplex flow-cytometric HCR (or FM-HCR) that allows for *in vivo* multiplexed DNA repair capacity measurements of different DNA repair pathways [29]. Here we build upon this method and report on the development of fluorescent plasmid reporters that can be used simultaneously, in a high-throughput manner to

measure *in vivo* the activity of at least four different DNA glycosylases and the abasic site endonuclease, APE1. These methods allow us to assess inter-individual variations in BER by measuring BER capacity in a panel of 24 cell lines derived from healthy individuals. Finally, we use these measurements to build mathematical models with the goal of predicting cellular sensitivity to clinically relevant DNA damaging agents know to elicit a BER response.

Materials and Methods

Plasmids

As described previously [29], AmCyan, EGFP, mOrange, mPlum and tagBFP reporter genes were subcloned into the pmaxCloning Vector (Lonza) between the KpnI and SacI restriction sites in the multiple cloning site. Plasmids were amplified in *E.coli* DH5α (Invitrogen) and purified using Qiagen endotoxin-free Maxi and Giga kits.

Substrates containing site-specific DNA damage reporting via transcriptional mutagenesis and transcriptional blockage

Non-fluorescent variants of the different reporter plasmids containing a single mutation in a site coding for their respective chromophores were identified and made via standard QuikChange site-directed mutagenesis (Agilent Technologies) (**Table 4.1**).

In order to produce single stranded plasmid (ssPlasmid) previously described methods were followed with minor modifications [29]. Reporter plasmids were nicked with either Nb.BtsI or Nt.BtsI (New England Biolabs, depending on the lesion containing strand, see **Table 4.1**). The nicked strand then was digested with exonuclease III, and the remaining ssPlasmid was purified by using a 1% agarose gel. Fifteen picomoles of the respective phosphorylated lesion-

containing-oligonucleotide (**Table 4.1**) were combined with 3.2 µg of the corresponding single-stranded plasmid (1/200, oligo/ssPlasmid molar ratio) in 1X Pfu polymerase AD buffer (Agilent Biotechnologies) in a final volume of 46 µL. The mixture was heated to 85 °C in a thermal cycler for 6 min, and then allowed to anneal by cooling to 40 °C at 1 °C per minute. To extend the primer, 5 units of Pfu polymerase AD (Agilent) and 0.4 µM dNTP were added. The incubation parameters used for each ssDNA-plasmids/oligonucleotide combination are shown in **Table 4.1**. The mixture was then cleaned up with a Qiagen PCR Purification kit column, eluted in EB buffer and subsequently combined with 1X Ligase Buffer (New England Biolabs - NEB), 0.4 µM dNTP, 1 µM ATP, 10ng/µL BSA, 1.5 units T4 DNA Polymerase and 80 units T4 DNA Ligase (both NEB) and incubated for an additional hour at 16 °C to yield closed circular plasmid. Finally, the product was purified from a 1% agarose gel using a Qiagen gel extraction kit.

Enzymatic plasmid treatment for lesion detection

To test for the presence of uracil in BFP-A191G-U and its negative control (BFP-WT), 150 ng of the plasmids were incubated with 15 units UDG (NEB) in 1X UDG buffer for 1 hour at 37 °C. Following cleanup with a Qiagen PCR Purification kit column, plasmids were incubated with 10 units APE1 (NEB) in 1X #4 buffer for 1 hour at 37 °C, followed by 20 min at 65 °C for heat-inactivation. Products were run in a 1% agarose gel for visualization. Appropriate no enzyme/no buffer controls were tested under the same conditions.

To test for the presence of Hypoxanthine in GFP-C289T-Hx and its negative controls (GFP-WT and GFP-C289T), 150 ng of the plasmids were incubated with 10 units ApaLI (NEB) in 1X Cutsmart buffer for 1 hour at 37 °C, followed by 20 min at 65 °C for heat-inactivation. Products were run in a 1% agarose gel for visualization.

To test for the presence of THF in GFP-617-THF and mOrange-A215C-THF and their negative controls (GFP-WT and mOrange-WT, respectively), 150 ng of the plasmids were incubated with 10 units APE1 (NEB) in 1X NEB#4 buffer for 1 hour at 37 °C, followed by 20 min at 65 °C for heat-inactivation. Products were run in a 1% agarose gel for visualization. To test for the presence of 8oxoG in mOrange-A215C-8oxoG and mPlum-T202WT-8oxoG (+) and their negative controls (mOrange-WT and mPlum-WT, respectively), 150 ng of the plasmids were incubated with 8 units Fpg (NEB) in 1X #1 buffer and 1X BSA for 1 hour at 37 °C, followed by 20 min at 65 °C for heat-inactivation. Products were run in a 1% agarose gel for visualization.

Cell culture

Cell lines and their respective culture conditions are detailed in **Table 4.2**.

MUTYH knockdowns

MUTYH was knockdown in HCT116 + Chromosome 3 cells as previously described [30]. shRNAs expressed in a lentiviral plasmid (pGIPZ) were purchased from Open Biosystems to target *MUTYH* transcript (v1:#RHS4430-98904053 and v2:#RHS4430-99140608). Knockdown cells were compared with cells expressing a nontargeting shRNA (#RHS4346). Virus was generated in 293T cells using packaging plasmids (psPAX2 and pMD2.G, Addgene). Cell line was infected with virus and stable clones selected using Puromycin.

MUTYH nuclear overexpression

Nuclear *MUTYH* isoform 4 cDNA was cloned into a retroviral plasmid (pBABE, C-terminal Flag-tagged). MUTYH Overexpressing HCT116 + Chromosome 3 cells were compared with cells expressing empty vector. Virus was generated in

Phoenix 293 cells. Cell line was infected with virus and stable clones selected using Puromycin.

In vitro MUTYH glycosylase assays

Cells were sonicated in MUTYH glycosylase dilution buffer (30 mM Tris pH 7.5, 1 mM EDTA, 1 mM β-mercaptoethanol, 50mM NaCl and 30% glycerol) with protease inhibitors. Protein concentration was measured using micro BCA Kit (Pierce). Glycosylase assays were performed as previously published [31]. A double-stranded oligonucleotide containing a [^{32}P]γ-labeled strand (5′-TTGGGGAATG**A**GTCAGGCCAC-3′) and a non-labeled strand (5′-GGTGGCCTGAC**8oxoG**CATTCCCCAA-3′) was incubated with an amount of extract determined to be in the linear range for activity at 37˚C for 60 min. The resulting AP sites were cleaved by incubation with 0.1 M NaOH at 75 ˚C for 15 min. The aliquots were then heat denatured and subjected to 7 M urea–10% polyacrylamide gel electrophoresis. A phosphorimager was used to visualize and quantitate MUTYH DNA glycosylase activity.

DNA Repair Assays Transfections

Electroporation

For suspension cells, 3 x 10^6 cells in 100 µL complete medium were combined with a reporter plasmid mixture (**Table 4.3**). Cells were electroporated using a 96-well Bio-Rad MXcell gene pulser, with an exponential waveform at 260 V and 950 µF. Following electroporation, 100 µL complete medium were added to each well in the electroporation plate and gently mixed. The electroporation mix was transferred to a 24-well cell culture plate prefilled with 1.3 mL of complete medium and incubated at 37 °C and 5% CO_2. Following 18 hours, cells were spun down for 5 min at 300 g, resuspended in 250 µL of complete media containing TO-PRO-3 and transferred to a 96-well plate for flow cytometry.

Lipofection

For adherent cells, 150,000 cells were plated in 6-well cell culture plates a day before transfection in order to achieve 50-80 % confluency the day of transfection. Cells were transfected with Lipofectamine LTX (Life Technologies) according to the manufacturer's instructions. Briefly 2.5 µg of total plasmid DNA (**Table 4.3**) were mixed with 2.5 µL Plus reagent and Opti-MEM, further mixed with 6.2 µL lipofectamine LTX in Opti-MEM and incubated at room temperature for 5 min before adding 200µL of the transfection reaction on top of the cells. Lipofection conditions were scaled down to 12- and 24-well plates when necessary. Transfected cells were incubated for 18 hours at 37 °C and 5% CO_2. Following incubation cells were trypsinized and resuspended in a total of 500 µL of complete media containing TO-PRO-3 and transferred to 75 mm Falcon tubes with Cell Strainer Caps (Fisher Scientific).

Flow cytometry

Cells suspended in culture media were analyzed for fluorescence on a BD LSR II cytometer running FACSDiva software. Cell debris, doublets, and aggregates were excluded based on their side- and forward-scatter properties. TO-PRO–3 was added to cells 5–10 min before analysis and was used to exclude dead cells from the analysis. The following fluorophores and their corresponding detectors (in parentheses) were used: tagBFP (Pacific Blue), AmCyan (AmCyan), EGFP (FITC), mOrange (phycoerythrin; PE), mPlum (PE-Cy5-5), and TO-PRO–3 (allophycocyanin; APC). Compensation was set by using single-color controls. Regions corresponding to cells positive for each of the five fluorescent proteins were established by using single-color dropout controls. For reporters that required compensation in more than one detector channel, fluorescence in the reporter channel was plotted separately against each of the channels requiring compensation. Using these plots, both single controls and the dropout control (in

which the reporter of interest was excluded from the transfection) were used to establish regions corresponding to positive cells. A threshold of at least 30 fluorescent events for each reporter was established. Samples with fewer than 30 events were not considered and the particular experiment was repeated.

Calculation of Percent Fluorescent Reporter Expression.

Every experimental setup consisted of two sets of transfections: A control transfection (CT) and a sample transfection (ST) containing one or more reporters with DNA lesions. Both transfections included the same color combination with the same undamaged reporter to normalize each set for transfection efficiency.

Fluorescence Index (FI) for a given reporter within one transfection was calculated as follows:

$$FI = \frac{C_F \times MFI}{C_L}$$

where C_F is the number of positive fluorescent cells for that given fluorophore, *MFI* is the mean fluorescence intensity of the C_F, and C_L is the total number of live cells.

The normalized fluorescence index for a given reporter FI^O was calculated as follows:

$$FI^O = \frac{FI^n}{FI^E}$$

where FI^n corresponds to the *FI* of a reporter normalized to the *FI* of the transfection efficiency normalization plasmid, FI^E.

Normalized reporter expression from a sample transfection, FI^{O}_{ST}, and that from the same reporter plasmid in control transfection, FI^{O}_{CT}, were used to compute the percent reporter expression (%R.E.) as follows:

$$\% R.E. = \frac{FI^{O}_{ST}}{FI^{O}_{CT}} \times 100$$

Sensitivity Assay

Drugs

Hydrogen Peroxide (H_2O_2) (Sigma) was prepared at working concentrations in complete media. 5-Fluorouracil (5-FU) (Sigma) was prepared as a 100 mM stock in DMSO, and diluted to working concentrations in complete media before addition to cells.

XTT Assay

Lymphoblastoids density was counted with a Beckman Coulter Counter and 25,000 cells were seeded in 100 µL in 96-well tissue culture plates. Cells were grown for 3 days with drugs added in a volume of 50 µL complete media immediately after plating (final concentration 25 µM H_2O_2 and 100 µM 5-FU). Each condition, including controls, was set up in three different replicate wells. On the third day, 2,3-bis(2-methoxy-4-nitro-5-sulfophenly)-5-[(phenylamino)-carbonyl]-2H-tetrazolium hydroxide (XTT) reagent (Cell Signaling Technology, in a volume of 50 µl), activated immediately prior to use by an electron coupling reagent according to the XTT kit protocol was added to each well (including blank wells containing complete medium alone). After 24 h of further incubation at 37 °C, the absorption was measured using a tunable microplate reader (Molecular Devices Versamax) at a wavelength of 450 nm. Results are shown as % Control Growth and represent the ratio of treated wells over untreated wells (previous

subtraction of no-cell controls). Each experiment was done at least in triplicate. MMS sensitivity was measured using a previously described trypan blue exclusion assay [30].

Multiple Linear Regression Models

Z-scored repair capacity values in each pathway reported for the 24 cell lines served as the 10 (independent) predictor variables (5 reporters developed here, **Table 4.4**; 5 previously developed reporters [29] **Table 4.5**), and sensitivity for the same 24 cell lines reported in **Table 4.6** and expressed in terms of % Control Growth were the (dependent) response variables. The resulting MLR models take the form:

$$Y = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \ldots + \beta_n x_n + b \qquad (1)$$

where *Y* represents the predicted response variable (sensitivity) reported as % Control Growth; x_i, are the DNA repair capacity predictor variables for the ten substrates reported as % Reporter Expression, β_i are the substrate slopes along the corresponding dimensions, and *b* is a constant that represents the *"y-axis"* intercept.

Z-scores for relative DNA repair capacity among the 24 cell lines were generated for each pathway as follows,

$$Z_{ij} = \frac{x_{ij} - \bar{x}_j}{\sigma_j} \qquad (X)$$

where Zij is the z-score for a DNA repair pathway in cell line i in pathway j, x_{ij} is the DNA repair capacity for a given pathway j in cell line i, $\bar{x}_j$ is the mean value of the DNA repair capacity in pathway j over the 24 lymphoblastoid cell lines (i = 1-24), and σj is the standard deviation of the DNA repair capacity in pathway j. MGMT scores were transformed to $\log_{MGMT}$ before z-scoring.

Multiple linear regression (MLR) models were generated by running MATLAB's *regress* function. Beta scores for each variable as well as the model fit to the data (R^2) were calculated for each generated model. Leave-one-out cross-validation (LOOCV), was used to assess the prediction power of each model. The correlation coefficient (R^2_{CV}) between the observed cellular sensitivities and the predicted ones following LOOCV as well as their Root Mean Squared Errors (RMSE) were calculated for each model. Initial models were generated by including all variables (Model 0). Subsequent models were generated by sequentially removing predictor variables with the lowest contribution to the model. In cases were the contribution of two or more variables were similarly low, all the combinations were tested.

Repair kinetics

In order to measure the dynamics of reporter expression, repair assays were set up and imaged with an Incucyte ZOOM® (Essen BioScience). 15,000 cells were seeded in 24-well cell culture plates a day before transfection. Cells were transfected with Lipofectamine as described above with either 300 ng of WT-GFP and WT-mCherry or 300 ng of GFP-C289T-Hx and WT-mCherry. Immediately after transfection, the plate was incubated in the Incucyte ZOOM at 37 °C and 5% CO2. Incucyte ZOOM software was used to setup acquisition conditions: a. scan type: only wells containing cells, 9 images per well; b. channel selection: green, red, phase; c. spectral unmixing: 0%; d. acquisition interval: 1 hour for 74 hours. Included software was used to identify cells and quantify integrated cellular fluorescence. Total field florescence for each channel was used to calculate % R.E.

Results

Generation of reporter plasmids to measure in vivo repair of BER substrates

When RNA Polymerase II (RNAPII) encounters a DNA lesion during transcription, it can **(a)** stall at the lesion and abort transcription, **(b)** bypass the lesion in an error-free manner or **(c)** bypass the lesion by misincorporating a ribonucleotide into the transcript opposite the lesion site. This last process is known as transcriptional mutagenesis (**Figure 4.2**). We have taken advantage of transcriptional properties of DNA lesions to design and engineer reporters for BER substrates. Specifically, we generated a variety of non-florescent plasmid reporter variants. Each plasmid bears a non-synonymous site-specific mutation that perturbs maturation of the chromophore of the fluorophore. To generate reporters for BER substrates, we have further modified these non-fluorescent variants by positioning a site-specific DNA lesion at the mutated site. The rationale behind such a design is that transcripts of the original fluorescent variant of the plasmid reporter can only be generated by transcriptional mutagenesis events in the presence of the lesion (a detailed example and rationale for this approach is presented in **Figure 4.3**). In order to corroborate the incorporation of the different DNA lesions into the plasmid reporters, we performed a series of *in vitro* enzymatic steps (shown below). Each plasmid variant is labeled according to the base pair that was modified. The bases shown before and after the number correspond to the identity of the mutation in the non-transcribed strand. As an example, GFP C289T corresponds to a non-fluorescent variant of GFP for which base pair 289 was mutated from a C in the coding strand to a T.

Hypoxanthine in GFP C289T.

The backbone of the pMAX vector used for all the reporter plasmids contains a single ApaLI restriction site, addition of the native WT GFP sequence does not contain additional ApaLI sites. The point mutation C289T adds an additional restriction site to the reporter. Accordingly, digestion with ApaLI of the native plasmid results in a single linear product whereas two digestion products are obtained following digestion of the C289T variant. Substituting a *hypoxanthine* into the transcribed strand of the mutated reporter, specifically in position 289, abolishes digestion by ApaLI at that site, resulting in a single cut (**Figure 4.4**).

8oxoG in mOrange A215C and mPlum T202WT.

Incorporation of an *8oxoG* into the transcribed strand at site 215 of the mutated mOrange A215C reporter and the non-transcribed strand at site 202 of the mPlum reporter were tested by Fpg treatment. In the presence of *8oxoG*, this bi-functional glycosylase removes the damaged base and nicks the remaining abasic site converting the closed circular plasmid into a nicked plasmid. Consequently, Fpg digestion of mOrange-A215C-8oxoG and mPlum-T202WT-8oxoG (+) resulted in a complete conversion of closed circular to nicked plasmid. In contrast, Fpg digestion of the native WT mOrange/mPlum reporters resulted only in some conversion to nicked products, likely due to the presence of random oxidized bases around the plasmid generated during storage or preparation (**Figures 4.5 and 4.6**). Others have also observed the presence of oxidative damage in untreated plasmids [32, 33]. Importantly, experimental variables such as this one are internally controlled in the methods by transfecting all cells of interest with the same plasmid cocktails.

Uracil in BFP A191G.

The presence of a site-specific Uracil at site 191 of the transcribed strand of the BFP reporter was tested with serial enzymatic treatments with the Uracil mono-functional glycosylase, UDG and the abasic site endonuclease, APE1. Therefore, only following both digestions in the BFP-A191G-U reporter but not the native WT BFP reporter converted the closed circular to the nicked circular plasmid. Single treatment with either UDG or APE1 did not result in a nicked product for either of the reporters (**Figure 4.7**).

THF in GFP 617 and mOrange A215C.

Native WT GFP and GFP-617-THF as well as WT mOrange and mOrange-A215C-THF reporters were digested with APE1 in order to assay for the presence of a *THF* positioned at site 617 and 215 in GFP and mOrange, respectively. As expected, only the GFP-617-THF and mOrange-A215C-THF reporters were converted from a closed circular conformation to a relaxed nicked form, confirming the presence of the *THF* lesion in the reporter (**Figure 4.8**).

Development of FM-HCR assays for measuring DNA glycosylase activity on alkylated, deaminated and oxidized bases.

Our group previously described a method to measure 8oxoG repair by the use of the FM-HCR assay [29] and here we further extend the capability to measure BER activity by developing a variety of plasmids containing known BER substrates at specific sites in a fluorescent protein reporter gene.

All of the reporters measuring DNA glycosylase activity are based on the transcriptional mutagenesis properties of the lesions of choice. As explained above, we have taken advantage of these properties so that a fluorescent protein will only be expressed as long as the DNA glycosylase has not excised its

substrate. Upon repair, the source of transcriptional mutagenesis is removed and a non-fluorescent protein is encoded. We reproduced our previous results regarding *8oxoG* repair in Ogg1 proficient and deficient MEFs. This plasmid reports on transcriptional mutagenesis events to *adenine* caused by the presence of an *8oxoG* lesion in the transcribed strand of the mOrange-A215C-8oxoG reporter. As expected, *Ogg1*$^{-/-}$ cells show approximately a 17-fold higher level of mOrange than that in WT MEFs indicating a much lower 8oxoG repair capacity in the deficient versus WT cells (**Figure 4.9A**, plasmid combination #5 and #6).

Similarly, we engineered two additional transcriptional mutagenesis-based reporters. A *uracil* repair reporter (BFP-A191G-U), which produces WT BFP transcripts only when RNA polymerase incorporates an *adenine* opposite *uracil* present in the transcribed strand. And a *hypoxanthine* repair reporter (GFP-C289T-Hx), which results in WT GFP only upon *cytosine* incorporation opposite *hypoxanthine*. The usability of these reporters was tested in MEFs deficient in the main Uracil DNA Glycosylase, Ung (plasmid combination #1 and #2), and in the only known *hypoxanthine* DNA glycosylase, Aag, respectively (plasmid combination #3 and #4). For both reporters, as predicted, WT MEFs exhibited low fluorescent reporter expression whereas *Ung*$^{-/-}$ and *Aag*$^{-/-}$ MEFs exhibited approximately an 8- and an 12-fold higher fluorescence, respectively (**Figure 4.9B** and **4.9C**).

When an *8oxoG* lesion evades repair and goes through a round of replication, DNA polymerases can incorporated and *adenine* opposite from it 10-70% of the times [34]. In this scenario, if the DNA repair machinery were to remove the *8oxoG* lesion and not *adenine*, upon completion of repair a point mutation would be fixed into DNA as a *thymine* would be incorporated opposite *adenine*. Mutyh is a unique glycosylases, in the sense that it recognizes the presence of a damaged base (*8oxoG*) but removes the undamaged base opposite from it (*adenine*). This gives the DNA repair machinery a new opportunity to repair

8oxoG and avoid mutation at this site. For the developed Mutyh reporter, removal of an *adenine* in the transcribed strand positioned opposite an *8oxoG* is measured in the same way as the previous three glycosylase reporters described. mPlum-T202WT-8oxoG (+) results in WT mPlum only upon incorporation of *uracil* opposite an *adenine*. As expected, MEFs deficient in Mutyh show approximately a 7-fold higher level of mPlum that WT MEFs following reporter transfection (**Figure 4.9D**, plasmid combination #7 and #8).

Validation of* in vivo *glycosylase activity assay by comparison to* in vitro *radiolabeled oligonucleotide activity assay.

In order to further validate the FM-HCR assay for glycosylase activity measurements, we compared our newly developed assay (plasmid combination #13 and #14) to the gold-standard method for measuring MUTYH activity, the *in vitro* cell extract radiolabeled oligonucleotide assay. This was done in a panel of cell lines with varying MUTYH activity levels following knockdown or overexpression of MUTYH (**Figure 4.10A**). A correlation plot between the *in vitro* and the reciprocal of the *in vivo* assay shows a very strong correlation (R^2=0.8953) between both assays (**Figure 4.10B**). The reciprocal is plotted for consistency between both axes so that larger numbers correspond in both cases to increased MUYTH activity.

Kinetics of reporter expression and repair.

The use of a live cell imaging system with a motorized stage allowed us to monitor Aag activity, in 30 min increments, simultaneously in WT and *Aag*$^{-/-}$ cells over the course of 74 hours. The Incucyte ZOOM ® allowed us to monitor the kinetics of DNA repair, by measuring fluorescent reporter expression over time. As showed for the flow cytometric approach, following an acclimation period, *hypoxanthine*-reporter signal is higher for repair deficient cells throughout the

length of the experiment. %R.E. for *Aag*$^{-/-}$ plateaus after approximately 20 hours whereas WT %R.E. decreases with time (**Figure 4.11**).

Simultaneous measurement of four DNA glycosylase activities.

The four DNA glycosylase reporters represent the repair of the majority of the lesion types recognized by BER (oxidation, *8oxoG*; deamination, *uracil and hypoxanthine*; and alkylation, indirectly by the action of Aag on *hypoxanthine* as will be discussed later). Furthermore, each one of them was designed with in a different fluorescent reporter with the ultimate goal of integrating them into one multiplexed assay. When the four glycosylase reporters were transfected simultaneously, together with the transfection efficiency control plasmid (for a total of five reporters in one transfection, plasmid combination #9 and #10, **Figure 4.12B**); nearly identical results where observed when compared to transfecting separately one lesion plasmid at a time (plasmid combination #1 through #8, **Figure 4.12A**). The dynamic ranges between proficient and deficient cells lines are maintained all along the WT MEFs and the four glycosylase deficient cell lines. Thus validating the capability of the FM-HCR assay to simultaneously assess the activity of four DNA glycosylases.

Development of FM-HCR assays for measuring Ape1 activity

DNA glycosylase activity determines the rate at which BER is initiated but the ability to measure downstream steps of this pathway can contribute in the understanding of the overall BER status of a cell, tissue or individual. With this in mind we engineered two different reporters for measuring APE1 activity, the abasic-site endonuclease immediately downstream of – mainly – mono-functional DNA glycosylases. Specifically we used as a substrate *THF*, which is more stable than a natural abasic site and is recognized and nicked by APE1 as well.

Both reporters are based on specific transcriptional properties of *THF*. The first reporter (mOrange-A215C-THF) takes advantage of the capacity of RNAPII to incorporate *adenine* opposite *THF* to cause transcriptional mutagenesis and generation of a WT transcript. Similarly to the 8oxoG:C reporter previously described, this reporter produces WT mOrange transcripts only when RNA polymerase incorporates an *adenine* opposite *THF*. Transfection of this plasmid (plasmid combination #17 and #18) into the Ape1 proficient/deficient cells resulted in no detectable fluorescent mOrange events for *Ape1* ++Δ cells and approximately a 14 % reporter expression for the deficient *Ape1* ΔΔΔ cells (**Figure 4.13B**). We also engineered a set of reporters to test for transcriptional mutagenesis of *THF* to *cytosine*, *guanine* and *uracil*, but none resulted in any measurable fluorescent events when transfected into *Ape1* ΔΔΔ cells (results not shown).

The second reporter (GFP-617-THF) takes advantage of transcriptional blocking properties of *THF* (see **Figure 4.2A**). As such, the presence of the lesion in the transcribed strand of this reporter, results in fluorescence only upon removal of the transcription blocking lesion and expression of a full length GFP WT transcript. We were able to test the usefulness of this reporter by using a unique pair of mouse B cell lines (CH12F3): APE1 proficient (*Ape1* ++Δ) and APE1 deficient (*Ape1* ΔΔΔ) [35]. As predicted, *Ape1* ++Δ cells show increased fluoresce representative of an increased repair capacity in comparison to *Ape1* ΔΔΔ cells following transfection with GFP-617-THF (**Figure 4.13A**, plasmid combination #15 and #16). Proficient cells show approximately a 4.6-fold higher repair capacity than deficient cells.

Analysis of BER Capacity and DNA damaging agent sensitivity in a panel of 24 cell lines derived from apparently healthy individuals.

The four DNA glycosylase transcriptional mutagenesis-based reporters (for repair of Hx:T, 8oxoG:C, A:8xoG and U:G), as well as the transcription blockage

reporter for APE1 were used to measure the BER capacity of a panel of 24 human B-lymphoblastoid cells lines derived from apparently healthy individuals of diverse ancestry [36] (plasmid combination #11 and #12, #15 and #16). A range of BER activity for the different substrates was observed across the 24-cell lines, specifically: **Hx:T**, 1.7-fold; **8oxoG:C**, 4.3-fold; **A:8xoG**, 7.2-fold; **U:G**, 5.3-fold and **THF:C**, 1.9-fold (**Figure 4.14** and detailed data in **Table 4.4**). Moreover, every cell line turns out to display a unique BER capacity fingerprint (**Figure 4.15A**).

Regression models based on BER capacity can predict sensitivity to DNA damaging agents.

Given that chemotherapeutic treatment has a direct effect on DNA by causing DNA damage, we sought to determine if our BER activity measurements could help predict sensitivity to a panel of DNA damaging agents known to form DNA lesions for which BER is known to play some role.

Sensitivity to 0.4mM MMS, 25μM H_2O_2 and 100μM 5-FU was measured for this cell panel. The 24-cell lines showed various degrees of sensitivity, in particular, a 9.5–fold, 4.7–fold and a 2.7–fold range were observed for MMS, H_2O_2 and 5-FU, respectively (data in **Figure 4.16** and **Table 4.6**).

Multiple Linear Regression (MLR) models seek to find the linear contribution that independent variables (z-scored plasmid reporter data, **Figure 4.15B**) have towards a dependent variable (sensitivity to DNA damaging agents). For this analysis, BER plasmid reporter measurements were complemented with previously published DRC plasmids for other DNA repair pathways [29]. More specifically NER, MMR, DR, NHEJ and HR (**Table 4.5**).

A variety of MLR models including different combinations of the data from the plasmid reporters were built for each agent (**Tables 4.7, 4.8** and **4.9**). The

models yielding the best correlation fit following leave-one-out cross-validation (R^2_{CV}) and the lowest RMSE (root mean square error) were further considered for discussion. Inclusion of all the variables in the model results in over-fitting of the data, and a low prediction power; as such R^2_{CV} is prioritized over model R^2 in order to minimize over-fitting of the data. The following models for predictions of % control growth (% C.G.) were selected:

$$\%\ C.G.\boldsymbol{MMS_{\#7}}\ (Table\ 4.7)$$
$$= -13.18\times\boldsymbol{8oxoG:C} + 8.74\times\boldsymbol{U:G} - 8.34\times\boldsymbol{THF} + 8.03\times\boldsymbol{HR}$$
$$+5.52\times\boldsymbol{NHEJ} + 44.22$$

$$\%\ C.G.\mathbf{5}-\boldsymbol{FU_{\#4}}\ (Table\ 4.8)$$
$$= -4.61\times\boldsymbol{8oxoG:C} + 3.81\times\boldsymbol{U:G} - 3.14\times\boldsymbol{THF} - 1.88\times\boldsymbol{HR} + 1.83\times\boldsymbol{NHEJ}$$
$$-1.12\times\boldsymbol{A:8oxoG} + 44.22$$

$$\%\ C.G.\boldsymbol{H_2O_{2\#4}}\ (Table\ 4.9)$$
$$= -14.50\times\boldsymbol{8oxoG:C} + 14.14\times\boldsymbol{U:G} - 10.49\times\boldsymbol{THF} - 5.95\times\boldsymbol{HR} + 4.75\times\boldsymbol{MGMT}$$
$$+ 4.34\times\boldsymbol{NER} + 52.20$$

Special attention should be given to contributions by transcriptional mutagenesis based reporters (Hx:T; U:G; 8oxoG:C; A:8oxoG and MGMT) for which a negative sign implies an increased repair activity for that particular lesion and a corresponding positive contribution towards resistance. The implications of these models are discussed below.

Discussion

Available methods to measure BER are cumbersome, time consuming, complicated to multiplex and primarily based on cell lysates, thus losing single-cell resolution. Our recently developed fluorescence-based multiplex flow-cytometric host cell reactivation assay (FM-HCR) provides an ideal framework to overcome most of these drawbacks. By building upon this method for rapid high-throughput measurements for DRC we expanded the number of assays to include lesions repaired by the BER pathway.

Even though, the validation for each lesion reporter was done in a MEF cell line deficient in a DNA glycosylase or Ape1, each site-specific plasmid reports on the overall cellular capacity to repair a particular lesion. This can involve either several glycosylases or even other repair pathways. While OGG1 has been shown, *in vitro* in human cell lines, to account for 90% of *8oxoG* repair [37], this lesion can also be a substrate for the NEIL1 and NEIL2 DNA glycosylases and to a lesser degree AAG [38, 39] (though not detected in our $Aag^{-/-}$ results, see **Figure 4.12**). *Uracil* found in DNA can be repaired by UNG, SMUG1, MBD4 and TDG as well as by the MMR machinery [40]. Abasic sites are primarily recognized and incised by APE1 but APE2 can also act on the lesion [41]. MUTYH mainly removes *adenines* across *8oxoG* in the BER pathway, but MMR can play a role in the repair of this base pair as well [42]. The reporter for *hypoxanthine* repair has several unique characteristics. Similarly to MUTYH, there are no other glycosylases with *hypoxanthine* as a substrate and only EndoV has been proposed to play a secondary role in its removal [43, 44]. Thus, in this particular case, *hypoxanthine* repair can be directly considered a proxy for AAG activity. Besides *hypoxanthine*, AAG has a wide lesion recognition and excision spectrum, which includes: *7-methylguanine*; *3-methyladenine*; *1-methyladenine*; *1-methylguanine*; *3-methylcytosine*; *8-oxoguanine*; *$1,N^6$-ethenoadenine*; *$1,N^2$-ethenoguanine* [45]. The use of *hypoxanthine* instead of *$1,N^6$-ethenoadenine*, as used for many AAG *in vitro* assays, is quite convenient

given that in our *in vivo* repair environment *1,N^6-ethenoadenine* would report on the activity of both AAG and ALKBH2 and possibly ALKBH3 [45-48]. For the *in vitro* assays this is not a problem as the reaction conditions are non-ALKBH2/3-permissive. The recognition and incision activity of AAG has been shown to be lesion [45] and sequence dependent [49]. Nevertheless, when *hypoxanthine* and *1,N^6-ethenoadenine* were used as AAG substrates, in sequence contexts that favored repair of each lesion, results were highly correlated [15, 50]. This suggests that our results for *hypoxanthine* repair reliably represent the activity of the AAG/Aag DNA glycosylase, regardless of its substrate.

By performing these functional assays in a physiologically relevant context (*in vivo*), one guarantees that molecular interactions between all players involved in the repair of each lesion are present at the time of repair. As such, any of the redundancies described above as well as any cellular stimulators or inhibitors of repair are represented in the final repair score. For example, APE1 [51] and separately NEIL1 [52] have been shown to stimulate OGG1's activity by increasing its turnover. Thus, an *in vitro* assay that doesn't recapitulate physiological conditions for OGG1, NEIL1 and APE1 could result in a misrepresentation of the actual OGG1 activity *in vivo*.

Importantly, the newly developed *in vivo* MUTYH assay recapitulates the results obtained for the *in vitro* radiolabeled oligonucleotides (**Figure 4.10**). Such result is extremely important as it shows that differences in glycosylase activity can be detected in human derived cell lines and that our new assay matches the gold standard assay in the field.

To our knowledge, the use of a multiplexed assay to measure repair activity of 4 BER lesions is unprecedented in the BER field, as no more than one human glycosylase had been measured simultaneously *in vivo*. By combining the assays shown here, together with our previously published methodologies [29], we have now have the ability to measure DRC in all the major DNA repair pathways.

Specifically: BER (including initiation by at least AAG, MUTYH, OGG1 and UNG, as well as APE1), HR, MGMT, MMR, NER, NHEJ. Put another way, we can measure repair of Hx:T, A:8oxoG, 8oxoG:C, U:G, THF:C, double strand breaks, O^6-methylguanine, base mismatches and thymine dimers

It is widely known that *hypoxanthine* preferentially pairs with *cytosine* during replication [53]. And while it is generally considered that the DNA mispairing events that occur during replication can also occur during transcription [54], it was only recently reported that *hypoxanthine* can mispair with *cytosine* during transcription in *E.coli* [55]. Our results show, for the first time, that this is also the case in mammalian cells, *in vivo*. Similarly, abasic sites (or its reduced and more stable analog THF) are known to follow the "A-rule" during replication, where an *adenine* is incorporated across the abasic site. *In vitro* systems have shown contrasting results for mammalian RNA Polymerase II describing incorporation of *cytosine* across THF [56] or *adenine* and to a lesser extent *guanine* [57]. We observed for the first time *in vivo*, that RNAPII incorporates *adenine* across from THF, thus recapitulating the "A-rule" but for transcription. Moreover, we complement *in vitro* results showing transcription blockage by abasic sites [58], with our *in vivo* THF transcription blockage reporter. We engineered a set of reporters to test for transcriptional mutagenesis where THF directs incorporation of C, G or U, but none of them resulted in any measurable fluorescent events (results not shown). It then appears that under *in vivo* conditions, abasic sites cause a strong transcription blockage that can be bypassed by the erroneous incorporation of *adenine* across from the lesion. The physiological implications of such events are certainly not negligible given that approximately 20,000 abasic sites are formed on average in a mammalian cell every day [59]. Unrepaired abasic sites may then not only give rise to mutations after replication but may also result in truncated or mutated transcripts that can have a direct impact on cellular physiology. Importantly, APE1 proficient cell lines did not give rise to any fluorescent events when transfected with the mOrange-A215C-THF reporter, implying that proficient APE1 cells might be very efficient at

removing any abasic sites present. In comparison, some fluorescent events always occur when probing glycosylase proficient cell lines with their respective transcriptional mutagenesis reporters.

The DNA glycosylase reporters, and the APE1 transcriptional mutagenesis reporter, specifically report on the glycosylase step or the endonuclease step of BER and do not report on any downstream events. This is in contrast with restored fluorescence for the APE1 transcription blockage reporter. Restored fluorescence is only possible once all the subsequent steps of BER, following incision by APE1, are completed. Thus, this reporter lumps together several BER steps. Given the fact that both 8oxoG and THF can cause transcriptional mutagenesis to *adenine*, we cannot rule out the possibility that, some of the mOrange reporter expression for the 8oxoG reporter is caused by the transient formation of an abasic site during the repair of this lesion. Nevertheless, this possibility seems unlikely as it is considered that the BER pathway functions similarly to "passing the baton" in a relay race where there is a handoff of lesion intermediates following base excision that allows to minimize exposure of even more toxic lesions [60, 61]. Furthermore, we didn't observe a single fluorescent mOrange event in Ape1 proficient cells (see **Figure 4.13B**) potentially indicating how efficiently APE1 acts on its substrate. It did not escape our attention that relatively low % R.E. values are obtained for the *8oxoG* and *uracil* reporters in comparison to those for *hypoxanthine* and A:8oxoG reporters. Redundancy in the repair of these two lesions in particular by a variety of DNA glycosylases could explain this phenomenon.

As an alternative to flow cytometric analysis, live cell imaging has the advantage that cells and their fluorescence can be tracked in time allowing the kinetics of repair to be assessed. The major drawback for this approach is that only a few lasers and filters can be used simultaneously, in this case only two, meaning that only two fluorescent reporters can be used in parallel. Given that one reporter is used as a transfection efficiency control, there is only availability for the analysis

one lesion specific reporter. Moreover, even with the use of a motorized stage to sample different fields of a single well, and several different wells, sampling time is relatively slow. Altogether we consider this method is lower throughput than flow cytometry. Nevertheless, it allowed us to track hypoxanthine repair with time in WT and *Aag*$^{-/-}$ MEFs. The fact that *% R.E.* in WT repair decreases with time can be explained by the fact that as more repair occurs (with time) less hypoxanthine is available for transcriptional mutagenesis. Thus, in comparison to the WT cells transfected with undamaged cocktail, where more GFP is constantly being transcribed, cells transfected with damaged cocktail express less WT-GFP as time progresses. This is not the case for *Aag*$^{-/-}$ cells.

The dramatic methodological improvement by the use of our newly developed BER reporters can be evidenced by the capacity to measure the repair of *hypoxanthine*, *uracil*, *8oxoG, adenine* paired with *8oxoG* and *THF* in 24 human B-lymphoblastoid cell lines derived from apparently health individuals. This was achieved with two different reporter cocktails given that with the current experimental setup, a maximum of 4 different lesions plus a single undamaged control can be assayed simultaneously.

Most analyses of inter-individual variation in DRC have been performed *in vitro* with PBMC lysates isolated from the blood of healthy donors. This complicates a direct comparison to the fold range observed with the use of our *in vivo* BER methods in human B-lymphoblastoid cell lines. Nevertheless, with the exception of AAG, for which a 3.3 to 10-fold variation has been reported [15, 50, 62, 63], the degree of inter-individual variation we report agrees with previously observed ranges for OGG1 in PBMCs (2 to 2.9-fold [16, 64, 65], compared to our 4.3-fold) and UNG in colon, stomach and liver (5.5, 3.2 and 3.1 [66], respectively, compared to our 5.3-fold). To our knowledge, only a single functional study of human MUTYH-associated variants has been described, [67] but an extensive study on the functional variation of MUTYH in healthy human populations (or cell lines derived from healthy humans) had not taken place so far (not to be

confused with genotypic *MUTYH* variants analysis [68]). The low ranges showed in AAG and APE1 activity amongst all 24-cell lines are particularly striking. Nevertheless both events can have different origins. All the AAG activities are on what seems to be the lower side of the spectrum (high *% R.E.*, compare to *% R.E.* by *Aag*$^{-/-}$ MEFs in **Figure 4.9C**), potentially indicating that the B-lymphoblastoids in general might not have particularly high AAG activity, making differences more difficult to pick up. In contrast, all APE1 activities are on the higher activity side of the spectrum (high *% R.E.*, compare to *% R.E.* by Ape1++Δ in **Figure 4.13A**). Considering that deficiencies in APE1 are detrimental [69, 70], and *Ape1*$^{-/-}$ mice are embryonic lethal, it is not particularly striking that "healthy" cells stay within a small range of variation. Nonetheless, a 4.9-fold range has been reported in a panel of a 100 human PBMCs for an *in vitro* APE1 assay [71].

Chemo- and radiotherapy are widely used as cancer treatments. The ability to determine ahead of time whether a particular treatment would be effective could improve therapy and avoid putting patients through unnecessary medical procedures. An increasing amount of evidence indicates that DRC can be an important factor in cancer treatment [72]. Consequently, if one were able to define a treatment by which tumor toxicity would be favored (by virtue of its impaired DRC), while minimizing toxicity for a patient's healthy tissues (by virtue of more proficient DRC compared to that of the tumor), therapy could be personalized in order to minimize side effects and decrease the risk of secondary cancers that can be caused by unrepaired lesions in healthy tissues as a consequence of treatment [73-75]. Here, we assess the contribution of 10 different functional DRC measurements towards predicting cellular sensitivity to a variety of DNA damaging agents.

When interpreting any of the contributions of the DNA glycosylase reporters to the prediction models, it should be taken into account that DRC for a particular lesion may represent the action of several glycosylases but also their activity on

several different substrates. As such, more than directly indicating the action of a single glycosylase or the relevance in repair of a single lesion, a large score for a given reporter indicates us that special attention should be given to the conglomerate of repair activities represented in that score.

MMS is an S_N2 alkylating agent known to form, amongst others: 7-meG, 3-meA, 1-meA and 3-meC. Our expectation was that AAG would emerge as the reporters with the highest contribution in the regression model. Nevertheless, the model (#7 in **Table 4.7**) indicates that repair activities associated with an oxidative stress response play the strongest role following alkylation treatment by MMS. Specifically, an increase in 8oxoG repair activity (or lesions represented by this repair measurement) is associated with resistance to MMS treatment. Interestingly, in accordance with this prediction, there is evidence in HCT116 cells overexpressing OGG1 that such an overexpression reduces sensitivity to MMS treatment [76].

The contribution of repair associated with uracil-excision can represent the activity of UNG and SMUG1 on substrates product of oxidative stress such as 5hU, isodialuric acid, alloxan and 5-hmU [20, 77], once again pointing out an important role for an oxidative response following MMS treatment. Sensitivity by an increase in APE1 and uracil-excision activity could be based on the formation of BER intermediates that can be more toxic than the initial damage. This scenario has been demonstrated for AAG-initiated BER [3]. Nevertheless, deficiencies in APE1 are generally associated with sensitivity to MMS [78]. A possibility that could reconcile these contrasting results could be that in this particular scenario we are not comparing APE1 proficient and deficient cell lines but proficient and "highly-proficient" cell lines. Moreover, DSBs formed directly by oxidative stress or by accumulation of closely positioned nicked-BER intermediates could help explain the presence of HR and NEHJ in the model [79]. It is not immediately obvious how to reconcile the idea that initiation by 8oxoG-excision glycosylases and uracil-excision glycosylases can have opposite effect

on sensitivity. One possibility is that 8oxoG is more toxic before being excised by a glycosylase than in any of it intermediate forms.

5-FU is used in the clinic to treat solid tumors including colorectal and breast cancer. 5-FU treatment has been shown to lead to incorporation of both 5-FU and uracil into DNA (reviewed in [80]). Moreover, it has been shown to induce ROS [81]. Our MLR model (#4 in **Table 4.8**) seems to accurately represent the need to repair DNA damage arising from both oxidative stress and genomic uracil/5-FU incorporation. The complexity of interpreting the role of uracil-excision measurements for 5-FU treatment can be rooted in the widely different and conflicting responses observed for the different glycosylases involved in their repair, both in mouse models and cancer cell lines. siRNA depletion of UNG in colon and ovarian cancer cell lines, but not SMUG1, MBD4 or TDG has been recently shown to profoundly sensitized cells to FdUrd [82]. In contrast, *Ung*$^{-/-}$ MEFs are not particularly sensitive to 5-FU and do not accumulate 5-FU in their genome, but rather uracil [83, 84]. In contrast, *Smug1*$^{-/-}$ MEFs are sensitive to 5-FU and accumulate it in DNA [83]. Mbd4 deficient mouse cells have been shown to be resistant to 5-FU [85], as have *S.cerevisiae* deficient in UNG [86]. For this latter example, repair intermediates appear crucial for toxicity. Furthermore, reincorporation of uracil into DNA following BER initiation during 5-FU treatment can lead to futile cycling, given the increased levels of dUTP during thymidylate synthase inhibition caused by the treatment. It appears that some cells overcome this cycling by downregulating UNG either at the transcriptional or posttranslational level but, this regulation does not happen in all cell lines. Indeed, in cell lines that do not downregulate UNG following treatment, decreased levels of UNG, achieved through siRNA, confer resistance to FdUrd [87]. This particular scenario would be in accordance with our model where increased uracil-excision activity confers 5-FU resistance. Even though MMR deficiencies have been shown to play a role in resistance to 5-FU treatment [88], in our experimental setup MMR does not appear to contribute towards sensitivity.

In comparison to the MMS and H_2O_2 models, it is interesting that for 5-FU several small contributions make up the MLR model. Even though, 5-FU has been used in the clinic for many years, its role from a DNA damage and repair perspective is still a subject of debate. All these small contribution argue in favor of several repair pathways being involved in the effect and outcome of 5-FU treatment.

The fact that the highest contributors in the H_2O_2 MLR model (#4 in **Table 4.9**) are similar to those of MMS and 5-FU helps support the postulate that sensitivity for these two agents is being driven, in part, by an oxidative stress response. The contribution of NER in the H_2O_2 model could be rooted in the extent to which oxidative stress takes place during this treatment where the BER machinery could be overwhelmed. Indeed, NER has been shown in mouse models to act as a back-up to BER in the repair of 8oxoG [89, 90]. The contribution of MGMT following H_2O_2 is not immediately obvious. Interestingly, H_2O_2 MLR model #9 (in **Table 4.9**) results in an almost equal prediction capacity as #4 but contains less predictor variables at the expense of a lower model fit. Further validation would be necessary to determine which of these models better represents the real data and predictions. But, the idea of achieving the same prediction with fewer measurements does speak in favor of a simpler model.

Even though it would be tempting to assume these models would apply for any cell type of interest, we must remain cautious when attempting to extrapolate these results and predictions to a context different than human B-lymphoblastoid cells lines. Similar analyses and validation in other cells lines will elucidate whether these predictions are conserved in a variety of cellular contexts.

Table 4.1 Site-specific plasmids and extension conditions used for their production.

Plasmid name	Nicking Enzyme	Sequence	X =		Extension Temp.	Extension Time (h)
BFP-A191G-U	Nb.BstI	5' GGT CTT GCT GCC G**X**A GAG GAA GCT AGT AGC	U	IDT	64.0 °C	3.5
GFP-C289T-dHx	Nb.BstI	5' GAA GAA GAT GGT GC**X** CTC CTG GAC GTA GCC	I	EGT	61.0 °C	2.0
GFP-617-THF	Nb.BstI	5' GCT CAG GGC G**X**A CTG GGT GCT CAG GTA GTG	THF	IDT	66.0 °C	1.0
mOrange-A215C-8oxoG	Nb.BstI	5' GTA GGC CTT GGA GCC G**X**A GGT GAA CTG AGG	8oxoG	EGT	64.0 °C	2.5
mOrange-A215C-THF	Nb.BstI	5' GCC TTG GAG CCG **X**AG GTG AAC TGA GGG GAC	THF	IDT	66.0 °C	1.5
mPlum-T202WT-8oxoG (+)	Nt.BstI	5' GTC CCC TCA GAT CAT G**X**A CGG CTC CAA GGC	8oxoG	EGT	61.0 °C	2.0

Nb./Nt.BstI (NEB); U (Uracil); I(Inosine); THF (Tetrahydrofuran); EGT (Eurogentech); IDT (Integrated DNA Techonologies)

Table 4.2 Cell lines used for this study

Cell Line	Species	Cell Type	Culture Medium
WT MEFs	*Mm*	MEFs	DMEM + 10% Fetal Bovice Serum (FBS) + 1% penicillin/streptomycin (pen-step)
Aag$^{-/-}$ MEFs	*Mm*		
Mutyh$^{-/-}$ MEFs	*Mm*		
Ogg1$^{-/-}$ MEFs	*Mm*		
Ung$^{-/-}$ MEFs *1	*Mm*		
Ape1 ++Δ *2	*Mm*	CH12F3	RPMI + 10% FBS + 1% pen-strep + 50% β-mercaptoethanol
Ape1 ΔΔΔ *2	*Mm*		
HCT116+3	*Hs*	Colorectal Carcinoma	DMEM/F12 + 10% FBS + 400 µg/mL G418
TK6	*Hs*	B-Lymphoblastoid	RPMI + 15% FBS + 1% pen-strep
GM15029 (#1)	*Hs*		
GM15036 (#2)	*Hs*		
GM15215 (#3)	*Hs*		
GM15223 (#4)	*Hs*		
GM15245 (#5)	*Hs*		
GM15224 (#6)	*Hs*		
GM15236 (#7)	*Hs*		
GM15510 (#8)	*Hs*		
GM15213 (#9)	*Hs*		
GM15221 (#10)	*Hs*		
GM15227 (#11)	*Hs*		
GM15385 (#12)	*Hs*		
GM15590 (#13)	*Hs*		
GM15038 (#14)	*Hs*		
GM15056 (#15)	*Hs*		
GM15072 (#16)	*Hs*		
GM15144 (#17)	*Hs*		
GM15216 (#18)	*Hs*		
GM15226 (#19)	*Hs*		
GM15242 (#20)	*Hs*		
GM15268 (#21)	*Hs*		

Table 4.3 Combinations of reporter plasmids and types of DNA damage used in each experiment.

Comb.	AmCyan	tagBFP	EGFP	mOrange	mPlum	empty vector
#1	WT 500 ng	WT 30 ng	-	-	-	1970 ng
#2	WT 500 ng	A191G-U 90 ng	-	-	-	1910 ng
#3	WT 500 ng	-	WT 25 ng	-	-	1975 ng
#4	WT 500 ng	-	C289T-Hx 25 ng	-	-	1975 ng
#5	WT 500 ng	-	-	WT 30 ng	-	1970 ng
#6	WT 500 ng	-	-	A215C-8oxoG 120 ng	-	1880 ng
#7	WT 500 ng	-	-	-	WT 90 ng	1910 ng
#8	WT 500 ng	-	-	-	T202WT-8oxoG(+) 135 ng	1865 ng
#9	WT 500 ng	WT 30 ng	WT 25 ng	WT 30 ng	WT 90 ng	1825 ng
#10	WT 500 ng	A191G-U 90 ng	C289T-Hx 25 ng	A215C-8oxoG 120 ng	T202WT-8oxoG(+) 135 ng	1630 ng
#11	WT 2500 ng	WT 100 ng	WT 50 ng	WT 75 ng	WT 200 ng	-
#12	WT 2500 ng	A191G-U 300 ng	C289T-Hx 50 ng	A215C-8oxoG 300 ng	T202WT-8oxoG(+) 300 ng	-
#13	-	WT 30 ng	-	-	WT 95 ng	2375 ng
#14	-	WT 30 ng	-	-	T202WT-8oxoG(+) 95 ng	2375 ng
#15	-	WT 100 ng	WT 50 ng	-	-	-
#16	-	WT 100 ng	617-THF 50 ng	-	-	-
#17	-	WT 100 ng	-	WT 50 ng	-	-
#18	-	WT 100 ng	-	A215C-THF 50 ng	-	-

Table 4.4 Repair scores for for a panel of 24-B-lymphoblostoids BER reporters Expressed as % reporter expression for each pathway.

Cell	Hx		U		A:OG		OG:C		THF	
#	MEAN	SD	MEAN	SD	MEAN	SD	MEAN	SD	MEAN	SD
1	31.7476	6.9272	0.6847	0.2921	2.5909	1.1191	0.3197	0.1293	42.6212	3.5520
2	28.0368	5.0375	0.6038	0.2152	2.1080	0.9264	0.3500	0.0978	31.8322	5.5165
3	28.5149	3.2271	0.2660	0.0808	0.6191	0.0535	0.2211	0.0374	40.8845	6.2096
4	29.3432	1.4032	0.5543	0.0917	1.0886	0.1558	0.4559	0.0332	39.9406	8.7235
5	26.6432	4.6090	0.4213	0.0367	1.0600	0.2496	0.2064	0.0449	42.0648	3.6058
6	22.5518	4.7798	0.3092	0.1203	0.9476	0.4177	0.2341	0.1042	55.1547	6.8666
7	29.0925	4.4079	0.8033	0.0647	1.6523	0.2523	0.2633	0.0711	34.1281	3.8519
8	25.2635	2.0170	0.7258	0.0478	2.9185	0.6384	0.4352	0.0356	39.5847	0.9325
9	29.6621	5.9970	0.6958	0.3072	2.8481	1.0719	0.3552	0.1334	42.1514	6.3947
10	37.4416	9.4921	0.7857	0.0782	4.4514	1.2587	0.4199	0.0902	39.1879	6.5322
11	36.8470	11.2565	0.3809	0.0607	0.7256	0.1397	0.2883	0.0787	41.4559	8.4169
12	26.6864	3.5607	0.3828	0.1404	1.5201	0.3392	0.2043	0.0553	47.0307	5.4261
13	30.3603	2.9735	1.4137	0.2209	2.2274	0.4145	0.6935	0.0974	31.3551	5.6757
14	27.6786	6.3855	0.4304	0.1247	2.2421	0.5120	0.2731	0.0522	34.6313	6.3408
15	31.2557	6.7372	0.5215	0.1097	3.7221	0.8418	0.3528	0.1006	29.5019	1.1636
16	30.6884	2.7655	0.5782	0.1057	1.3387	0.1958	0.3130	0.0220	37.3524	4.7722
17	23.8122	6.8422	0.5472	0.1340	1.1472	0.2865	0.4092	0.1160	32.2018	4.8508
18	32.8206	4.8745	0.2438	0.0784	1.8045	0.6163	0.4177	0.2995	35.0768	5.1886
19	32.9597	6.8523	0.4481	0.2653	1.3836	0.5879	0.3076	0.0981	38.4652	2.4529
20	31.6107	1.5855	0.8218	0.1000	2.5626	0.1841	0.4428	0.0250	44.5430	4.9815
21	29.9678	6.1604	0.5334	0.1147	2.2397	0.3510	0.2325	0.0994	29.8717	2.6342
22	30.9048	2.3298	0.5289	0.0573	1.9590	0.1420	0.3514	0.0317	37.4110	3.6321
23	27.8542	2.0999	0.5835	0.1083	3.0599	0.6451	0.2891	0.0380	46.3986	4.5515
24	25.5898	1.8596	0.4104	0.1015	1.9716	0.4970	0.1627	0.0142	34.4424	5.5373

Table 4.5 Previous reporter repair scores for a panel of 24-B-lymphoblostoids for NER, MMR, MGMT, NHEJ and HR [29].

Expressed as % reporter expression for each pathway.

Cell #	NER		MMR		MGMT		NHEJ		HR	
	MEAN	SD	MEAN	SD	MEAN	SD	MEAN	SD	MEAN	SD
1	9.2781	2.0013	3.5140	0.9979	0.7009	0.5158	28.2047	0.6901	2.0685	0.1692
2	8.9573	1.8238	3.2380	1.1680	0.3866	0.4400	22.5242	1.3439	2.0938	0.3434
3	12.1743	3.2365	2.1832	0.4050	0.7279	0.1988	29.1756	2.6739	2.3198	0.4855
4	7.2693	1.3390	2.0054	0.5851	29.0662	3.6264	39.6789	0.9895	1.4082	0.0888
5	7.9498	2.2210	2.4278	0.7849	26.1840	1.3063	20.2051	1.8451	1.1997	0.1996
6	12.4463	1.1527	4.6820	2.0444	4.5758	2.9702	27.0968	2.3740	1.8608	0.0917
7	17.7320	1.4970	3.2990	0.9137	1.6421	0.7307	32.0921	0.1787	3.9250	0.5625
8	10.8772	1.7438	2.1019	0.1826	0.2396	0.1233	32.2305	1.8653	2.5813	0.3936
9	12.7878	1.5503	4.5340	1.0609	20.1442	4.7766	22.6693	3.2582	2.4509	0.5054
10	15.5415	3.6853	6.5304	3.6435	16.1823	7.8170	20.9192	1.1513	2.3589	0.0809
11	9.9035	1.2448	3.2382	2.1219	5.3888	2.0785	35.4267	1.6237	1.9425	0.3411
12	10.0329	0.6228	1.5705	0.5843	1.8077	0.8396	22.8900	1.7561	2.8113	0.2910
13	11.3239	0.7554	3.2314	1.0126	0.2375	0.1866	27.7017	0.8553	1.3920	0.2933
14	19.0151	1.6827	4.3005	0.9408	0.1264	0.2189	20.1158	0.1862	2.3722	0.2745
15	9.0475	1.5176	2.2953	1.3776	0.2541	0.3152	28.3219	0.8671	1.8440	0.1944
16	7.0868	0.8046	2.0760	0.2480	0.1161	0.0618	35.0393	4.4419	2.3304	0.3704
17	22.2200	2.3549	2.6914	0.2939	8.1276	1.4629	31.2115	1.5730	2.4570	0.2109
18	13.9422	3.9006	2.8573	0.3695	16.1317	4.4949	28.4611	2.4450	2.7856	0.3916
19	10.3605	1.1757	2.9133	0.5001	7.9507	2.1730	32.8561	2.1547	4.4729	1.0861
20	12.2124	1.9547	2.8482	0.8955	1.5818	0.3061	28.3748	3.0673	2.4522	0.2267
21	11.6051	3.2563	2.3013	0.4534	0.5304	0.4478	40.6062	1.9954	2.2167	0.2429
22	9.1889	1.8501	1.9676	0.0238	0.1009	0.0882	41.5611	0.7180	1.6927	0.2519
23	14.5713	0.7546	4.7510	0.1358	1.0128	0.8827	28.3893	0.5601	2.8489	0.2345
24	12.1954	3.3484	2.9780	0.9670	0.1329	0.0590	34.8476	3.4392	1.3581	0.0880

Table 4.6 Sensitivity scores for a panel of 24-B-lymphoblostoids for MMS, H_2O_2 and 5-FU

Expressed as % control growth for each agent.

Cell #	MMS (0.4 mM)		H_2O_2 (25 µM)		5-FU (100 µM)	
	% CG	SEM	% CG	SEM	% CG	SEM
1	9.0967	0.3047	70.5907	2.7271	18.7780	1.8852
2	25.1637	2.2501	47.6543	5.6195	24.2160	2.4286
3	52.2147	5.6967	48.1357	2.7593	26.6790	2.7063
4	9.6616	0.8028	46.0637	7.7415	16.1457	1.3657
5	33.6439	3.3999	52.7867	2.2656	22.5523	1.7534
6	15.3978	0.7470	17.7153	1.1121	15.1233	0.3090
7	75.3995	3.1024	74.8363	1.1078	28.7777	1.5253
8	33.3795	2.7304	20.3710	1.7752	14.0807	0.2214
9	57.3506	2.7559	56.8233	0.7099	23.9643	2.1158
10	47.8478	2.9437	65.8023	3.5998	17.5057	1.3908
11	28.9190	4.6732	52.8730	4.7147	16.0713	1.2137
12	30.7978	3.1453	55.9020	3.2226	21.5987	1.0618
13	34.9105	5.0543	67.9667	4.8097	21.6720	1.6267
14	46.5299	2.2637	53.0513	2.2782	20.9303	1.6654
15	49.3690	1.7173	68.0097	1.6215	21.8110	2.4271
16	61.7204	4.8186	44.0875	7.0657	11.9528	1.0838
17	47.4784	2.3231	80.6340	6.4990	21.3110	2.4776
18	33.8546	3.7448	24.9633	2.8176	12.6343	1.5327
19	67.7150	0.4311	26.5020	7.6616	23.2783	2.0430
20	30.6582	0.8645	34.5293	3.2944	13.7620	0.6918
21	86.4849	5.2408	83.4443	2.2869	23.0897	0.7286
22	64.0126	4.0846	57.7197	4.4027	10.8347	0.6482
23	67.6894	2.2139	48.2663	2.3131	16.0113	1.4902
24	51.9472	2.1955	53.9923	4.6619	21.4353	1.2456

Table 4.7 MLR Models for MMS sensitivity prediction.

R^2, model fit to the data; R^2_{CV}, correlation coefficient to observed data following leave-one-out cross-validation (LOOCV); RMSE, root mean squared error following LOOCV. b, "y-axis" intercept. GO, 8oxoG. Each reporter column shows *beta* scores for each z-scored reporter for every regression model. Scores with green backgrounds contribute towards resistance to the damaging agent, whereas scores with red backgrounds contribute towards sensitivity. Interpretation of transcriptional mutagenesis reporter scores is counterintuitive given that increased activity results in low reporter expression. Since the data is z-scored, low reporter expression results in a large negative z-scored value, which for a transcriptional mutagenesis reporter that has a contribution with a negative sign in the model (such as GO:C) the product becomes a large positive contributor towards resistance.

#	R^2	R^2_{CV}	RMSE	Hx	U	A:GO	GO:C	THF	NER	MMR	MGMT	NHEJ	HR	b
0	0.65	0.24	18.14	0.62	6.97	3.70	-13.17	-7.20	3.38	0.69	-0.19	7.10	6.90	44.22
1	0.65	0.30	17.41	0.56	7.04	3.76	-13.24	-7.27	3.34	0.71		7.14	6.90	44.22
2	0.65	0.37	17.01		6.92	3.97	-13.07	-7.31	3.19	0.68		7.24	7.02	44.22
3	0.64	0.31	17.28		7.06	3.93	-13.14	-7.09	3.42		-0.05	7.06	7.04	44.22
4	0.64	0.37	16.59		7.08	3.94	-13.17	-7.11	3.42			7.06	7.04	44.22
5	0.62	0.37	16.52		6.95	4.23	-13.04	-7.79				6.29	8.14	44.22
6	0.62	0.38	16.35		8.75		-13.31	-7.56	3.73			6.42	6.83	44.22
7	0.59	0.40	16.19		8.74		-13.18	-8.34				5.52	8.03	44.22
8	0.52	0.38	16.40		7.82		-12.55	-9.67					7.95	44.22
9	0.45	0.30	17.38				-6.85	-10.13					8.07	44.22
10	0.37	0.20	18.60		8.10		-14.00	-9.79						44.22
11	0.30	0.16	19.14				-8.11	-10.28						44.22

Table 4.8 MLR Models for 5-FU sensitivity prediction.

R^2, model fit to the data; R^2_{CV}, correlation coefficient to observed data following leave-one-out cross-validation (LOOCV); RMSE, root mean squared error following LOOCV. b, "y-axis" intercept. GO, 8oxoG. Each reporter column shows *beta* scores for each z-scored reporter for every regression model. Scores with green backgrounds contribute towards resistance to the damaging agent, whereas scores with red backgrounds contribute towards sensitivity. Interpretation of transcriptional mutagenesis reporter scores is counterintuitive given that increased activity results in low reporter expression. Since the data is z-scored, low reporter expression results in a large negative z-scored value, which for a transcriptional mutagenesis reporter that has a contribution with a negative sign in the model (such as GO:C) the product becomes a large positive contributor towards resistance.

#	R^2	R^2_{CV}	RMSE	Hx	U	A:GO	GO:C	THF	NER	MMR	MGMT	NHEJ	HR	b
0	0.72	0.33	4.05	-0.49	3.71	-0.92	-4.37	-3.13	-0.08	-0.10	1.85	-1.79	0.70	19.34
1	0.72	0.36	3.96	-0.47	3.72	-0.94	-4.37	-3.10		-0.12	1.83	-1.79	0.68	19.34
2	0.72	0.40	3.85	-0.45	3.69	-0.95	-4.36	-3.13			1.83	-1.75	0.66	19.34
3	0.72	0.41	3.81		3.73	-1.12	-4.43	-3.08			1.71	-1.88	0.62	19.34
4	0.70	0.45	3.66		3.81	-1.12	-4.61	-3.14			1.83	-1.88		19.34
5	0.66	0.41	3.80		3.44		-4.68	-3.06			2.04	-1.65		19.34
6	0.60	0.41	3.81		3.03	-1.45	-3.70	-2.57				-2.16		19.34
7	0.57	0.27	4.23		4.09	-0.67	-4.98	-2.78			2.19			19.34
8	0.53	0.31	4.10		2.42		-3.66	-2.38				-1.90		19.34
9	0.39	0.21	4.41		2.73		-3.88	-1.92						19.34

Table 4.9 MLR Models for H_2O_2 sensitivity prediction.

R^2, model fit to the data; R^2_{CV}, correlation coefficient to observed data following leave-one-out cross-validation (LOOCV); RMSE, root mean squared error following LOOCV. b, "y-axis" intercept. GO, 8oxoG. Each reporter column shows *beta* scores for each z-scored reporter for every regression model. Scores with green backgrounds contribute towards resistance to the damaging agent, whereas scores with red backgrounds contribute towards sensitivity. Interpretation of transcriptional mutagenesis reporter scores is counterintuitive given that increased activity results in low reporter expression. Since the data is z-scored, low reporter expression results in a large negative z-scored value, which for a transcriptional mutagenesis reporter that has a contribution with a negative sign in the model (such as GO:C) the product becomes a large positive contributor towards resistance.

#	R^2	R^2_{CV}	RMSE	Hx	U	A:GO	GO:C	THF	NER	MMR	MGMT	NHEJ	HR	b
0	0.69	0.18	16.69	3.18	15.11	-0.63	-15.76	-8.51	6.69	-2.40	4.02	1.28	-6.65	52.20
1	0.69	0.22	16.30	2.93	14.92		-15.75	-8.55	6.55	-2.37	4.21	1.44	-6.59	52.20
2	0.69	0.25	15.95	3.14	14.71		-15.61	-8.67	6.43	-2.80	3.98		-6.50	52.20
3	0.67	0.27	15.72	3.36	14.04		-15.22	-9.69	5.39		4.04		-6.58	52.20
4	0.64	0.33	15.04		14.14		-14.50	-10.49	4.34		4.75		-5.95	52.20
5	0.59	0.32	15.16		14.61		-14.80	-11.39			5.41		-4.64	52.20
6	0.59	0.30	15.36		11.80		-12.07	-8.67	4.95				-5.44	52.20
7	0.55	0.27	15.74		13.64		-13.13	-10.41	2.53		3.98			52.20
8	0.52	0.30	15.35		11.98		-12.02	-9.42					-3.83	52.20
9	0.48	0.33	15.10		11.84		-11.33	-9.36						52.20
10	0.23	0.10	17.43		12.99		-9.59							52.20

Figures

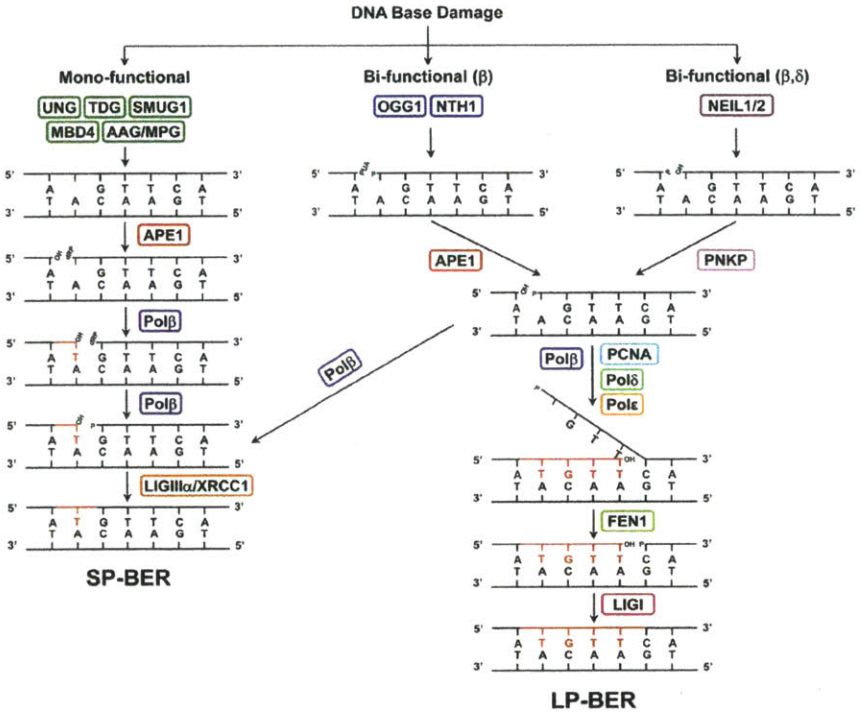


Figure 4.1 The DNA Base Excision Repair Pathway and its components.
BER is initiated by a DNA glycosylase. Mono-functional glycosylases create abasic sites that can be incised by APE1 resulting in a 5'-deoxyribosephosphase (5'-dRP) moiety that is cleared by the lyase activity of POLβ. Bi-functional glycosylases can both cleave and incise at the damaged base site. The DNA termini left by bi-functional glycosylases can be further processed by APE1 or PNPK. Short-Patch BER (SP-BER) involves incorporation of a single nucleotide by POLβ and nick sealing by LIGIII. Long-Patch BER (LP-BER) involves incorporation of up to 12 nucleotides by the action of POLβ, POLδ, or POLε, accompanied by PCNA. The flap created by the polymerase is cleaved by the FEN1 and the nick is usually ligated by LIGI. NEIL3 although not shown in the figure is technically considered a Bi-functional glycosylase (β-elimination) but its base excision activity is much more efficient than its incision activity and is functionally considered mono-functional [91]. Adapted from [2].

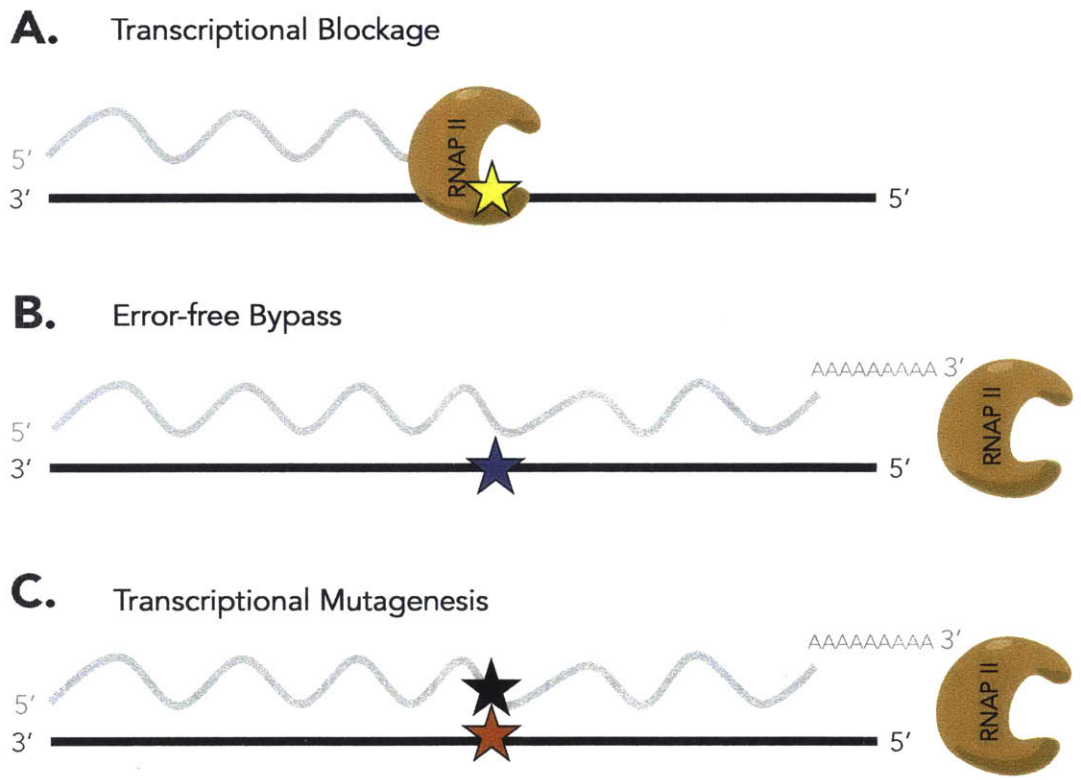


Figure 4.2 DNA lesions from an RNA Polymerase II perspective.
When RNA Polymerase II (RNAPII) encounters a DNA lesion during transcription, it can **(A)** stall at the lesion and abort transcription, **(B)** bypass the lesion in an error-free manner or **(C)** bypass the lesion by misincorporating a ribonucleotide into the transcript across the lesion site. This last process is known as transcriptional mutagenesis.

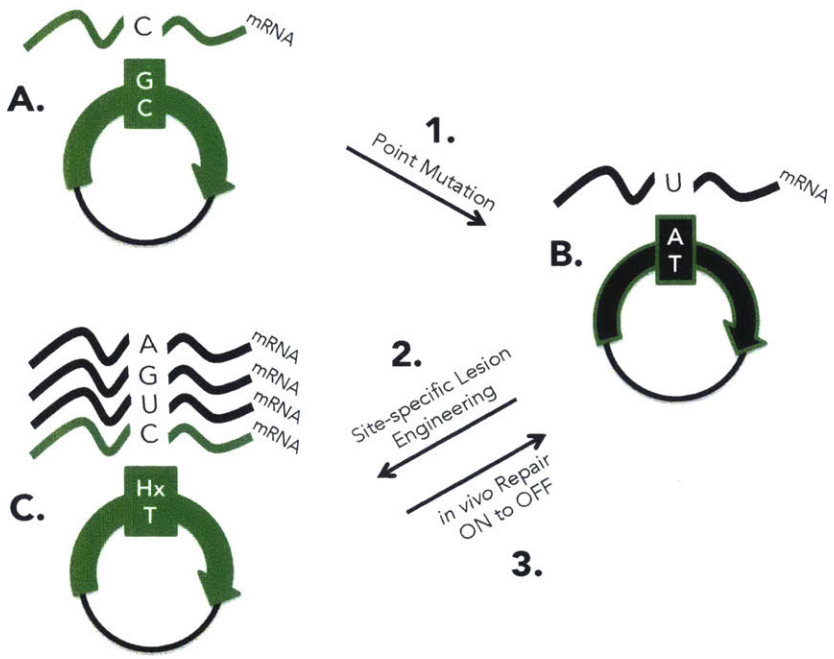


Figure 4.3 Generalized rationale for a transcriptional mutagenesis based fluorescent reporter.

The base pair shown corresponds to a site that codes for a key amino acid of the chromophore of the fluorescent protein. Transcribed strand is always shown on top. The base shown in mRNA corresponds to the ribonucleotide incorporated by RNAPII across the specific base pair in the plasmid.

A. Wildtype (WT) GFP reporter codes for a fluorescent reporter, incorporation of a C at the shown site is necessary for fluorescence to occur. WT reporter was **1.** mutated through QuikChange site-directed mutagenesis (see materials and methods) in order to produce a **B.** non fluorescent reporter variant. **C.** Hypoxanthine is incorporated on the transcribed strand at the mutated site of the non-fluorescent variant **2.** through primer extension (see materials and methods). Only transcriptional mutagenesis events to cytosine, caused by the presence of hypoxanthine will result in the expression of a fluorescent variant of the reporter. Incorporation of any other ribonucleotides result in non-fluorescent proteins **3.** Repair by the cellular DNA repair machinery leads to removal of the hypoxanthine, the source of transcriptional mutagenesis events, so that only non-fluorescent variants are expressed.

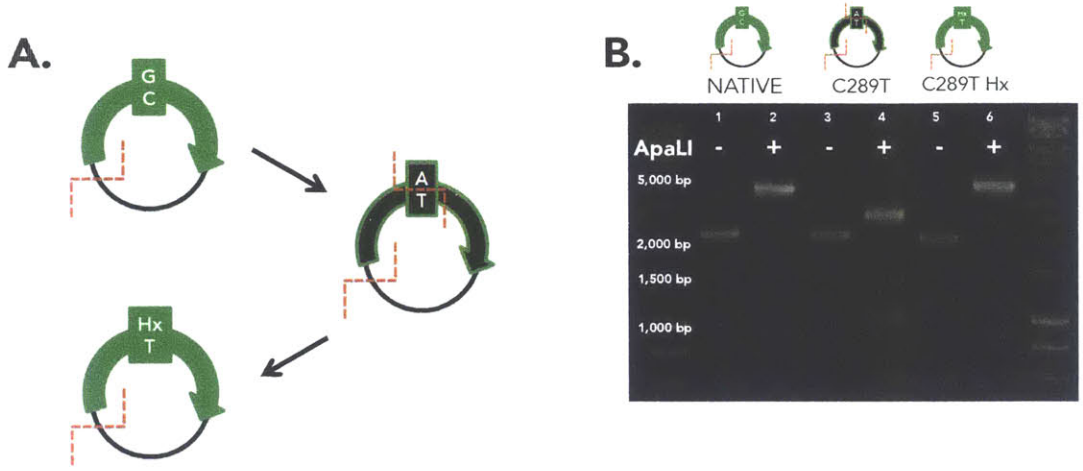


Figure 4.4 ApaLI digenstion for confirmation of hypoxanthine presence in GFP-C289T-Hx plasmid.

A. Native GFP reporter has a single ApaLI restriction site (red dashed line) in the plasmid backbone. The creation of a point mutation (rationale explained in **Figure 4.2**) generates an addition ApaLI, whereas incorporation of hypoxanthine at that same site blocks recognition by ApaLI at this secondary restriction site. **B.** Agarose (1%) gel electrophoresis. WT (native), C289T or C289T-Hx GFP plasmids were left untreated (lanes 1,3,5) or treated (lanes 2,4,6) with ApaLI. The presence of a single Hx at site 289 prevent cutting at this site (but not the one in the backbone) by the restriction enzyme thus generating a single DNA band (lane 6).

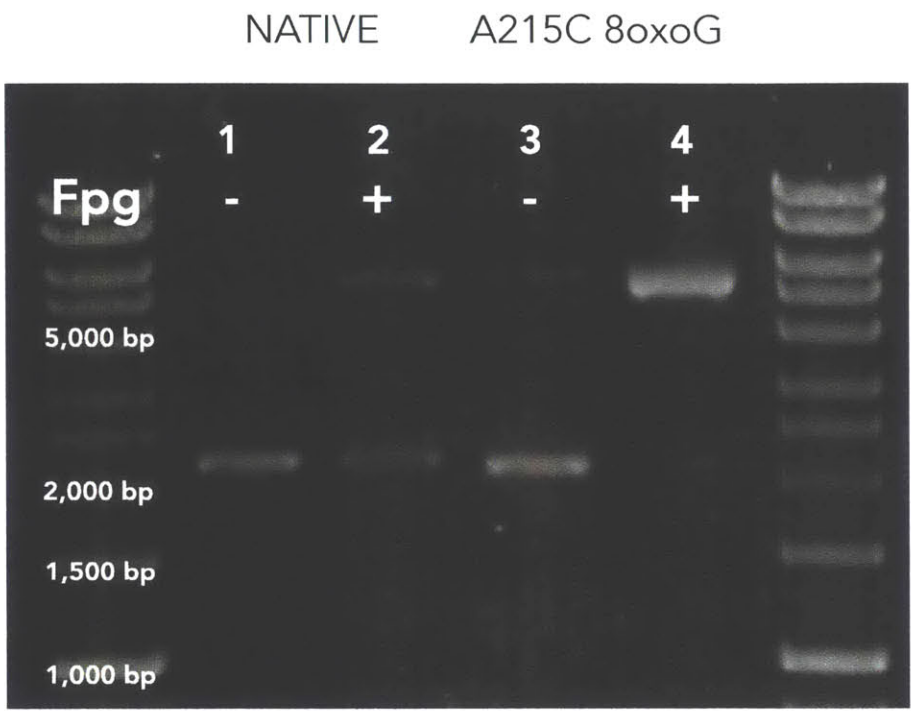


Figure 4.5 Fpg treatment for confirmation of 8oxoGuanine presence in mOrange-A215C-8oxoG plasmid.

Agarose (1%) gel electrophoresis. WT (native) or A215C-8oxoG mOrange plasmids were incubated in Fpg buffer and left untreated (lanes 1,3) or treated with Fpg (lanes 2,4). The presence of a single 8oxoG at site 215 allows for complete nicking of lesion-bearing plasmid by the bi-functional glycosylase.

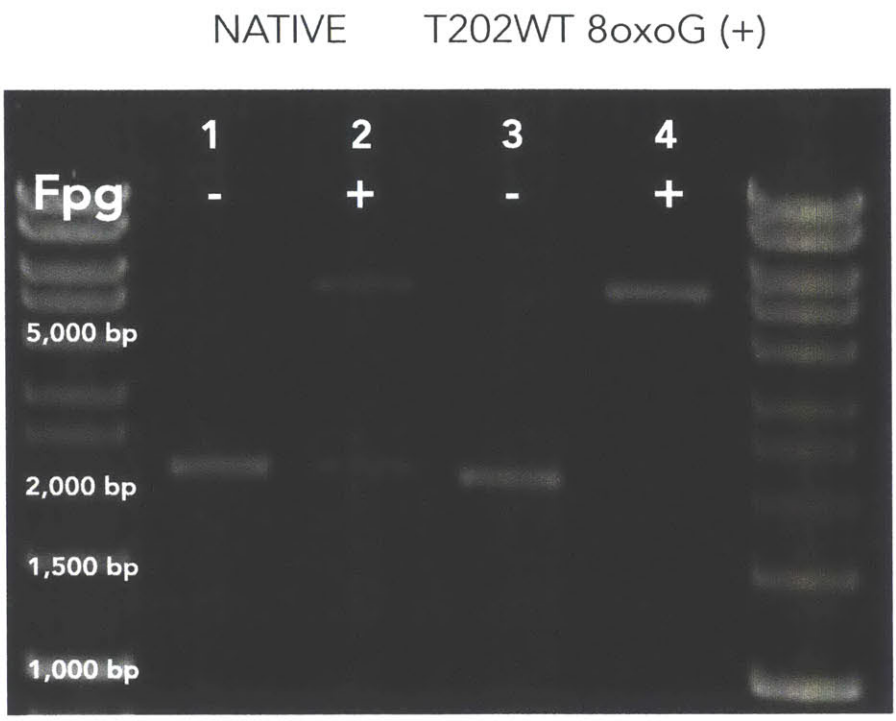


Figure 4.6 Fpg treatment for confirmation of 8oxoGuanine presence in mPlum-T202WT-8oxoG (+) plasmid.

Agarose (1%) gel electrophoresis. WT (native) or T202WT-8oxoG (+) mPlum plasmids were incubated in Fpg buffer and left untreated (lanes 1,3) or treated with Fpg (lanes 2,4). The presence of a single 8oxoG at site 202 allows for complete nicking of lesion-bearing plasmid by the bi-functional glycosylase.

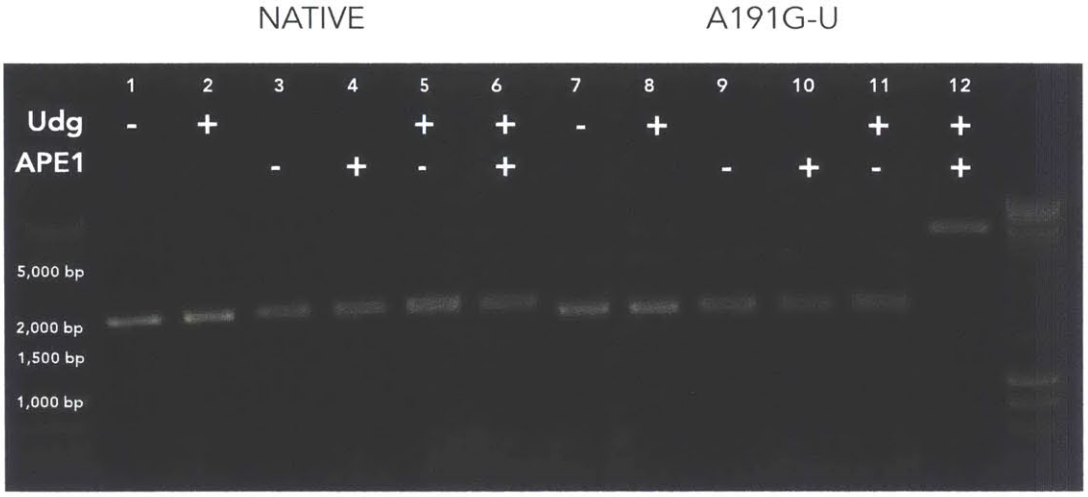


Figure 4.7 Udg and APE1 treatment for confirmation of Uracil presence in BFP-A191G-U plasmid.

Agarose (1%) gel electrophoresis. WT (native) or A191G-U BFP plasmids were incubated in Ung buffer and left untreated (lanes 1,7) or treated with Udg only (lanes 2,8); incubated in NEBuffer 4 and left untreated (lanes 3,9) or treated with APE1 only (lanes 4,10); incubated in Ung buffer, treated with Ung followed by incubation in NEBuffer 4 and left untreated (lanes 5,11) or treated with APE1 (lanes 6,12). The presence of a single Uracil at site 191 allows for complete nicking of lesion-bearing plasmid by the sequential action of Udg and APE1.

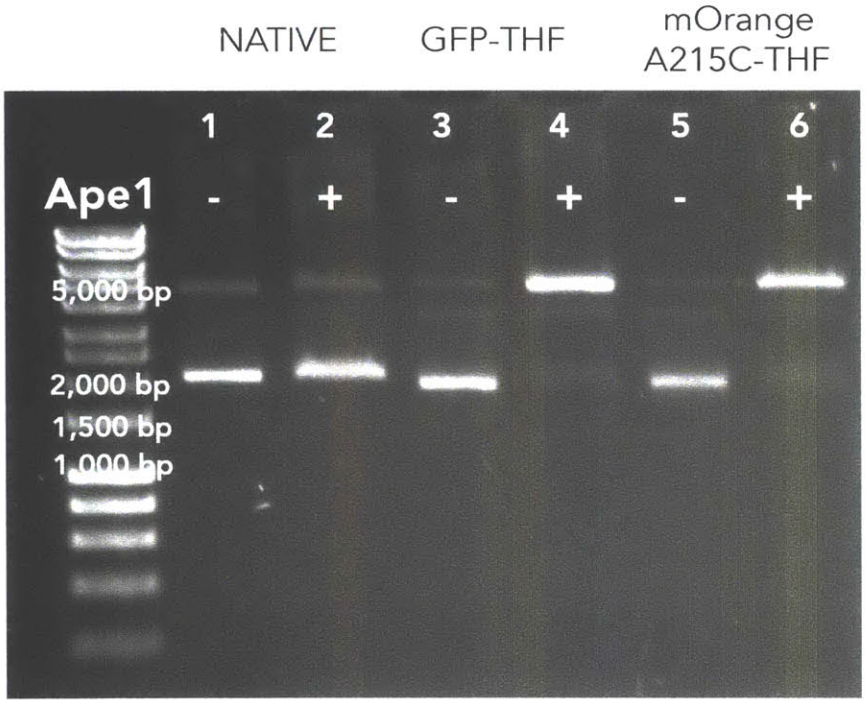


Figure 4.8 APE1 treatment for confirmation of tetrahydrofuran presence in GFP-THF and mOrange-A215C-THF plasmids.

Agarose (1%) gel electrophoresis. WT (native), GFP-THF GFP or mOrange-A215C-THF plasmids were incubated in NEBuffer 4 and left untreated (lanes 1,3,5) or treated with APE1 (lanes 2,4,6). The presence of a single THF allows for complete nicking of lesion-bearing plasmid the endonuclease.

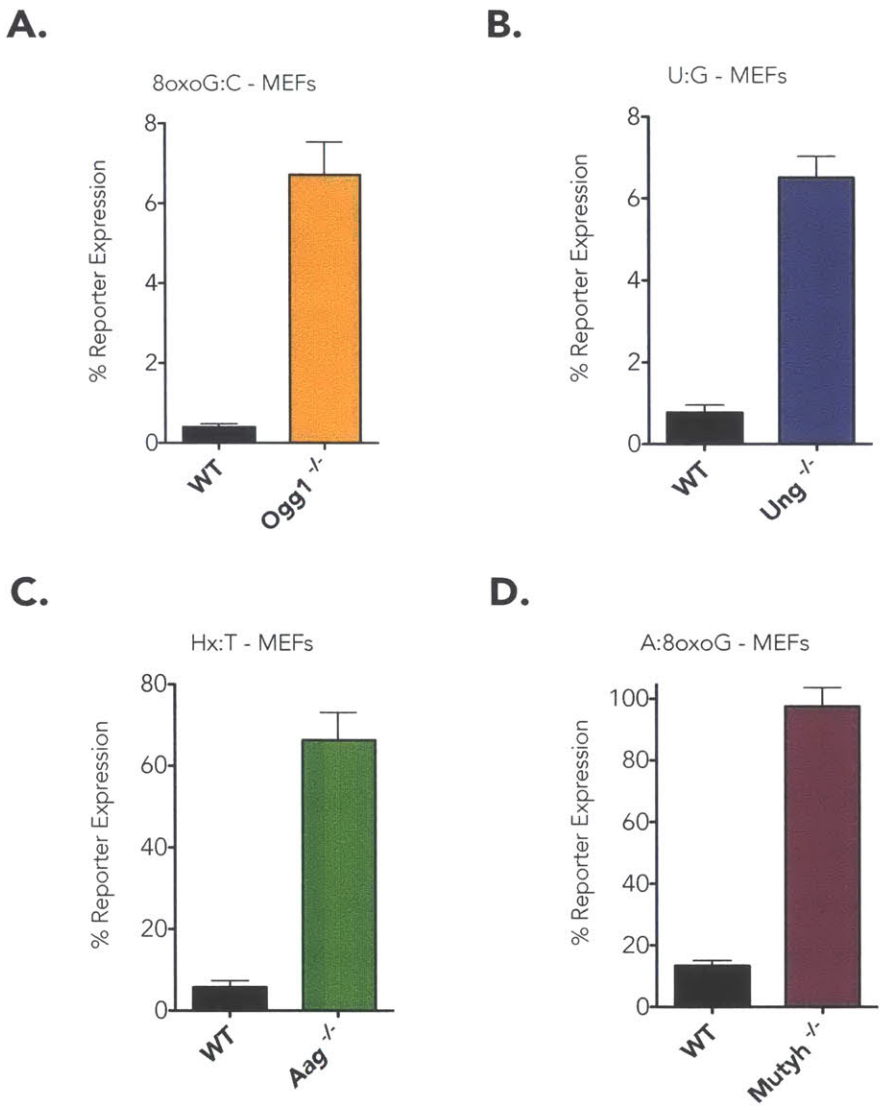


Figure 4.9 FM-HCR assays for measuring DNA glycosylase activity on alkylated, deaminated, oxidized and misincorporated bases.

Proof-of-concept BER repair capacity measurements comparing WT MEFs and MEFs deficient in glycosylases known to repair the lesion of choice. **A.** 8oxoG:C repair reporter in *Ogg1*$^{-/-}$ cells; **B.** Uracil repair reporter in *Ung*$^{-/-}$ cells; **C.** Hx repair reporter in *Aag*$^{-/-}$ cells and **D.** A:8oxoG repair reporter in *Mutyh*$^{-/-}$ cells. Error bars represent the standard deviation calculated from at least biological triplicates

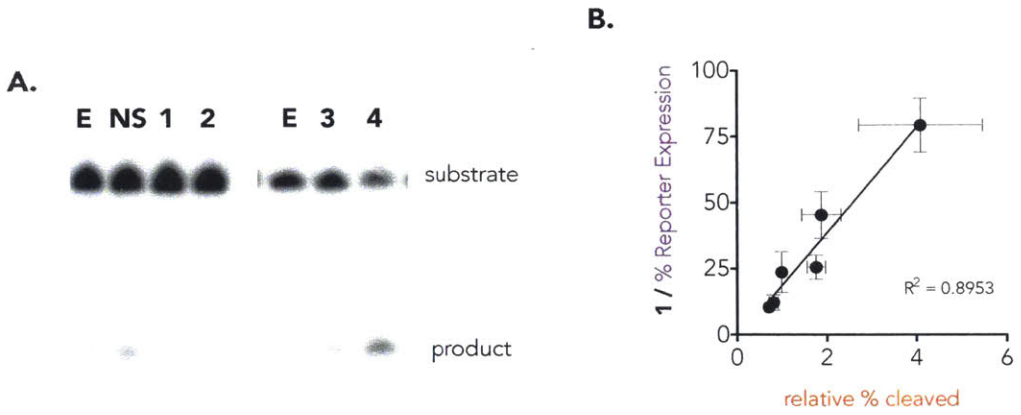


Figure 4.10 *in vitro* and *in vivo* MUTYH activity comparison.

For simplicity, HCT116 + Chromosome 3 cell lines were renamed as follows: pBABE empty vector (E), pGIPZ non-silencing (NS), *Mutyh* KD-A (1), *Mutyh* KD-B (2), nuclear *Mutyh*-A (3), nuclear *Mutyh*-B (4). **A.** Cell lysates from labeled cell lines were incubated with a radiolabeled oligonucleotide containing an A:8oxoG base pair. The MUYTH activity of the lysate is determined by the ratio of product to substrate. **B.** Correlation plot between *in vivo* MUTYH assay with plasmid mPlum-T202WT-8oxoG (+) and *in vitro* MUTYH activity assay showed in section A of this figure. Reciprocal of % R.E. is plotted for ease of comparison. Error bars represent the standard deviation calculated from biological triplicates for *in vivo* assay and technical duplicates for *in vitro* assay.

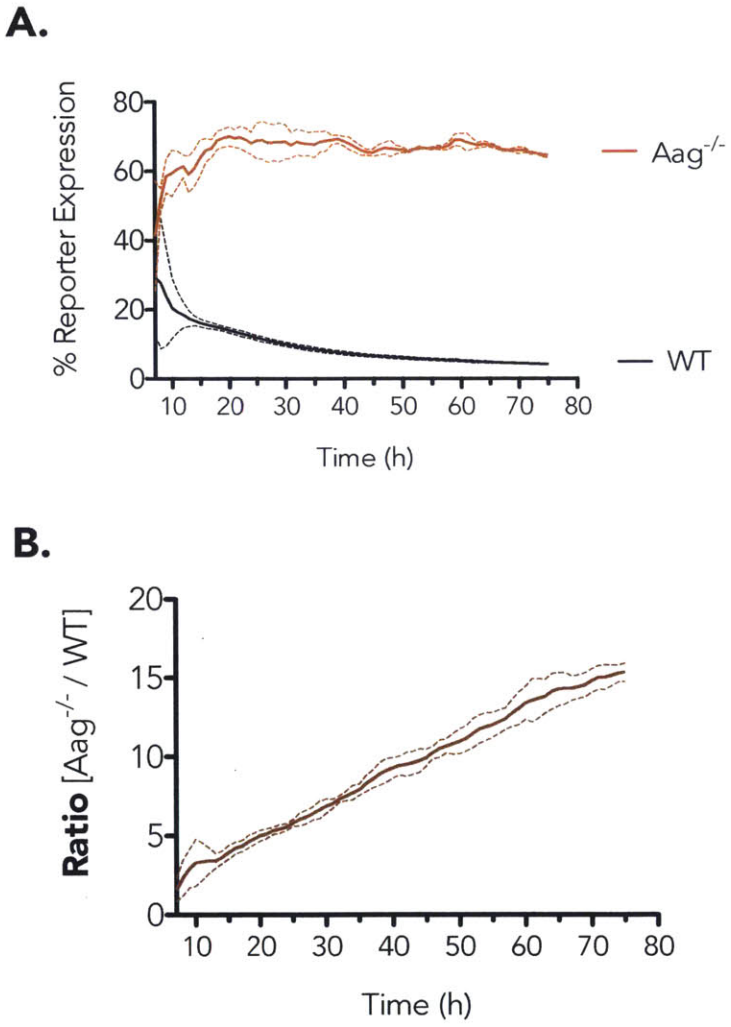


Figure 4.11 Kinetics of hypoxanthine repair in WT and *Aag*$^{-/-}$ MEFs.
Time-lapse, live-cell imaging over 74 hours, in 30 min increments. Integrated cellular fluorescence per field was used for reporter expression quantification. **A.** Hypoxanthine repair activity over time for *Aag*$^{-/-}$ and WT MEFs. **B.** Ratio of % R.E. between *Aag*$^{-/-}$ and WT MEFs over the course of the experiment. Solid lines denote mean while dashed lines indicate standard deviation from the mean of technical triplicates.

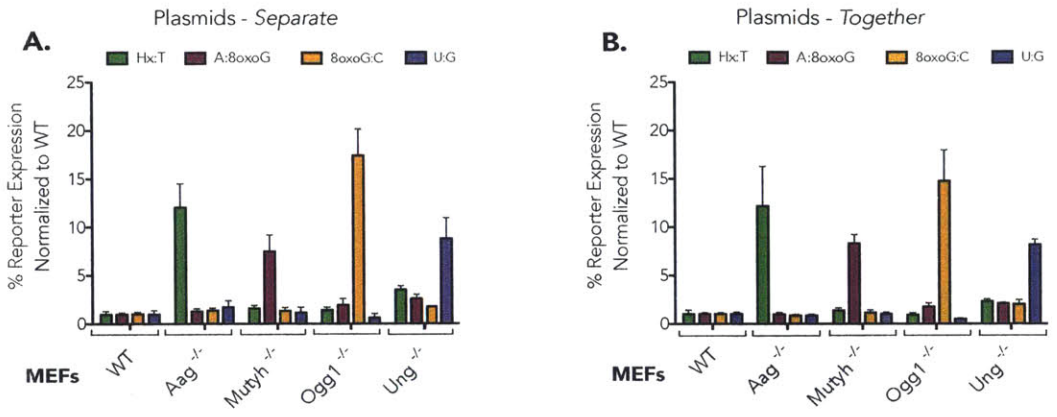


Figure 4.12 Comparison of single and multiplexed measurements of four BER Glycosylase substrates.

A. Separate- or **B.** Together plasmid reporter combinations were transfected into repair proficient (WT) cells or MEFs deficient for one glycosylase know to repair each one of the tested lesions at a time. The same transfection efficiency control plasmid was used for all the experiments (AmCyan). % R.E. was normalized for each reporter to WT cells. Error bars represent the standard deviation calculated from at least biological triplicates.

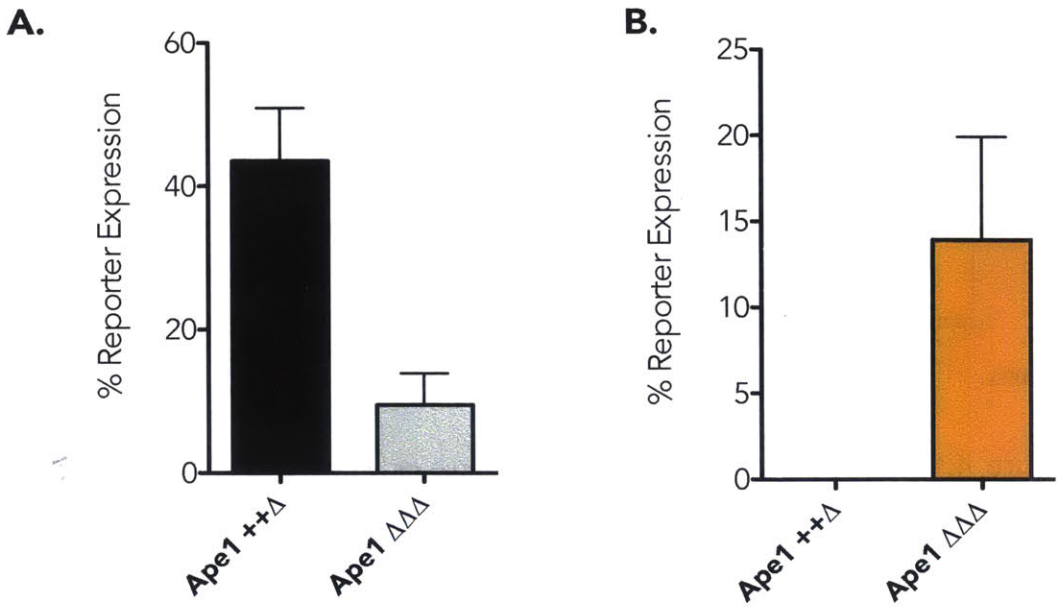


Figure 4.13 FM-HCR assays for measuring APE1 activity.

Reporters are based on the different effects abasic site have on transcription. **A.** Transcription blockage by THF results in decreased reporter expression. **B.** Transcriptional mutagenesis to Adenine across THF results in a fluorescent mOrange signal in the absence of APE1 activity.

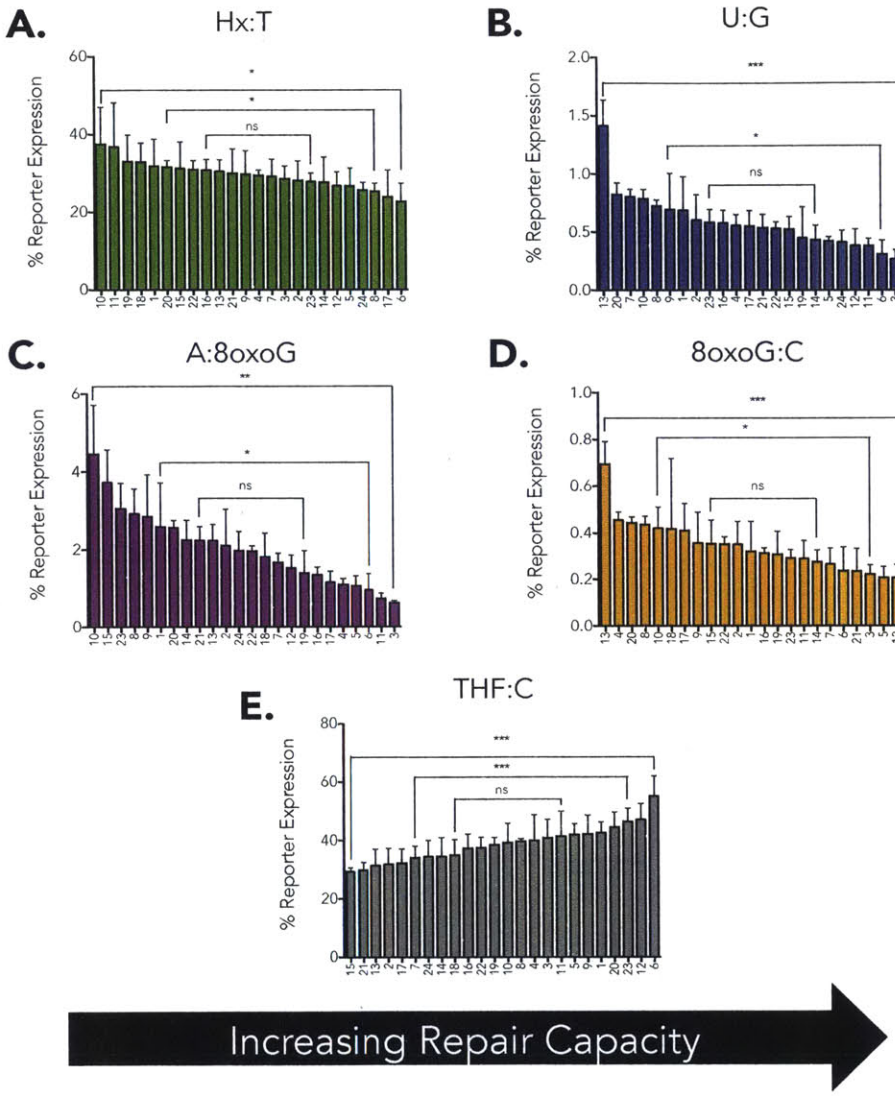
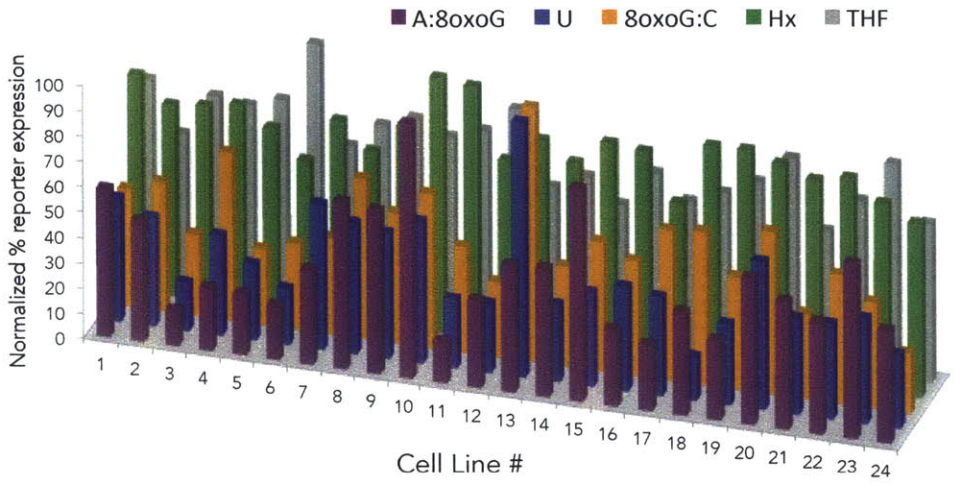


Figure 4.14 Inter-individual variability in BER of a panel of 24 B-lymphoblastoid cell lines ranked individually.

FM-HCR was assessed in all cell lines with the following substrates: **A.** Hx:T, **B.** U:G, **C.** A:8oxoG, **D.** 8oxoG:C and **E.** THF:C Cells are ranked from left to right based on increasing repair capacity. Error bars show standard deviation of at least 3 biological replicates. *, $p \leq 0.05$; **, $p \leq 0.01$; ***, $p \leq 0.001$; ns, not significant.

A.



B.

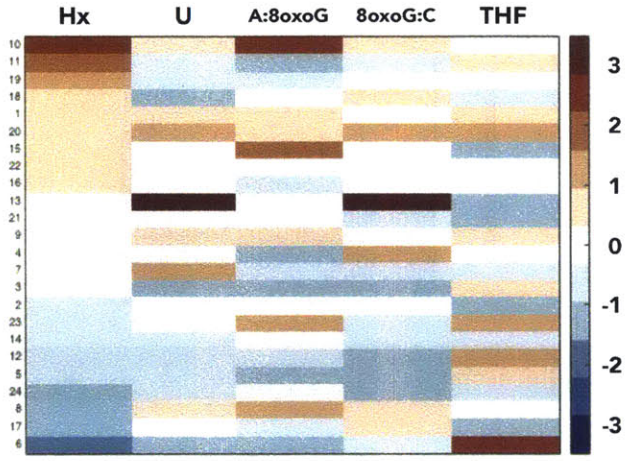


Figure 4.15 Inter-individual variability in BER of a panel of 24 B-lymphoblastoid cell lines.

FM-HCR with BER substrates was assessed in all cell lines **A.** % R.E. was normalized to the highest reporter expressing cell line for each plasmid reporter **B.** The data has been z-scored and organized by lower to higher Hx excision activity. Warm colors represent lower excision activity for all reporters but for THF. Cooler colors represent higher excision activity for all reporters but for THF.

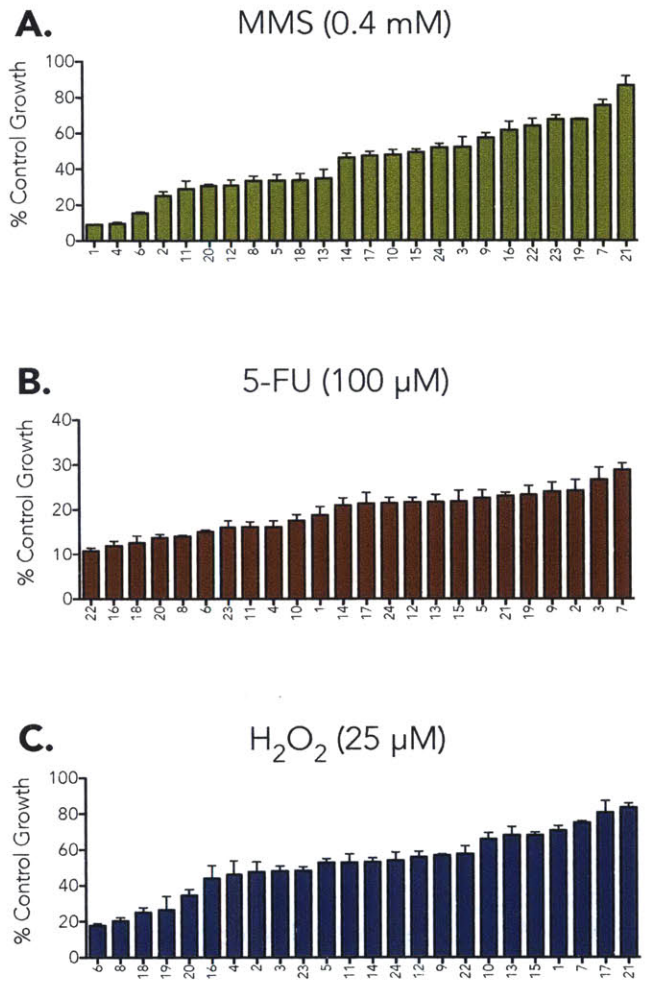


Figure 4.16 Inter-individual variability in sensitivity to DNA damaging agents of a panel of 24 B-lymphoblastoid cell lines.

Cellular sensitivity to **A.** 0.4mM MMS was measured by the trypan blue exclusion assay. Sensitivity to **B.** 100µM 5-FU and **C.** 25µM H_2O_2 were measured with the XTT assay. Error bars show standard error of the mean of at least 3 biological replicates.

References

1. Ciccia, A. and S.J. Elledge, *The DNA damage response: making it safe to play with knives.* Molecular cell, 2010. **40**(2): p. 179-204.
2. Kim, Y.-J. and D.M. Wilson III, *Overview of base excision repair biochemistry.* Current molecular pharmacology, 2012. **5**(1): p. 3.
3. Fu, D., J.A. Calvo, and L.D. Samson, *Balancing repair and tolerance of DNA damage caused by alkylating agents.* Nature Reviews Cancer, 2012. **12**(2): p. 104-120.
4. Hoeijmakers, J.H.J., *Genome maintenance mechanisms for preventing cancer.* Nature, 2001. **411**(6835): p. 366-374.
5. Jackson, S.P. and J. Bartek, *The DNA-damage response in human biology and disease.* Nature, 2009. **461**(7267): p. 1071-1078.
6. Wallace, S.S., *Base excision repair: A critical player in many games.* DNA repair, 2014.
7. Jones, S., P. Emmerson, J. Maynard, J.M. Best, S. Jordan, G.T. Williams, J.R. Sampson, and J.P. Cheadle, *Biallelic germline mutations in MYH predispose to multiple colorectal adenoma and somatic G: C→ T: A mutations.* Human Molecular Genetics, 2002. **11**(23): p. 2961-2967.
8. Imai, K., G. Slupphaug, W.-I. Lee, P. Revy, S. Nonoyama, N. Catalan, L. Yel, M. Forveille, B. Kavli, H.E. Krokan, H.D. Ochs, A. Fischer, and A. Durandy, *Human uracil-DNA glycosylase deficiency associated with profoundly impaired immunoglobulin class-switch recombination.* Nature immunology, 2003. **4**(10): p. 1023-1028.
9. Shao, C., S. Xiong, G.-M. Li, L. Gu, G. Mao, W.R. Markesbery, and M.A. Lovell, *Altered 8-oxoguanine glycosylase in mild cognitive impairment and late-stage Alzheimer's disease brain.* Free Radical Biology and Medicine, 2008. **45**(6): p. 813-819.
10. Gackowski, D., E. Speina, M. Zielinska, J. Kowalewski, R. Rozalski, A. Siomek, T. Paciorek, B. Tudek, and R. Olinski, *Products of oxidative DNA damage and repair as possible biomarkers of susceptibility to lung cancer.* Cancer research, 2003. **63**(16): p. 4899-4902.
11. Paz-Elizur, T., R. Ben-Yosef, D. Elinger, A. Vexler, M. Krupsky, A. Berrebi, A. Shani, E. Schechtman, L. Freedman, and Z. Livneh, *Reduced repair of the oxidative 8-oxoguanine DNA damage and risk of head and neck cancer.* Cancer research, 2006. **66**(24): p. 11683-11689.
12. Obtułowicz, T., M. Swoboda, E. Speina, D. Gackowski, R. Rozalski, A. Siomek, J. Janik, B. Janowska, J.M. Cieśla, and A. Jawien, *Oxidative stress and 8-oxoguanine repair are enhanced in colon adenoma and carcinoma patients.* Mutagenesis, 2010: p. geq028.
13. Sevilya, Z., Y. Leitner-Dagan, M. Pinchev, R. Kremer, D. Elinger, H.S. Rennert, E. Schechtman, L.S. Freedman, G. Rennert, and T. Paz-Elizur, *Low integrated DNA repair score and lung cancer risk.* Cancer Prevention Research, 2013: p. canprevres. 0318.2013.

14. Matuo, R., F.G. Sousa, A.E. Escargueil, D.G. Soares, I. Grivicich, J. Saffi, A.K. Larsen, and J.A.P. Henriques, *DNA repair pathways involved in repair of lesions induced by 5-fluorouracil and its active metabolite FdUMP.* Biochemical pharmacology, 2010. **79**(2): p. 147-153.
15. Leitner-Dagan, Y., Z. Sevilya, M. Pinchev, R. Kramer, D. Elinger, L.C. Roisman, H.S. Rennert, E. Schechtman, L. Freedman, G. Rennert, Z. Livneh, and T. Paz-Elizur, *N-Methylpurine DNA Glycosylase and OGG1 DNA Repair Activities: Opposite Associations With Lung Cancer Risk.* Journal of the National Cancer Institute, 2012. **104**(22): p. 1765-1769.
16. Paz-Elizur, T., D. Elinger, Y. Leitner-Dagan, S. Blumenstein, M. Krupsky, A. Berrebi, E. Schechtman, and Z. Livneh, *Development of an enzymatic DNA repair assay for molecular epidemiology studies: distribution of OGG activity in healthy individuals.* DNA repair, 2007. **6**(1): p. 45-60.
17. Goto, M., K. Shinmura, Y. Nakabeppu, H. Tao, H. Yamada, T. Tsuneyoshi, and H. Sugimura, *Adenine DNA glycosylase activity of 14 human MutY homolog (MUTYH) variant proteins found in patients with colorectal polyposis and cancer.* Human mutation, 2010. **31**(11): p. E1861.
18. Molatore, S., M.T. Russo, V.G. D'Agostino, F. Barone, Y. Matsumoto, A.M. Albertini, A. Minoprio, P. Degan, F. Mazzei, and M. Bignami, *MUTYH mutations associated with familial adenomatous polyposis: functional characterization by a mammalian cell - based assay.* Human mutation, 2010. **31**(2): p. 159-166.
19. Liu, M., V. Bandaru, A. Holmes, A.M. Averill, W. Cannan, and S.S. Wallace, *Expression and purification of active mouse and human NEIL3 proteins.* Protein expression and purification, 2012. **84**(1): p. 130-139.
20. Kavli, B., O. Sundheim, M. Akbari, M. Otterlei, H. Nilsen, F. Skorpen, P.A. Aas, L. Hagen, H.E. Krokan, and G. Slupphaug, *hUNG2 is the major repair enzyme for removal of uracil from U: A matches, U: G mismatches, and U in single-stranded DNA, with hSMUG1 as a broad specificity backup.* Journal of Biological Chemistry, 2002. **277**(42): p. 39926-39936.
21. Yoon, J.H., S. Iwai, T.R. O' Connor, and G.P. Pfeifer, *Human thymine DNA glycosylase (TDG) and methyl - CpG - binding protein 4 (MBD4) excise thymine glycol (Tg) from a Tg: G mispair.* Nucleic acids research, 2003. **31**(18): p. 5399-5404.
22. Sevilya, Z., Y. Leitner-Dagan, M. Pinchev, R. Kremer, D. Elinger, F. Lejbkowicz, H.S. Rennert, L.S. Freedman, G. Rennert, T. Paz-Elizur, and Z. Livneh, *Development of APE1 enzymatic DNA repair assays: low APE1 activity is associated with increase lung cancer risk.* Carcinogenesis, 2015: p. bgv082.
23. McNeill, D.R., A. Narayana, H.-K. Wong, and D.M. Wilson 3rd, *Inhibition of Ape1 nuclease activity by lead, iron, and cadmium.* Environmental Health Perspectives, 2004. **112**(7): p. 799.
24. Edwards, S.K., T. Ono, S. Wang, W. Jiang, R.M. Franzini, J.W. Jung, K.M. Chan, and E.T. Kool, *In Vitro Fluorogenic Real - Time Assay of the Repair of Oxidative DNA Damage.* ChemBioChem, 2015. **16**(11): p. 1637-1646.
25. Maksimenko, A., A.A. Ishchenko, G. Sanz, J. Laval, R.H. Elder, and M.K. Saparbaev, *A molecular beacon assay for measuring base excision repair*

activities. Biochemical and biophysical research communications, 2004. **319**(1): p. 240-246.
26. Ono, T., S.K. Edwards, S. Wang, W. Jiang, and E.T. Kool, *Monitoring eukaryotic and bacterial UDG repair activity with DNA-multifluorophore sensors.* Nucleic acids research, 2013: p. gkt309.
27. Guerniou, V., D. Rapin, J.-F. Millau, E. Bufflier, A. Favier, J. Cadet, and S. Sauvaigo, *Repair of oxidative damage of thymine by HeLa whole-cell extracts: simultaneous analysis using a microsupport and comparison with traditional PAGE analysis.* Biochimie, 2005. **87**(2): p. 151-159.
28. Sauvaigo, S., V. Guerniou, D. Rapin, D. Gasparutto, S. Caillat, and A. Favier, *An oligonucleotide microarray for the monitoring of repair enzyme activity toward different DNA base damage.* Analytical biochemistry, 2004. **333**(1): p. 182-192.
29. Nagel, Z.D., C.M. Margulies, I.A. Chaim, S.K. McRee, P. Mazzucato, A. Ahmad, R.P. Abo, V.L. Butty, A.L. Forget, and L.D. Samson, *Multiplexed DNA repair assays for multiple lesions and multiple doses via transcription inhibition and transcriptional mutagenesis.* Proceedings of the National Academy of Sciences, 2014. **111**(18): p. E1823-E1832.
30. Fry, R.C., J.P. Svensson, C. Valiathan, E. Wang, B.J. Hogan, S. Bhattacharya, J.M. Bugni, C.A. Whittaker, and L.D. Samson, *Genomic predictors of interindividual differences in response to DNA damaging agents.* Genes & development, 2008. **22**(19): p. 2621-2626.
31. Shinmura, K., S. Yamaguchi, T. Saitoh, M. Takeuchi-Sasaki, S.-R. Kim, T. Nohmi, and J. Yokota, *Adenine excisional repair function of MYH protein on the adenine: 8-hydroxyguanine base pair in double-stranded DNA.* Nucleic acids research, 2000. **28**(24): p. 4912-4918.
32. Khobta, A., N. Kitsera, B. Speckmann, and B. Epe, *8-Oxoguanine DNA glycosylase (Ogg1) causes a transcriptional inactivation of damaged DNA in the absence of functional Cockayne syndrome B (Csb) protein.* DNA repair, 2009. **8**(3): p. 309-317.
33. Kitsera, N., D. Stathis, B. Lühnsdorf, H. Müller, T. Carell, B. Epe, and A. Khobta, *8-Oxo-7, 8-dihydroguanine in DNA does not constitute a barrier to transcription, but is converted into transcription-blocking damage by OGG1.* Nucleic Acids Research, 2011.
34. van Loon, B., E. Markkanen, and U. Hübscher, *Oxygen as a friend and enemy: how to combat the mutational potential of 8-oxo-guanine.* DNA repair, 2010. **9**(6): p. 604-616.
35. Masani, S., L. Han, and K. Yu, *Apurinic/apyrimidinic endonuclease 1 is the essential nuclease during immunoglobulin class switch recombination.* Molecular and cellular biology, 2013. **33**(7): p. 1468-1473.
36. Collins, F.S., L.D. Brooks, and A. Chakravarti, *A DNA polymorphism discovery resource for research on human genetic variation.* Genome research, 1998. **8**(12): p. 1229-1231.
37. Monden, Y., T. Arai, M. Asano, E. Ohtsuka, H. Aburatani, and S. Nishimura, *Human MMH (OGG1) type 1a protein is a major enzyme for repair of 8-*

hydroxyguanine lesions in human cells. Biochemical and biophysical research communications, 1999. **258**(3): p. 605-610.
38. Bessho, T., R. Roy, K. Yamamoto, H. Kasai, S. Nishimura, K. Tano, and S. Mitra, *Repair of 8-hydroxyguanine in DNA by mammalian N-methylpurine-DNA glycosylase.* Proceedings of the National Academy of Sciences, 1993. **90**(19): p. 8901-8904.
39. Dou, H., S. Mitra, and T.K. Hazra, *Repair of oxidized bases in DNA bubble structures by human DNA glycosylases NEIL1 and NEIL2.* Journal of Biological Chemistry, 2003. **278**(50): p. 49679-49684.
40. Wilson, T.M., A. Vaisman, S.A. Martomo, P. Sullivan, L. Lan, F. Hanaoka, A. Yasui, R. Woodgate, and P.J. Gearhart, *MSH2–MSH6 stimulates DNA polymerase η, suggesting a role for A: T mutations in antibody genes.* The Journal of experimental medicine, 2005. **201**(4): p. 637-645.
41. Hadi, M.Z. and D.M. Wilson, *Second human protein with homology to the Escherichia coli abasic endonuclease exonuclease III.* Environmental and molecular mutagenesis, 2000. **36**(4): p. 312-324.
42. Fortini, P., B. Pascucci, E. Parlanti, M. D'errico, V. Simonelli, and E. Dogliotti, *8-Oxoguanine DNA damage: at the crossroad of alternative repair pathways.* Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis, 2003. **531**(1): p. 127-139.
43. Mi, R., M. Alford-Zappala, Y.W. Kow, R.P. Cunningham, and W. Cao, *Human endonuclease V as a repair enzyme for DNA deamination.* Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis, 2012. **735**(1): p. 12-18.
44. Moe, A., J. Ringvoll, L.M. Nordstrand, L. Eide, M. Bjørås, E. Seeberg, T. Rognes, and A. Klungland, *Incision at hypoxanthine residues in DNA by a mammalian homologue of the Escherichia coli antimutator enzyme endonuclease V.* Nucleic acids research, 2003. **31**(14): p. 3893-3900.
45. Lee, C.Y.I., J.C. Delaney, M. Kartalou, G.M. Lingaraju, A. Maor-Shoshani, J.M. Essigmann, and L.D. Samson, *Recognition and Processing of a New Repertoire of DNA Substrates by Human 3-Methyladenine DNA Glycosylase (AAG).* Biochemistry, 2009. **48**(9): p. 1850-1861.
46. Saparbaev, M., K. Kleibl, and J. Laval, *Escherichia coil, Saccharomyces cerevisiae, rat and human 3-methyladenine DNA glycosylases repair 1, N6-ethenoadenine when present in DNA.* Nucleic Acids Research, 1995. **23**(18): p. 3750-3755.
47. Ringvoll, J., M.N. Moen, L.M. Nordstrand, L.B. Meira, B. Pang, A. Bekkelund, P.C. Dedon, S. Bjelland, L.D. Samson, P. Falnes, and A. Klungland, *AlkB Homologue 2-Mediated Repair of Ethenoadenine Lesions in Mammalian DNA.* Cancer Research, 2008. **68**(11): p. 4142.
48. Mishina, Y., C.G. Yang, and C. He, *Direct repair of the exocyclic DNA adduct 1, N6-ethenoadenine by the DNA repair AlkB proteins.* Journal of the American Chemical Society, 2005. **127**(42): p. 14594-14595.
49. Wyatt, M.D. and L.D. Samson, *Influence of DNA structure on hypoxanthine and 1, N6-ethenoadenine removal by murine 3-methyladenine DNA glycosylase.* Carcinogenesis, 2000. **21**(5): p. 901-908.

50. Leitner-Dagan, Y., Z. Sevilya, M. Pinchev, R. Kremer, D. Elinger, H.S. Rennert, E. Schechtman, L. Freedman, G. Rennert, and Z. Livneh, *Enzymatic MPG DNA repair assays for two different oxidative DNA lesions reveal associations with increased lung cancer risk.* Carcinogenesis, 2014. **35**(12): p. 2763-2770.
51. Hill, J.W., T.K. Hazra, T. Izumi, and S. Mitra, *Stimulation of human 8-oxoguanine-DNA glycosylase by AP-endonuclease: potential coordination of the initial steps in base excision repair.* Nucleic acids research, 2001. **29**(2): p. 430-438.
52. Mokkapati, S.K., L. Wiederhold, T.K. Hazra, and S. Mitra, *Stimulation of DNA glycosylase activity of OGG1 by NEIL1: functional collaboration between two human DNA glycosylases.* Biochemistry, 2004. **43**(36): p. 11596-11604.
53. Nordmann, P.L., J.C. Makris, and W.S. Reznikoff, *Inosine induced mutations.* Molecular and General Genetics MGG, 1988. **214**(1): p. 62-67.
54. Brégeon, D. and P.W. Doetsch, *Transcriptional mutagenesis: causes and involvement in tumour development.* Nature Reviews Cancer, 2011. **11**(3): p. 218-227.
55. Morreall, J., A. Kim, Y. Liu, N. Degtyareva, B. Weiss, and P.W. Doetsch, *Evidence for Retromutagenesis as a Mechanism for Adaptive Mutation in Escherichia coli.* PLoS Genet, 2015. **11**(8): p. e1005477.
56. Kuraoka, I., M. Endou, Y. Yamaguchi, T. Wada, H. Handa, and K. Tanaka, *Effects of Endogenous DNA Base Lesions on Transcription Elongation by Mammalian RNA Polymerase II.* Journal of Biological Chemistry, 2003. **278**(9): p. 7294-7299.
57. Damsma, G.E., A. Alt, F. Brueckner, T. Carell, and P. Cramer, *Mechanism of transcriptional stalling at cisplatin-damaged DNA.* Nature structural & molecular biology, 2007. **14**(12): p. 1127-1133.
58. Tornaletti, S., L.S. Maeda, and P.C. Hanawalt, *Transcription arrest at an abasic site in the transcribed strand of template DNA.* Chemical research in toxicology, 2006. **19**(9): p. 1215-1220.
59. Brenerman, B.M., J.L. Illuzzi, and D.M. Wilson, *Base excision repair capacity in informing healthspan.* Carcinogenesis, 2014: p. bgu225.
60. Mol, C.D., T. Izumi, S. Mitra, and J.A. Tainer, *DNA-bound structures and mutants reveal abasic DNA binding by APE1 DNA repair and coordination.* Nature, 2000. **403**(6768): p. 451-456.
61. Prasad, R., D.D. Shock, W.A. Beard, and S.H. Wilson, *Substrate channeling in mammalian base excision repair pathways: passing the baton.* Journal of Biological Chemistry, 2010. **285**(52): p. 40479-40488.
62. Calvo, J.A., C.A. Moroski-Erkul, A. Lake, L.W. Eichinger, D. Shah, I. Jhun, P. Limsirichai, R.T. Bronson, D.C. Christiani, L.B. Meira, and L.D. Samson, *Aag DNA Glycosylase Promotes Alkylation-Induced Tissue Damage Mediated by Parp1.* Plos Genetics, 2013. **9**(4).
63. Crosbie, P.A.J., A.J. Watson, R. Agius, P.V. Barber, G.P. Margison, and A.C. Povey, *Elevated N3-methylpurine-DNA glycosylase DNA repair activity is associated with lung cancer.* Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis, 2012.

64. Janssen, K., K. Schlink, W. Götte, B. Hippler, B. Kaina, and F. Oesch, *DNA repair activity of 8-oxoguanine DNA glycosylase 1 (OGG1) in human lymphocytes is not dependent on genetic polymorphism Ser 326/Cys 326.* Mutation Research/DNA Repair, 2001. **486**(3): p. 207-216.
65. Paz-Elizur, T., M. Krupsky, S. Blumenstein, D. Elinger, E. Schechtman, and Z. Livneh, *DNA repair activity for oxidative damage and risk of lung cancer.* Journal of the National Cancer Institute, 2003. **95**(17): p. 1312-1319.
66. Myrnes, B., K.-E. Giercksky, and H. Krokan, *Interindividual variation in the activity of O6-methyl guanine-DNA methyltransferase and uracil-DNA glycosylase in human organs.* Carcinogenesis, 1983. **4**(12): p. 1565-1568.
67. Raetz, A.G., Y. Xie, S. Kundu, M.K. Brinkmeyer, C. Chang, and S.S. David, *Cancer-associated variants and a common polymorphism of MUTYH exhibit reduced repair of oxidative DNA damage using a GFP-based assay in mammalian cells.* Carcinogenesis, 2012. **33**(11): p. 2301-2309.
68. Theodoratou, E., H. Campbell, A. Tenesa, R. Houlston, E. Webb, S. Lubbe, P. Broderick, S. Gallinger, E. Croitoru, and M. Jenkins, *A large-scale meta-analysis to refine colorectal cancer risk estimates associated with MUTYH variants.* British journal of cancer, 2010. **103**(12): p. 1875-1884.
69. Vasko, M.R., C. Guo, and M.R. Kelley, *The multifunctional DNA repair/redox enzyme Ape1/Ref-1 promotes survival of neurons after oxidative stress.* DNA repair, 2005. **4**(3): p. 367-379.
70. Fung, H. and B. Demple, *A vital role for Ape1/Ref1 protein in repairing spontaneous DNA damage in human cells.* Molecular cell, 2005. **17**(3): p. 463-470.
71. Sevilya, Z., Y. Leitner-Dagan, M. Pinchev, R. Kremer, D. Elinger, H.S. Rennert, E. Schechtman, L.S. Freedman, G. Rennert, T. Paz-Elizur, and Z. Livneh, *Low integrated DNA repair score and lung cancer risk.* Cancer Prevention Research, 2013. **7**(4): p. 398-406.
72. Jalal, S., J.N. Earley, and J.J. Turchi, *DNA repair: From genome maintenance to biomarker and therapeutic target.* Clinical Cancer Research, 2011.
73. Leone, G., M.T. Voso, S. Sica, R. Morosetti, and L. Pagano, *Therapy related leukemias: susceptibility, prevention and treatment.* Leukemia & lymphoma, 2001. **41**(3-4): p. 255-276.
74. Oeffinger, K.C. and S. Bhatia, *Second primary cancers in survivors of childhood cancer.* The Lancet, 2009. **374**(9700): p. 1484-1485.
75. Wood, M.E., V. Vogel, A. Ng, L. Foxhall, P. Goodwin, and L.B. Travis, *Second malignant neoplasms: assessment and strategies for risk reduction.* Journal of Clinical Oncology, 2012. **30**(30): p. 3734-3745.
76. Lee, M.-R., S.-H. Kim, H.-J. Cho, K.-Y. Lee, A.R. Moon, H.G. Jeong, J.-S. Lee, J.-W. Hyun, M.-H. Chung, and H.J. You, *Transcription factors NF-YA regulate the induction of human OGG1 following DNA-alkylating agent methylmethane sulfonate (MMS) treatment.* Journal of Biological Chemistry, 2004. **279**(11): p. 9857-9866.
77. Dizdaroglu, M., A. Karakaya, P. Jaruga, G. Slupphaug, and H.E. Krokan, *Novel activities of human uracil DNA N-glycosylase for cytosine-derived products of oxidative DNA damage.* Nucleic acids research, 1996. **24**(3): p. 418-422.

78. Wang, D., M. Luo, and M.R. Kelley, *Human apurinic endonuclease 1 (APE1) expression and prognostic significance in osteosarcoma: enhanced sensitivity of osteosarcoma to DNA damaging agents using silencing RNA APE1 expression inhibition.* Molecular cancer therapeutics, 2004. **3**(6): p. 679-686.
79. Kryston, T.B., A.B. Georgiev, P. Pissis, and A.G. Georgakilas, *Role of oxidative stress and DNA damage in human carcinogenesis.* Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis, 2011. **711**(1): p. 193-201.
80. Wyatt, M.D. and D.M. Wilson III, *Participation of DNA repair in the response to 5-fluorouracil.* Cellular and molecular life sciences, 2009. **66**(5): p. 788-799.
81. Hwang, I.T., Y.M. Chung, J.J. Kim, J.S. Chung, B.S. Kim, H.J. Kim, J.S. Kim, and Y. Do Yoo, *Drug resistance to 5-FU linked to reactive oxygen species modulator 1.* Biochemical and biophysical research communications, 2007. **359**(2): p. 304-310.
82. Huehls, A.M., C.J. Huntoon, P.M. Joshi, C.A. Baehr, J.M. Wagner, X. Wang, M.Y. Lee, and L.M. Karnitz, *Genomically Incorporated 5-Fluorouracil that Escapes UNG-Initiated Base Excision Repair Blocks DNA Replication and Activates Homologous Recombination.* Molecular pharmacology, 2016. **89**(1): p. 53-62.
83. An, Q., P. Robins, T. Lindahl, and D.E. Barnes, *5-Fluorouracil incorporated into DNA is excised by the Smug1 DNA glycosylase to reduce drug cytotoxicity.* Cancer research, 2007. **67**(3): p. 940-945.
84. Andersen, S., T. Heine, R. Sneve, I. König, H.E. Krokan, B. Epe, and H. Nilsen, *Incorporation of dUMP into DNA is a major source of spontaneous DNA damage, while excision of uracil is not required for cytotoxicity of fluoropyrimidines in mouse embryonic fibroblasts.* Carcinogenesis, 2005. **26**(3): p. 547-555.
85. Sansom, O.J., J. Zabkiewicz, S.M. Bishop, J. Guy, A. Bird, and A.R. Clarke, *MBD4 deficiency reduces the apoptotic response to DNA-damaging agents in the murine small intestine.* Oncogene, 2003. **22**(46): p. 7130-7136.
86. Seiple, L., P. Jaruga, M. Dizdaroglu, and J.T. Stivers, *Linking uracil base excision repair and 5-fluorouracil toxicity in yeast.* Nucleic acids research, 2006. **34**(1): p. 140-151.
87. Fischer, J.A., S. Muller-Weeks, and S.J. Caradonna, *Fluorodeoxyuridine modulates cellular expression of the DNA base excision repair enzyme uracil-DNA glycosylase.* Cancer research, 2006. **66**(17): p. 8829-8837.
88. Devaud, N. and S. Gallinger, *Chemotherapy of MMR-deficient colorectal cancer.* Familial cancer, 2013. **12**(2): p. 301-306.
89. David, S.S., V.L. O'Shea, and S. Kundu, *Base-excision repair of oxidative DNA damage.* Nature, 2007. **447**(7147): p. 941-950.
90. Guo, J., P.C. Hanawalt, and G. Spivak, *Comet-FISH with strand-specific probes reveals transcription-coupled repair of 8-oxoGuanine in human cells.* Nucleic Acids Research, 2013.
91. Krokeide, S.Z., J.K. Laerdahl, M. Salah, L. Luna, F.H. Cederkvist, A.M. Fleming, C.J. Burrows, B. Dalhus, and M. Bjørås, *Human NEIL3 is mainly a monofunctional DNA glycosylase removing spiroimindiohydantoin and guanidinohydantoin.* DNA repair, 2013. **12**(12): p. 1159-1164

Chapter V: The novel *in vivo* role of Transcription-Coupled Nucleotide Excision Repair of the lipid-peroxidation product, 3,N^4-ethenocytosine

Isaac A. Chaim, Alycia M. Gardner, Teruaki Iyama,
David M Wilson III, Leona D. Samson

Contributions: I.A.C., A.M.G and L.D.S. designed experiments. I.A.C. and A.M.G. conducted all experiments. I.A.C. and A.M.G analyzed data. T.I. and D.M.WIII provided isogenic TC-NER proficient and deficient cell lines and advice.

Table of Contents

Introduction

Etheno(ε)-adducts are highly mutagenic DNA base lesions characterized by an exocyclic (imidazole) ring and are produced endogenously by lipid peroxidation (LPO) products that are generated under oxidative stress conditions [1-4] (**Figure 5.1**). ε-adducts were first identified as products that result from exposure of DNA bases to chloroacetaldehyde (CAA) [5]. The role of these lesions in carcinogenesis became evident after occupational exposure to vinyl chloride (VC), a precursor of chloroethylene oxide and CAA [6], was correlated to development of a very rare tumor, angiosarcoma of the liver (ASL) [7-9]. Studies in rodents have shown comparable effects after both VC [10] and urethane [11] exposure. A series of studies followed, demonstrating the toxicity and mutagenicity of these lesions in both bacteria [12-16] and mammalian cells [13, 17, 18].

With the development of more sensitive techniques to detect ε-adducts, background levels of εA and εC were first measured in rodents [19, 20]. Interestingly, a range of adduct levels were detected for different tissues suggesting variation in tissue DNA repair capacity and/or differential tissue exposure to damaging agents [19]. It was later shown that LPO products, in particular *trans*-4-hydroxy-2-nonenal (HNE), and its epoxide intermediate, are a major endogenous source of ε-adducts [1-3]. Lipid peroxides are formed from the reaction of reactive oxygen and nitrogen species (RONS) with polyunsaturated fatty acid residues of phospholipids. Conditions such as inflammation, impaired metal transport, or dietary imbalance can induce persistent oxidative stress and LPO excess, which in turn causes an accumulation of ε-adducts in different tissues. Accordingly, many of these conditions are characteristic of cancer-prone diseases (**Table 1**, reviewed in Nair *et al* 2007 [21]).

Two DNA repair pathways are known to repair of ε-adducts: Base Excision Repair (BER) and Direct Reversal (DR). BER (**Figure 5.2**) is a finely tuned

process that begins with the excision of a damaged base by a DNA glycosylase generating an abasic (AP) site. For monofunctional DNA glycosylases this AP site is processed by the apurinic/apyriminidic endonuclease 1 (APE1) generating a single strand break with a 3'-OH and 5'-deoxyribosephosphate (5'-dRP). DNA polymerase β (Pol β) removes the 5'-dRP moiety and fills in the gap by inserting the complementary base to the undamaged strand. Repair is completed by the action of DNA ligase which seals the nick [22]. DR (**Figure 5.3**) by the ALKBH proteins do not involve the activity of any other proteins and repairs the lesion directly. The repair reaction begins with epoxidation of the alkyl chain. Addition of water then opens the ring, forming a glycol intermediate that is spontaneously released as a glyoxal [23].

In mammalian cells, εA is repaired by the Alkyladenine DNA glycosylase (AAG/MPG) [24, 25] and the AlkB homolog 2 (ALKBH2) [26]. There are conflicting reports in the literature regarding the role of ALKBH3 in repairing εA present in ssDNA [26-28]. Interestingly, εC can be bound by AAG, yet it is not excised by it [25]. This interaction has been shown to block the repair of εC by ALKBH2 *in vitro* [29]. Counterintuitively, $Aag^{-/-}$ mice show accumulation of εC in the absence of Aag. This might implicate the recruitment of additional repair proteins by the εC-AAG complex likely through transcription coupled repair [30]. εC has also been shown to be repaired *in vitro* by other DNA glycosylases such as SMUG1 [31] and TDG [32] and weakly by MBD4 [33]. 1,N^2-εG is bound and excised by AAG but in a weaker and less efficient manner than εA [25]; its repair by ALKBH3 was also recently described [34]. No repair mechanisms for N^2,3-εG have been described in mammals so far.

Another DNA repair pathway that has been proposed to play a role in the repair of some DNA lesions formed by LPO products, such as bulky HNE-DNA adducts, is Nucleotide Excision Repair (NER) [4, 35] (**Figure 5.4**). NER has two distinct sub-pathways: Transcription Coupled NER (TC-NER), which recognizes lesions that stall RNA Polymerase II (RNAP II) during active transcription, and Global

Genome NER (GG-NER), which can act on bulky DNA lesions regardless of their transcription status [3]. These sub-pathways only differ in their method of lesion recognition, after which they converge into the same downstream enzymatic steps. Deficiencies in TC-NER result in neurodegenerative disorders such as Cockayne's Syndrome (CS) [2,3], while deficiencies in GG-NER and the downstream steps result Xeroderma Pigmentosum (XP), a disease characterized by a 2,000 fold increased risk of developing skin cancer [36]. Interestingly, cells deficient in TC-NER (specifically, CSB) were hypersensitive to physiological concentrations of HNE and developed a higher level of sister chromatid exchanges in comparison to the wild type cells, suggesting that HNE-DNA adducts can be formed endogenously in mammalian cells, block transcription and are processed by the TC-NER [4, 35].

Here we show that εC, a DNA lesion produced endogenously by HNE, can cause transcriptional blockage and transcriptional mutagenesis *in vivo.* Furthermore, we demonstrate that unexpectedly, εC is repaired by TC-NER. We achieve this by taking advantage of our recently described methods on using modified fluorescence multiplexed host cell reactivation (FM-HCR) assays for DNA repair capacity measurements.

Materials and Methods

Plasmids

As described previously [37], EGFP, mOrange and tagBFP reporter genes subcloned into the pmaxCloning Vector (Lonza) between the KpnI and SacI restriction sites in the multiple cloning site were employed. Plasmids were amplified in *E.coli* DH5α (Invitrogen) and purified using Qiagen endotoxin-free Maxi and Giga kits.

UV treatment

Plasmids were irradiated in TE buffer (10 mM Tris·HCl, 1 mM EDTA, pH 7.0) at a DNA concentration of 50 ng/μL in a volume of 1.5 mL in 10-cm polystyrene Petri dishes (without lids) with 800 J/m^2 UV-C light generated by a Stratalinker 2000 box.

Site-Specific Thymine Dimer Reporter

pMax:GFP 615AA plasmid (described previously [37]) was prepared using a Giga Prep kit according to manufacturer instructions (Qiagen). The plasmid was nicked with Nb.BtsI (New England Biolabs) to generate a single-strand break in the transcribed strand. The nicked strand was then digested with ExoIII (New England Biolabs), and the remaining single-stranded circular DNA (ssDNA) was purified using a 1% agarose gel. The ssDNA was extracted from the agarose gel using a Gel Extraction Kit (Qiagen). 20 μg of the ssDNA were combined with 9 μL of a 10 mM oligonucleotide sequence containing a thymine dimer localized at bases 614 and 615 of the plasmid [5' TCAGGGCGGAXGGTTGC 3', where X denotes a thymine dimer] (Trilink Biotech) (1/200, oligo/ssDNA molar ratio) and 1x Pfu Polymerase Buffer (Thermo Scientific). The mixture was heated to 85°C in a thermocycler for six minutes, and then allowed to anneal by cooling to 37°C at 1°C per minute. Following annealing, 12 μL of 10 mM dNTPs and 30 U Pfu Polymerase AD were added to the mixture. The oligo was extended at 61°C for 1.5 hours. Following extension, the plasmid was purified using a PCR Purification kit following manufacturer instructions (Qiagen). After elution in 105.4 μL EB Buffer (10 mM Tris-Cl, pH 8.5), 5 μL T4 DNA Ligase buffer (New England Biolabs), 2 μL 10 mM dNTPs, 2 μL 25 mM ATP, 0.5 μL 10mg/mL BSA, 1.5 U T4 DNA Polymerase (New England Biolabs) and 8 U T4 DNA Ligase (New England Biolabs) were added, and the solution was incubated at 16°C for 1 hour. The solution was subjected to electrophoresis in a 1% Agarose gel to purify the ligated band. Ligated plasmid was extracted from the gel using a Gel Extraction

kit (Qiagen). Gel extraction was performed according to manufacturer instructions except that plasmids were eluted in TE buffer (10mM Tris-Cl, pH 8.0, 1mM EDTA) (Invitrogen)

Site-Specific εA (GFP-616-εA) and εC (GFP-615-εC) reporter

pMax:GFP WT plasmid was prepared using a Giga Prep kit according to manufacturer instructions. Single stranded plasmid DNA was prepared as described above. Annealing of a site-specific oligo was performed as described above, using an oligonucleotide containing εA [5'GCTCAGGGCGGXCTGGGTGCTCAGGTAGTG, where X denotes εA] or εC [GCTCAGGGCGGAXTGGGTGCTCAGGTAGTG, where X denotes εC] (Life Technologies). Following annealing, 12 µL of 10 mM dNTPs and 30 U of Pfu Polymerase AD were added to the mixture, and the oligo was extended at 68°C for 1.5 hours. After extension, the plasmid was purified using a phenol/chloroform/isoamyl alcohol extraction and ethanol precipitation. The plasmid DNA was resuspended in a mixture containing 105.4 µL EB Buffer (10 mM Tris-Cl, pH 8.5), 5 µL T4 DNA Ligase buffer (New England Biolabs), 2 µL 10 mM dNTPs, 2 µL 25 mM ATP, and 0.5 µL 100x BSA. Once the DNA was resuspended, 1.5 U T4 DNA Polymerase and 8 U T4 DNA Ligase (New England Biolabs) were added, and the solution was incubated at 16°C for 1 hour. In order to remove excess salts from the solution, the buffer was changed following ligation using 30,000 NMWL Filtration columns (Millipore). The plasmid-containing solution was concentrated three times and finally resuspended in 150 µL TE for purification by gel electrophoresis. In order to purify the plasmid, a 1.5-2% NuSieve GTG low melting point agarose gel (Lonza) was used. The plasmid DNA was extracted from the gel according to the manufacturer's instructions. Briefly, the gel was melted and a series of phenol/chloroform extractions were performed, followed by ethanol precipitation and resuspension of the plasmid in TE buffer.

Primary Cells and Cell Lines

Animals

Aag$^{-/-}$ and Triple Knockout (TKO, *Aag*$^{-/-}$,*Alkbh2*$^{-/-}$*Alkbh3*$^{-/-}$) mice were described previously [38]. All mice used for experiments were on a mixed B6/129 background. Age- and gender-matched mice were 6–10 weeks old at the time of euthanasia. Mice were euthanized by CO_2 asphyxiation and theirs spleen was subsequently removed. Single-cell suspensions of splenocytes are prepared by pushing tissue pieces through a 70µm pore size nylon mesh screen. Erythrocytes are lysed in ACK lysis buffer.

Mice were housed in an Association of Assessment and Accreditation of Laboratory Animal Care–accredited facility. All procedures were approved by the Massachusetts Institute of Technology Committee on Animal Care.

Human Cells

B-lymphoblastoid cells (GM01712, GM01857, GM01953, GM02246, GM02345, GM02253, GM07752 and GM21148) were maintained in RPMI 1640 medium (Gibco) supplemented with 15% FBS, 2 mM L-glutamine, and 1% penicillin and streptomycin. SV40-transformed CSA-deficient CS3BE and CSB-deficient CS1AN cell lines (Coriell Institute) transfected with pcDNA-CSA or pcDNA-CSB, respectively (previously described [39]) were maintained in DMEM medium (Gibco) supplemented with 10% HI-FBS, 2 mM L-glutamine, and 1% penicillin and streptomycin. Adherent cell lines were sub-cultured by trypsinization.

Rodent Cells

Mouse embryonic fibroblasts (MEFs) were prepared from wild-type C57BL/6J and previously described Aag$^{-/-}$ mouse models[16]. Primary MEFs were previously

immortalized by transfecting MEFs with a pSV3-neo plasmid expressing a large T-antigen. Immortalized MEFs were cultured in DMEM medium (Gibco) supplemented with 10% HI-FBS, 2 mM L-glutamine, and 1% penicillin and streptomycin, and were subcultured by trypsinization. Cells were maintained at 37°C and 5% CO_2.

Splenic lymphocytes were cultured in supplemented RPMI Glutamax: 20 % heat-inactivated FBS, 1% penicillin/streptomycin. B-lymphocytes are stimulated with 10 µg/mL Kdo2-Lipid A (Avanti Polar Lipids) for 48 hours before transfection.

DNA Repair Assays Transfections

Electroporation

Lymphoblastoid cell lines were transfected by electroporation. 2 x 10^6 lymphoblastoid cells in 100 µL of complete medium were combined with a reporter plasmid mixture (**Table 5.2**). Cells were electroporated using a 96-well electroporation plate and gene pulser (Bio-Rad) with an exponential waveform at 260 V and 950 µF. Following electroporation, 100 µL of complete medium were added to each well in the plate, and the 200 µL volume was transferred to a 12-well plate containing 800µL of equilibrated medium.

Lipofection

MEFs and human fribroblasts were transfected by lipofection using Lipofectamine LTX (LifeTechnologies), following manufacturer's instructions. Briefly, plasmid reporter DNA mixture (**Table 5.2**) combined with Opti-MEM, Lipofectamine and Plus reagents, and 200 µL of the total mixture were added to each well of the 6-well plate.

Flow Cytometry analysis of lesions

Eighteen hours after transfection, cells suspended in culture media were analyzed for fluorescence on a BD LSR II cytometer running FACSDiva software. Cell debris, doublets, and aggregates were excluded based on their side- and forward-scatter properties. TO-PRO–3 was added to cells 5–10 min before analysis and was used to exclude dead cells from the analysis. The following fluorophores and their corresponding detectors (in parentheses) were used: tagBFP (Pacific Blue), EGFP (FITC), mOrange (phycoerythrin; PE), and TO-PRO–3 (allophycocyanin; APC). Compensation was set by using single-color controls. Regions corresponding to cells positive for each of the five fluorescent proteins were established by using single-color dropout controls. For reporters that required compensation in more than one detector channel, fluorescence in the reporter channel was plotted separately against each of the channels requiring compensation. Using these plots, both single controls and the dropout control (in which the reporter of interest was excluded from the transfection) were used to establish regions corresponding to positive cells.

Calculation of Percent Fluorescent Reporter Expression.

Every experimental setup consisted of two sets of transfections: A control transfection (CT) and a sample transfection (ST) containing one or more reporters with DNA lesions. Both transfections included the same color combination with the same undamaged reporter to normalize each set for transfection efficiency.

Fluorescence Index (FI) for a given reporter within one transfection was calculated as follows:

$$FI = \frac{C_F \times MFI}{C_L}$$

where C_F is the number of positive fluorescent cells for that given fluorophore, *MFI* is the mean fluorescence intensity of the C_F, and C_L is the total number of live cells.

The normalized fluorescence index for a given reporter FI^O was calculated as follows:

$$FI^O = \frac{FI^n}{FI^E}$$

where FI^n corresponds to the *FI* of a reporter normalized to the *FI* of the transfection efficiency normalization plasmid, FI^E.

Normalized reporter expression from a sample transfection, $FI^O{}_{ST}$, and that from the same reporter plasmid in control transfection, $FI^O{}_{CT}$, were used to compute the percent reporter expression (%R.E.) as follows:

$$\%\, R.E. = \frac{FI^O{}_{ST}}{FI^O{}_{CT}} \times 100$$

RNA isolation for mRNA-Seq

At 18 h, a fraction of transfected cells were harvested by centrifugation, washed three times with PBS, and resuspended in 1 mL TRIzol reagent. The suspension was extracted with 200µL chloroform. The aqueous phase was removed, combined with one volume of absolute ethanol, and applied to a Qiagen RNeasy miniprep spin column. The column then was washed two times with 500 µL Buffer RPE (Qiagen) and finally eluted in 40µL diethylpyrocarbonate (DEPC)-treated water. The quality of the RNA preparation was determined by using a bioanalyzer to confirm an RNA integrity number (RIN) of at least 9.0. At least

190ng of total RNA was stored in TE buffer at −80 °C until submission for mRNA sequencing (mRNA-Seq).

Total RNA samples were submitted to the Massachusetts Institute of Technology BioMicroCenter for preparation and sequencing. Briefly, total RNA was poly-A purified, fragmented, and converted to cDNA by using the Clontech protocol. Library construction from cDNA was performed using the Beckman Coulter SPRIworks system. During library amplification, a unique bar code was introduced for each of the twelve samples corresponding to the four transfections performed in triplicate (**Table 5.3**) and from which total RNA was generated. All samples were combined and run on a NextSeq instrument.

Next generation sequencing data analysis

RSEM [40] (version 1.2.15) was used to estimate gene expression based on hg19 UCSC known gene annotations. The count table was imported into DESeq [41] (version 1.10.1) for differential gene expression analysis. BWA[42] (version 0.7.10) was used to map the raw reads to the plasmids. Then all properly paired, uniquely and perfectly mapped reads were counted to estimate expression of the reporters. Relative expression levels of the reporters were computed by normalizing their counts to BFP expression in the same sample. SAMtools [43] (version 1.3) mpileup (options: -d1000000 -B) was used to process the bam alignments, followed by VarScan [44] (version 2.3.6) to call mutations on the reporters.

Results

εA is repaired in vivo by AAG but not ALKBH2 or ALKBH3; εC is not repaired by any of them.

In order to test the role of known DNA repair proteins on εA and εC *in vivo*, we performed an HCR assay by transfecting our site-specific reporters into repair proficient (WT^{B6} and $WT^{129/B6}$) and deficient MEFs (*Aag*$^{-/-}$ or *Aag*$^{-/-}$/*Alkbh2*$^{-/-}$/*Alkbh3*$^{-/-}$, **T**riple **K**nock**o**ut, TKO). Two different WT MEFs were used as controls since the repair deficient cells that were derived from different mouse backgrounds (C57BL/6 and mixed B6/129s). The presence of either εA or εC results in decreased % reporter expression (%R.E.) compared to undamaged controls. This decrease is stronger for εC than for εA lesions (**Figure 5.5**, plasmid combinations #1,2,3 in **Table 5.2**). For the εA reporter *Aag*$^{-/-}$ MEFs exhibit significantly lower %R.E. compared to WT, agreeing with previously published work demonstrating εA repair by AAG initiated BER [45]; interestingly, there appears to be no difference in the repair of εC in WT versus *Aag*$^{-/-}$ cells. Moreover, TKO MEFs while deficient in the repair of εA, they do not appear more deficient that *Aag*$^{-/-}$ cells, but given their different backgrounds a direct comparison should be avoided. Moreover TKO MEFs show no difference in εC repair compared to WT cells (**Figure 5.6**, plasmid combinations #1,2,3 in **Table 5.2**).

The same repair relationships where observed following transfection of site-specific εA and εC plasmids into Kdo2-Lipid-A-stimulated mouse primary splenic-B-lymphocytes (**Figure 5.7**, plasmid combinations #4,5,6 in **Table 5.2**). In the absence of AAG, repair of εA is reduced, while further deficiencies in ALKBH2 and ALKBH3 do not exacerbate this repair deficit. εC repair remains constant in all genotypes.

TC-NER and some downstream NER components are involved in the repair of εC.

Given the strong decrease in %R.E. observed for site-specific εC lesions we sought to determine if this decrease was a result of a transcription block that could be repaired by the NER pathway. To this end, we used a panel of B-lymphoblastoid cell lines derived from individuals with Xeroderma Pigmentosum (XP) and Cockayne Syndrome (CS), diseases characterized by deficiencies in NER and sensitivity to UV light. To verify defects in NER, we first tested the ability of each cell line to repair UV-induced lesions. Plasmids treated with 800 J/m^2 UV-C were chosen as this dose had been previously reported to show significant differences between NER proficient and deficient cells [37]. As expected, deficiencies in CSA, XPA, XPB, XPC and XPD resulted in decreased reactivation of the damaged plasmid compared to the reactivation in healthy control (**Figure 5.8**, plasmid combinations #7,8 in **Table 5.2**). While plasmid reactivation appeared lower in CSB-deficient cells compared to cells from the healthy control, the difference did not quite reach significance ($p = 0.056$).

We had previously shown that XPA deficient lymphoblastoids could not efficiently repair plasmids containing a site-specific thymine dimer (T^T), the most common lesion caused by UV light [37]. Here we test whether the NER deficiencies of the rest of the XP and CS cell lines could be phenotyped using a known site-specific transcription-blocking lesion, namely a thymine dimer. Following transfection of a GFP reporter bearing a site-specific T^T, we reproduced our previous XPA deficient results and further report similar repair deficits in cells deficient for TC-NER (CSA and CSB) as well as other downstream NER components (XPB and XPD). Interestingly, deficiencies in the GG-NER sub-pathway (cells deficient in XPC) do not appear to affect the repair of this site-specific lesion (**Figure 5.9**, plasmid combinations #7,9 in **Table 5.2**) even though the repair of randomly induced UV damage is clearly deficient (**Figure 5.8**)

After validating that most NER deficient cell lines are deficient in the repair of a site-specific transcription blocking base lesion, we transfected the panel of XP and CS cell lines with either a site-specific εA or εC plasmid. None of the cells were deficient in the repair of εA; curiously the XPC deficient cell line displayed higher εA repair capacity that the healthy control (**Figure 5.10**, plasmid combinations #7,10 in **Table 5.2**). In stark contrast, repair of εC was deficient in 5 out of 6 NER deficient cell lines, closely recapitulating the repair phenotypes observed for the site-specific T^T (**Figure 5.11**, plasmid combinations #7,11 in **Table 5.2**). TC-NER as well as downstream NER deficiencies displayed significantly diminished εC repair capacity, with the exception of XPC. Again, the GG-NER deficient XPC line appears more repair proficient than the healthy control, this time for εC repair. It should be noted that "healthy" cell line GM07752 happens to be a particularly low εC repair cell line; if compared with another "healthy" lymphoblastoid, GM01953 (**Figure 5.12**, plasmid combinations #7,10 in **Table 5.2**).

To further investigate the role of TC-NER in the repair of εC, we utilized two pairs of isogenic human immortalized fibroblasts derived from Cockayne Syndrome patients deficient in either CSA (CS3BE) or CSB (CS1AN) and their respective repair proficient complemented counterparts. We determined the repair capacity of these four cell lines for UV-C treated plasmids and for plasmids with site-specific T^T, εA and εC lesions. Consistent with our previous results in non-isogenic lymphoblastoid cells, both CSA and CSB deficient cells showed impaired repair of UV-C treated plasmids as well as site-specific T^T and εC plasmids in comparison to their complemented equivalents. Furthermore, no differences in the repair of εA were detected in either CSA or CSB deficient cells (**Figure 5.13A and 5.13B**, plasmid combinations #12-16 in **Table 5.2**). Altogether, these results show that cells deficient in TC-NER are deficient in εC repair.

εC can be bypassed by RNAPII and is a source of transcriptional mutagenesis

Given this newly discovered *in vivo* blockage of transcription by εC and its repair through TC-NER, we sought to determine whether this lesion could be some times bypassed *in vivo* by RNA Pol II during transcription resulting in transcriptional mutagenesis. εC is positioned at GFP base-pair 615 of the coding strand, which is at the wobble position and codes for a glutamine. Point mutations to T also result in glutamine and fluorescence, whereas point mutations to either G or A result in histidine and a non-fluorescent variant (results not shown). As such, given our fluorescent experimental setup, we cannot discern if fluorescence is due to εC repair or due to the misincorporation of an adenine during transcription. Similarly, reduced fluorescence could be due to either transcriptional blockage or transcriptional mutagenesis to cytosine or uracil. To address these possibilities, we performed high-throughput sequencing of the reporter transcripts. We performed a multiplexed HCR assay that included a GFP site-specific εC plasmid and a UV-C treated (800 J/m^2) mOrange (plasmid combinations #17,18 in **Table 5.2**) in CSA deficient and CSA complemented fibroblasts. At the time of analysis, the samples were split in two; one half was assayed through flow cytometry, whereas the other half was used for RNA isolation and submitted for mRNA-Seq. Both analyses yielded remarkably similar results (**Figure 5.14**), the differences between proficient and deficient cells in repair capacity of both UV-irradiated and εC plasmids are maintained for this multiplexed transfection. Furthermore, sequence level analysis of mRNA-Seq data at and around site 615 of GFP where εC was positioned, reveals that the lesion can be bypassed by RNAPII (**Figure 5.15**). Transcripts from control plasmids containing an undamaged cytosine at site 615 had a very low mutagenesis frequency (0.08% in average), thus validating the observations in the presence of the lesion. This bypass nevertheless, can result in the incorporation of an erroneous base across the lesion, more specifically A > T(U) > C (5.6%, 3.3% and 0.6%, respectively - in CSA deficient cells). This

relationship is maintained in both CSA-proficient and –deficient cells. As expected, repair proficient cells show a lower percentage of transcriptional mutagenesis events given that the lesion is being actively repaired.

Discussion

Using of a modified HCR assay, we were able to elucidate a novel role for the NER pathway, specifically TC-NER, in the repair of the highly mutagenic DNA lesion, εC. Given that this lesion is a product of endogenous LPO products, characteristic of cancer-prone inflammatory diseases, we propose that TC-NER may play an important role in protection against inflammation associated carcinogenesis.

It is noteworthy that we do not observe a mayor contribution by ALKBH2 and ALKBH3 in the repair of εA, *in vivo*, in either MEFs or mouse splenic-B-lymphocytes. Both enzymes have been shown *in vitro* to repair εA DNA lesions[26-28]. Deletion of both enzymes together with AAG, in chronic-inflammation mouse models, have been shown lo lead to increased accumulation of ε-adducts in colon epithelium (specifically 1,N^2-εG, while εA did not reach significance) compared to AAG deletion alone [38]. It is then possible that the contribution of these enzymes in εA repair is cell-type specific and not expressed or active in MEFs and splenic-B-lymphocytes. Alternatively, *in vivo*, they might indeed not play a significant role in the repair of εA.

Interestingly, εC can be bound by AAG, yet it is not excised by it [25]. *In vitro*, this interaction blocks the repair of εC by ALKBH2 [29]. Surprisingly, *Aag*$^{-/-}$ mice show increase accumulation of εC in the absence of AAG in the colonic mucosa of IBD mouse models [46]. As such, we hypothesized that if AAG binds to εC and recruits downstream repair enzymes, we would observe decreased εC plasmid reactivation in Aag$^{-/-}$ cells. Alternatively, we postulated increased εC repair in *Aag*$^{-/-}$ cells if the absence of AAG allows ALKBH2 to access the lesion.

Nevertheless, neither was the case for our *in vivo* experimental set up. Given the lack of a clear role from AAG and ALKBH2 on εC repair, the strong decrease in reporter expression observed, lead us to propose that the presence of the lesion causes a transcription blockage in a similar fashion to our previous results regarding site-specific T^T reporters [37], and that it could in consequence be repaired by the NER machinery.

Validation of the repair phenotype of NER deficient cell lines by the use of a UV irradiated plasmid resulted in the expected decrease reporter expression. It is quite interesting that the repair of a site-specific T^T is not affected in XPC deficient cells but that this is not the case for the repair of UV-C irradiated plasmids. Perhaps, the absence of GG-NER has no effect on a single transcription-blocking lesion in an actively transcribed gene that is repaired by TC-NER. But, accumulation of lesions in the plasmid (as for UV-treated plasmids) results in an overwhelmed TC-NER machinery that is unable to cope with the amount of damage present in the absence of GG-NER.

Our *in vivo* results validate the capacity of our site-specific HCR assay to reflect the phenotype of deficiencies in TC-NER and the downstream steps of NER. Based on the results obtained for XPC deficient cells, we can't differentiate between our method being unable to pick up deficiencies in GG-NER in the repair of site-specific transcription-blocking lesions or cells simply not utilizing GG-NER for the repair of single DNA lesions in a highly transcribed gene. Given that the plasmid used is non-replicative and that transcription is highly favored by a CMV promoter, it is possible that the observed roles for TC- and GG-NER would change if the plasmid would be replicative. Another indication that repair of site-specific transcription-blocking lesions is occurring through TC-NER and not GG-NER is that TC-NER deficiencies show similar repair phenotypes for site-specific T^T repair as downstream-NER deficient cells, even though this wasn't the case with UV-irradiated plasmids. It calls our attention the fact that a single site-specific transcription-blocking lesions has almost the same effect on %R.E.

as a highly UV irradiated plasmid in TC-NER deficient cells but not in any others. A rationale for this observation is not immediately obvious.

The fact that repair phenotype of a site-specific T^T by human lymphoblastoid cell lines deficient in NER is very closely recapitulated by that of a site-specific εC, directly links the repair of εC to the NER pathway, in particular TC-NER and downstream NER components. This observation was further validated in a different cell type, human fibroblasts, in cell lines deficient or complemented in either CSA or CSB. Importantly, each cell pair was isogenic minimizing the contribution of any inter-individual variations in NER capacity.

Involvement of TC-NER has been postulated in the repair of BER intermediates. In particular, the contribution of CSB to the repair of 8oxoG and uracil have been shown to be DNA glycosylase dependent (OGG1 and UNG, respectively) in mouse cells [47, 48] and human fibroblasts [49]. Similarly, CSB has been implicated in the repair of single strand breaks (SSBs) in actively transcribed genes. SSBs decrease transcription rates when present in either the transcribed or non-transcribed strand [50]. These SSBs could also be formed as BER intermediates downstream of a DNA glycosylase. Nevertheless, we do not observe at any point an AAG-dependent TC-NER involvement for εA repair due to the potential formation of BER intermediates. We cannot rule out the possibility that εC repair by TC-NER involves to some degree formation of BER intermediates created by the action of a DNA glycosylase (such as TDG, MBD4 or SMUG1) on εC. Similarly, even though AAG does not seem to play a role in the repair of εC under our experimental conditions, it is possible that when present, it binds tightly to εC and blocks transcription past the lesion, thus promoting TC-NER.

mRNA-Seq analysis of our reporter transcripts allowed us to corroborate that, in fact, fewer transcripts are being produced in the presence of εC, validating the transcription blocking effect of the lesion observed by flow cytometric analysis of

the FM-HCR assay. Moreover, we were able to describe the transcriptional mutagenesis properties of εC *in vivo*. The only *in vitro* studies on the effect of εC on transcription date back to the 1980s where CAA treated poly(dC) oligos were shown to block transcription and induce misincorporations by calf thymus RNAPII [51]. To our knowledge, the transcription blockage and mutagenic properties of εC described here are novel *in vivo* features for this lesion. RNAPII transcriptional mutagenesis tends to resemble misincorporations across from DNA lesions by mammalian DNA polymerases during DNA replication [52]. In the case of εC, misincorporations during replication have the following relationship T>A>>C [17]; however we do not observe the same relationship for transcription, A incorporation is significantly higher than T (U) incorporation in cells were εC is not efficiently repaired.

In the broader context of LPO induced DNA damage in mammals, while bulky HNE-DNA adducts are widely known to block transcription and their repair to be processed through TC-NER [4, 35], for the non-bulky ε-adducts, only 1,N^2-εG had been reporter to block transcription *in vitro* [53]. Altogether, our *in vivo* results showing transcriptional blockage and transcriptional mutagenesis by εC complement the notion that DNA damage and repair play a pivotal role in the etiology of cancer-prone inflammatory diseases and other diseases characterized by increased in oxidative stress that can lead to the formation of LPO products. Repair of εC then becomes highly relevant in both a replicative and non-replicative context. On the one hand, it can contribute in the accumulation of potentially deleterious mutations during replication; and on the other hand it can perturb cellular homeostasis via RNAPII blockage or production of mutated transcripts via transcriptional mutagenesis.

Tables

Table 5.1 ε-lesions in disease associated with increased cancer risk (modified from Nair *et al* 2007 [21]).

Predisposing condition	Site affected	Fold increase in εdA or εdC levels
Wilson's disease **[54]**	*Liver*	3
Chronic pancreatitis **[55]**	*Pancreas*	3-28
Crohn's disease **[55]**	*Colon*	2-20
Ulcerative colitis **[55]**	*Colon*	4 (εdC only)
Alcohol related hepatitis/ cirrhosis **[56]**	*Liver*	16 (εdA)
H. pylori infection and high salt intake **[57]**	*Stomach*	εdA increased (urine)

Table 5.2 Combinations of reporter plasmids and types of DNA damage used in each experiment.

Combination	*tagBFP*	*EGFP*	*mOrange*	*Carrier*
1	WT 90 ng	WT 30 ng		
2	WT 90 ng	616-εA 30 ng		
3	WT 90 ng	615-εC 30 ng		
4	WT 400 ng	WT 100 ng		
5	WT 400 ng	616-εA 100 ng		
6	WT 400 ng	615-εC 100 ng		
7	WT 100 ng	WT 100 ng		
8	WT 100 ng	UV (800 J/m^2) 100 ng		
9	WT 100 ng	614-T^T 100 ng		
10	WT 100 ng	616-εA 100 ng		
11	WT 100 ng	615-εC 100 ng		
12	WT 25 ng	WT 25 ng		450 ng
13	WT 25 ng	UV (800 J/m^2) 25 ng		450 ng
14	WT 25 ng	614-T^T 25 ng		450 ng
15	WT 25 ng	616-εA 25 ng		450 ng
16	WT 25 ng	615-εC 25 ng		450 ng
17	WT 25 ng	WT 25 ng	WT 25 ng	450 ng
18	WT 25 ng	615-εC 25 ng	UV (800 J/m^2) 25 ng	450 ng

Table 5.3 Samples submitted for next-generation sequencing.
A total of twelve samples were submitted for complete mRNA-Seq using 40- bp paired-end libraries. These included damaged and undamaged samples for the two cell lines and three replicates for each cell line.

Sample ID	Cell Line	Status	Replicate
D15-8383	CS3BE	Undamaged	1
D15-8384	CS3BE	Damaged	1
D15-8385	CS3BE + CSA	Undamaged	1
D15-8386	CS3BE + CSA	Damaged	1
D15-8387	CS3BE	Undamaged	2
D15-8388	CS3BE	Damaged	2
D15-8389	CS3BE + CSA	Undamaged	2
D15-8390	CS3BE + CSA	Damaged	2
D15-8391	CS3BE	Undamaged	3
D15-8392	CS3BE	Damaged	3
D15-8393	CS3BE + CSA	Undamaged	3
D15-8394	CS3BE + CSA	Damaged	3

Figures

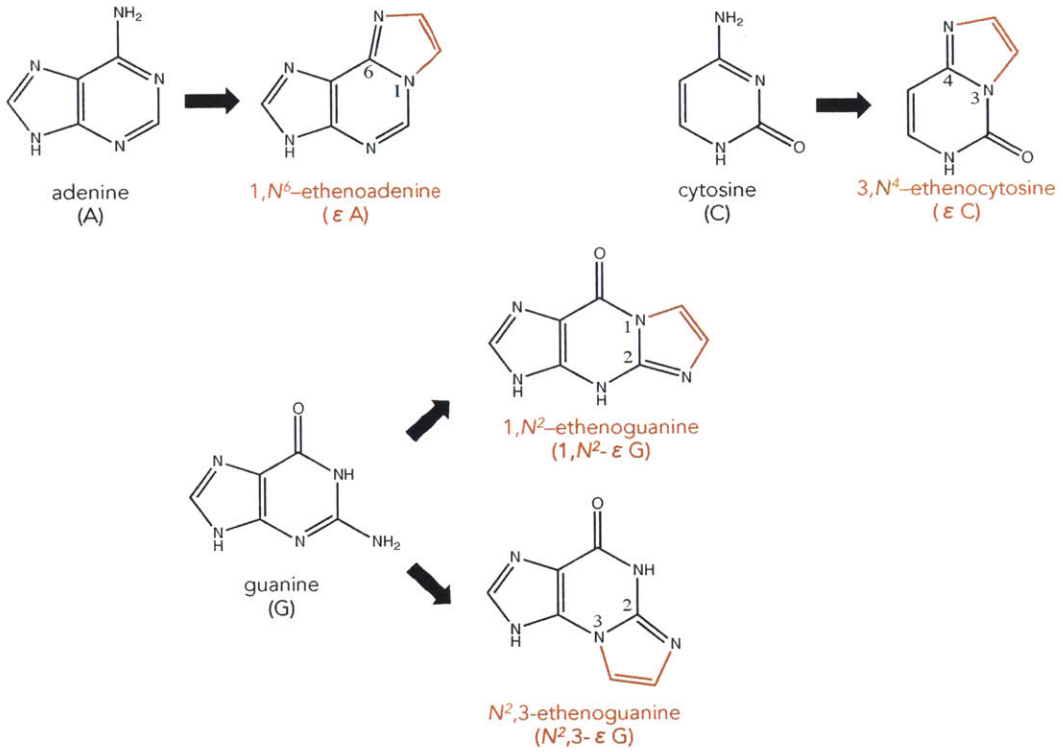


Figure 5.1 DNA Bases and their etheno(ε)-adduct derivatives.
ε-adducts are DNA base lesions characterized by the formation of and exocyclic (imidazole) ring (shown in red) in adenine, cytosine and guanine.

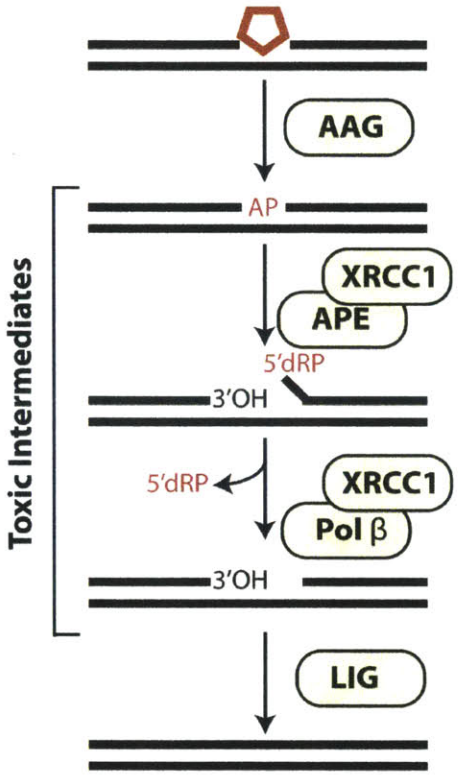


Figure 5.2 The Base Excision Repair (BER) Pathway.

The simplest version of BER begins with the excision of a damaged base by a DNA glycosylase generating an abasic (AP) site. For monofunctional DNA glycosylases this AP sites is processed by the apurinic/apyriminidic endonuclease 1 (APE1) generating a single strand break with a 3'-OH and 5'-deoxyribosephosphate (5'-dRP). DNA polymerase β (Pol β) removes the 5'-dRP moiety and fills in the gap by inserting the complementary base to the undamaged strand. Repair is completed by the action of DNA ligase that seals the nick [22]. Red pentagon represents the exocyclic imidazole ring found in ε-adducts.

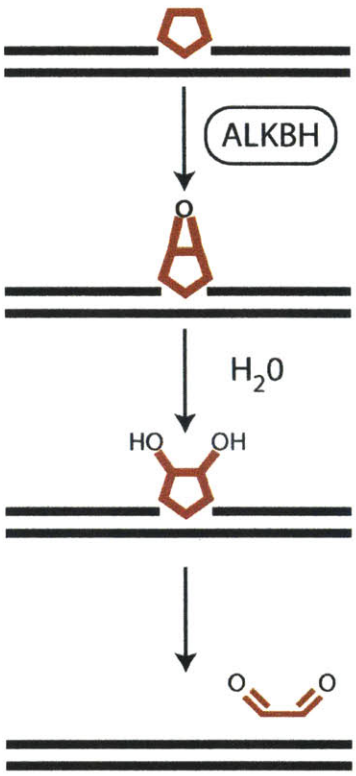


Figure 5.3 Direct Reversal Repair by the ALKBH enzymes.

Direct Reversal by they ALKBH enzymes begins with epoxidation of the alkyl chain of the lesion. Addition of water then opens the ring, forming a glycol intermediate that is spontaneously released as a glyoxal [23]. Red pentagon represents the exocyclic imidazole ring found in ε-adducts.

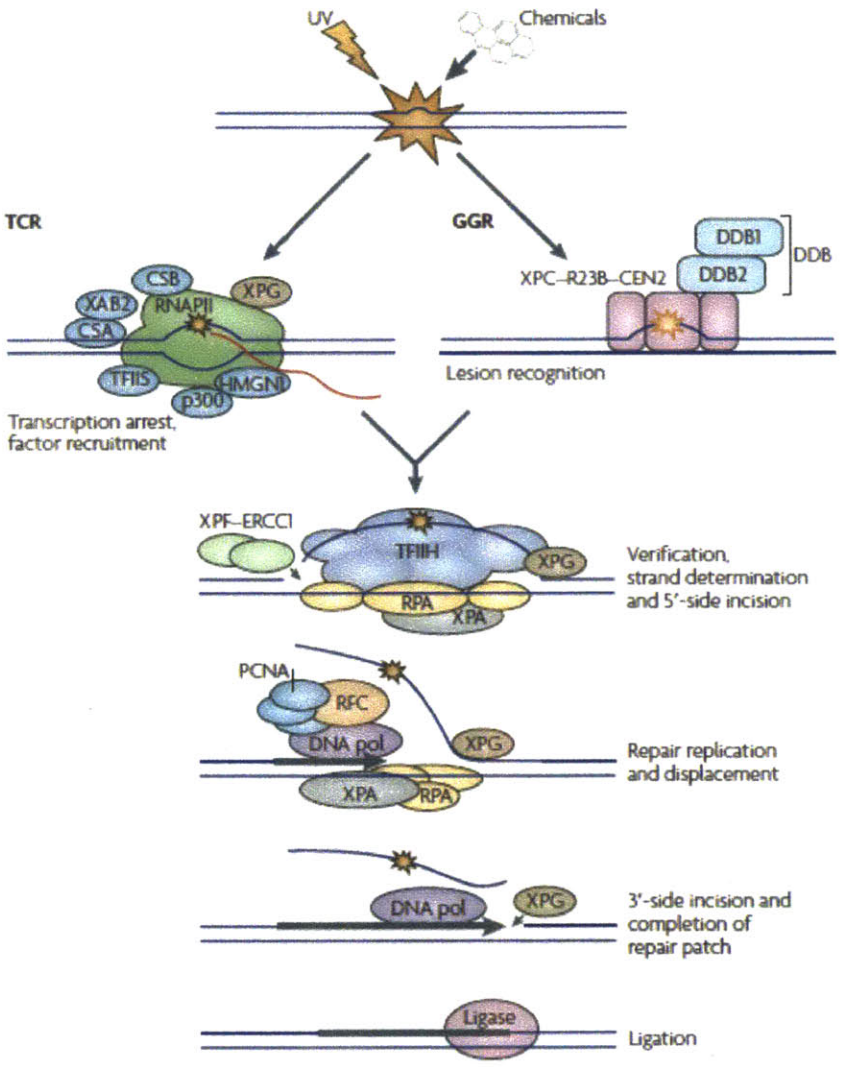


Figure 5.4 Two subpathways of Nucleotide Excision Repair.

Lesions are initially recognized, either by a translocating RNA polymerase (RNAP) (for transcription coupled NER (TCR); left) or through the binding of the lesion sensor DNA damage-binding-2 (DDB2; the product of xeroderma pigmentosum complementation group E (XPE)) — which forms a heterodimer with DDB1 to constitute the DDB complex— and/or XPC in complex with RAD23B and centrin-2 (CEN2) (for global genomic NER (GGR); right). The subpathways converge to the following steps for nucleotide excision repair

(NER): transcription factor TFIIH is recruited (along with XPG, which stabilizes TFIIH); and the helicase and ATPase activities of its subunits XPD and XPB, respectively, are stimulated for further opening of the damaged DNA (not shown). TTDA, another subunit of TFIIH, is required for NER, but its role has not been clarified. Replication protein A (RPA) and XPA might be present before and/or after the appearance of TFIIH, as they have lesion-verification roles and protect the single-stranded DNA in the denatured bubble and stabilize the pre-incision complex. The XPF–ERCC1 (excision repair cross complementing-1) endonuclease complex is recruited and incises the damaged DNA strand at the 5′ side of the bubble, whereas XPG incises on the 3′ side. Replication factor C (RFC) loads the processivity factor proliferating cell nuclear antigen (PCNA) to accommodate DNA polymerases (DNA pol) δ, ε and/or κ that have been implicated in repair replication. The final ligation step can be carried out by ligase-I and flap endonuclease-1 or by the ligase-III–XRCC1 complex. (Taken from Hanawalt and Spivak, 2008 [58])

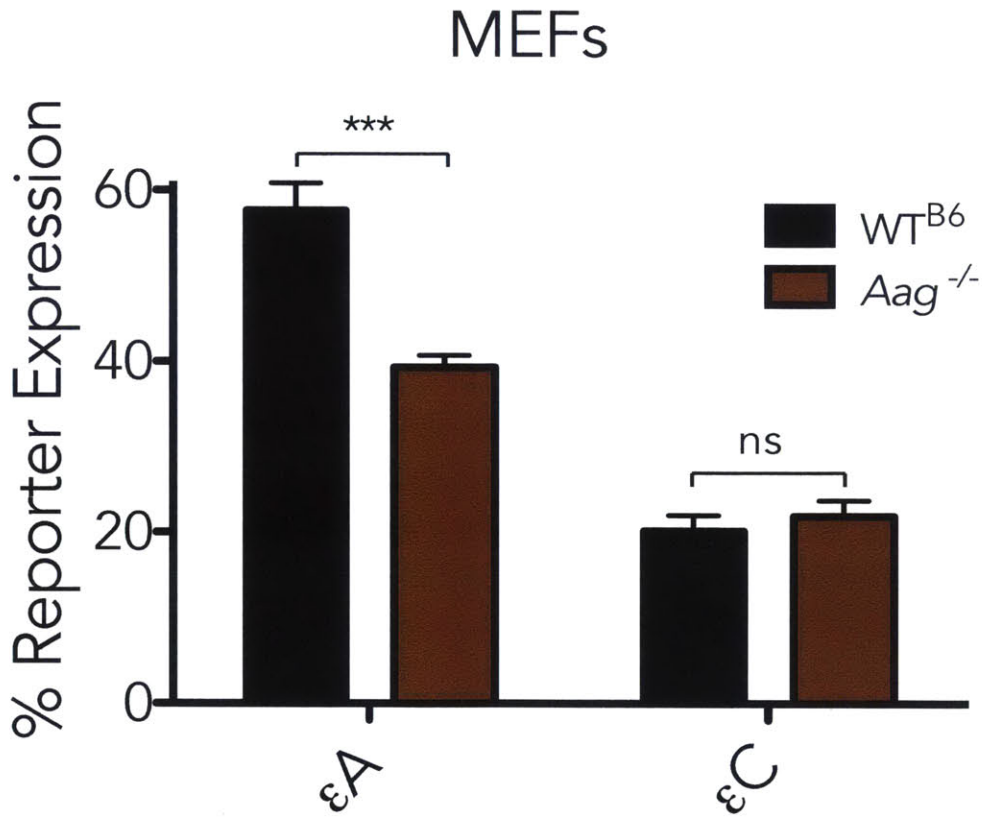


Figure 5.5 Repair of εA and εC in WT and *Aag*$^{-/-}$ MEFs.

Cell lines were transfected with site-specific εA or εC GFP repair reporters. Both cell lines are derived from C57BL/6 background mice. Error bars show standard deviation of at least 3 biological replicates. ***, $p \leq 0.001$; ns, not significant.

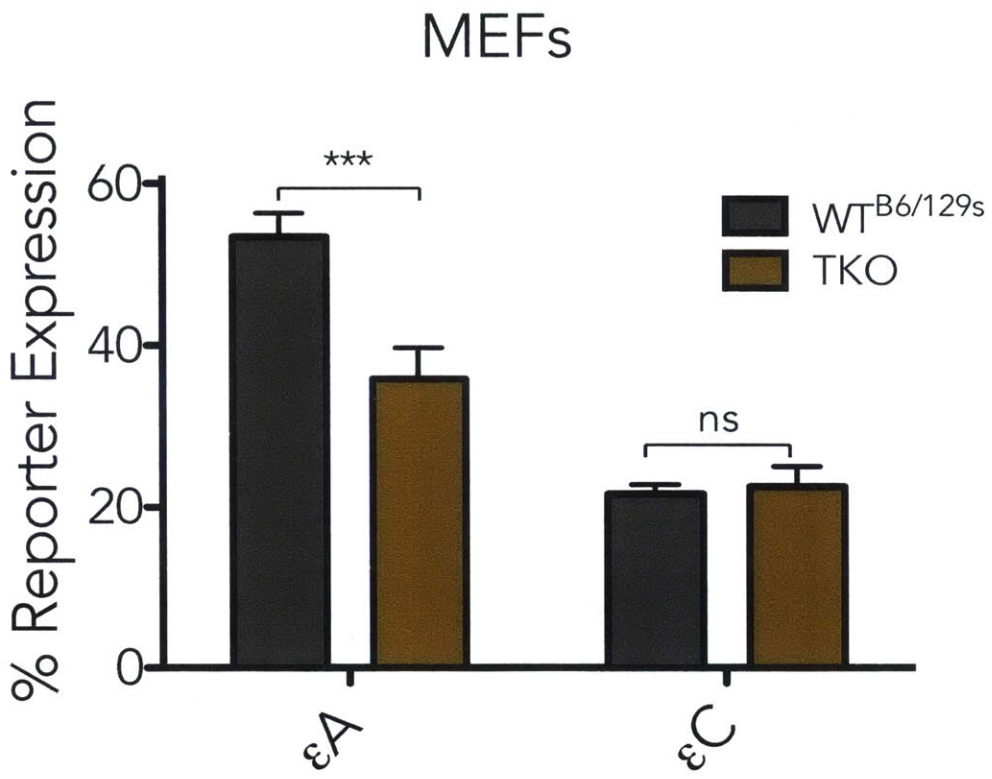


Figure 5.6 Repair of εA and εC in WT and Triple Knockout MEFs.
Cell lines were transfected with site-specific εA or εC GFP repair reporters. TKO = Triple Knockout, *Aag*$^{-/-}$/*Alkbh2*$^{-/-}$/*Alkbh3*$^{-/-}$). Both cell lines are derived from mixed background mice (C57BL/6 and 129s). Error bars show standard deviation of at least 3 biological replicates. ***, $p \leq 0.001$; ns, not significant.

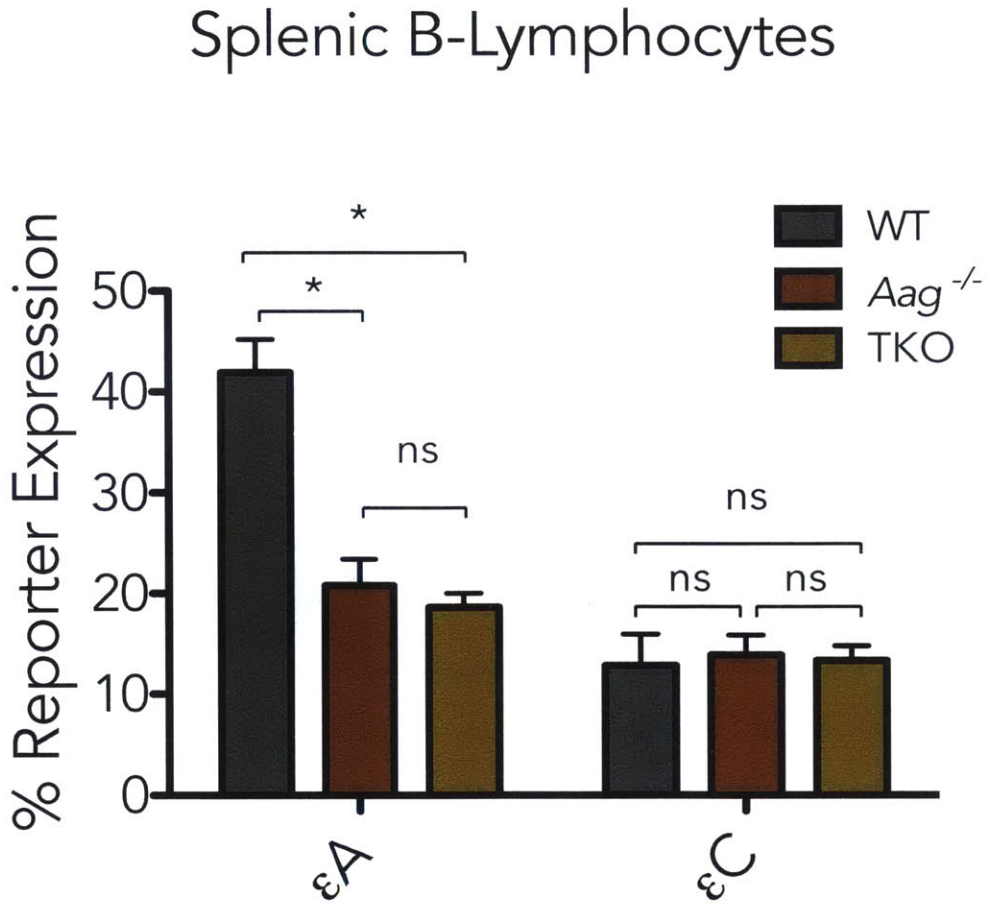


Figure 5.7 Repair of εA and εC in B-splenic-lymphocytes from WT, *Aag*$^{-/-}$ and Triple Knockout MEFs.

Following spleen removal, lymphocytes were stimulated with 10 μM LPS for 48 hours before transfection. Cell were transfected with site-specific εA or εC GFP repair reporters. TKO = Triple Knockout, *Aag*$^{-/-}$/*Alkbh2*$^{-/-}$/*Alkbh3*$^{-/-}$). All mice were in a mixed background (C57BL/6 and 129s). Error bars show standard deviation of 2 technical replicates. *, $p \leq 0.05$; ns, not significant.

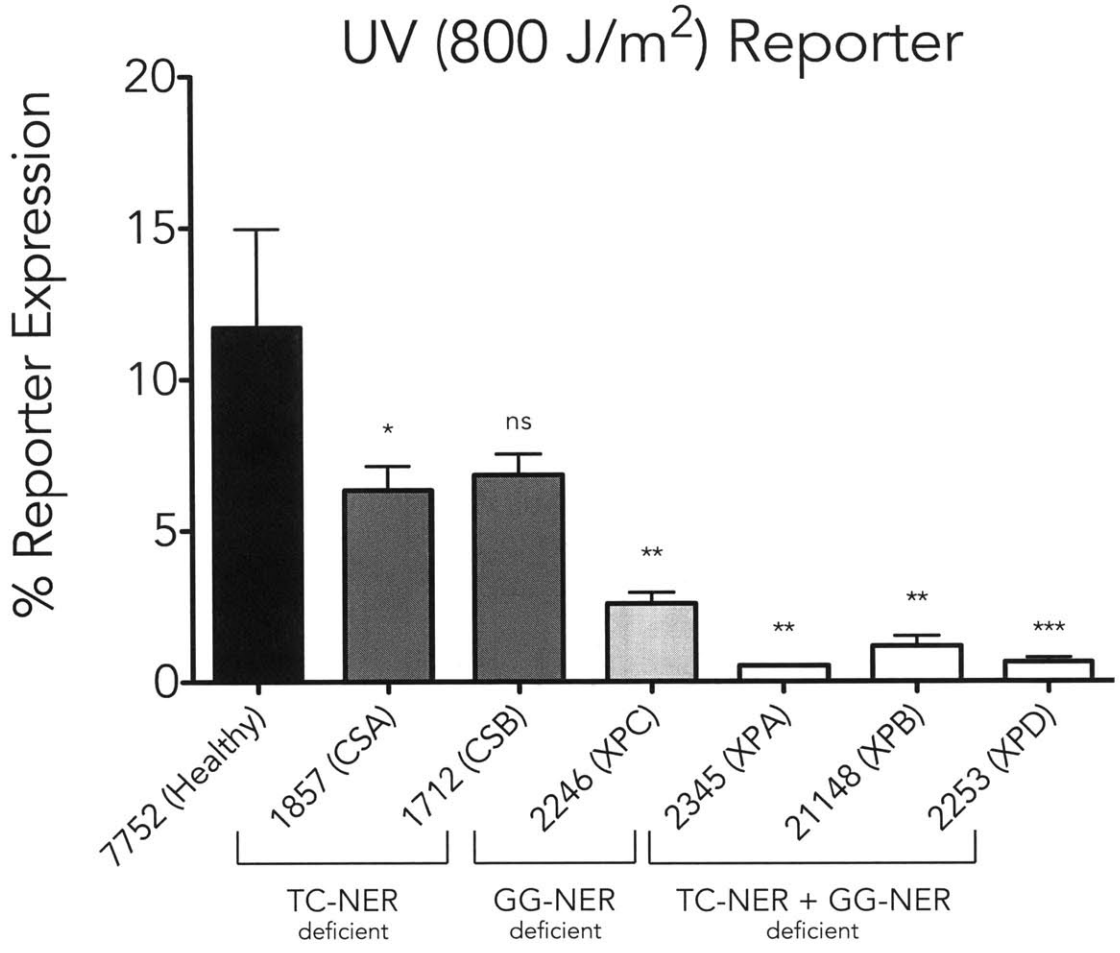


Figure 5.8 Repair of UV-irradiated plasmids by a panel of B-lymphoblastoid cell lines with varying NER deficiencies.

Cell lines were transfected with GFP repair reporters irradiated with 800 J/m^2 UV-C. Black box corresponds to NER proficient cells; dark grey, TC-NER deficient; light grey, GG-NER deficient; white, downstream-NER deficient. Error bars show standard deviation of at least 3 biological replicates. Statistics between healthy control and each deficient cell line *, $p \leq 0.05$; **, $p \leq 0.01$; ***, $p \leq 0.001$; ns, not significant.

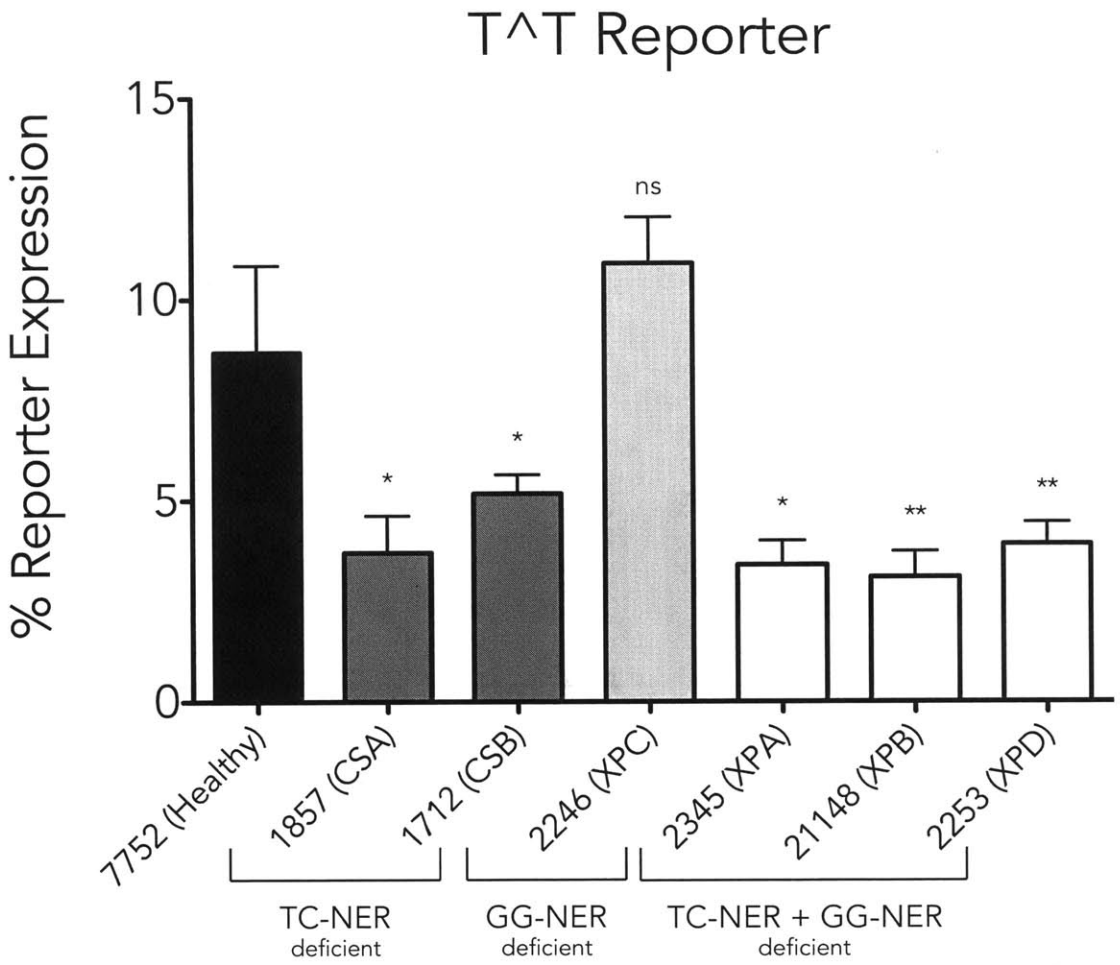


Figure 5.9 Repair of thymine dimers by a panel of B-lymphoblastoid cell lines with varying NER deficiencies.

Cell lines were transfected with a site-specific T^T GFP repair reporter. Black box corresponds to NER proficient cells; dark grey, TC-NER deficient; light grey, GG-NER deficient; white, downstream-NER deficient. Error bars show standard deviation of at least 3 biological replicates. Statistics between healthy control and each deficient cell line *, $p \leq 0.05$; **, $p \leq 0.01$; ns, not significant.

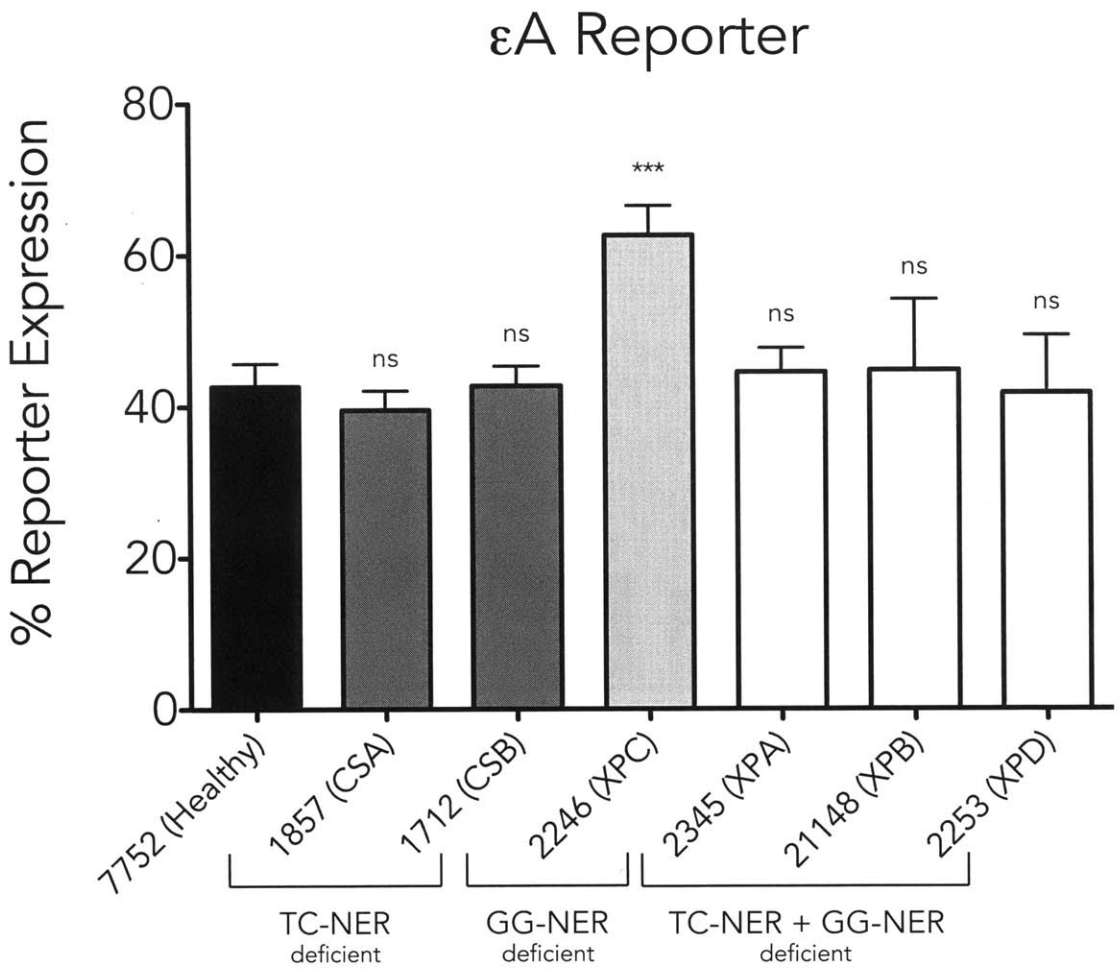


Figure 5.10 Repair of εA by a panel of B-lymphoblastoid cell lines with varying NER deficiencies.

Cell lines were transfected with a site-specific εA GFP repair reporter. Black box corresponds to NER proficient cells; dark grey, TC-NER deficient; light grey, GG-NER deficient; white, downstream-NER deficient. Error bars show standard deviation of at least 3 biological replicates. Statistics between healthy control and each deficient cell line ***, $p \leq 0.001$; ns, not significant.

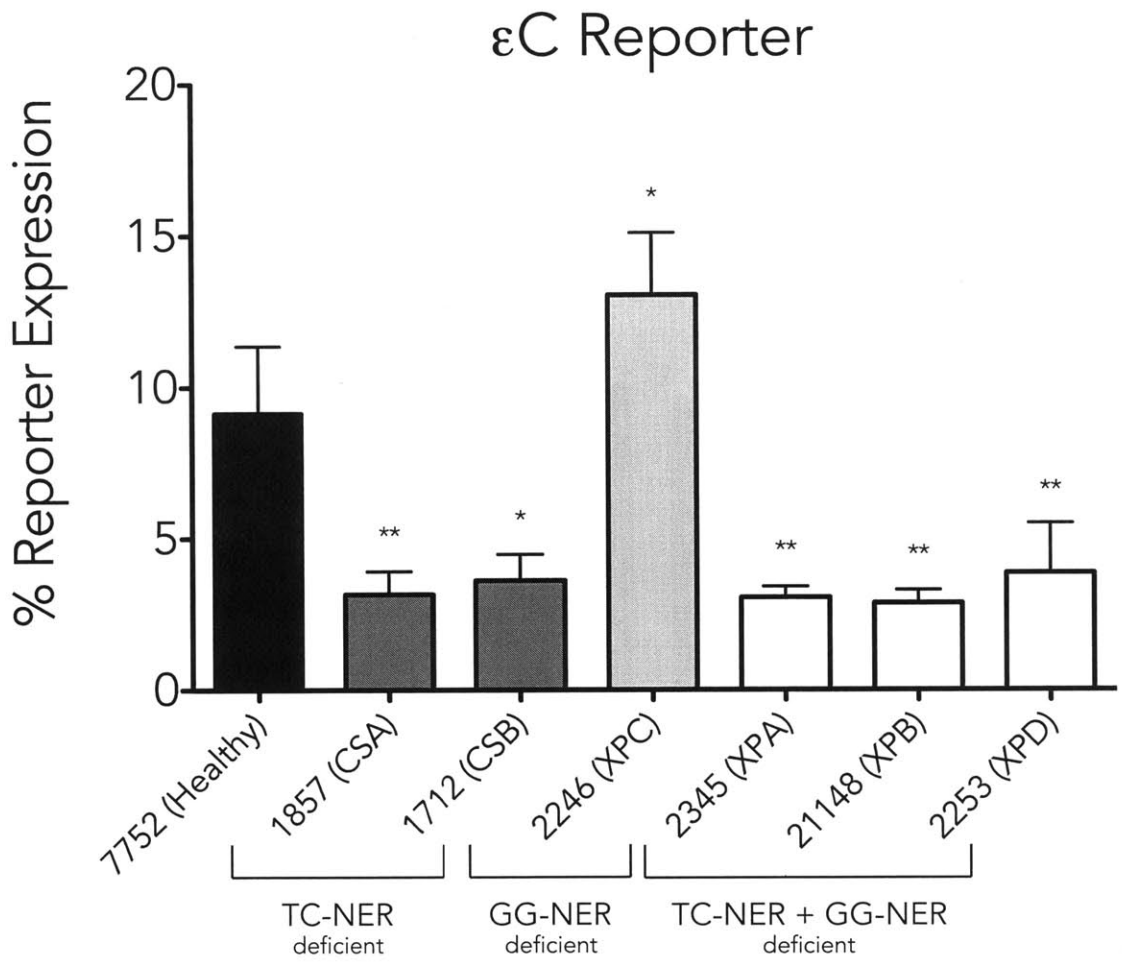


Figure 5.11 Repair of εC by a panel of B-lymphoblastoid cell lines with varying NER deficiencies.

Cell lines were transfected with a site-specific εC GFP repair reporter. Black box corresponds to NER proficient cells; dark grey, TC-NER deficient; light grey, GG-NER deficient; white, downstream-NER deficient. Error bars show standard deviation of at least 3 biological replicates. Statistics between healthy control and each deficient cell line *, $p \leq 0.05$; **, $p \leq 0.01$; ns, not significant.

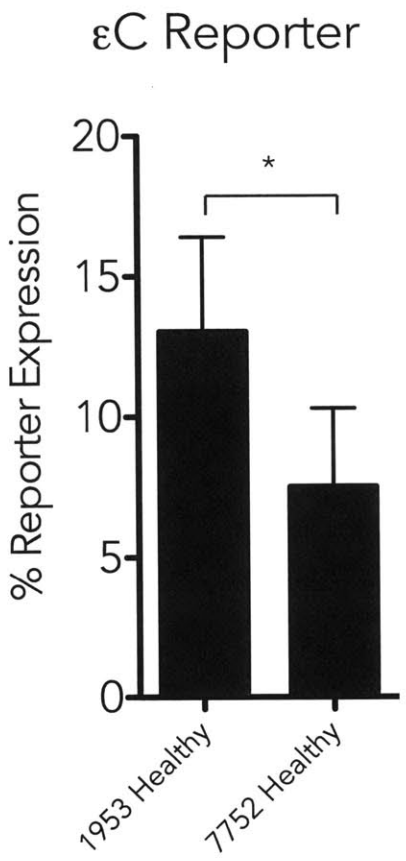


Figure 5.12 Repair of εC by two "healthy" B-lymphoblastoid cell lines.
Cell lines were transfected with a site-specific εC GFP repair reporter. Error bars show standard deviation of at least 3 biological replicates. *, $p \leq 0.05$.

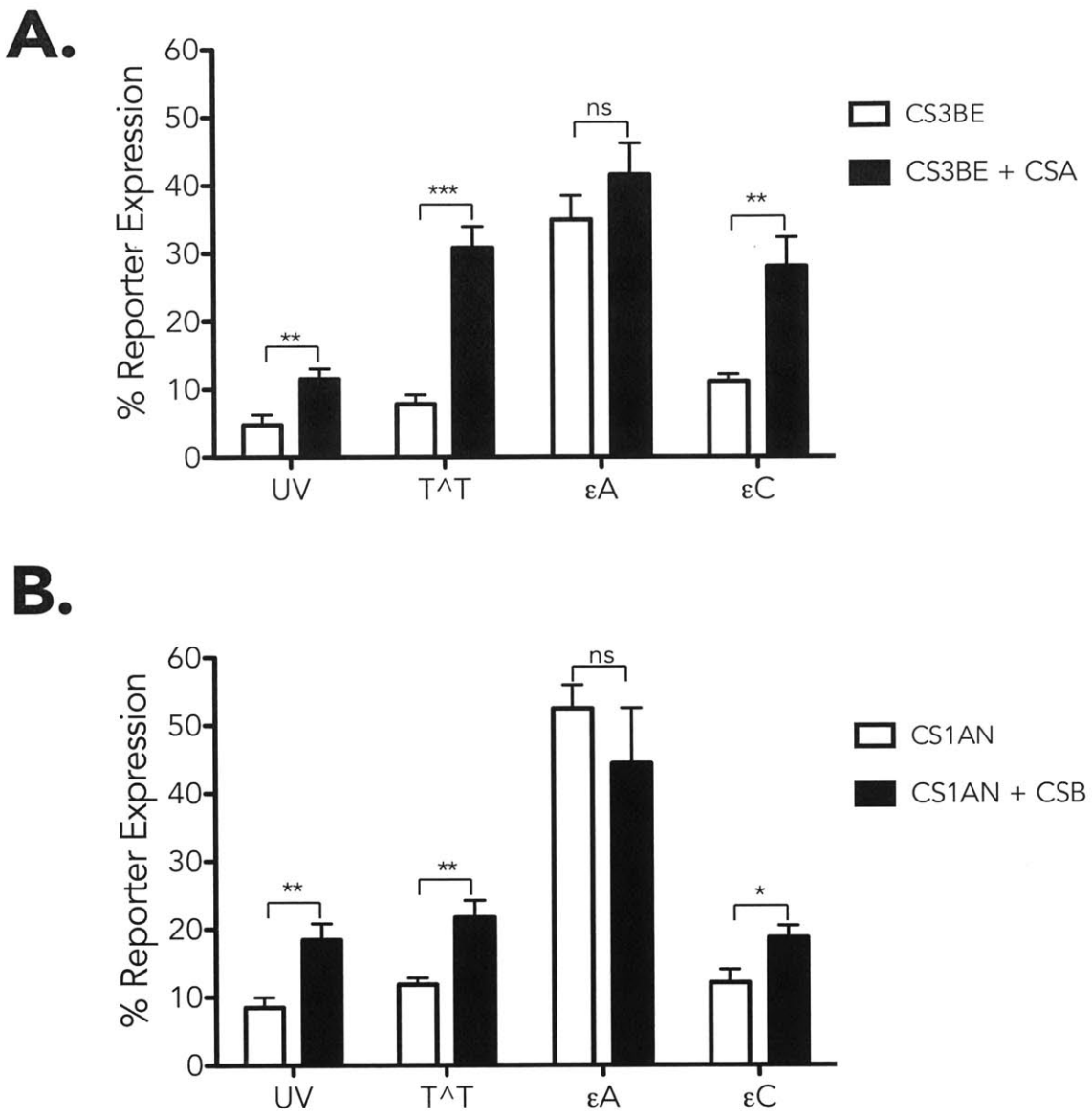


Figure 5.13 DNA Repair Capacity measurements of isogenic patient derived fibroblasts.

DNA Repair Capacity of two pairs of isogenic patient derived fibroblasts (**A.** CSA, **B.**CSB) deficient (white boxes) or complemented with the respective deficient protein (dark boxes) was measured. They were transfected with GFP repair reporters that were (1) irradiated with 800 J/m^2 UV-C or that contained a site-specific (2) T^T (3) εA or (4) εC. Error bars show standard deviation of at least 3 biological replicates. *, $p \leq 0.05$; **, $p \leq 0.01$; ns, not significant.

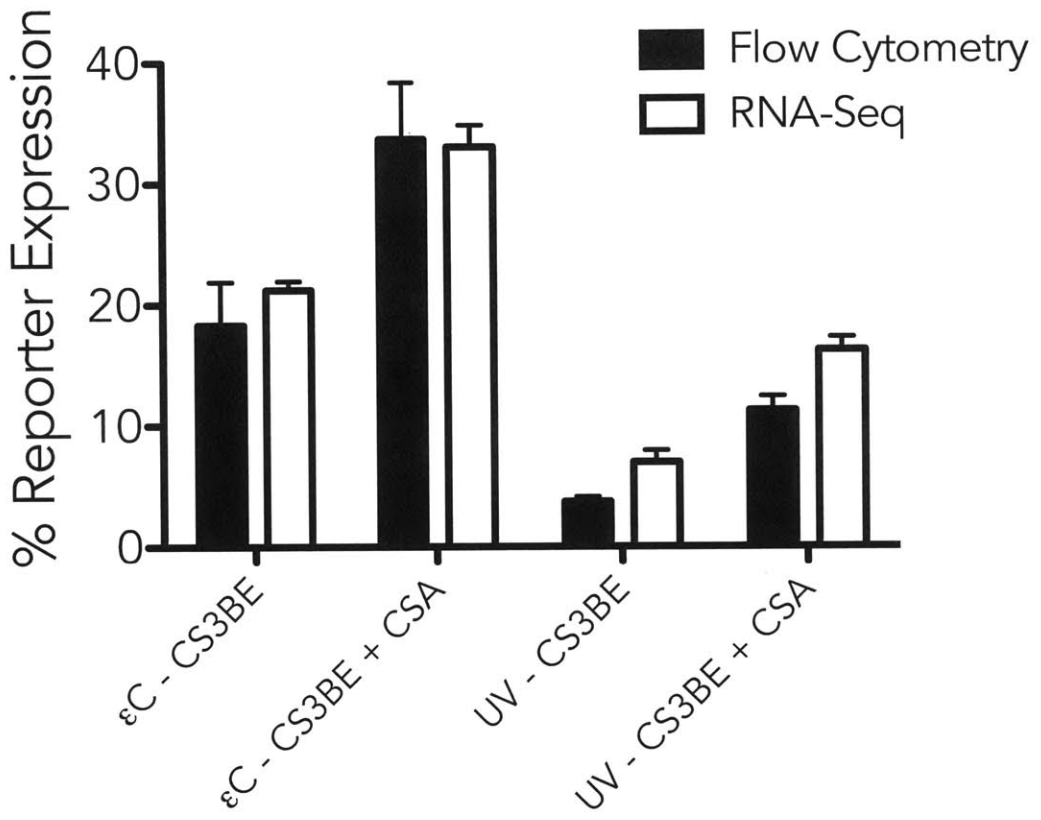


Figure 5.14 Comparison of FM-HCR results by flow cytometry and RNA-Seq. DNA Repair Capacity of patient derived fibroblasts deficient in CSA (CS3BE) and their CSA complemented counterpart (CS3BE + CSA). Cells were simulateously transfected with mOrange repair reporters irradiated with 800 J/m^2 UV-C and with GFP site-specific εC reporters. % R.E. for flow cytometry data was calculated as explained in materials and methods. % R.E. RNA-Seq data is calculated directly from transcript counts and normalized to efficiency normalization plasmid counts and the respective control transfection. Error bars show standard deviation for 3 biological replicates.

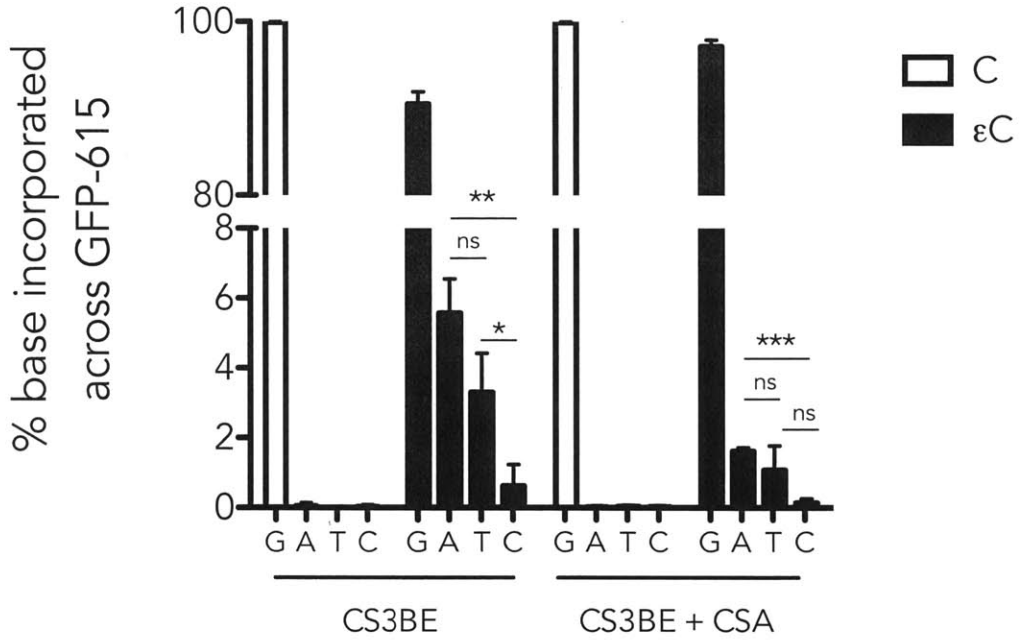


Figure 5.15 Transcriptional mutagenesis events by RNAPII across εC.
RNA-Seq sequence variants detected in transcripts at position 615 in a GFP reporter containing either a C or an εC on the transcribed strand. Frequencies are reported for the expected variant (G) as well as for the reminding 3 possible variants (C, G, T/U). Error bars show standard deviation for 3 biological replicates. *, $p \leq 0.05$; **, $p \leq 0.01$; ***, $p \leq 0.001$; ns, not significant.

References

1. El Ghissassi, F., A. Barbin, J. Nair, and H. Bartsch, *Formation of 1, N6-ethenoadenine and 3, N4-ethenocytosine by lipid peroxidation products and nucleic acid bases.* Chemical research in toxicology, 1995. **8**(2): p. 278-283.
2. Sodum, R.S. and F.L. Chung, *1, N2-ethenodeoxyguanosine as a potential marker for DNA adduct formation by trans-4-hydroxy-2-nonenal.* Cancer Research, 1988. **48**(2): p. 320.
3. Sodum, R.S. and F.L. Chung, *Stereoselective formation of in vitro nucleic acid adducts by 2, 3-epoxy-4-hydroxynonanal.* Cancer Research, 1991. **51**(1): p. 137-143.
4. Winczura, A., D. Zdzalik, and B. Tudek, *Damage of DNA and proteins by major lipid peroxidation products in genome stability.* Free Radic Res, 2012. **46**(4): p. 442-59.
5. Kochetkov, N., V. Shibaev, and A. Kost, *New reaction of adenine and cytosine derivatives, potentially useful for nucleic acids modification.* Tetrahedron Letters, 1971. **12**(22): p. 1993-1996.
6. El Ghissassi, F., A. Barbin, and H. Bartsch, *Metabolic activation of vinyl chloride by rat liver microsomes: low-dose kinetics and involvement of cytochrome P450 2E1.* Biochemical pharmacology, 1998. **55**(9): p. 1445-1452.
7. Block, J.B., *Angiosarcoma of the liver following vinyl chloride exposure.* Journal of the American Medical Association, 1974. **229**(1): p. 53.
8. Creech Jr, J. and M. Johnson, *Angiosarcoma of liver in the manufacture of polyvinyl chloride.* Journal of occupational medicine, 1974. **16**(3): p. 150.
9. Lee, F. and D. Harry, *Angiosarcoma of the liver in a vinyl-chloride worker.* The Lancet, 1974. **303**(7870): p. 1316-1318.
10. Maltoni, C. and G. Cotti, *Carcinogenicity of Vinyl Chloride in Sprague-Dawley Rats after Prenatal and Postnatal Exposure.* Annals of the New York Academy of Sciences, 1988. **534**(1): p. 145-159.
11. Zimmerli, B. and J. Schlatter, *Ethyl carbamate: analytical methodology, occurrence, formation, biological activity and risk assessment.* Mutation Research/Genetic Toxicology, 1991. **259**(3-4): p. 325-350.
12. Basu, A.K., M.L. Wood, L.J. Niedernhofer, L.A. Ramos, and J.M. Essigmann, *Mutagenic and genotoxic effects of three vinyl chloride-induced DNA lesions: 1, N6-ethenoadenine, 3, N4-ethenocytosine, and 4-amino-5-(imidazol-2-yl) imidazole.* Biochemistry, 1993. **32**(47): p. 12793-12801.
13. Pandya, G.A. and M. Moriya, *1,N6-Ethenodeoxyadenosine, a DNA adduct highly mutagenic in mammalian cells.* Biochemistry, 1996. **35**(35): p. 11487-11492.
14. Palejwala, V.A., R.W. Rzepka, D. Simha, and M.Z. Humayun, *Quantitative multiplex sequence analysis of mutational hot spots. Frequency and specificity of mutations induced by a site-specific ethenocytosine in M13 viral DNA.* Biochemistry, 1993. **32**(15): p. 4105-4111.
15. Cheng, K., B. Preston, D. Cahill, M. Dosanjh, B. Singer, and L. Loeb, *The vinyl chloride DNA derivative N2,3-ethenoguanine produces G-A transitions in*

Escherichia coli. Proceedings of the National Academy of Sciences of the United States of America, 1991. **88**(22): p. 9974.
16. Langouët, S., A.N. Mican, M. Müller, S.P. Fink, L.J. Marnett, S.A. Muhle, and F.P. Guengerich, *Misincorporation of Nucleotides opposite Five-Membered Exocyclic Ring Guanine Derivatives by Escherichia c oli Polymerases in Vitro and in Vivo: 1, N 2-Ethenoguanine, 5, 6, 7, 9-Tetrahydro-9-oxoimidazo [1, 2-a] purine, and 5, 6, 7, 9-Tetrahydro-7-hydroxy-9-oxoimidazo [1, 2-a] purine.* Biochemistry, 1998. **37**(15): p. 5184-5193.
17. Moriya, M., W. Zhang, F. Johnson, and A. Grollman, *Mutagenic potency of exocyclic DNA adducts: marked differences between Escherichia coli and simian kidney cells.* Proceedings of the National Academy of Sciences, 1994. **91**(25): p. 11899.
18. Akasaka, S. and F.P. Guengerich, *Mutagenicity of site-specifically located 1,N 2-ethenoguanine in Chinese hamster ovary cell chromosomal DNA.* Chemical research in toxicology, 1999. **12**(6): p. 501-507.
19. Barbin, A., *Role of etheno DNA adducts in carcinogenesis induced by vinyl chloride in rats.* IARC scientific publications, 1999(150): p. 303.
20. Bartsch, H., A. Barbin, M. Marion, J. Nair, and Y. Guichard, *Formation, detection, and role in carcinogenesis of ethenobases in DNA.* Drug metabolism reviews, 1994. **26**(1-2): p. 349.
21. Nair, U., H. Bartsch, and J. Nair, *Lipid peroxidation-induced DNA damage in cancer-prone inflammatory diseases: a review of published adduct types and levels in humans.* Free Radical Biology and Medicine, 2007. **43**(8): p. 1109-1120.
22. Fu, D., J.A. Calvo, and L.D. Samson, *Balancing repair and tolerance of DNA damage caused by alkylating agents.* Nature Reviews Cancer, 2012. **12**(2): p. 104-120.
23. Yi, C., G. Jia, G. Hou, Q. Dai, W. Zhang, G. Zheng, X. Jian, C.G. Yang, Q. Cui, and C. He, *Iron-catalysed oxidation intermediates captured in a DNA repair dioxygenase.* Nature, 2010. **468**(7321): p. 330-333.
24. Saparbaev, M., K. Kleibl, and J. Laval, *Escherichia coil, Saccharomyces cerevisiae, rat and human 3-methyladenine DNA glycosylases repair 1, N6-ethenoadenine when present in DNA.* Nucleic Acids Research, 1995. **23**(18): p. 3750-3755.
25. Lee, C.Y.I., J.C. Delaney, M. Kartalou, G.M. Lingaraju, A. Maor-Shoshani, J.M. Essigmann, and L.D. Samson, *Recognition and Processing of a New Repertoire of DNA Substrates by Human 3-Methyladenine DNA Glycosylase (AAG).* Biochemistry, 2009. **48**(9): p. 1850-1861.
26. Ringvoll, J., M.N. Moen, L.M. Nordstrand, L.B. Meira, B. Pang, A. Bekkelund, P.C. Dedon, S. Bjelland, L.D. Samson, P. Falnes, and A. Klungland, *AlkB Homologue 2-Mediated Repair of Ethenoadenine Lesions in Mammalian DNA.* Cancer Research, 2008. **68**(11): p. 4142.
27. Aas, P.A., M. Otterlei, P. Falnes, C.B. Vagbo, F. Skorpen, M. Akbari, O. Sundheim, M. Bjoras, G. Slupphaug, and E. Seeberg, *Human and bacterial oxidative demethylases repair alkylation damage in both RNA and DNA.* Nature, 2003. **421**(6925): p. 859-863.

28. Mishina, Y., C.G. Yang, and C. He, *Direct repair of the exocyclic DNA adduct 1, N6-ethenoadenine by the DNA repair AlkB proteins.* Journal of the American Chemical Society, 2005. **127**(42): p. 14594-14595.
29. Fu, D. and L.D. Samson, *Direct repair of 3,N4-ethenocytosine by the human ALKBH2 dioxygenase is blocked by the AAG/MPG glycosylase.* DNA repair, 2011.
30. Meira, L.B., C.A. Moroski-Erkul, S.L. Green, J.A. Calvo, R.T. Bronson, D. Shah, and L.D. Samson, *Aag-initiated base excision repair drives alkylation-induced retinal degeneration in mice.* Proceedings of the National Academy of Sciences, 2009. **106**(3): p. 888.
31. Kavli, B., O. Sundheim, M. Akbari, M. Otterlei, H. Nilsen, F. Skorpen, P.A. Aas, L. Hagen, H.E. Krokan, and G. Slupphaug, *hUNG2 is the major repair enzyme for removal of uracil from U: A matches, U: G mismatches, and U in single-stranded DNA, with hSMUG1 as a broad specificity backup.* Journal of Biological Chemistry, 2002. **277**(42): p. 39926-39936.
32. Saparbaev, M. and J. Laval, *3, N4-ethenocytosine, a highly mutagenic adduct, is a primary substrate for Escherichia coli double-stranded uracil-DNA glycosylase and human mismatch-specific thymine-DNA glycosylase.* Proceedings of the National Academy of Sciences, 1998. **95**(15): p. 8508.
33. Petronzelli, F., A. Riccio, G.D. Markham, S.H. Seeholzer, M. Genuardi, M. Karbowski, A.T. Yeung, Y. Matsumoto, and A. Bellacosa, *Investigation of the substrate spectrum of the human mismatch specific DNA glycosylase MED1 (MBD4): Fundamental role of the catalytic domain.* Journal of cellular physiology, 2000. **185**(3): p. 473-480.
34. Zdżalik, D., A. Domańska, P. Prorok, K. Kosicki, E. van den Born, P.Ø. Falnes, C.J. Rizzo, F.P. Guengerich, and B. Tudek, *Differential repair of etheno-DNA adducts by bacterial and human AlkB proteins.* DNA repair, 2015. **30**: p. 1-10.
35. Maddukuri, L., E. Speina, M. Christiansen, D. Dudzińska, J. Zaim, T. Obtułowicz, S. Kabaczyk, M. Komisarski, Z. Bukowy, and J. Szczegielniak, *Cockayne syndrome group B protein is engaged in processing of DNA adducts of lipid peroxidation product trans-4-hydroxy-2-nonenal.* Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis, 2009. **666**(1): p. 23-31.
36. Li, C., L.E. Wang, and Q. Wei, *DNA repair phenotype and cancer susceptibility--a mini review.* Int J Cancer, 2009. **124**(5): p. 999-1007.
37. Nagel, Z.D., C.M. Margulies, I.A. Chaim, S.K. McRee, P. Mazzucato, A. Ahmad, R.P. Abo, V.L. Butty, A.L. Forget, and L.D. Samson, *Multiplexed DNA repair assays for multiple lesions and multiple doses via transcription inhibition and transcriptional mutagenesis.* Proceedings of the National Academy of Sciences, 2014. **111**(18): p. E1823-E1832.
38. Calvo, J.A., L.B. Meira, C.-Y.I. Lee, C.A. Moroski-Erkul, N. Abolhassani, K. Taghizadeh, L.W. Eichinger, S. Muthupalani, L.M. Nordstrand, and A. Klungland, *DNA repair is indispensable for survival after acute inflammation.* The Journal of clinical investigation, 2012. **122**(7): p. 2680-2689.

39. Iyama, T. and D.M. Wilson III, *Elements That Regulate the DNA Damage Response of Proteins Defective in Cockayne Syndrome.* Journal of Molecular Biology, 2015.
40. Li, B. and C.N. Dewey, *RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome.* BMC Bioinformatics, 2011. **12**: p. 323.
41. Anders, S. and W. Huber, *Differential expression analysis for sequence count data.* Genome Biol, 2010. **11**(10): p. R106.
42. Li, H. and R. Durbin, *Fast and accurate short read alignment with Burrows-Wheeler transform.* Bioinformatics, 2009. **25**(14): p. 1754-60.
43. Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, and R. Durbin, *The Sequence Alignment/Map format and SAMtools.* Bioinformatics, 2009. **25**(16): p. 2078-9.
44. Koboldt, D.C., Q. Zhang, D.E. Larson, D. Shen, M.D. McLellan, L. Lin, C.A. Miller, E.R. Mardis, L. Ding, and R.K. Wilson, *VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing.* Genome Res, 2012. **22**(3): p. 568-76.
45. Engelward, B.P., G. Weeda, M.D. Wyatt, J.L.M. Broekhof, J. De Wit, I. Donker, J.M. Allan, B. Gold, J.H.J. Hoeijmakers, and L.D. Samson, *Base excision repair deficient mice lacking the Aag alkyladenine DNA glycosylase.* Proceedings of the National Academy of Sciences, 1997. **94**(24): p. 13087.
46. Meira, L.B., J.M. Bugni, S.L. Green, C.W. Lee, B. Pang, D. Borenshtein, B.H. Rickman, A.B. Rogers, C.A. Moroski-Erkul, and J.L. McFaline, *DNA damage induced by chronic inflammation contributes to colon carcinogenesis in mice.* The Journal of clinical investigation, 2008. **118**(7): p. 2516.
47. Khobta, A., N. Kitsera, B. Speckmann, and B. Epe, *8-Oxoguanine DNA glycosylase (Ogg1) causes a transcriptional inactivation of damaged DNA in the absence of functional Cockayne syndrome B (Csb) protein.* DNA repair, 2009. **8**(3): p. 309-317.
48. Kitsera, N., D. Stathis, B. Lühnsdorf, H. Müller, T. Carell, B. Epe, and A. Khobta, *8-Oxo-7, 8-dihydroguanine in DNA does not constitute a barrier to transcription, but is converted into transcription-blocking damage by OGG1.* Nucleic Acids Research, 2011.
49. Guo, J., P.C. Hanawalt, and G. Spivak, *Comet-FISH with strand-specific probes reveals transcription-coupled repair of 8-oxoGuanine in human cells.* Nucleic Acids Research, 2013.
50. Khobta, A., T. Lingg, I. Schulz, D. Warken, N. Kitsera, and B. Epe, *Mouse CSB protein is important for gene expression in the presence of a single-strand break in the non-transcribed DNA strand.* DNA repair, 2010. **9**(9): p. 985-993.
51. Kusmierek, J. and B. Singer, *Chloroacetaldehyde-treated ribo-and deoxyribopolynucleotides. 2. Errors in transcription by different polymerases resulting from ethenocytosine and its hydrated intermediate.* Biochemistry, 1982. **21**(22): p. 5723-5728.
52. Brégeon, D. and P.W. Doetsch, *Transcriptional mutagenesis: causes and involvement in tumour development.* Nature Reviews Cancer, 2011. **11**(3): p. 218-227.

53. Dimitri, A., A.K. Goodenough, F.P. Guengerich, S. Broyde, and D.A. Scicchitano, *Transcription processing at 1,N2-ethenoguanine by human RNA polymerase II and bacteriophage T7 RNA polymerase.* Journal of molecular biology, 2008. **375**(2): p. 353-366.
54. Nair, J., P.L. Carmichael, R.C. Fernando, D.H. Phillips, A.J. Strain, and H. Bartsch, *Lipid peroxidation-induced etheno-DNA adducts in the liver of patients with the genetic metal storage disorders Wilson's disease and primary hemochromatosis.* Cancer Epidemiology Biomarkers & Prevention, 1998. **7**(5): p. 435-440.
55. Nair, J., F. Gansauge, H. Beger, P. Dolara, G. Winde, and H. Bartsch, *Increased etheno-DNA adducts in affected tissues of patients suffering from Crohn,Äôs disease, ulcerative colitis, and chronic pancreatitis.* Antioxidants & redox signaling, 2006. **8**(5-6): p. 1003-1010.
56. Frank, A., H.K. Seitz, H. Bartsch, N. Frank, and J. Nair, *Immunohistochemical detection of 1, N6-ethenodeoxyadenosine in nuclei of human liver affected by diseases predisposing to hepato-carcinogenesis.* Carcinogenesis, 2004. **25**(6): p. 1027-1031.
57. Hanaoka, T., J. Nair, Y. Takahashi, S. Sasaki, H. Bartsch, and S. Tsugane, *Urinary level of 1, N6,ethenodeoxyadenosine, a marker of oxidative stress, is associated with salt excretion and polyunsaturated fatty acid intake in postmenopausal Japanese women.* International Journal of Cancer, 2002. **100**(1): p. 71-75.
58. Hanawalt, P.C. and G. Spivak, *Transcription-coupled DNA repair: two decades of progress and surprises.* Nature Reviews Molecular Cell Biology, 2008. **9**(12): p. 958-970.

Chapter VI: Inter- and Intra-individual variability in all major DNA Repair Pathways: A Pilot Study

Isaac A. Chaim, Zachary D. Nagel, Patrizia Mazzucato, Lizzie Ngo, Leona D. Samson

Contributions: I.A.C., Z.D.N and L.D.S. designed experiments. I.A.C., Z.D.N., P.M. and L.N. conducted all experiments. I.A.C. and Z.D.N analyzed data.

Table of Contents

Introduction

This introduction is an edited version of our review titled "Inter-individual variation in DNA repair capacity: A need for multi-pathway functional assays to promote translational DNA repair research" [1].

The integrity of our DNA is challenged on the order of 100,000 lesions per cell per day [2, 3]. Left unrepaired DNA damage has the potential to lead to mutant cells, dead cells and ensuing disease. The precise number and type of DNA lesions formed varies from one individual to the next in part because of differences in exposure and lifestyle, and also because of variation in metabolism and other cellular processes. Many types of DNA damage, such as abasic sites, alkylation damage, oxidative damage, mismatches, single and double strand breaks, result from normal metabolic processes. Others are induced upon exposure to environmental agents. Among the environmentally induced lesions are bulky DNA adducts, including heterocyclic amines induced by compounds in cooked foods, cyclobutane pyrimidine dimers induced by sunlight, alkylation damage from nitroso compounds in combustion products, and oxidative damage and DNA strand breaks induced by ionizing radiation [4-8].

Fortunately, human cells mount a robust response to DNA damage that includes at least 6 major DNA repair pathways that specialize in the repair of subsets of DNA lesions, namely direct reversal (DR), mismatch repair (MMR), nucleotide excision repair (NER), base excision repair (BER), homologous recombination (HR) and non-homologous end joining (NHEJ). The relationship between DNA damage and DNA repair is complex; no single pathway efficiently repairs all types of DNA lesions, some lesions are substrates for more than one pathway, and evidence for extensive interactions among proteins involved in distinct pathways continues to emerge [9-14]. Mutations in DNA repair genes can have profound consequences for disease risk. The classic example is that individuals with the disease Xeroderma Pigmentosum (XP) are highly prone to skin cancer

because they have mutations in genes required for NER, which repairs bulky lesions such as those induced by UV light. These individuals are at a 2000-fold higher risk of skin cancer in sun-exposed skin [15]. A variety of other diseases including neurological, developmental and immunological disorders, as well as premature aging, are associated with aberrant DNA repair in humans [16]. Thus, it is clear that defective DNA repair caused by mutations in repair genes represents a major disease risk factor, and genetic tests are now available for the most common disease-associated mutations in DNA repair genes [17].

Besides distinctive mutations in repair genes being associated with disease risk, increasing evidence suggests that small variations in DNA repair capacity (DRC) of apparently healthy individuals can be correlated with disease risk. Particularly, for those cases that fall at the extremes of a DRC distribution of apparently healthy individuals. It has been over two decades since these ideas were articulated by Hsu [18], and by Grossman and Wei [19], and although a wealth of evidence for inter-individual DRC differences has since emerged from multiple laboratories using various methods (reviewed in [1]), it seems fair to say that the intervening studies have not yet resulted in personalized prevention or treatment efforts. This being the case, there is a clear need in the field for assays that can reliably measure BER capacity and detect these small inter-individual differences, preferably in a fast and high-throughout manner that would allow for epidemiological studies.

A major strength of *in vivo* functional DNA repair assays is that, because they measure levels of functional protein, they integrate much of the biological complexity that might confound indirect measures of DRC; indeed a low correlation between enzymatic activity and mRNA levels has been documented in some cases [20].

The Host Cell Reactivation (HCR) assay offers a powerful way to measure DRC in living cells. The foundation of the assay lies in the ability of transcription

blocking DNA damage to impede expression of a transiently transfected reporter gene; repair restores transcription of the reporter gene, which may encode enzymes such as chloramphenicol acetyltransferase (CAT) and luciferase, or a fluorescent protein [21]. A major strength of HCR assays, stemming from the *in vitro* generation of damaged reporter plasmid DNA, is the ability to measure the *in vivo* repair of specific DNA lesions in intact cells. A fluorescence-based multiplexed flow cytometric HCR assay (FM-HCR) that uses different colored fluorescent reporter plasmids to measure repair of multiple doses or multiple types of DNA damage in a single assay was recently developed [22]. FM-HCR is less labor intensive than HCR assays that require cell lysate preparation, and uses DNA lesion-induced transcriptional mutagenesis to measure repair of specific DNA lesions, such as O^6-methylguanine, 8-oxoguanine, hypoxanthine, uracil and A:8-oxoguanine base-pairs that are bypassed by RNA polymerase and thus refractory to conventional HCR assays [22].

The majority of epidemiological studies that apply functional DRC assays have focused on a single DNA repair pathway, namely NER. However, data continue to emerge in support of the notion that DRC for more than one pathway will be required to gain maximal biological insight. With this in mind, the present pilot study sought to apply the recently developed FM-HCR assays on peripheral blood mononuclear cells (PBMCs) isolated from 56 apparently healthy individuals. With the goal of simultaneously measuring 6 different DNA repair pathways, including 10 different DNA lesions, *in vivo*, in primary cells. The results obtained in this pilot study have helped us better understand and measure inter- and intra-individual differences in DRC as well as to gain a better understanding of the experimental challenges that these kind of studies represent. To our knowledge, such an effort is unprecedented in the field.

Materials and Methods

Ethics statement

The MIT Committee on the Use of Humans as Experimental Subjects reviewed and approved the research involving human subjects. Written informed consent was obtained from all participants.

Study Subjects

They study population consisted of 56 healthy volunteers from the Massachusetts Institute of Technology, including students, employees and the surrounding communities. Participants completed demographics and life-style surveys. In order to study DNA Repair Capacity intra-individual variation (change in time), 10 subjects came back up to 4 times. Detailed demographics data is shown in **Table 6.1** and **Figure 6.1**. 40 mL peripheral blood samples were obtained from healthy individuals at the MIT Catalyst Clinical Research Center and collected in 4, 10 mL (10cc) sodium heparin vacutainers. A generalized scheme of the complete experimental protocol is shown in **Figure 6.2**.

Peripheral Blood Mononuclear Cells (PBMCs)

PBMCs were isolated using standard Ficoll-Paque (GE) density gradient. Briefly, each 10 cc of blood were mixed with 1 volume RPMI 1640 media in a 50 mL conical tube. 8 mL Ficoll-Paque were slowly injected onto the bottom of the tube without disturbing upper blood layer. Samples were centrifuged (with brakes off) at 400g for 40 min at 18˚C. Buffy coat was collected and transferred to new 50 mL conical tubes. At this point 2 (out of 4) samples were pooled together and RMPI 1640 media was added to complete to 50 mL and spun at 600g for 20 min at 18˚C. Supernatant was aspirated and 20 mL RPMI 1640 media added before

spinning at 400g for 15 min at 18˚C. Pellets equivalent to 4, 10 cc units of the same subject, were resuspended in 4 mL freezing media (40% RMPI 1640 + 50% Heat Inactivated FBS + 10% DMSO) and transferred in 1mL aliquots to 4 cryovials. Vials were transferred to Styrofoam boxes and left at -80˚C for 24h before being transferred to Liquid Nitrogen for extended storage.

PBMCs were thawed by transferring cryovials to a 37˚C water bath immediately after being removed from storage. Thawed cells were transferred to a 15 mL conical tube containing 9 mL thawing media (40% RMPI 1640 + 50% Heat Inactivated FBS + 10% Dextrose) and spun at 600g for 5 min at room temperature. Supernatant was aspirated and PBMCs resuspended in 10 mL T-Lymphocyte stimulation media (RPMI 1640 + 20 % Heat Inactivated FBS + 1% penicillin/streptomycin + 5µg/mL PHA-L (L4144, Sigma)). PBMCs were stimulated for 72 hours before being transfected with DNA repair reporters. 72 hours for stimulation was chosen based on previous optimization for cell density and diameter upon PHA stimulation (**Figure 6.3**).

Plasmids

As described previously [23], AmCyan, EGFP, mOrange, mPlum and tagBFP reporter genes subcloned into the pmaxCloning Vector (Lonza) between the KpnI and SacI restriction sites in the multiple cloning site were employed. Plasmids were amplified using *E.coli* DH5α (Invitrogen) and purified using Qiagen endotoxin-free Maxi and Giga kits.

Substrates containing site-specific DNA damage

Non-fluorescent variants of the different reporter plasmids containing a single mutation in a site coding for their respective chromophores were identified via QuikChange mutagenesis (**Table 6.2**). The only two exceptions are mPlum-T202WT, which also reports via transcriptional mutagenesis but has no point

mutations; and the substrate reporting via transcriptional blockage, which is engineered from GFP-WT.

In order to produce ssDNA previously described methods were followed with minor modifications [23]. Reporter plasmids were nicked with either Nb.BtsI or Nt.BtsI (New England Biolabs, depending on the lesion containing strand, see **Table 6.2**). The nicked strand then was digested with exonuclease III, and the remaining single-stranded circular DNA (ssDNA) was purified by using a 1% agarose gel.15 picomoles of the respective phosphorylated lesion-containing-oligonucleotide (**Table 6.2**) were combined with 3.2 µg of the corresponding single-stranded plasmid in 1X Pfu polymerase AD buffer (Agilent Biotechnologies) in a final volume of 46 µL (1/200, oligo/ssDNA molar ratio). The mixture was heated to 85 °C in a thermal cycler for 6 min, and then allowed to anneal by cooling to 40 °C at 1 °C per minute. To extend the primer, 5 units of Pfu polymerase AD (Agilent) and 0.4 µM dNTP were added. The incubation parameters used for each ssDNA-plasmids/oligonucleotide combination are shown in **Table 6.2**. The reaction is then cleaned up with a Qiagen PCR Purification kit column, eluted in EB buffer and subsequently combined with 1X Ligase Buffer (New England Biolabs - NEB), 0.4 µM dNTP, 1 µM ATP, 1X BSA, 1.5 units T4 DNA Polymerase and 80 units T4 DNA Ligase (both NEB) and incubated for an additional hour at 16 °C to yield closed circular plasmid. Finally, the product was purified from a 1% agarose gel using a Qiagen gel extraction kit.

Substrates containing a G:G mismatch

Substrates were prepared using a method based on a previously published protocol [24]. The pmax:mOrange plasmid was nicked with *Nb.BstI* (New England Biolabs) to generate a single strand break in the transcribed strand. The nicked strand was then digested with exonuclease III, and the remaining single stranded circular DNA (ssDNA) purified using a 1% agarose gel. 20 µg of the ssDNA was combined with 40 µg of G299C mutant pmax:mOrange plasmid

linearized with *NheI* (New England Biolabs); the mixture was denatured by addition of 0.3N sodium hydroxide, and then returned to neutral pH to facilitate annealing between wild type ssDNA and the complementary strand of the linearized mutant sequence to yield a heteroduplex containing a G:G mismatch at position 299 of the mOrange gene. Subsequently, reactions were cleaned up using a Qiagen PCR cleanup kit, and unwanted linear and single stranded DNA side products were digested with Plasmid Safe ATP dependent DNAse (Epicentre). Nicked plasmid was purified using a 1% agarose gel, and ligated using 800 units T4 DNA ligase (New England Biolabs).

UV-irradiated substrates

mOrange plasmids were irradiated in TE buffer (10 mM Tris·HCl, 1 mM EDTA, pH 7.0) at a DNA concentration of 50 ng/μL in a volume of 1.5 mL in 10-cm polystyrene Petri dishes (without lids) with 800 J/m^2 UV-C light generated by a Stratalinker 2000 box.

Substrates containing a blunt-end double strand break

A unique *ScaI* restriction site was inserted into the 5' untranslated region of the pmax:BFP reporter plasmid, immediately 5' of the reporter gene. Plasmids were linearized with *ScaI* restriction enzyme, purified by phenol/chloroform extraction and ethanol precipitation, and digest completeness was confirmed by gel electrophoresis. The uncut pmax:BFP reporter with the *ScaI* restriction site was used as the undamaged control in experiments measuring repair of the linearized reporter.

Substrates and methods for measuring homologous recombination

GFP-based HR reporter plasmids have been described previously [25], and were a generous gift from Prof. Bevin Engelward. The D5G plasmid expresses a non-

fluorescent GFP reporter that is truncated at the 5' end of the gene. This plasmid was modified to include a StuI restriction site so that a blunt-end double strand break could be introduced into the plasmid with the goal of reducing the likelihood of unwanted re-ligation of the plasmid, because this process does not yield fluorescent signal. Cells were transfected with 0.5 micrograms of StuI-linearized D5G, 5 micrograms of D3G, plus 0.5 µg of an undamaged plasmid that was used as a control for transfection efficiency. Homology directed repair of the DSB in the D5G reporter plasmid that uses the D3G plasmid as a donor sequence leads to a full length GFP-encoding gene, and results GFP expression.

DNA Repair Assays Transfections

2×10^6 cells stimulated T-Lymphocytes in 100µL complete medium (RPMI 1640 + 20% Heat Inactivated FBS + 1% penicillin/streptomycin) were combined with a reporter plasmid mixture (**Table 6.3** and **Figure 6.4**). Cells were electroporated using a 96-well Bio-Rad MXcell gene pulser, with an exponential waveform at 260 V and 950 µF. Following electroporation, 100µL complete medium were added to each well in the electroporation plate and gently mixed. The electroporation mix was transferred to a 24-well cell culture plate prefilled with 800 µL of complete medium and incubated at 37 °C and 5% CO_2. Following 24 hours, 750 µL of supernatant was gently removed and 50µL complete media containing 6X TO-PRO-3 were added to each well (for a total of 300µL per well) and transferred to 75 mm Falcon tubes with Cell Strainer Caps (Fisher Scientific).

Flow cytometry

Cells suspended in culture media were analyzed for fluorescence on a BD LSR II cytometer running FACSDiva software. Cell debris, doublets, and aggregates were excluded based on their side- and forward-scatter properties. TO-PRO–3 was added to cells 5–10 min before analysis and was used to exclude dead cells from the analysis. The following fluorophores and their corresponding detectors

(in parentheses) were used: tagBFP (Pacific Blue), AmCyan (AmCyan), EGFP (FITC), mOrange (phycoerythrin; PE), mPlum (PE-Cy5-5), and TO-PRO–3 (allophycocyanin; APC). Compensation was set by using single-color controls. Regions corresponding to cells positive for each of the five fluorescent proteins were established by using single-color dropout controls. For reporters that required compensation in more than one detector channel, fluorescence in the reporter channel was plotted separately against each of the channels requiring compensation. Using these plots, both single controls and the dropout control (in which the reporter of interest was excluded from the transfection) were used to establish regions corresponding to positive cells. A threshold of at least 30 fluorescent events for each reporter was established. Samples with fewer than 30 events were not considered and the particular experiment was repeated.

Calculation of Percent Fluorescent Reporter Expression.

Every experimental setup consisted of two sets of transfections: A control transfection (CT) and a sample transfection (ST) containing one or more reporters with DNA lesions. Both transfections included the same color combination with the same undamaged reporter to normalize each set for transfection efficiency.

Fluorescence Index (FI) for a given reporter within one transfection was calculated as follows:

$$FI = \frac{C_F \times MFI}{C_L}$$

where C_F is the number of positive fluorescent cells for that given fluorophore, *MFI* is the mean fluorescence intensity of the C_F, and C_L is the total number of live cells.

The normalized fluorescence index for a given reporter FI^O was calculated as follows:

$$FI^O = \frac{FI^n}{FI^E}$$

where FI^n corresponds to the *FI* of a reporter normalized to the *FI* of the transfection efficiency normalization plasmid, FI^E.

Normalized reporter expression from a sample transfection, $FI^O{}_{ST}$, and that from the same reporter plasmid in control transfection, $FI^O{}_{CT}$, were used to compute the percent reporter expression (%R.E.) as follows:

$$\%\, R.E. = \frac{FI^O{}_{ST}}{FI^O{}_{CT}} \times 100$$

Statistics

Statistical analyses were performed using GraphPad Prism software. Data are presented as mean ± SEM or mean ± SD (as stated in figure legends). Statistical significance was determined using unpaired t-test or one-way ANOVA. A p-value is considered significant if less than 0.05.

Results

Overview on DNA Repair Assays in PBMCs

Blood samples from all subjects were collected and frozen before beginning DNA Repair Capacity (DRC) measurements. Samples were analyzed in 37 different batches; triplicates for each subject were spread out along the duration of the analysis. Each batch included two cell controls: TK6 cell line and a PBMC control (henceforth called 000). 000 corresponds to PBMCs from 450 mL of peripheral

blood, purchased from a local provider, which were isolated and frozen in the same way as subjects' PBMCs.

In order to measure DRC for 10 different lesions for each subject, 5 different transfection cocktails were prepared (**Figure 6.4**), which translates to 5 different transfections per replicate per subject. This represented the best combination of plasmid colors (excitation and emission spectra) and lesion types that we had previously established to work well with each other. Two different undamaged cocktails were prepared, one for 4-color transfections (U#1-2) and another for 5-color transfections (U#3). The 4-color plasmid combination allowed us to take advantage of one undamaged transfection (U#1-2) for 2 damaged ones (D#1 and D#2).

Time-associated drift on DRC scores

By the end of flow cytometric analysis of all subjects, which extended for two months, we calculated "raw" DRC scores in the form of %R.E. for 10 DNA lesions that included: 3 replicates for each subject and 37 replicates for each control. When the 37 replicates for 000 for each reporter were plotted against time (days after beginning flow cytometric analysis), a batch effect became obvious as DRC drifted with time for several reporters (**Panel A, Figures 6.5-6.15**). Linear regressions to calculate the slope of the trend as well as a p-value for the slope being different than 0 are shown in **Table 6.4**. Six reporters showed a significant drift with time (Hx:T; 8oxoG:C; THF:C; MMR; NHEJ and HR). In order to determine if comparable drift occurred with the subject samples, we performed a similar linear regression analysis by plotting the single replicates of all 56 subjects (for a total of 168 data-points per reporter) against time (**Panel B, Figures 6.5-6.14**). In this case, 9 reporters showed significant drift in time, specifically: U:G; Hx:T; 8oxoG:C; THF:C; A:8oxoG; MMR; NHEJ; HR and NER (**Table 6.5**). An alternative way to visualize time-associated drifts is to group subjects by replicate number (**Panel D, Figures 6.5-6.14**). Given that

measurements for each replicate were spread out in time, replicate #1 was always measured before #2, and #2 always before #3. As such, if time is somehow associated with changes in DRC, the mean of each replicate number for all grouped subjects will change with replicate number. One-way ANOVA showed differences in %R.E. between replicates for 8 reporters: Hx:T; 8oxoG:C; THF:C; A:8oxoG; MMR; NHEJ; HR and NER (**Table 6.6**).

Time-associated drift correction

Given that significant drift was observed for both 000 controls and all subject samples, but that linear regression to subject samples showed significant drift for more reporters than 000 (compare **Tables 6.4** and **6.5**), we decided on a normalization strategy involving all subject samples linear regression, just not for 000. Each data-point was normalized by diving to a calculated %R.E. corresponding to the linear regression score of all subject data-points in time for the particular day the measurement was taken for each DNA repair reporter (an example of the normalization strategy for NER is shown in **Figure 6.15**).

By applying this normalization strategy, time-associated drifts were effectively suppressed for all pathways when linear regressions were considered (**Panel C, Figures 6.16-6.25**, all p-values not significant, data not shown). As for replicate associated drifts, normalization also suppressed the differences in mean %R.E. for all but two reporters (8oxoG:C and A:8oxoG, **Table 6.6** and **Panel E, Figures 6.16-6.25**).

DRC inter-individual variability of 10 DNA lesions in 56 healthy individuals

DRC measurements by lesion, organized by increasing repair activity are shown in **Panel A, Figures 6.16-6.25**. It should be kept in mind that for all reporters based on transcriptional mutagenesis (U:G; Hx:T, 8oxoG:C; A:8oxoG and O^6MeG/MGMT) increased %R.E. reflects lower repair activity. Three subjects

chosen on **Figure 6.26A** with low, medium and high uracil repair were colored (yellow, blue and red, respectively) and retained their identity for all the subsequent figures in order to give a sense of repair capacity variability for each subject for all different lesions. Average Coefficient of Variation (CV) of the CVs of all subjects for each reporter is shown in **Table 6.7**, while the fold-range obtained for each reporter is shown in **Table 6.8**.

Interestingly, the activities of several reporter pairs are significantly correlated with each other in this data set (**Table 6.9**). This was not the case in any of our previous studies with B-lymphoblastoid cells (with the only exception of a U:G repair activity with 8oxoG:C and A:8oxoG activity, data not shown). All DRC measurements were standardized by z-scoring, in order to allow direct visual and numerical comparison between reporters. **Figure 6.26** shows all z-scored data together, organized by increasing Hx:T repair; some positive and negative correlations can be visually appreciated (i.e. positive correlation between Hx:T and THF:C or negative correlation between Hx:T and NER).

MGMT measurements should be taken into special consideration. Throughout the screen explained above most of the subjects' PBMCs presented a very high repair activity for O^6MeG, which resulted in a reduced number of mPlum events (including no events for some subjects), and very low %R.E. as a consequence. The MGMT results in **Figure 6.24** were shown for consistency but the effect of such low number of fluorescent cells is reflected in the large CV (86%) obtained for this reporter. Given that we realized this issue after having run a single replicate from each subject and duplicates for some subjects, we decided to incorporate an extra transfection containing 4-times more MGMT reporter (2 µg) than that used for cocktail D#3, and to run duplicates for 8 subjects spanning different MGMT activities, based on the information gathered for the single run of those individuals. This substantial increase in reporter transfected did allow us to reliably measure MGMT activity on these 8 subjects, resulting in an observed 15-fold range (**Figure 6.27A**). The correlation between both plasmid quantities was

better than expected (R^2 = 0.70). Nevertheless, it still needs to be determined if the use of 2µg of plasmid is feasible for a greater scale population study. As it will be discussed later, production of such high quantities of plasmid reporter do require substantial time.

Correlations between DRC and demographics

A question that has been a subject of debate in the field is whether DRC changes in healthy individuals with age and whether there are any differences between genders. Here we observe a lower U:G and 8oxoG:C repair in females than in males (**Table 6.10**). Moreover, when subjects where subdivided by age (<40 or ≥40), 8oxoG:C repair decreased with age whereas NHEJ increased (**Table 6.11**). It should be taken into consideration that the "older" population is made up by almost 70% female subjects (8 out of 12). As such, given the lower 8oxoG:C repair observed for females, it is possible that these age-related observation is being skewed by the gender interaction. When subjects where subdivided by a different age range (<30 or ≥30) no differences where observed for any pathway (**Table 6.11**).

DRC intra-individual variability of 10 DNA lesions in 10 healthy individuals

It is of utmost importance to understand if and/or how stable DRC is with time for a given individual. If DRC for a given lesion is highly variable, then direct comparison between individuals on a given day might not reflect the average (or the extremes) of those individuals' DRC. On the contrary, if DRC for a given lesion is stable in time then assays that seek to determine DRC should aim to (and be able to) reliably measure this stability. We measured DRC with our 10 reporters for 10 subjects that came back for blood draws up to 4 times (for a total of 5 time-points). Results are shown in **Panel B, Figures 6.16-6.25**. In general it appears that measurements for a given individual vary with time. Nevertheless, there are some particular cases where, either DRC appears stable in time or

where, even given some variability, a particular subject stays in the top/bottom group of DRCs. Some good examples are: **(a)** for U:G repair, subject J is consistently in the top group whereas subject D is at the bottom **(b)** for Hx:T repair, subject D is consistently the highest %R.E. subject whereas subject G is consistently in the bottom 2 **(c)** for A:8oxoG repair, subjects G and I present clearly and consistently the highest %R.E. **(d)** for NER, subjects C and H are consistently in the top group whereas subject E is consistently at the bottom.

Discussion

To our knowledge this pilot study is the first of its kind in the field as no more than one DNA repair pathway or DNA lesion have been attempted to be measured simultaneously, *in vivo*, in primary human cells. The closest approximation to this effort comes from the Livneh group where the activity of AAG, OGG1 and APE1 where measured in parallel (not simultaneously) *in vitro* [26].

As it could be expected, such a novel approach has been accompanied with a variety of obstacles. The first one was manpower; almost every step of the experimental protocol required 2 (and sometimes even 3) researchers to work in parallel in order to multiplex all the steps. These included, PBMC isolation and freezing of up to 7 donors per day; simultaneous transfection of approximately 50 conditions (5 DNA cocktails for 8 study subjects plus two controls); and flow cytometric analysis for which one researcher was running a third of the samples while another one was preparing the next third to be run that day and so on. Second, sources of variation had to be thought of in advance and attempted to be minimized as much as possible. Large plasmid repair cocktails were prepared for up to 500 transfections. Preparation of these amounts of plasmid reporters took months of work by different researchers. 110 mg of lyophilized PHA from the same lot were purchased, resuspended, aliquoted and frozen in proper volumes to avoid freeze-thaw cycles. Enough Heat-Inactivated FBS from the same lot was aliquoted and frozen to be used exclusively for this experiment. Moreover, as it

would be expected for such a relatively large-scale experiment, each step resulted in a very large dataset for which proper subject labeling and database management was invaluable. Finally, the time-associated drift in DRC measurements was likely the largest obstacle encountered.

Days from the beginning of flow cytometric analysis was chosen as the "x-axis" for a linear-regression normalization. But, there are many variables that could be contributing to the observed drift. Some examples that we considered are: time PBMC were frozen, PBMC counts in whole blood, PBMC counts following isolation, time that PHA was stored frozen, time since the plasmid cocktails were prepared, cell density/size/viability following stimulation, viability at time of flow cytometry and transfection efficiency. Some other variables that could play a role but are less straightforward and more complicated to quantify are: researcher that performed the transfection, potential small but incremental drifts of lasers in the flow cytometer or even rack in the incubator where the samples where placed before and after transfection, just to mention two examples. Any of these variables (and more) could be contributing to the observed drift. Moreover a combination of two or more of them could be contributing in non-linear ways as well. For now, and for this pilot study we decided to normalize to the linear-regression of each DRC measurement in "real-time", since the start of the first flow cytometric analysis. We considered and tested other simpler normalization approaches, such as normalizing to the 000 control repair score for each day but, the approach chosen was the one that showed the lowest average CV of all (data not shown). This does not mean that it is the best normalization method, but it is the best one we have tested so far.

It is difficult to determine if the large CVs observed for some repair measurements arise **(a)** as a consequence of flaws/undetermined variations in the methodology or experimental execution by the researchers or alternatively **(b)** as a product of non-ideal/incorrect normalization. For consistency we decided to normalize all reporters the same way, but it is possible that each reporter

should be considered as a single unit and normalized using different strategies. For example, in the cases for which the R^2 for the linear regression was significant but small (i.e. for U:G and HR), normalization by this methods could be introducing more noise to the data.

Another consideration for some of the large CVs obtained (from highest to lowest, MGMT, HR, A:8oxoG and U:G) is that all these assays report on rare events. As such, some subjects with particularly high repair activity (in the case of MGMT, A:8oxoG and U:G) or low repair activity (in the case of HR) could result in large inter-replicate variation arising from statistics of small numbers. A strategy we have considered, but not yet executed, to asses this potential issue is to check one by one the number of events obtained for each replicate of these reporters and look for inconsistencies in the cases where flurorescent events were particularly low.

An ideal case scenario that could help minimize variability would be to include cells with low, medium and large repair capacity together with each transfection. This would allow to normalize each analyzed batch by interpolation to a repair score that will be composed of cells with known relative DRCs. The challenge with this approach is that each lesion/pathway might require different cells. This would increase the number of transfections and conditions required for each experiment to even higher multiplexing and higher throughput approaches than what we currently have available.

Regardless of all the potential caveats discussed, the inter-individual variability fold ranges among the 56 study subjects obtained here for the different DRC reporters are actually very similar to the ones we had previously reported for a panel of 24-B-Lymphoblastoids derived from apparently healthy individuals [23] and **Chapter IV** (compare Hx:T, 2.4-fold; 8oxoG:C, 2.1-fold; A:8xoG, 9.0-fold; U:G, 3.7-fold; THF:C, 1.8-fold; MGMT, 100-fold; MMR, 2.3-fold; HR, 6.3-fold; NER, 3.6-fold, and NHEJ, 1.7-fold to our previous measurements Hx:T, 1.7-fold;

8oxoG:C, 4.3-fold; A:8xoG, 7.2-fold; U:G, 5.3-fold; THF:C, 1.9-fold; MGMT, 285-fold; MMR, 4.4-fold; HR, 3.7-fold; NER, 3.2-fold, and NHEJ, 2.1-fold).

Indeed, a direct comparison between the DRC results reported here and any previously published data is complicated given that inter-individual differences in the human population have not been measured, to our knowledge, for many of the pathways/lesions. Moreover, only variations in NER capacity have been evaluated using an *in vivo* HCR assay analogous to ours; whereas most other pathways/lesions have been measured either *in vitro* by using cell protein extracts or *in vivo* with the comet assay.

Nevertheless, our numbers agree with (or are close to) previous reporter ranges for: Hx:T repair from PBMC extracts (3.5-fold [27] compared to our 2.4-fold); 8oxoG:C repair from PBMC extracts (2- to 2.9-fold [20, 28, 29], compared to our 2.1-fold) or 8oxoG repair measured by comet assay (4-fold [30]); U:G repair from extracts from colon, stomach and liver (5.5, 3.2 and 3.1-fold, respectively [31], compared to our 3.7-fold); THF:C repair from PBMC extracts (2.5- to 4.9-fold [26, 32], compared to our 1.8-fold) and MGMT activity from PBMC extracts (approximately 10-fold [33], compared to our 15-fold for increased MGMT reporter transfection). Finally, in comparison to our 3.6-fold range observed for NER, early HCR assays using UV-irradiated CAT reporter plasmids showed approximately a 10-fold range of inter-individual differences in NER capacity [34] while more recent adaptations of the method have also shown a 5.6 to 11-fold range of inter-individual variation [35, 36]. Moreover, a 10-fold variation has also been reported when using the comet assay to measure NER of UV-induced damage [30]. To our knowledge, extensive studies on the functional variation of repair of A:8oxoG as well MMR, NHEJ and HR in healthy human populations has not been reported so far. Thus making this pilot study the first to do so for these pathways.

There have been relatively few longitudinal studies of DRC. In one study, BER and NER capacity were measured using comet assays for 35 subjects at 6 time-points over a period of 5 months. The correlation coefficient (R) for the different monthly comparisons varied from 0.12 to 0.62 for BER and from 0.18 to 0.64 for NER. The mean CV reported was 27% for BER and 49% for NER, suggesting that there may be significant intra-individual variability with time [30]. A different study, found that 8oxoG:C activity from PBMC extracts was stable when measured in cancer patients up to 4 months before and over 1 year after treatment, surgery and remission [29]. Given the relatively little evidence on DRC stability in time (or lack thereof), it is challenging to determine if our intra-individual observations accurately represent DRC variability or if they are are a product of technical and experimental variation. The possibility that each lesion or repair pathway have different activities at different times for a given individual is also a very likely scenario that will only be tested with more studies such as this one.

Considering the relatively high CVs observed for some of the reporters, it seems unlikely, at least for the present results, to be able to reliably tell apart individuals whose repair capacities are not spread apart enough. Alternatively, focusing on the extremes by subdividing repair scores in tertiles or quartiles can be used to establish more reliable comparisons between high and low repair capacity.

The present pilot study is the first of its kind, as no more than one DNA repair pathway had been measured simultaneously, *in vivo*. This is in part because the functional assays utilized here were recently developed [23] (and **Chapter IV**). As such, this pioneering task has been accompanied by several challenges that we have tried to address in detail. Future studies similar to this one should aim to reduce sources of variation, likely by the introduction of more controls. Lower CVs for all the assays will help determine the stability of DRC for a given lesion/pathway in time. A better understanding of variation in repair capacity,

accompanied by precise measurements of DRC capacity have the real potential of aiding in personalizing disease prevention and treatment.

Tables

Table 6.1 Demographics of the study population.

		N	%
Gender			
	Females	31	55.4
	Males	25	44.6
	Total	56	100.0
Age			
	20-24	8	
	25-29	27	48.2
	30-34	6	10.7
	35-39	1	1.8
	40-44	2	3.6
	45-49	3	5.4
	50-54	3	5.4
	55-59	3	5.4
	60-64	0	0.0
	65-69	1	1.8
	Undisclosed	2	3.6
	Total	56	100.0
Race			
	American Indian or Alaskan Native	1	1.8
	Asian	12	21.4
	Black or African American	2	3.6
	White	41	73.2
		56	100.0
Place of Birth			
	US	36	64.3
	Outside US	20	35.7
		56	100.0

Table 6.2 Site-specific plasmids and extension conditions used for their production.

Plasmid name	Nicking Enzyme	Sequence	X =		Extension Temp.	Extension Time (h)
BFP-A191G-U	Nb.BtsI	5' GGT CTT GCT GCC G**X**A GAG GAA GCT AGT AGC	U	IDT	64.0 °C	3.5
GFP-C289T-dHx	Nb.BtsI	5' GAA GAA GAT GGT GC**X** CTC CTG GAC GTA GCC	I	EGT	61.0 °C	2.0
GFP-617-THF	Nb.BtsI	5' GCT CAG GGC G**X**A CTG GGT GCT CAG GTA GTG	THF	IDT	66.0 °C	1.0
mOrange-A215C-8oxoG	Nb.BtsI	5' GTA GGC CTT GGA GCC G**X**A GGT GAA CTG AGG	8oxoG	EGT	64.0 °C	2.5
mOrange-A215C-THF	Nb.BtsI	5' GCC TTG GAG CCG **X**AG GTG AAC TGA GGG GAC	THF	IDT	66.0 °C	1.5
mPlum-T202C-C207G-O^{6}MeG	Nb.BtsI	5' CAC GTA GGC CTT GG**X** CCC GTA CAT GAT CTG	O^{6}MeG	EGT	68.0 °C	1.0
mPlum-T202WT-8oxoG (+)	Nt.BtsI	5' GTC CCC TCA GAT CAT G**X**A CGG CTC CAA GGC	8oxoG	EGT	61.0 °C	2.0

Nb./Nt.BstI (NEB); U (Uracil); I(Inosine); THF (Tetrahydrofuran); EGT (Eurogentech); IDT (Integrated DNA Techonologies)

Table 6.3 Combinations of reporter plasmids and types of DNA damage used in each experiment.

Undamaged Cocktail 1-2 **(U#1-2)**

	Plasmid	ng per assay
Control	BFP	100
Control	GFP	100
Control	mOrange	100
Control	mPlum	250
Carrier	pMAX empty	8050

Undamaged Cocktail 3 **(U#3)**

	Plasmid	ng per assay
Control	AmCyan	2000
Control	ScaI_BFP	100
Control	pCX-NNX-GFP	500
	D3 GFP	5000
Control	mOrange	100
Control	C207G	500
Carrier	pMAX empty	400

Damaged Cocktail 1 **(D#1)**

	Plasmid	ng per assay
U:G	BFP A191G-U	300
Hx:T	GFP C289T-Hx	100
8oxoG:C	mOrange A215C-8oxoG	300
Control	mPlum	250
Carrier	pMAX empty	7650

Damaged Cocktail 3 **(D#3)**

	Plasmid	ng per assay
Control	AmCyan	2000
NHEJ	ScaI linearized	300
HR	D5G StuI Linear	500
	D3 GFP	5000
NER	UV 800 J/m^2	300
MGMT	mPlum T202C-C207G-O^{6}MeG	500
Carrier	pMAX empty	0

Damaged Cocktail 2 **(D#2)**

	Plasmid	ng per assay
Control	BPF	100
THF:C	GFP 617-THF	100
G:G	mOrange G299C-G	100
A:8oxoG	mPlum T202WT-8oxoG(+)	500
Carrier	pMAX empty	7800

Damaged Cocktail MGMT **(MGMT)**

	Plasmid	ng per assay
Control	BFP	100
Control	GFP	100
Control	mOrange	100
MGMT	mPlum T202C-C207G-O^{6}MeG	2000
Carrier	pMAX empty	6300

Table 6.4 Linear regression analyses for 000 control samples between each repair reporter and time.

Time corresponds to days after beginning flow cytometric analysis. **Bold** p-values are statistically significant for the respective slope being different than zero.

	Slope	R^2	p-value
U:G	0.002524	0.0602	0.1433
Hx:T	-0.1572	0.3032	**0.0004**
8oxoG:C	0.004329	0.2514	**0.0016**
THF:C	-0.6537	0.2771	**0.0008**
A:8oxoG	0.0023	0.0007794	0.8697
MMR	0.04933	0.199	**0.0057**
NHEJ	-0.1138	0.1337	**0.026**
HR	-0.03913	0.2802	**0.0008**
MGMT	-8.73E-05	0.005442	0.6643
NER	0.07851	0.02175	0.3838

Table 6.5 Linear regression analyses for all subject replicates between each repair reporter and time.

Time corresponds to days after beginning flow cytometric analysis. **Bold** p-values are statistically significant for the respective slope being different than zero.

	Slope	R^2	p-value
U:G	0.002154	0.028	**0.0044**
Hx:T	-0.2275	0.3194	**< 0.0001**
8oxoG:C	0.002728	0.1067	**< 0.0001**
THF:C	-0.5654	0.1642	**< 0.0001**
A:8oxoG	0.009873	0.08344	**< 0.0001**
MMR	0.07691	0.2081	**< 0.0001**
NHEJ	-0.1369	0.2597	**< 0.0001**
HR	-0.01779	0.02349	**0.0091**
MGMT	6.46E-05	0.006956	0.1573
NER	0.4356	0.3417	**< 0.0001**

Table 6.6 ANOVA for comparison between % Reporter Expression of all subject grouped by replicated number.

Left columns shows p-values for each raw repair reporter measurements whereas right columns shows the same relationships following normalization. **Bold** p-values correspond to statistically significant differences between replicates.

	p-value raw	**p-value normalized**
U:G	0.5066	0.5036
Hx:T	**< 0.0001**	0.469
8oxoG:C	**0.0019**	**0.0226**
THF:C	**< 0.0001**	0.4143
A:8oxoG	**0.0004**	**0.0304**
MMR	**< 0.0001**	0.9323
NHEJ	**< 0.0001**	0.2776
HR	**0.007**	0.0663
MGMT	0.4414	0.0851
NER	**< 0.0001**	0.6626

Table 6.7 Average Coefficient of Variation (CV) of the CVs of all subjects for each reporter.

	Average CV (%)
U:G	31.52
Hx:T	15.96
8oxoG:C	15.74
THF:G	13.84
A:8oxoG	32.46
MMR	19.00
NHEJ	12.25
HR	44.97
MGMT	86.89
NER	23.34

Table 6.8 Fold-range for each reporter measured.

	Fold-range
U:G	3.71
Hx:T	2.40
8oxoG:C	2.07
THF:G	1.82
A:8oxoG	9.02
MMR	2.27
NHEJ	1.65
HR	6.13
MGMT	100.62
NER	3.59

Table 6.9 Pairwise linear relationships between DRC measurements.
Top shows correlation coefficient (R) values, bottom shows p-values for each relationship. In **bold** are values for which the relationship is significant.

R	U:G								
Hx:T	**0.395**	**Hx:T**							
8oxoG:C	0.205	-0.214	**8oxoG:C**						
THF:C	0.184	**0.602**	-0.091	**THF:C**					
A:8oxoG	-0.080	-0.188	**0.276**	**-0.298**	**A:8oxoG**				
MMR	-0.127	**-0.329**	**0.375**	0.023	0.095	**MMR**			
NHEJ	0.172	0.193	-0.202	0.180	-0.089	**-0.571**	**NHEJ**		
HR	-0.186	-0.235	0.107	-0.158	**0.455**	**0.324**	-0.250	**HR**	
MGMT	0.137	0.112	-0.001	0.113	0.049	-0.122	**0.375**	-0.135	**MGMT**
NER	**-0.282**	**-0.413**	0.095	-0.120	-0.079	**0.527**	**-0.439**	0.154	-0.179

p-value	U:G								
Hx:T	**0.002**	**Hx:T**							
8oxoG:C	0.127	0.111	**8oxoG:C**						
THF:C	0.171	**0.000**	0.499	**THF:C**					
A:8oxoG	0.555	0.161	**0.038**	**0.025**	**A:8oxoG**				
MMR	0.345	**0.013**	**0.004**	0.867	0.482	**MMR**			
NHEJ	0.202	0.151	0.132	0.180	0.513	**0.000**	**NHEJ**		
HR	0.165	0.078	0.429	0.239	**0.000**	**0.014**	0.060	**HR**	
MGMT	0.310	0.407	0.997	0.402	0.716	0.365	**0.004**	0.315	**MGMT**
NER	**0.033**	**0.001**	0.484	0.375	0.558	**0.000**	**0.001**	0.252	0.184

Table 6.10 Gender-associated variations in DRC.
Significant p-values are shown in **bold**.

	Female (n=31) Mean	Male (n=25) Mean	Student t-test
U:G	113.049	96.688	**0.042**
Hx:T	89.928	92.331	0.584
8oxoG:C	103.300	96.109	**0.031**
THF:C	98.973	95.544	0.326
A:8oxoG	103.462	108.461	0.720
MMR	107.039	97.799	0.074
NHEJ	98.205	99.541	0.667
HR	104.761	105.365	0.959
MGMT	132.942	86.464	0.217
NER	106.297	108.465	0.802

Table 6.11 Age-associated variations in DRC.
Significant p-values are shown in **bold**.

	< 40 (n=42) Mean	> 40 (n=12) Mean	Student t-test
U:G	104.751	111.525	0.495
Hx:T	92.706	84.430	0.116
8oxoG:C	98.445	107.295	**0.029**
THF:C	98.182	94.105	0.330
A:8oxoG	97.803	130.824	0.051
MMR	102.447	105.658	0.618
NHEJ	96.975	104.009	**0.049**
HR	101.807	119.281	0.222
MGMT	92.134	116.376	0.433
NER	106.470	113.299	0.519

	< 30 (n=35) Mean	> 30 (n=19) Mean	Student t-test
U:G	102.544	113.094	0.220
Hx:T	92.963	87.005	0.195
8oxoG:C	98.356	104.200	0.100
THF:C	98.929	94.231	0.196
A:8oxoG	94.817	124.159	0.046
MMR	103.104	103.265	0.977
NHEJ	97.581	100.302	0.388
HR	104.848	107.241	0.849
MGMT	86.831	117.214	0.258
NER	103.805	115.692	0.195

Figures

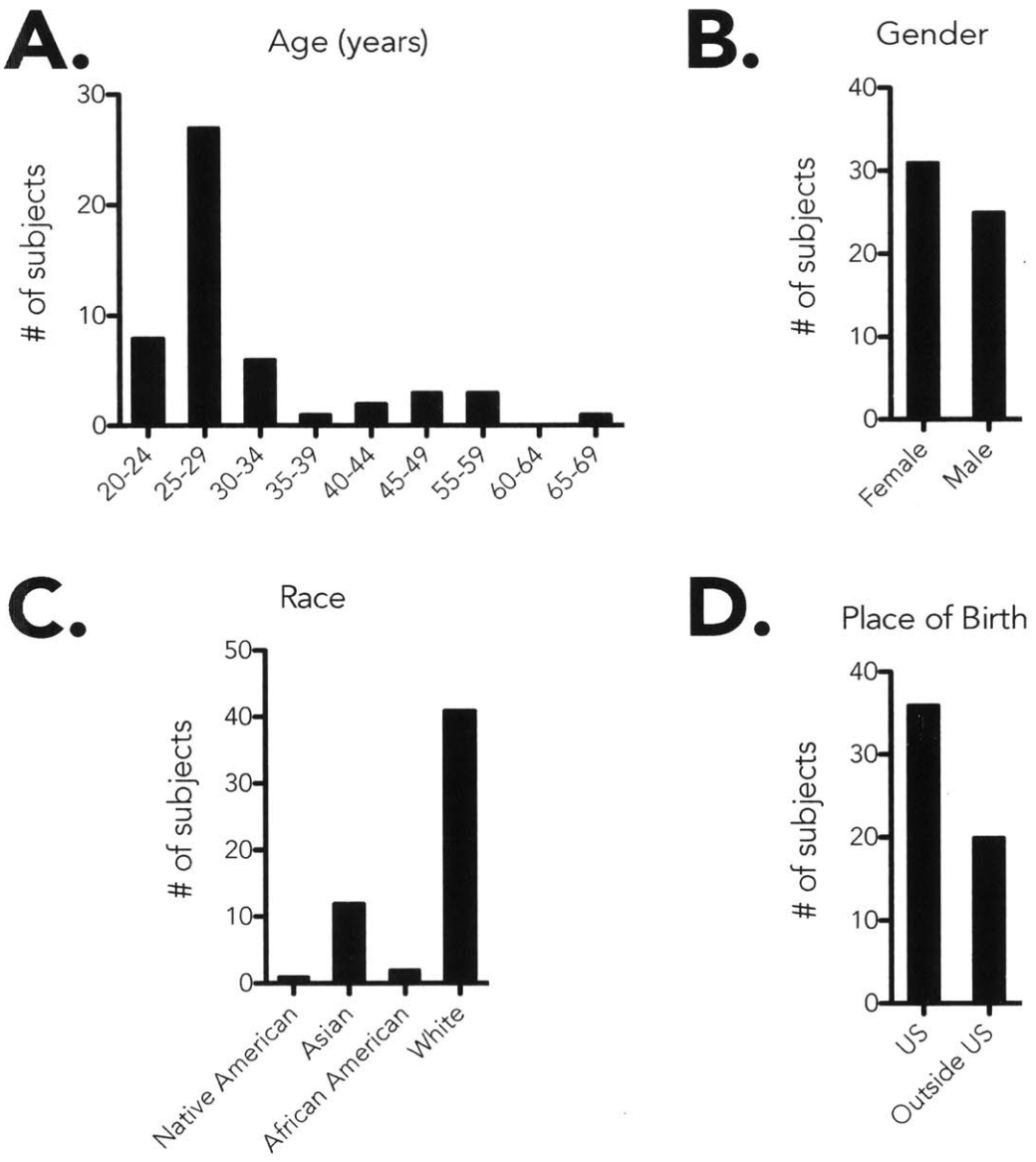


Figure 6.1 Demographics of the study population.

A. Histogram by age groups at time of blood drawing. **B**. Gender. **C.** Race and **D.** Place of birth.

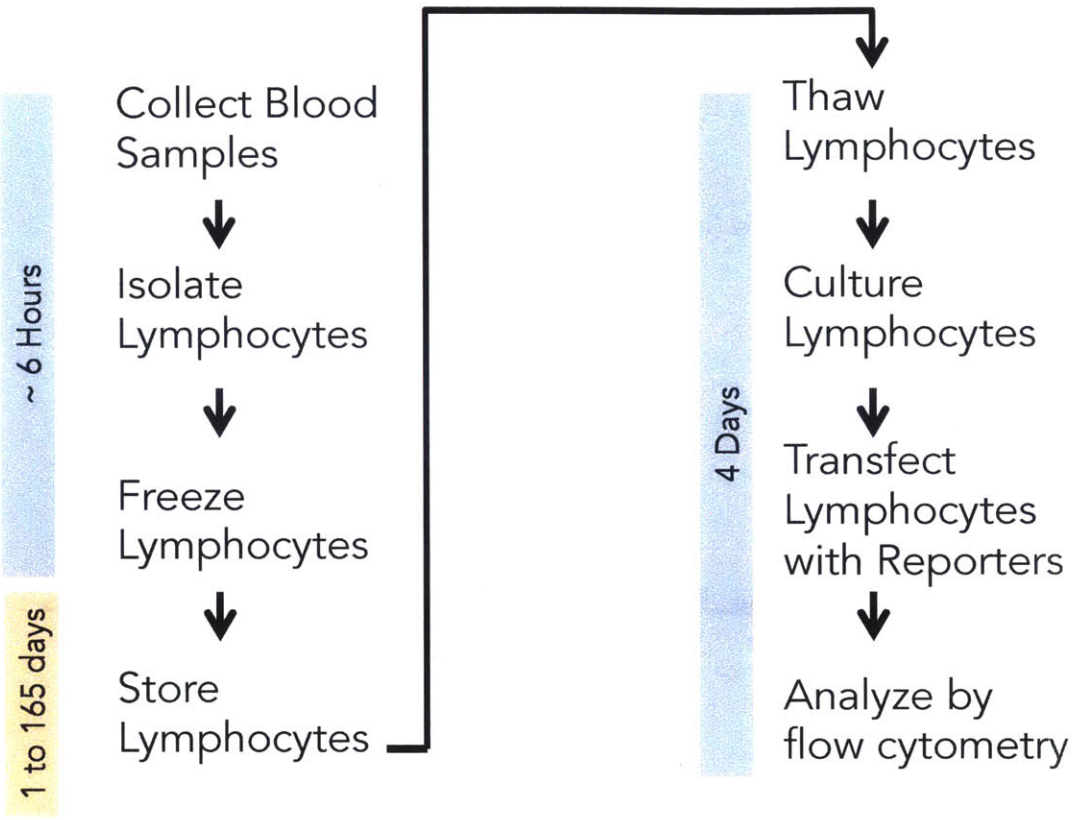


Figure 6.2 Generalized scheme of the complete experimental protocol
Each step is explained in detail in the Materials and Methods section.

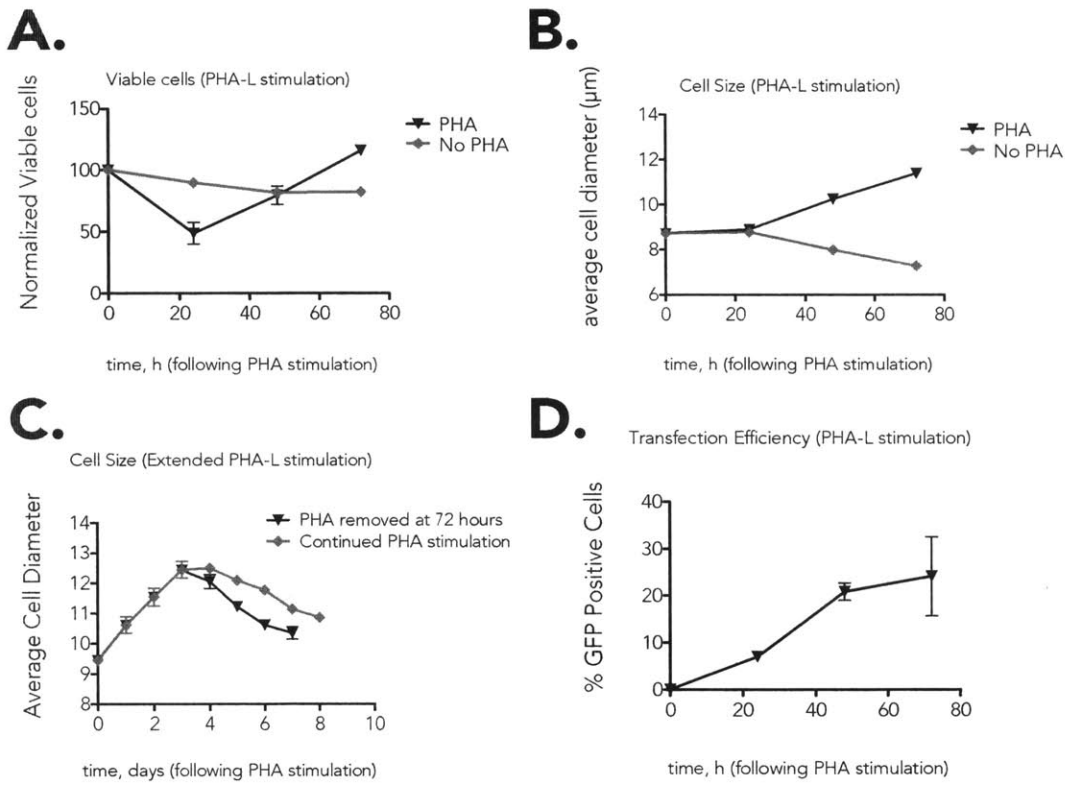


Figure 6.3 T-lymphocyte PHA (5µg/mL) stimulation optimization

Comparison between no stimulation and PHA stimulation for **A.** viability or **B.** cell size. PHA as a mitogen is expected to make cell replicate and grow in size. Given that stimulated cells will have to be incubated for an extra 24 hours following transfection, **C.** Cell size as a proxy for stimulation was compared between leaving cells in PHA for 72 hours or for an extended period. **D.** Effect of PHA stimulation on transfection efficiency. Altogether, 72 hours stimulation was chosen as an ideal time-point for transfection.

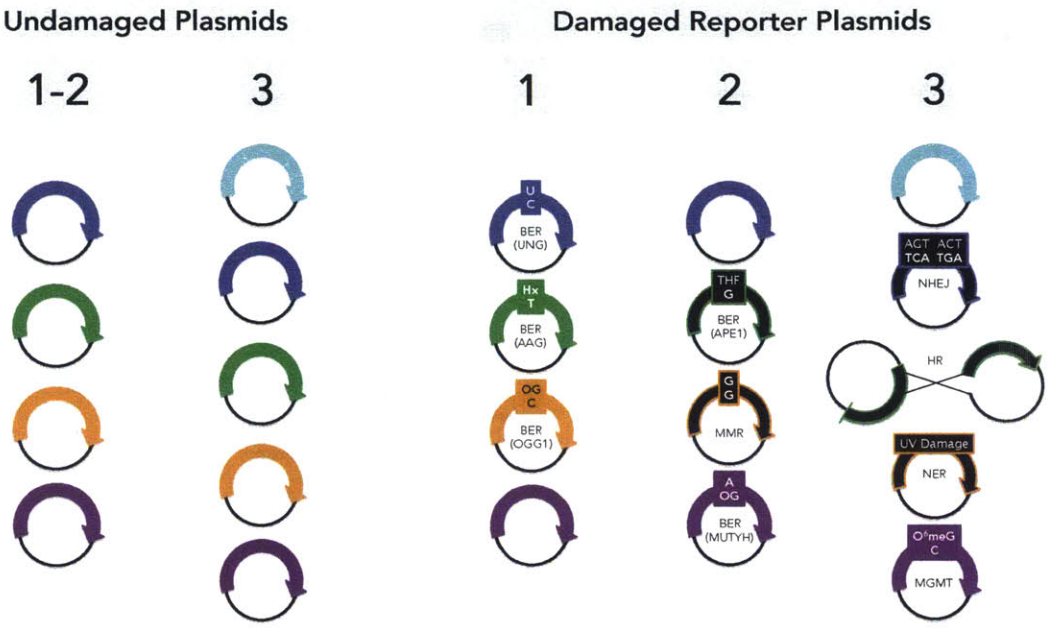


Figure 6.4. Plasmid Cocktails for DNA Repair Capacity Assays.

5 different plasmid reporter cocktails were necessary in order to measure repair of 10 different lesions. Quantity details are shown in **Table 6.3**.

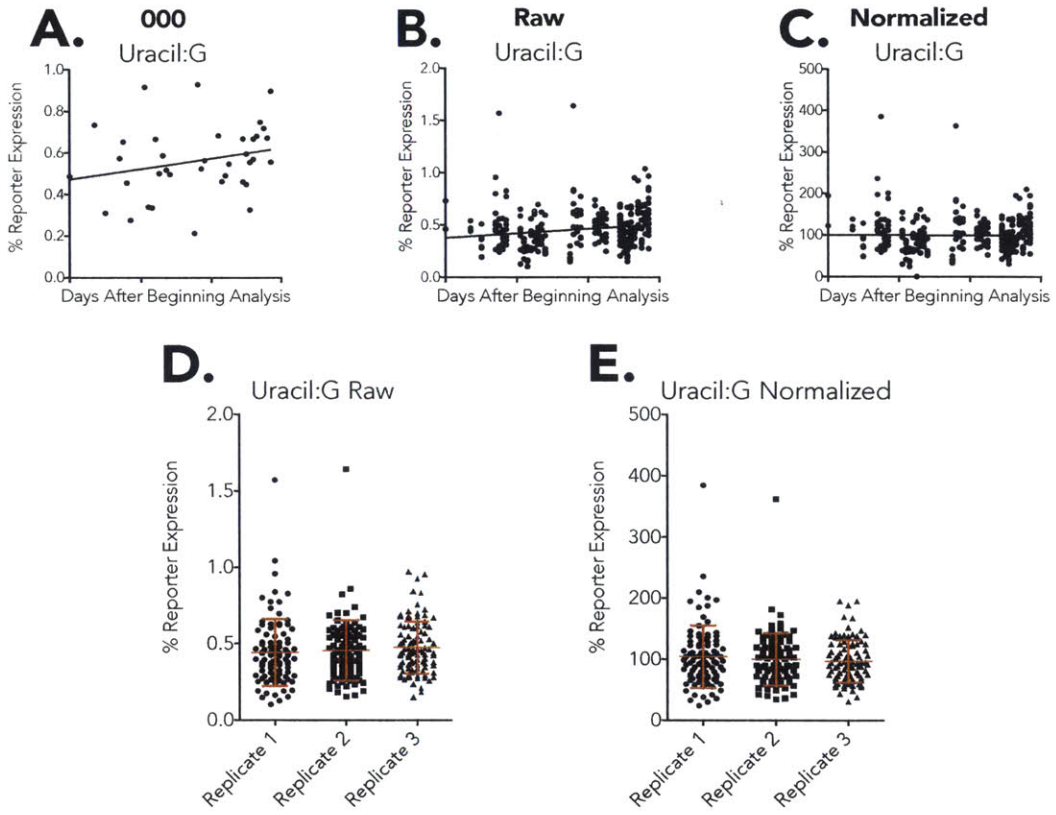


Figure 6.5 Evidence of time-associated drift and its normalization for the U:G reporter.

Scatter plots for U:G measurements in time (days after beginning flow cytometric analysis) and associated linear regression trend-line for **A.** Raw 000 controls, **B.** Each raw subject time-point and **C.** each subject time-point following normalization with strategy showed in **Figure 6.15**. U:G measurements for all study subjects grouped by replicate number **D.** Raw and **E.** Normalized. Red error bars correspond to standard deviation from the mean.

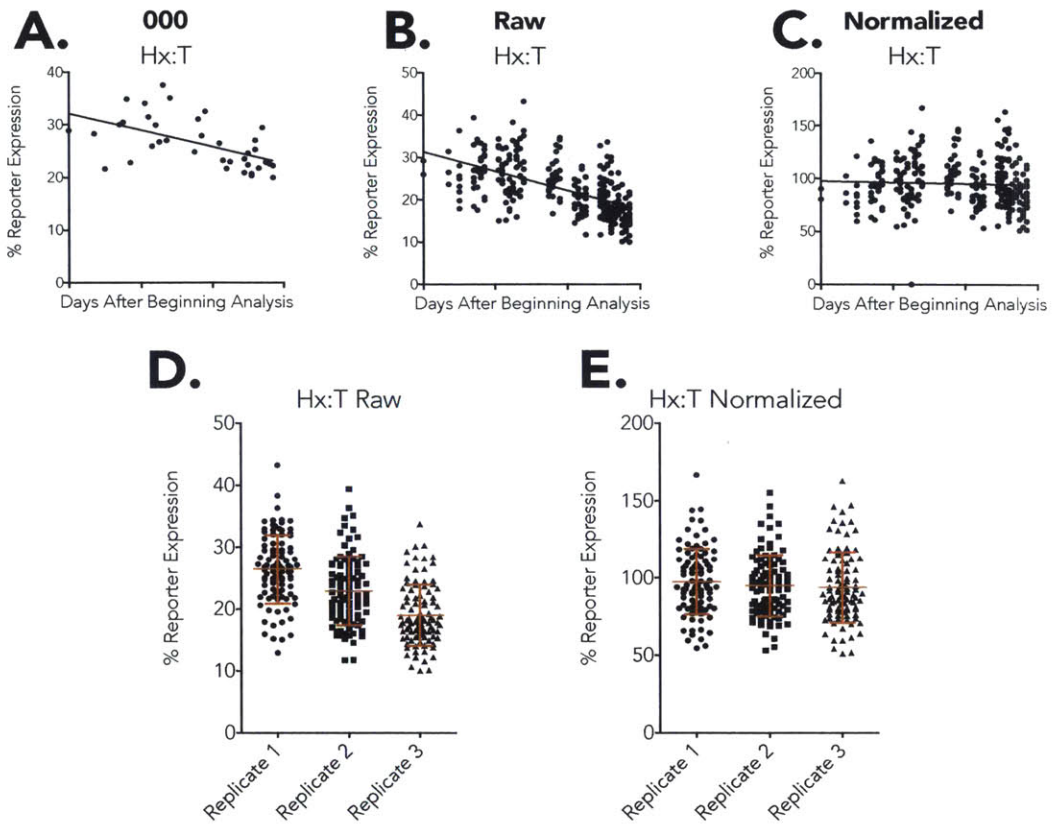


Figure 6.6 Evidence of time-associated drift and its normalization for the Hx:T reporter.

Scatter plots for Hx:T measurements in time (days after beginning flow cytometric analysis) and associated linear regression trend-line for **A.** Raw 000 controls, **B.** Each raw subject time-point and **C.** each subject time-point following normalization with strategy showed in **Figure 6.15**. Hx:T measurements for all study subjects grouped by replicate number **D.** Raw and **E.** Normalized. Red error bars correspond to standard deviation from the mean.

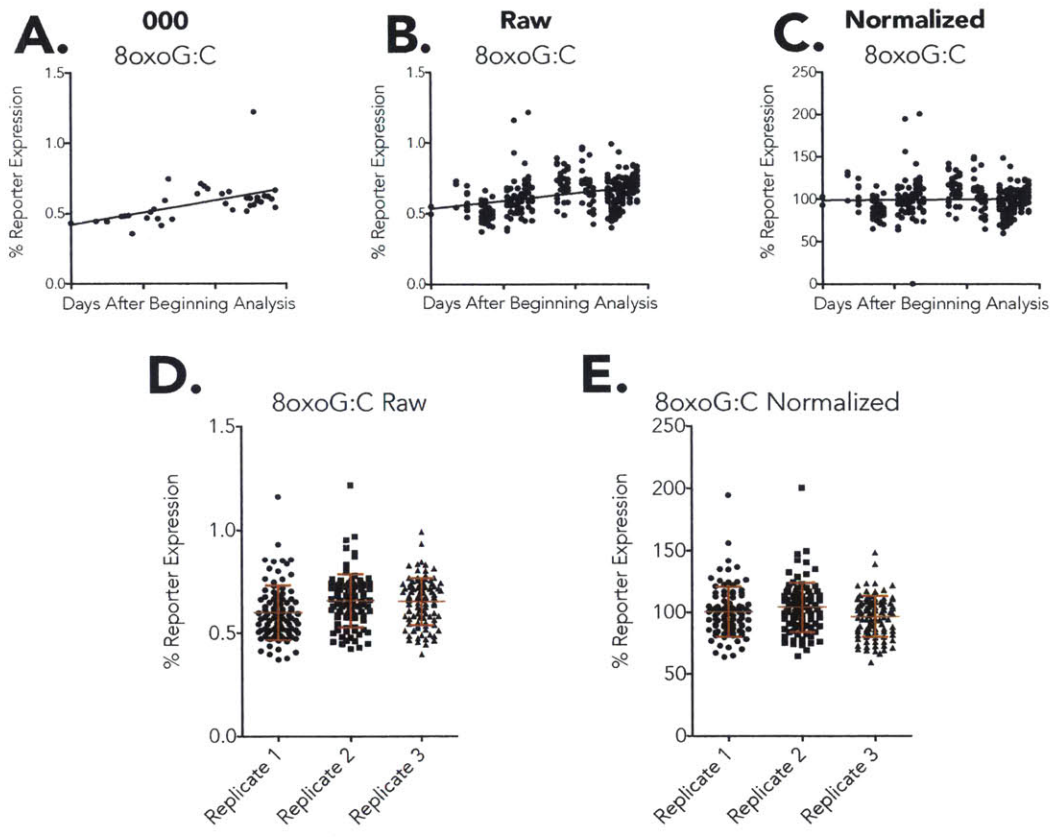


Figure 6.7 Evidence of time-associated drift and its normalization for the 8oxoG:C reporter.

Scatter plots for 8oxoG:C measurements in time (days after beginning flow cytometric analysis) and associated linear regression trend-line for **A.** Raw 000 controls, **B.** Each raw subject time-point and **C.** each subject time-point following normalization with strategy showed in **Figure 6.15**. 8oxoG:C measurements for all study subjects grouped by replicate number **D.** Raw and **E.** Normalized. Red error bars correspond to standard deviation from the mean.

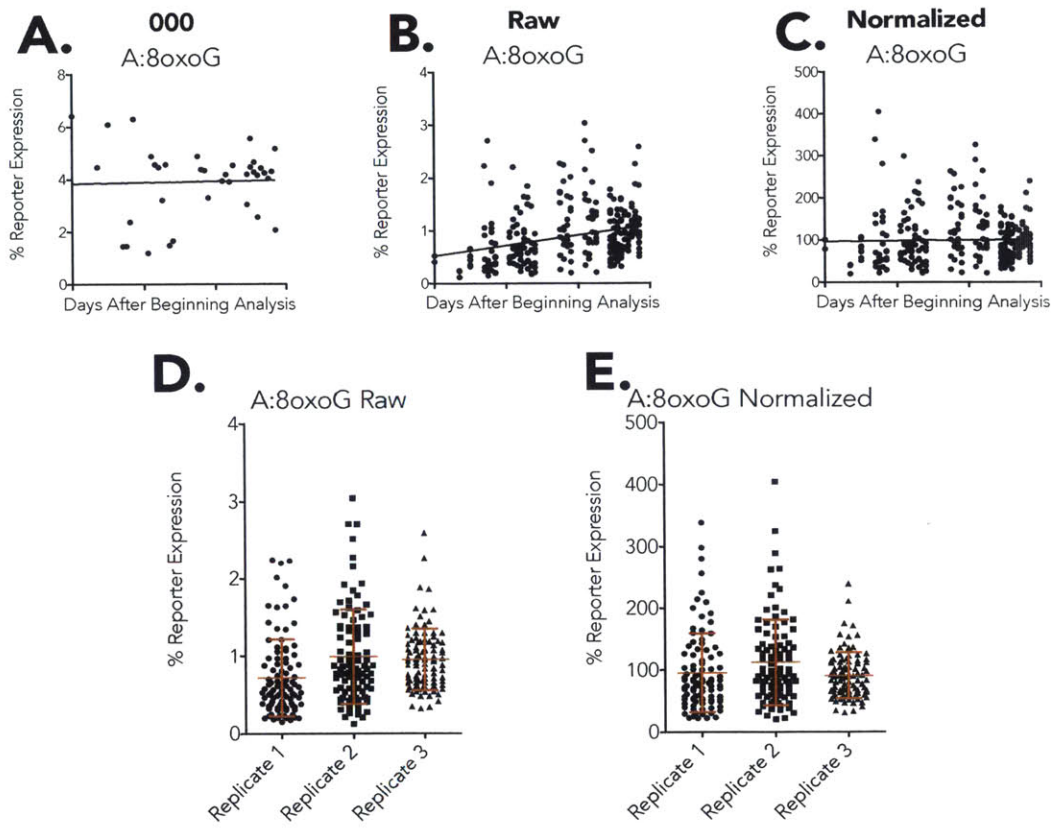


Figure 6.8 Evidence of time-associated drift and its normalization for the A:8oxoG reporter.

Scatter plots for A:8oxoG measurements in time (days after beginning flow cytometric analysis) and associated linear regression trend-line for **A.** Raw 000 controls, **B.** Each raw subject time-point and **C.** each subject time-point following normalization with strategy showed in **Figure 6.15**. A:8oxoG measurements for all study subjects grouped by replicate number **D.** Raw and **E.** Normalized. Red error bars correspond to standard deviation from the mean.

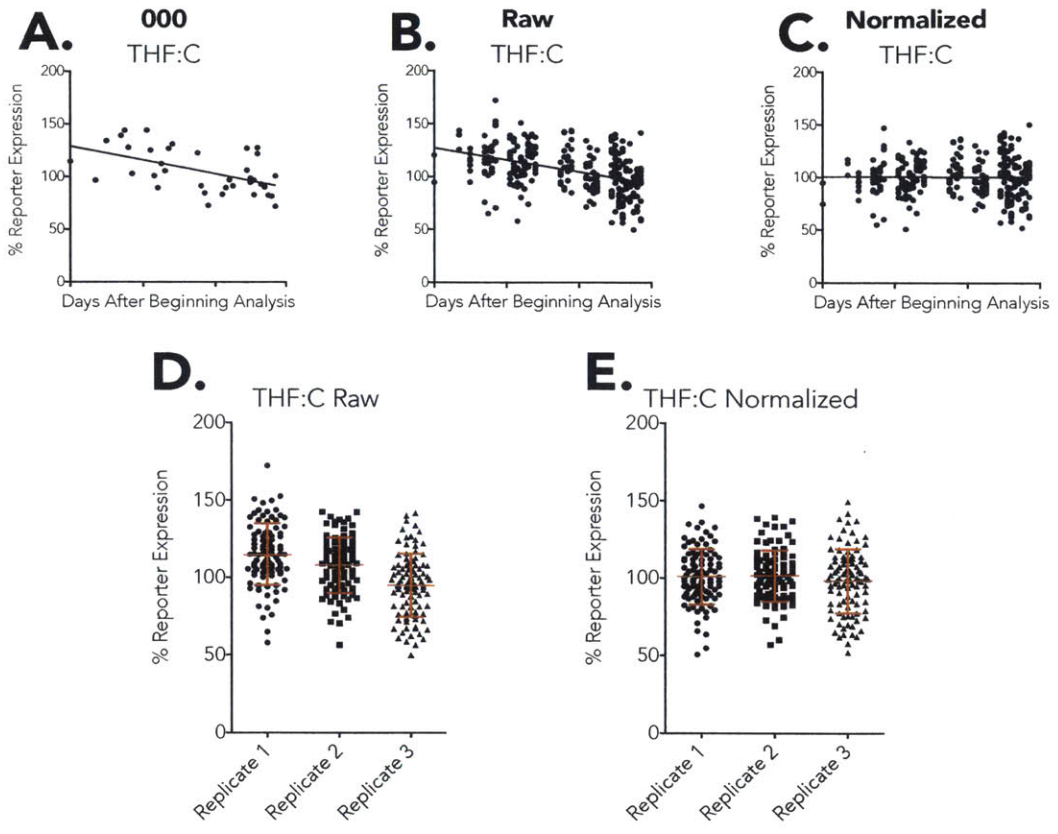


Figure 6.9 Evidence of time-associated drift and its normalization for the THF:C reporter.

Scatter plots for THF:C measurements in time (days after beginning flow cytometric analysis) and associated linear regression trend-line for **A.** Raw 000 controls, **B.** Each raw subject time-point and **C.** each subject time-point following normalization with strategy showed in **Figure 6.15**. THF:C measurements for all study subjects grouped by replicate number **D.** Raw and **E.** Normalized. Red error bars correspond to standard deviation from the mean.

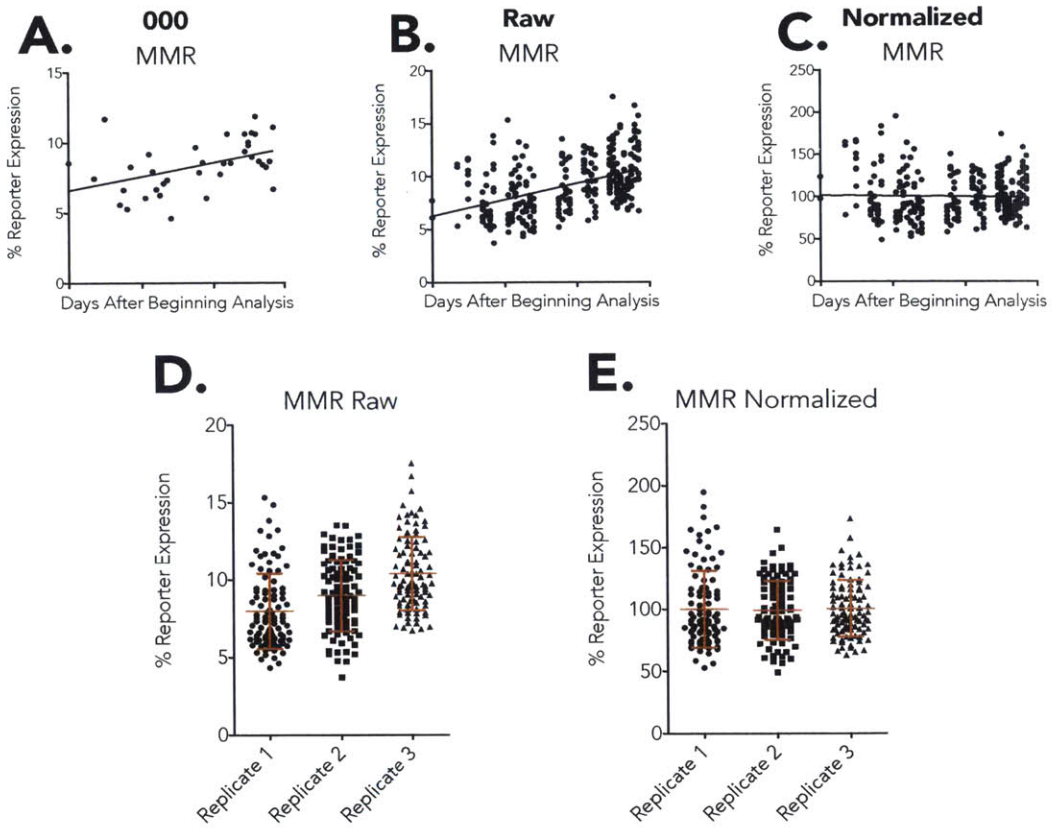


Figure 6.10 Evidence of time-associated drift and its normalization for the MMR reporter.

Scatter plots for MMR measurements in time (days after beginning flow cytometric analysis) and associated linear regression trend-line for **A.** Raw 000 controls, **B.** Each raw subject time-point and **C.** each subject time-point following normalization with strategy showed in **Figure 6.15**. MMR measurements for all study subjects grouped by replicate number **D.** Raw and **E.** Normalized. Red error bars correspond to standard deviation from the mean.

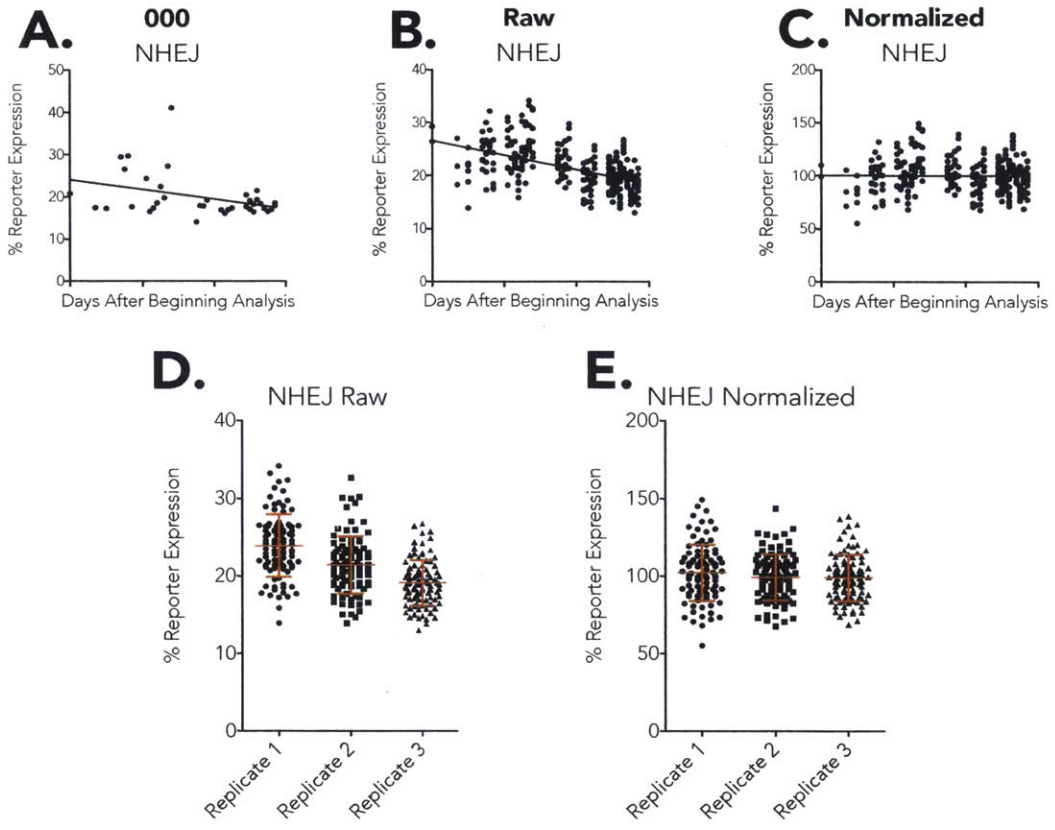


Figure 6.11 Evidence of time-associated drift and its normalization for the NHEJ reporter.

Scatter plots for NHEJ measurements in time (days after beginning flow cytometric analysis) and associated linear regression trend-line for **A.** Raw 000 controls, **B.** Each raw subject time-point and **C.** each subject time-point following normalization with strategy showed in **Figure 6.15**. NHEJ measurements for all study subjects grouped by replicate number **D.** Raw and **E.** Normalized. Red error bars correspond to standard deviation from the mean.

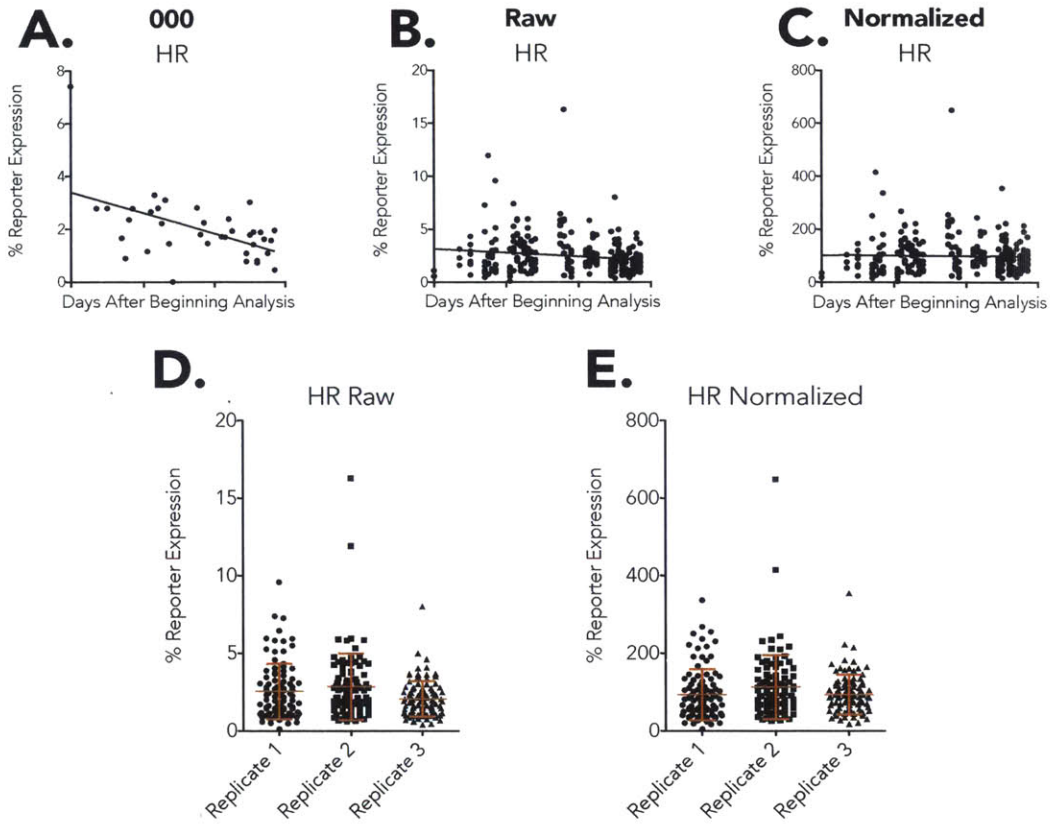


Figure 6.12 Evidence of time-associated drift and its normalization for the HR reporter.

Scatter plots for HR measurements in time (days after beginning flow cytometric analysis) and associated linear regression trend-line for **A.** Raw 000 controls, **B.** Each raw subject time-point and **C.** each subject time-point following normalization with strategy showed in **Figure 6.15**. HR measurements for all study subjects grouped by replicate number **D.** Raw and **E.** Normalized. Red error bars correspond to standard deviation from the mean.

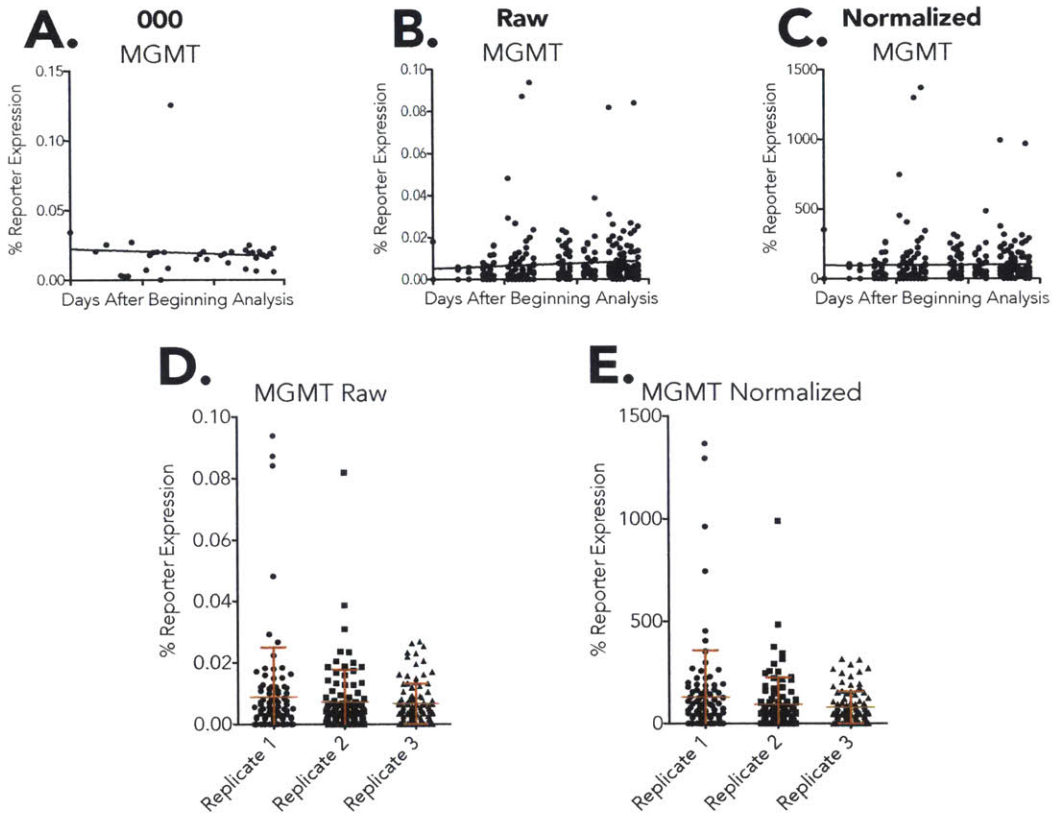


Figure 6.13 Evidence of time-associated drift and its normalization for the MGMT reporter.

Scatter plots for MGMT measurements in time (days after beginning flow cytometric analysis) and associated linear regression trend-line for **A.** Raw 000 controls, **B.** Each raw subject time-point and **C.** each subject time-point following normalization with strategy showed in **Figure 6.15**. MGMT measurements for all study subjects grouped by replicate number **D.** Raw and **E.** Normalized. Red error bars correspond to standard deviation from the mean.

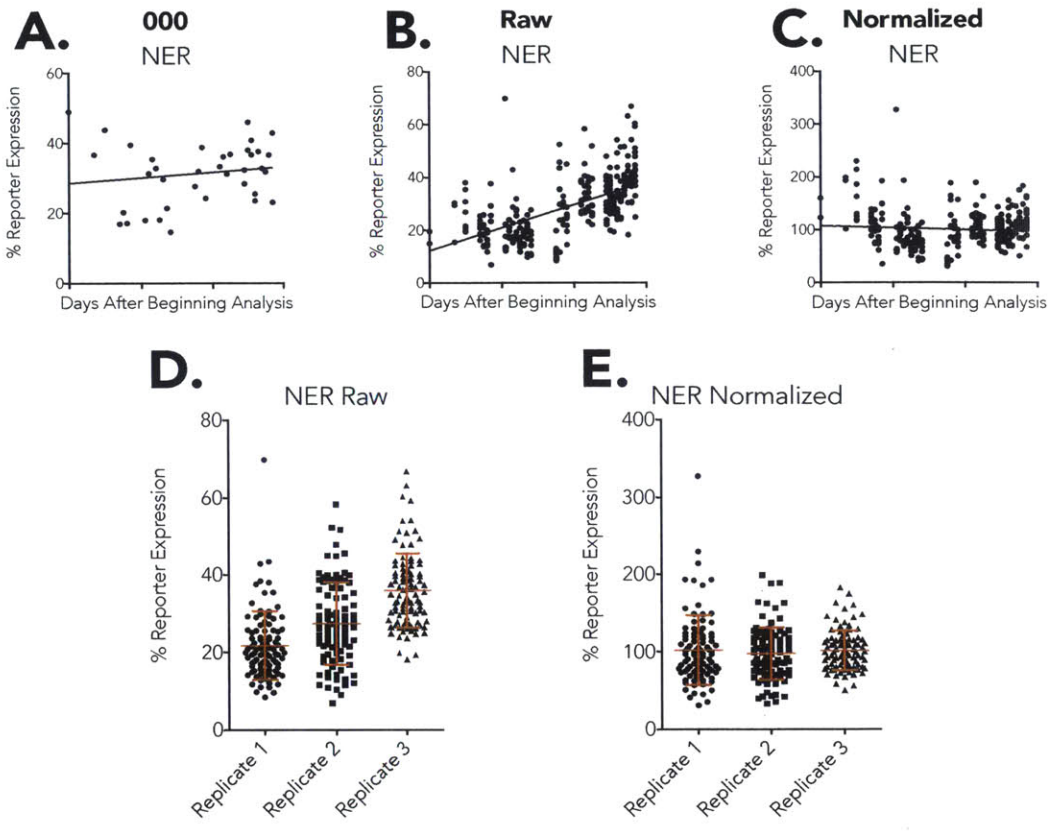
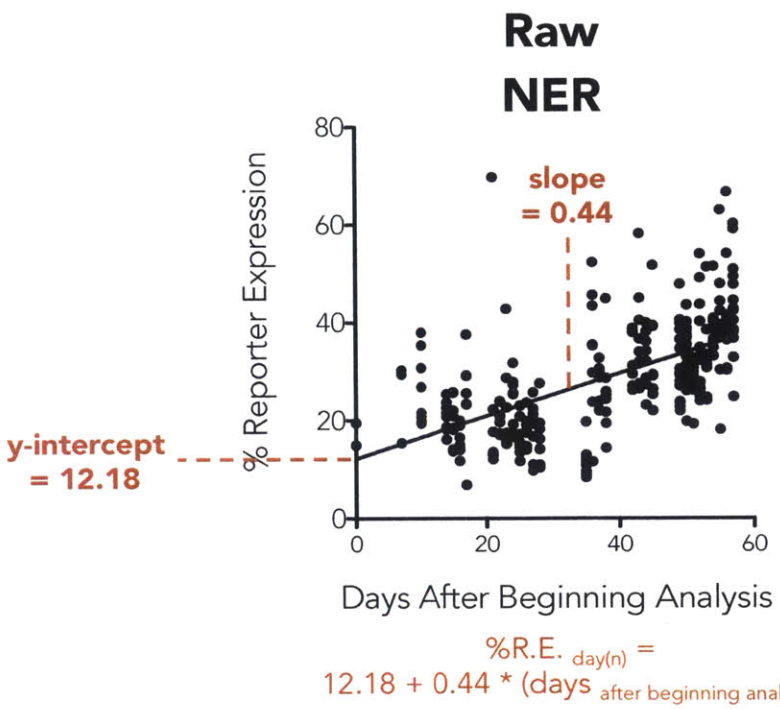


Figure 6.14 Evidence of time-associated drift and its normalization for the NER reporter.

Scatter plots for NER measurements in time (days after beginning flow cytometric analysis) and associated linear regression trend-line for **A.** Raw 000 controls, **B.** Each raw subject time-point and **C.** each subject time-point following normalization with strategy showed in **Figure 6.15**. NER measurements for all study subjects grouped by replicate number **D.** Raw and **E.** Normalized. Red error bars correspond to standard deviation from the mean.



Normalized %R.E. $_{\text{Subject(i),Replicate(ii)}}$ **= (%R.E.** $_{\text{Subject(i),Replicate(ii)}}$**) / (%R.E.** $_{\text{day(n)}}$**)**

Figure 6.15 Generalized normalization strategy for time-associated drift.
Each data-point was normalized by dividing to a calculated %R.E. corresponding to the linear regression score of all subject data-points in time for the particular day the measurement was taken for each DNA repair reporter.

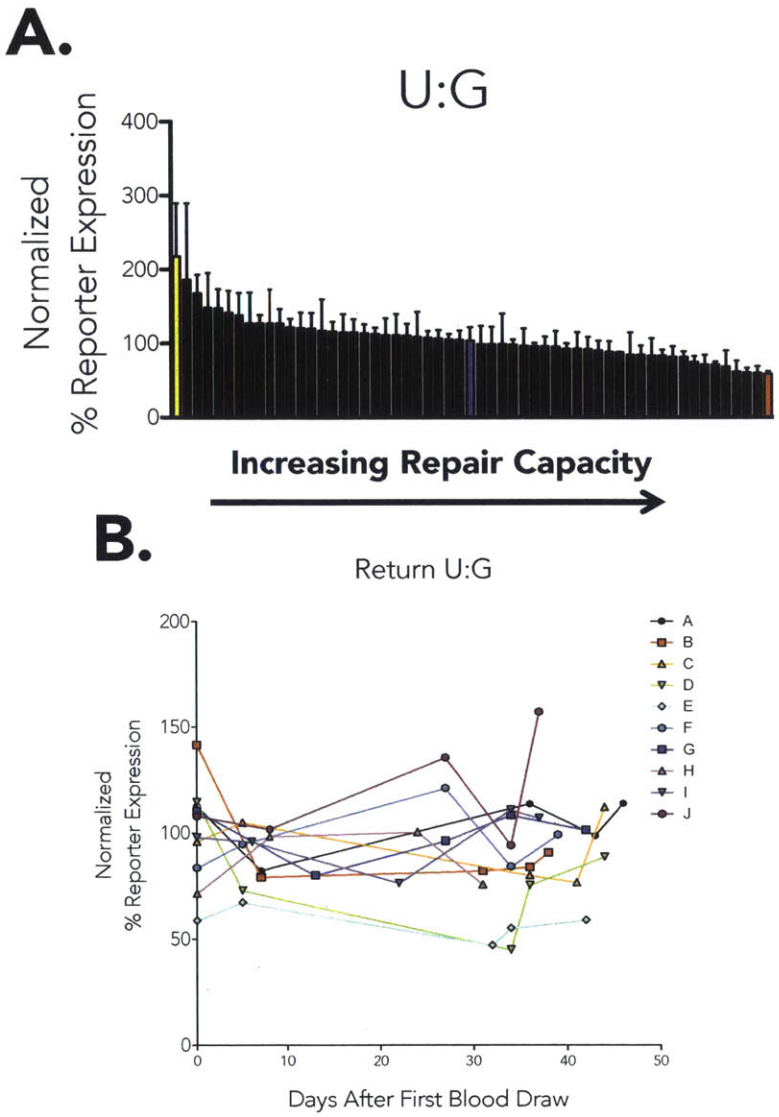


Figure 6.16 Inter- and intra-individual U:G repair capacity.

A. Normalized U:G repair capacity measured for 56 study subjects. Each bar corresponds to a different subject. Yellow, blue and red bars correspond to single individuals labeled for each different reporter for comparison. Error bars represent the standard error of the mean calculated from biological triplicates. **B.** Intra-individual variability, each color represents a different study subject for whom U:G repair capacity was measured from PBMCs isolated from blood drew on different days.

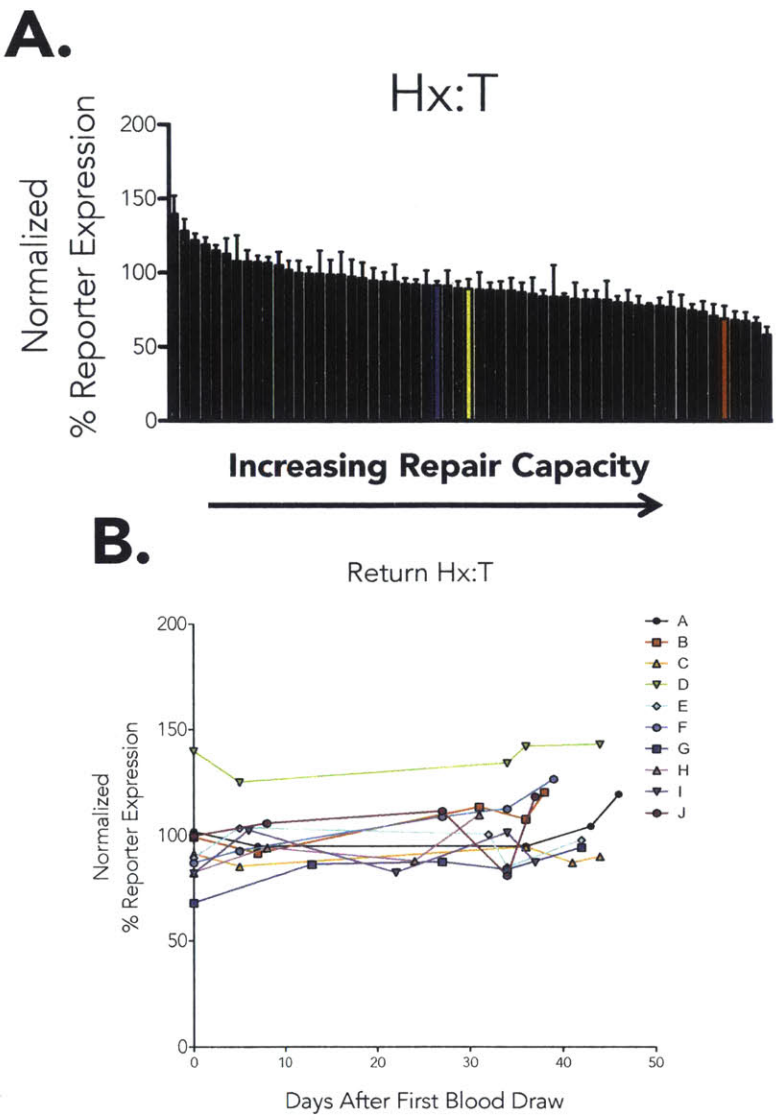


Figure 6.17 Inter- and intra-individual Hx:T repair capacity.

A. Normalized Hx:T repair capacity measured for 56 study subjects. Each bar corresponds to a different subject. Yellow, blue and red bars correspond to single individuals labeled for each different reporter for comparison. Error bars represent the standard error of the mean calculated from biological triplicates. **B.** Intra-individual variability, each color represents a different study subject for whom Hx:T repair capacity was measured from PBMCs isolated from blood drew on different days.

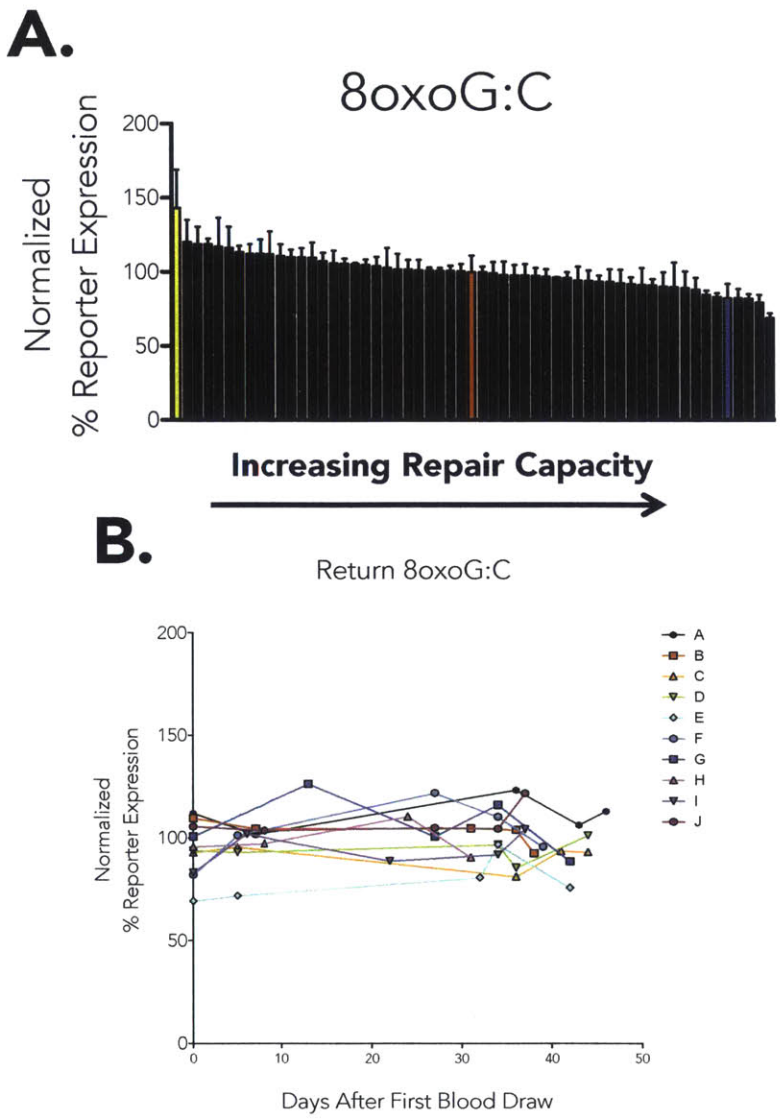


Figure 6.18 Inter- and intra-individual 8oxoG:C repair capacity.

A. Normalized 8oxoG:C repair capacity measured for 56 study subjects. Each bar corresponds to a different subject. Yellow, blue and red bars correspond to single individuals labeled for each different reporter for comparison. Error bars represent the standard error of the mean calculated from biological triplicates. **B.** Intra-individual variability, each color represents a different study subject for whom 8oxoG:C repair capacity was measured from PBMCs isolated from blood drew on different days.

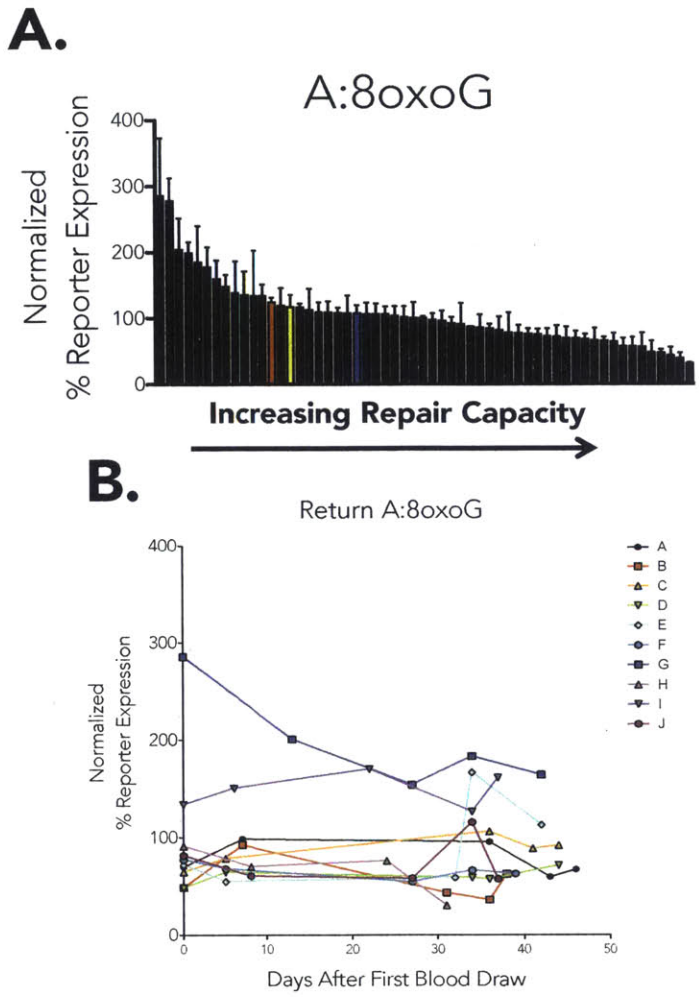


Figure 6.19 Inter- and intra-individual A:8oxoG repair capacity.

A. Normalized A:8oxoG repair capacity measured for 56 study subjects. Each bar corresponds to a different subject. Yellow, blue and red bars correspond to single individuals labeled for each different reporter for comparison. Error bars represent the standard error of the mean calculated from biological triplicates. **B.** Intra-individual variability, each color represents a different study subject for whom A:8oxoG repair capacity was measured from PBMCs isolated from blood drew on different days.

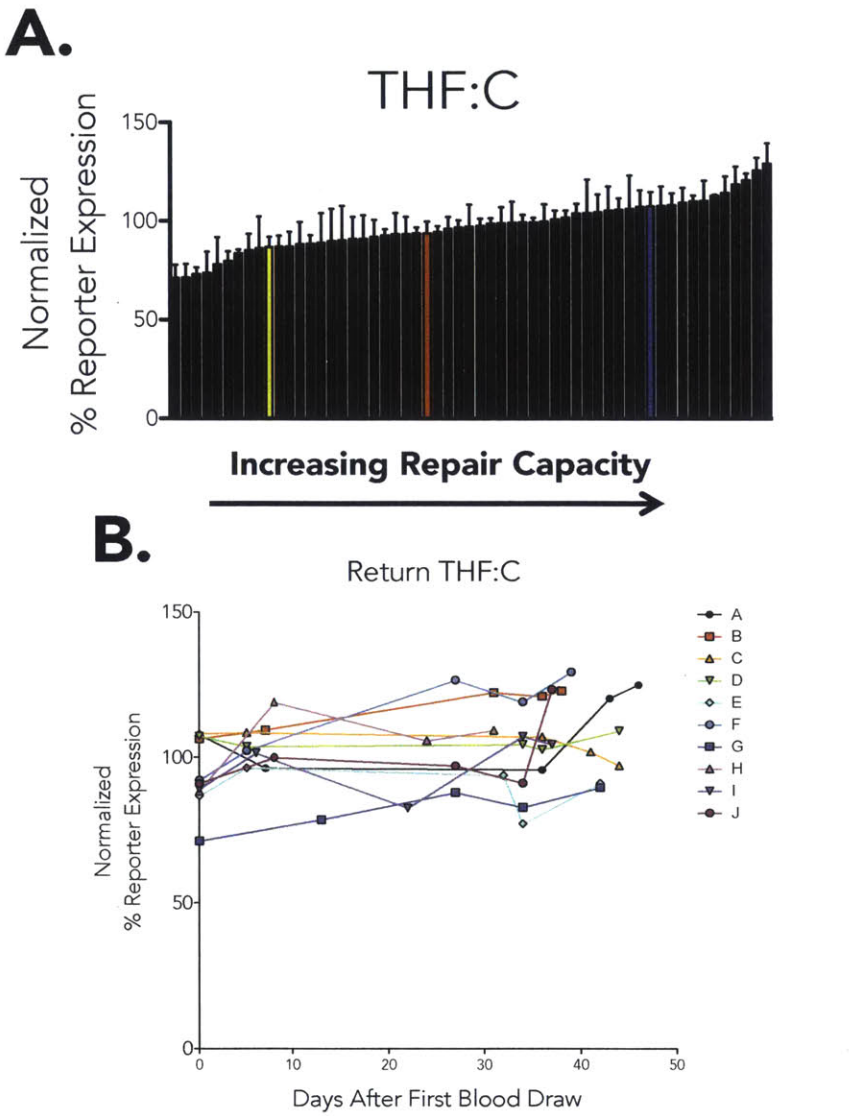


Figure 6.20 Inter- and intra-individual THF:C repair capacity.

A. Normalized THF:C repair capacity measured for 56 study subjects. Each bar corresponds to a different subject. Yellow, blue and red bars correspond to single individuals labeled for each different reporter for comparison. Error bars represent the standard error of the mean calculated from biological triplicates. **B.** Intra-individual variability, each color represents a different study subject for whom THF:C repair capacity was measured from PBMCs isolated from blood drew on different days.

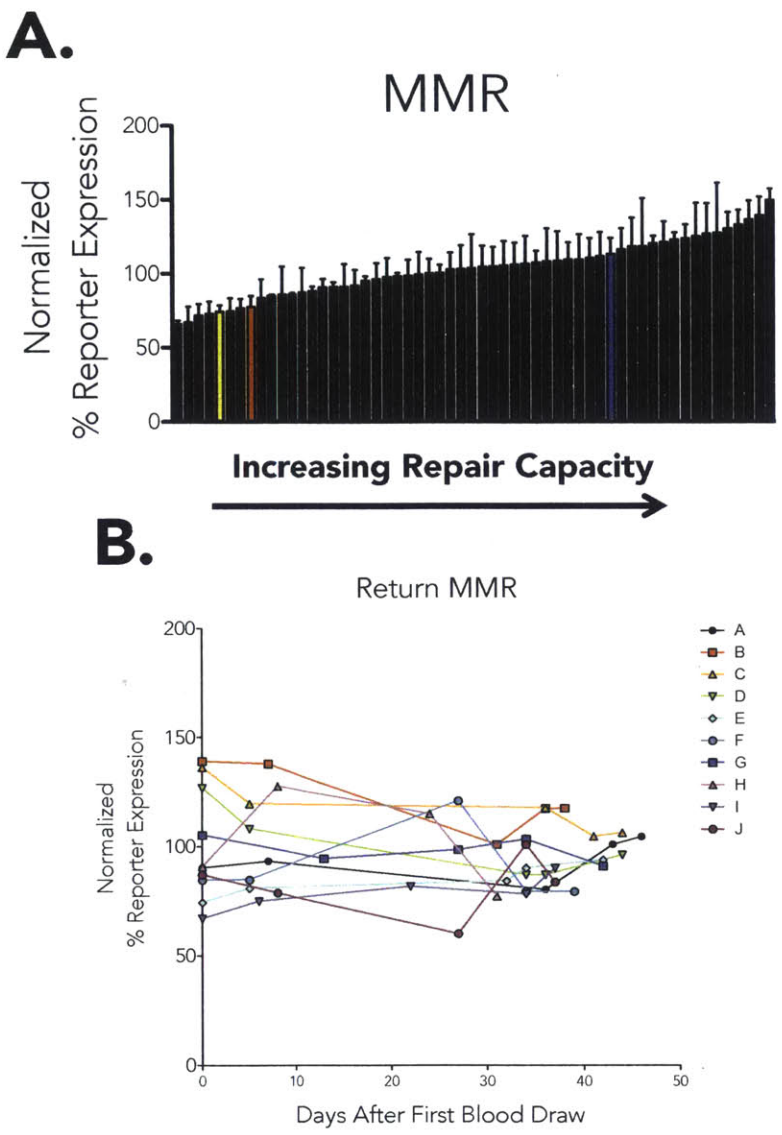


Figure 6.21 Inter- and intra-individual MMR capacity.

A. Normalized MMR capacity measured for 56 study subjects. Each bar corresponds to a different subject. Yellow, blue and red bars correspond to single individuals labeled for each different reporter for comparison. Error bars represent the standard error of the mean calculated from biological triplicates. **B.** Intra-individual variability, each color represents a different study subject for whom MMR capacity was measured from PBMCs isolated from blood drew on different days.

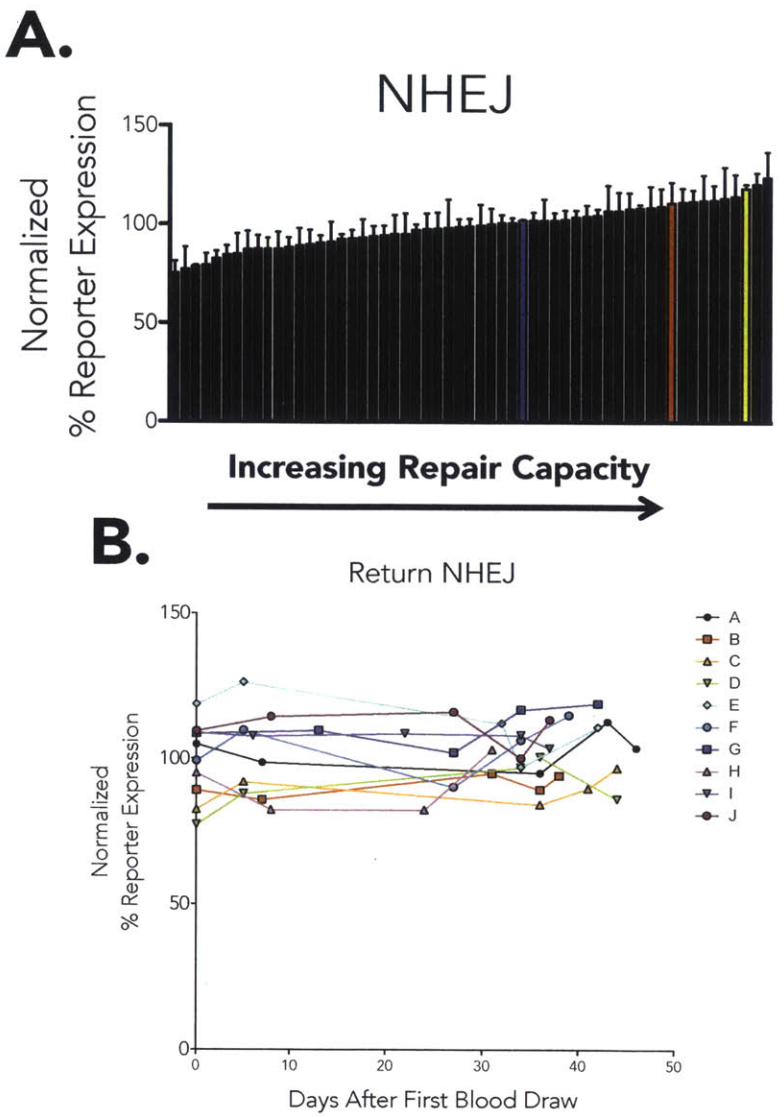


Figure 6.22 Inter- and intra-individual NHEJ capacity.

A. Normalized NHEJ capacity measured for 56 study subjects. Each bar corresponds to a different subject. Yellow, blue and red bars correspond to single individuals labeled for each different reporter for comparison. Error bars represent the standard error of the mean calculated from biological triplicates. **B.** Intra-individual variability, each color represents a different study subject for whom NHEJ capacity was measured from PBMCs isolated from blood drew on different days.

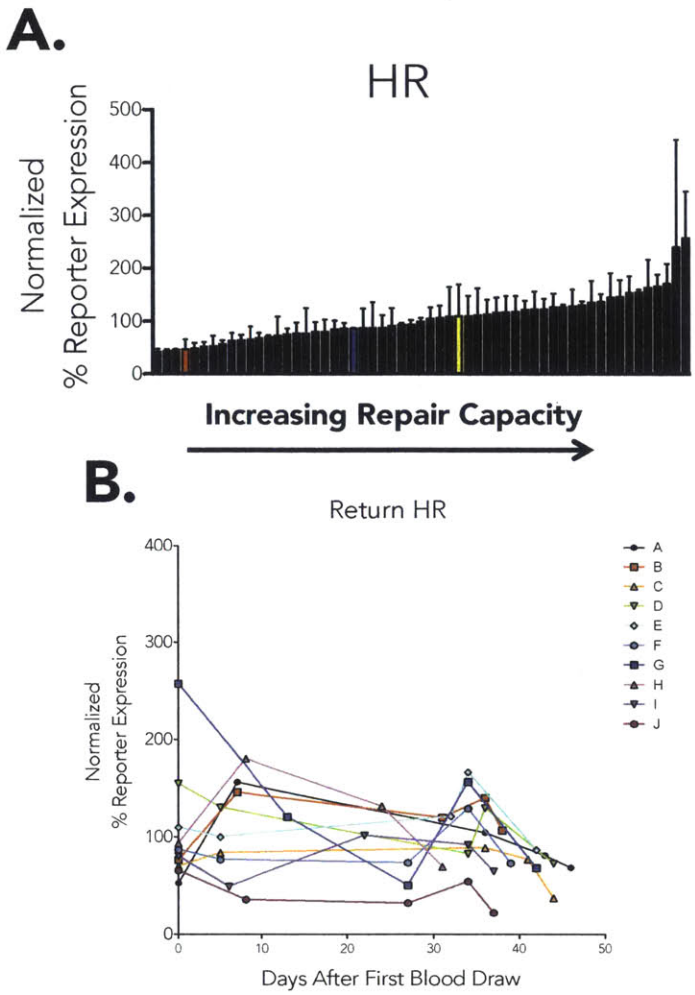


Figure 6.23 Inter- and intra-individual HR capacity.

A. Normalized HR capacity measured for 56 study subjects. Each bar corresponds to a different subject. Yellow, blue and red bars correspond to single individuals labeled for each different reporter for comparison. Error bars represent the standard error of the mean calculated from biological triplicates. **B.** Intra-individual variability, each color represents a different study subject for whom HR capacity was measured from PBMCs isolated from blood drew on different days.

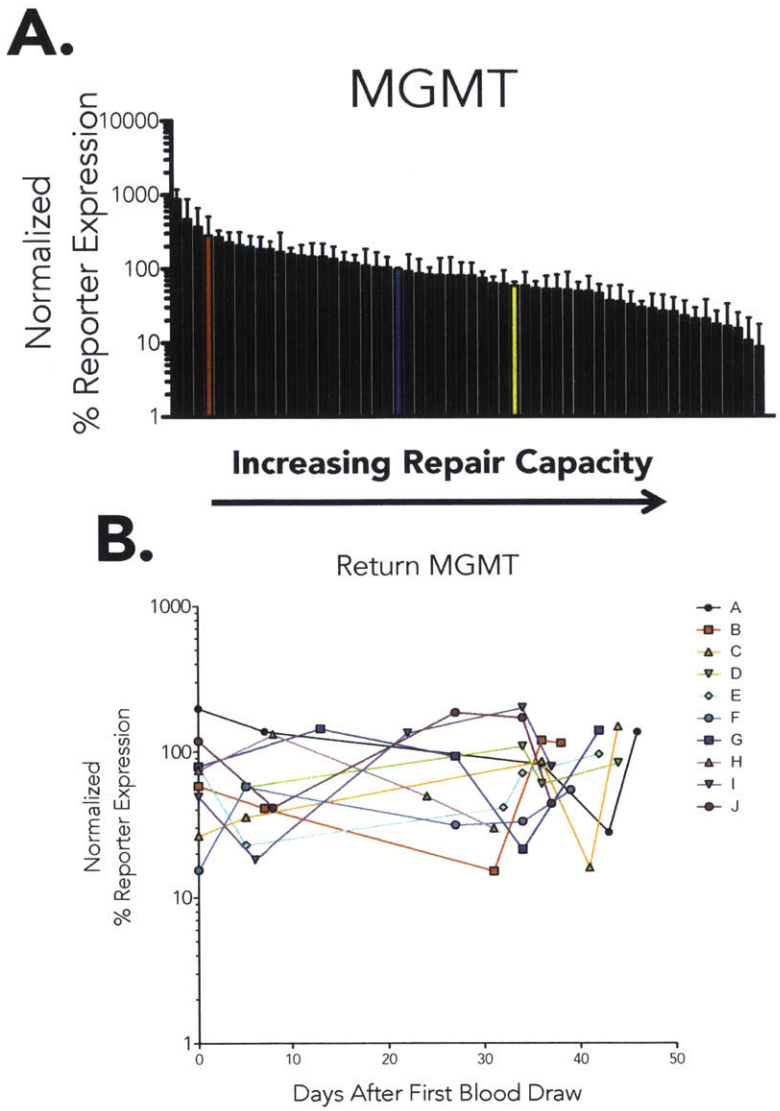


Figure 6.24 Inter- and intra-individual MGMT capacity.

A. Normalized MGMT capacity measured for 56 study subjects. Each bar corresponds to a different subject. Yellow, blue and red bars correspond to single individuals labeled for each different reporter for comparison. Error bars represent the standard error of the mean calculated from biological triplicates. **B.** Intra-individual variability, each color represents a different study subject for whom MGMT capacity was measured from PBMCs isolated from blood drew on different days.

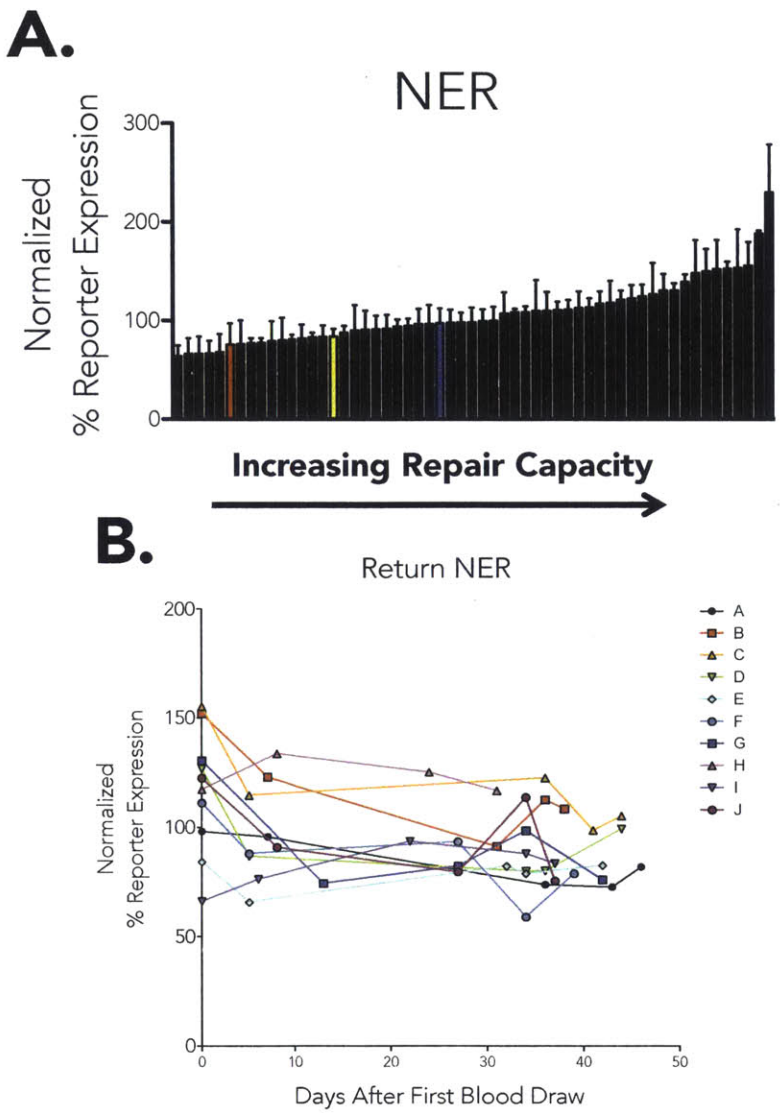


Figure 6.25 Inter- and intra-individual NER capacity.

A. Normalized NER capacity measured for 56 study subjects. Each bar corresponds to a different subject. Yellow, blue and red bars correspond to single individuals labeled for each different reporter for comparison. Error bars represent the standard error of the mean calculated from biological triplicates. **B.** Intra-individual variability, each color represents a different study subject for whom NER capacity was measured from PBMCs isolated from blood drew on different days.

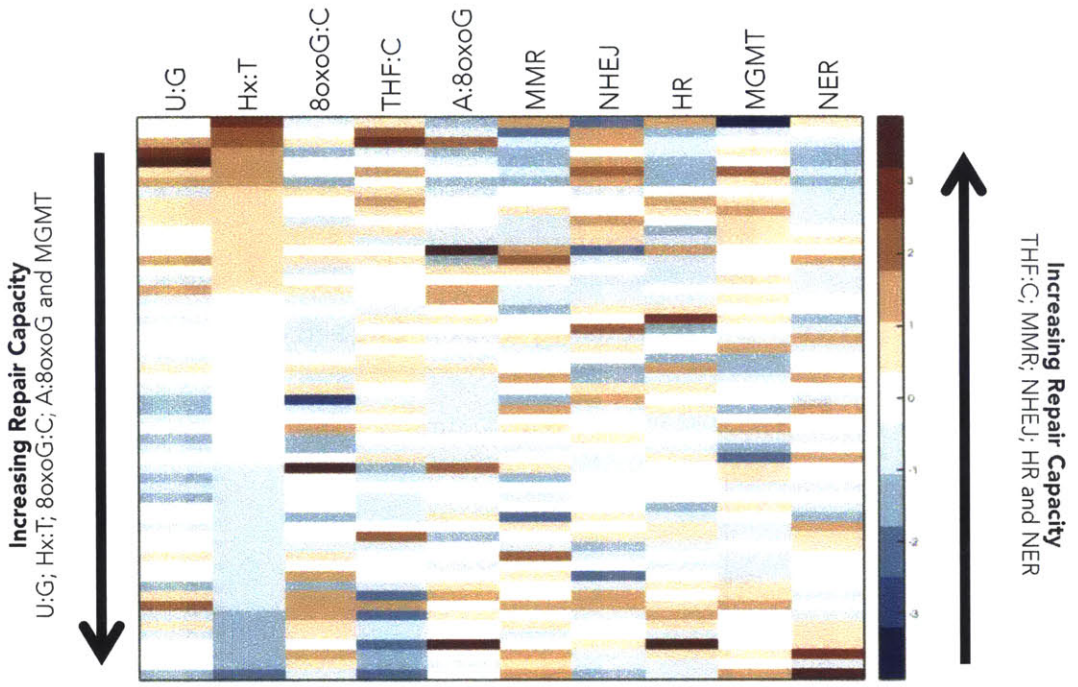


Figure 6.26 Inter-individual variability in DRC of a panel of PBMCs derived from 56 apparently healthy individuals.

Normalized DRC measurements for each repair reporter was z-scored and organized by lower to higher Hx excision activity. Warm colors represent lower repair for all reporters transcriptional mutagenesis reporter (U:G; Hx:T; 8oxoG:C; A:8oxoG and MGMT) and higher repair for the rest (THF:C; MMR; NHEJ; HR and NER).

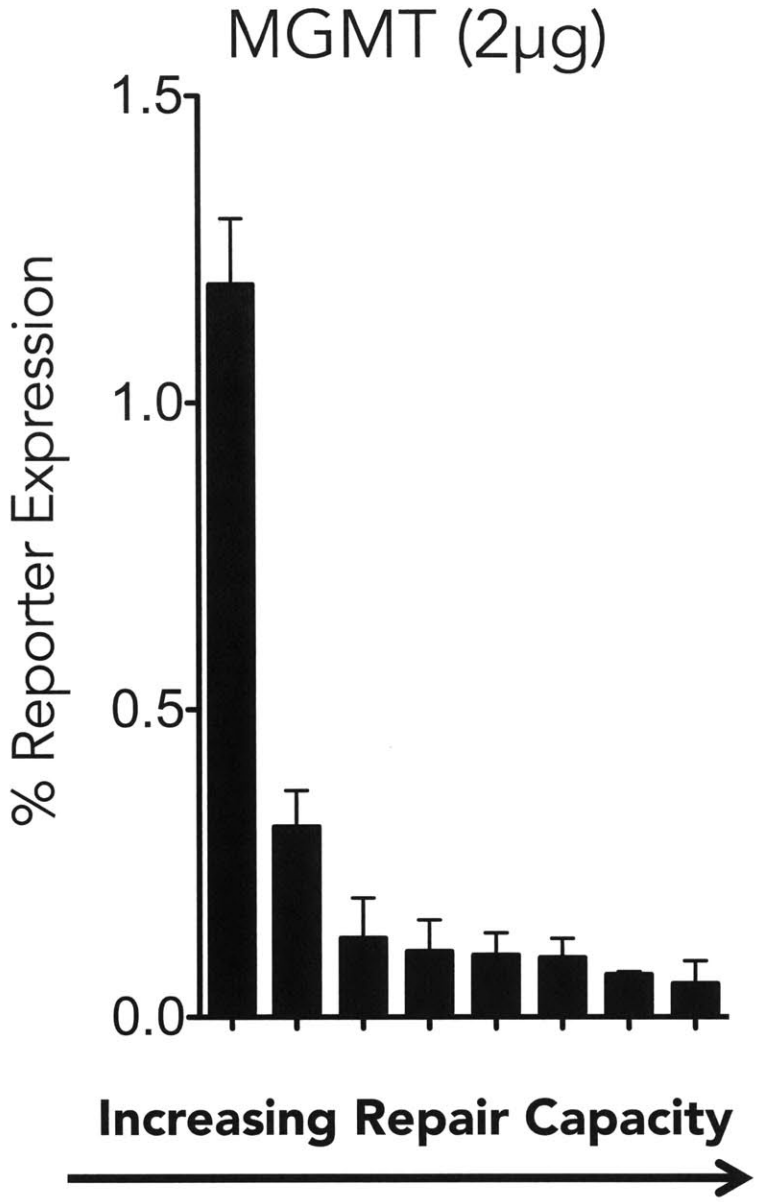


Figure 6.27 MGMT measurements for 8 individuals with increased reporter transfection.

2µg of the MGMT reporter where included in each transfection. Each bar corresponds to a different subject. Error bars represent the standard deviation calculated from biological duplicates.

References

1. Nagel, Z.D., I.A. Chaim, and L.D. Samson, *Inter-individual variation in DNA repair capacity: A need for multi-pathway functional assays to promote translational DNA repair research.* DNA repair, 2014. **19**: p. 199–213.
2. Ciccia, A. and S.J. Elledge, *The DNA damage response: making it safe to play with knives.* Molecular cell, 2010. **40**(2): p. 179-204.
3. Kim, Y.-J. and D.M. Wilson III, *Overview of base excision repair biochemistry.* Current molecular pharmacology, 2012. **5**(1): p. 3.
4. Hartwig, A., H. Blessing, T. Schwerdtle, and I. Walter, *Modulation of DNA repair processes by arsenic and selenium compounds.* Toxicology, 2003. **193**(1-2): p. 161-169.
5. Andrew, A.S., J.L. Burgess, M.M. Meza, E. Demidenko, M.G. Waugh, J.W. Hamilton, and M.R. Karagas, *Arsenic exposure is associated with decreased DNA repair in vitro and in individuals exposed to drinking water arsenic.* Environmental Health Perspectives, 2006. **114**(8): p. 1193-1198.
6. Ahmed, S., S.M.E. Khoda, R.S. Rekha, R.M. Gardner, S.S. Ameer, S. Moore, E.C. Ekstrom, M. Vahter, and R. Raqib, *Arsenic-Associated Oxidative Stress, Inflammation, and Immune Disruption in Human Placenta and Cord Blood.* Environmental Health Perspectives, 2011. **119**(2): p. 258-264.
7. Hengstler, J.G., U. Bolm-Audorff, A. Faldum, K. Janssen, M. Reifenrath, W. Gotte, D.L. Jung, O. Mayer-Popken, J. Fuchs, S. Gebhard, H.G. Bienfait, K. Schlink, C. Dietrich, D. Faust, B. Epe, and F. Oesch, *Occupational exposure to heavy metals: DNA damage induction and DNA repair inhibition prove co-exposures to cadmium, cobalt and lead as more dangerous than hitherto expected.* Carcinogenesis, 2003. **24**(1): p. 63-73.
8. Mattern, J., R. Koomagi, and M. Volm, *Smoking-related increase of O-6-methylguanine-DNA methyltransferase expression in human lung carcinomas.* Carcinogenesis, 1998. **19**(7): p. 1247-1250.
9. Kothandapani, A., V.S.M.N. Dangeti, A.R. Brown, L.A. Banze, X.-H. Wang, R.W. Sobol, and S.M. Patrick, *Novel role of base excision repair in mediating cisplatin cytotoxicity.* Journal of Biological Chemistry, 2011. **286**(16): p. 14564-14574.
10. Enoiu, M., J. Jiricny, and O.D. Schärer, *Repair of cisplatin-induced DNA interstrand crosslinks by a replication-independent pathway involving transcription-coupled repair and translesion synthesis.* Nucleic Acids Research, 2012. **40**(18): p. 8953-8964.
11. Kim, H. and A.D. D'Andrea, *Regulation of DNA cross-link repair by the Fanconi anemia/BRCA pathway.* Genes & Development, 2012. **26**(13): p. 1393-1408.
12. Guo, J., P.C. Hanawalt, and G. Spivak, *Comet-FISH with strand-specific probes reveals transcription-coupled repair of 8-oxoGuanine in human cells.* Nucleic Acids Research, 2013.
13. Kothandapani, A., A. Sawant, V.S.M.N. Dangeti, R.W. Sobol, and S.M. Patrick, *Epistatic role of base excision repair and mismatch repair pathways in mediating cisplatin cytotoxicity.* Nucleic Acids Research, 2013.

14. Lv, L., F. Wang, X. Ma, Y. Yang, Z. Wang, H. Liu, X. Li, Z. Liu, T. Zhang, M. Huang, E.C. Friedberg, T.-S. Tang, and C. Guo, *Mismatch repair protein MSH2 regulates translesion DNA synthesis following exposure of cells to UV radiation.* Nucleic Acids Research, 2013. **41**(22): p. 10312-10322.
15. Kraemer, K.H., M.M. Lee, and J. Scotto, *DNA repair protects against cutaneous and internal neoplasia - evidence from xeroderma pigmentosum.* Carcinogenesis, 1984. **5**(4): p. 511-514.
16. O'Driscoll, M., *Diseases Associated with Defective Responses to DNA Damage.* Cold Spring Harbor Perspectives in Biology, 2012. **4**(12).
17. Ellis, N.C., *Obtaining and Using Genetic Information.* Inherited Cancer Syndromes: Current Clinical Management, ed. N.C. Ellis. 2003, New York: Springer.
18. Hsu, T.C., *Genetic instability in the human population - a working hypothesis* Hereditas, 1983. **98**(1): p. 1-9.
19. Grossman, L. and Q. Wei, *DNA repair and epidemiology of basal cell carcinoma.* Clinical chemistry, 1995. **41**(12): p. 1854-1863.
20. Paz-Elizur, T., D. Elinger, Y. Leitner-Dagan, S. Blumenstein, M. Krupsky, A. Berrebi, E. Schechtman, and Z. Livneh, *Development of an enzymatic DNA repair assay for molecular epidemiology studies: distribution of OGG activity in healthy individuals.* DNA repair, 2007. **6**(1): p. 45-60.
21. Johnson, J.M. and J.J. Latimer, *Analysis of DNA Repair Using Transfection-Based Host Cell Reactivation*, in *Molecular Toxicology Protocols*, P. Keohavong and G.G. Grant, Editors. 2005, Humana Press: Totowa, NJ.
22. Nagel, Z.D., C.M. Thompson, I.A. Chaim, S.K. McRee, P. Mazzucato, A. Ahmad, R.P. Abo, V.L. Butty, A.L. Forget, and L.D. Samson, *Multiplexed DNA repair assays for multiple lesions and multiple doses via transcription inhibition and transcriptional mutagenesis.* Proceedings of the National Academy of Sciences, 2014. ***In Press***.
23. Nagel, Z.D., C.M. Margulies, I.A. Chaim, S.K. McRee, P. Mazzucato, A. Ahmad, R.P. Abo, V.L. Butty, A.L. Forget, and L.D. Samson, *Multiplexed DNA repair assays for multiple lesions and multiple doses via transcription inhibition and transcriptional mutagenesis.* Proceedings of the National Academy of Sciences, 2014. **111**(18): p. E1823-E1832.
24. Baerenfaller, K., F. Fischer, and J. Jiricny, *Characterization of the "mismatch repairosome" and its role in the processing of modified nucleosides in vitro*, in *DNA Repair, Pt A*, J. Campbell and P. Modrich, Editors. 2006. p. 285-303.
25. Kiziltepe, T., A. Yan, M. Dong, V.S. Jonnalagadda, P.C. Dedon, and B.P. Engelward, *Delineation of the chemical pathways underlying nitric oxide-induced homologous recombination in mammalian cells.* Chemistry & Biology, 2005. **12**(3): p. 357-369.
26. Sevilya, Z., Y. Leitner-Dagan, M. Pinchev, R. Kremer, D. Elinger, H.S. Rennert, E. Schechtman, L.S. Freedman, G. Rennert, T. Paz-Elizur, and Z. Livneh, *Low integrated DNA repair score and lung cancer risk.* Cancer Prevention Research, 2013. **7**(4): p. 398-406.
27. Leitner-Dagan, Y., Z. Sevilya, M. Pinchev, R. Kramer, D. Elinger, L.C. Roisman, H.S. Rennert, E. Schechtman, L. Freedman, G. Rennert, Z. Livneh, and T. Paz-

Elizur, *N-Methylpurine DNA Glycosylase and OGG1 DNA Repair Activities: Opposite Associations With Lung Cancer Risk.* Journal of the National Cancer Institute, 2012. **104**(22): p. 1765-1769.
28. Janssen, K., K. Schlink, W. Götte, B. Hippler, B. Kaina, and F. Oesch, *DNA repair activity of 8-oxoguanine DNA glycosylase 1 (OGG1) in human lymphocytes is not dependent on genetic polymorphism Ser 326/Cys 326.* Mutation Research/DNA Repair, 2001. **486**(3): p. 207-216.
29. Paz-Elizur, T., M. Krupsky, S. Blumenstein, D. Elinger, E. Schechtman, and Z. Livneh, *DNA repair activity for oxidative damage and risk of lung cancer.* Journal of the National Cancer Institute, 2003. **95**(17): p. 1312-1319.
30. Gaivão, I., A. Piasek, A. Brevik, S. Shaposhnikov, and A.R. Collins, *Comet assay-based methods for measuring DNA repair in vitro; estimates of inter-and intra-individual variation.* Cell biology and toxicology, 2009. **25**(1): p. 45-52.
31. Myrnes, B., K.-E. Giercksky, and H. Krokan, *Interindividual variation in the activity of O6-methyl guanine-DNA methyltransferase and uracil-DNA glycosylase in human organs.* Carcinogenesis, 1983. **4**(12): p. 1565-1568.
32. Redaelli, A., R. Magrassi, S. Bonassi, A. Abbondandolo, and G. Frosina, *AP endonuclease activity in humans: development of a simple assay and analysis of ten normal individuals.* Teratogenesis, carcinogenesis, and mutagenesis, 1998. **18**(1): p. 17-26.
33. O'Donnell, P.N.S., P.V. Barber, G.P. Margison, and A.C. Povey, *Association between O-6-alkylguanine-DNA- alkyltransferase activity in peripheral blood lymphocytes and bronchial epithelial cells.* Cancer Epidemiology Biomarkers & Prevention, 1999. **8**(7): p. 641-645.
34. Athas, W.F., M.A. Hedayati, G.M. Matanoski, E.R. Farmer, and L. Grossman, *Development and field-test validation of an assay for DNA-repair in circulating human lymphocytes.* Cancer Research, 1991. **51**(21): p. 5786-5793.
35. Tyson, J., F. Caple, A. Spiers, B. Burtle, A.K. Daly, E.A. Williams, J.E. Hesketh, and J.C. Mathers, *Inter-individual variation in nucleotide excision repair in young adults: effects of age, adiposity, micronutrient supplementation and genotype.* British Journal of Nutrition, 2009. **101**(9): p. 1316.
36. Mendez, P., M. Taron, T. Moran, M.A. Fernandez, G. Requena, and R. Rosell, *A modified host-cell reactivation assay to quantify DNA repair capacity in cryopreserved peripheral lymphocytes.* DNA Repair, 2011. **10**(6): p. 603-610.

Chapter VII: Discussion

Table of Contents

Key Concepts and Conclusions

The main accomplishment of the present work carried out by a team of researchers assembled in the Samson lab has been the development of novel tools for the *in vivo* assessment of DNA repair capacity of up to 6 major DNA repair pathways and 10 different DNA lesions, 5 of which were developed by me. Such methods have allowed for the measurement of inter-individual differences in DNA repair capacity (DRC) that could help provide a better understanding of disease prevention and treatment. The focus of this thesis has been on a subset of these lesions, namely those repaired by the base excision repair pathway (BER). BER is characterized by the recognition and repair of small non-bulky DNA lesions that are products of DNA base oxidation, deamination and alkylation. Importantly, deficiencies in the repair of some of these lesions have been correlated with increased risk of cancer development as well as with neurodegenerative disease and immunodeficiency [1-4].

We have chosen to develop functional *in vivo* methods for the assessment of DRC in order to integrate the different levels of complexity that take place in the cellular environment. Accordingly, the information contained in genes, their methylation status and expression, translation rate, post-translational modifications, sub-cellular localization and protein-protein interactions are all represented within the functional assay readout. Current functional methods used to measure BER though effective, are based on the use of protein cell extracts which have as a caveat that protein compartmentalization, physiological reaction conditions and single-cell resolution are immediately lost. Moreover, they are in general cumbersome and time consuming and involve the use of radioactive labeling.

The method developed throughout this work is now published and is known as the fluorescence-based multiplex flow-cytometric host cell reactivation assay (FM-HCR) [5]. It is an adaptation of the original HCR method used for the

measurement of nucleotide excision repair (NER), which was based on the capacity of a host cell to repair a transiently transfected UV-irradiated plasmid reporter (originally containing a CAT gene, chloramphenicol acetyltransferase) [6]. Our novel method has the capacity to measure simultaneously and in a high-throughput manner, the ability of mammalian cells to repair plasmid reporters each bearing different doses of the same type of DNA damage or different types of DNA damage. This multiplexing is achieved by the use of a combination of 5 fluorescent plasmid reporters that have minimally overlapping excitation and emission spectra and can be measured through flow cytometry. Importantly, as fluorescence is used as readout, DNA repair capacity can be measured *in vivo*, without the need for cell lysis. We have shown that FM-HCR can be used to simultaneously measure repair capacity in any four of the following pathways: NER, MMR (mismatch repair), BER, NHEJ (non homologous end joining), HR (homologous recombination) and the direct reversal protein, MGMT.

Besides the lesion spectrum on which it acts, the BER pathway differs from other DNA repair pathways in that it can be initiated by a variety of DNA glycosylases, specifically 11 of them in human cells; some DNA glycosylases have a broad lesion recognition spectrum and show some degree of redundancy. The choice of DNA glycosylase to initiate the pathway is largely determined by the lesion to be repaired. This implies that using a single repair reporter bearing one of these lesions wouldn't properly represent the complexity of BER as a whole. To address this issue we have adapted the FM-HCR assay to simultaneously report on the activity of at least four different DNA glycosylases and their immediately downstream enzyme APE1. The biggest challenge in development of FM-HCR substrates for BER was that, for the most part, lesions repaired by this pathway are characterized by being small and non-bulky, which generally implies that they don't block transcription. This problem was overcome by taking advantage of the transcriptional mutagenesis properties of these lesions; many of which were previously unknown (transcriptional mutagenesis will be discussed later)

An important consideration for the BER substrates is that the repair of each lesion represents the action of several redundant BER proteins or even other DNA repair pathways that can act on it. As an example, even though the 8oxoG substrate was validated by the use of *Ogg1*$^{-/-}$ MEFs, the 8oxoG-containing plasmid is actually reporting on the overall cellular capacity to repair 8oxoG. This not only includes the activity of OGG1, but also on that of NEIL1, NEIL2 and AAG [7-10]. Similarly, uracil found in DNA can be repaired by UNG, SMUG1, MBD4 and TDG as well as by the MMR machinery [11]. Abasic sites are primarily recognized and incised by APE1 but APE2 can also act on the lesion [12]. MUTYH mainly removes adenines across *8oxoG* in the BER pathway, but MMR can also play a role in the repair of this base pair [13]. Finally, hypoxanthine repair by AAG can be considered as a unique case, as no other DNA glycosylases can repair this lesion. As such hypoxanthine repair can be considered as a proxy for AAG activity.

The capability to assess DRC in multiple DNA repair pathways simultaneously, *in vivo*, in a high-throughput manner is unprecedented in the DNA repair field. By using combinations of the five BER substrates and those for NER, MMR, MGMT, NHEJ and HR we demonstrated the potential applications of FM-HCR for population studies by measuring inter-individual DRC differences in a panel of 24 lymphoblastoid cell lines derived from apparently healthy individuals. We observed a range of DRC for each different lesion, which for the most cases agreed with previously reported measurements of inter-individual variation. Moreover, each cell line turned out to display a unique DRC fingerprint, highlighting the need of measuring multiple pathways to properly inform on an individual's DRC. Based on the positive results obtained from this work on cell lines, we performed a pilot population study for which PBMCs from 56 apparently healthy individuals were isolated and transfected with our 10 DNA repair reporters to assess inter- and intra-individual variability in all major DNA repair pathways in the human population. Overall, inter-individual variability measured by our method agreed with previously reported data, with the main difference that

we performed these measurements simultaneously, and that inter-individual variability for some of the pathways or lesions we measured had, to our knowledge, never been assessed before. Nevertheless, the pilot study brought to our attention several key challenges that included: **(a)** the need for multiple researchers to work in parallel, **(b)** the necessity to produce, aliquote and store at a relatively large scale, almost every single source of potential experimental variation (i.e. plasmid cocktails, serum and mitogens for T-lymphocyte stimulation) and **(c)** time-associated drifts in DRC measurements and the need to find a proper source of normalization for its correction. These experiences will help outperform this study in the future.

Transcriptional mutagenesis considerations

Transcriptional mutagenesis is the process by which a DNA lesion is bypassed by an RNA polymerase through the incorporation of the incorrect ribonucleotide into the nascent transcript. A major breakthrough in the development of many of the DNA reporters presented in this work, particularly for lesions repaired by BER, was the utilization of transcriptional mutagenesis properties of DNA lesions to measure their repair. This made it possible to overcome the restrictions posed by the traditional HCR assay for which transcription blocking lesions are needed.

An added advantage to the methods developed in the present work is that they provide an easily adaptable framework for the testing and discovery of transcriptional mutagenesis events for any site-specific lesions, *in vivo*. Notably, we have engineered a set of non-fluorescent reporter variants each of which can express a fluorescent protein with different excitation emission spectra upon occurrence of each possible mutagenesis event (A, C, G, U); thus allowing for their simultaneous analysis *in vivo*. These methods allowed us to take advantage of known transcriptional mutagenesis properties of O^6-meG and 8oxoG but also to show for the first time, to our knowledge, the *in vivo* transcriptional

mutagenesis properties in mammalian cells of *hypoxanthine*, *uracil, ε-cytosine* and *abasic sites.*

The fact that endogenously formed DNA lesions, many of which are repaired through BER, are potential *in vivo* sources of transcriptional mutagenesis has gained increased interest over the past decade [14, 15]. Taking into consideration the generation on average, on a daily basis per cell, of about 200 O^6-meGs, 10 hypoxanthines, 200 ε-cytosines, 500 uracils, 1,500 8oxoGs and 10,000 abasic sites [16, 17] then, the conceivable sources of *in vivo* transcriptional mutagenesis events are not negligible. Indeed, if one considers that each transcript can give rise to hundreds of proteins [18] then, an unrepaired lesion that caused transcriptional mutagenesis could result in an increased number of mutated transcripts that can consequently give rise to a large number of mutated proteins (**Figure 7.1**, [14]). These mutated transcripts/proteins could alter cellular physiology in such a way that transient cellular phenotypes different from that in the absence of the lesion could be favored (**Figure 7.2**, [14]). An extreme but plausible example of this series of events could be the transient expression of otherwise unwanted proliferation signals that would favor mitosis in non-mitotic cells. Moreover, if the original lesion that initiated this transient phenotype is not repaired by the time the cells divide, the lesion could now cause a permanent point mutation through a round of replication [14].

Transcriptional mutagenesis of site-specific 8oxoG to adenine of a transfected HRAS oncogene reporter results in the expression of a constitutively active (dominant) HRAS-Q61K mutant protein whose effect is reflected in activation of the different components of the MAPK pathway, including ERK phosphorylation. Importantly, ERK activation could only be readily detected when this reporter was transfected into *Ogg1*$^{-/-}$ and not WT MEFs [19]. Even though it is unknown if these cells would have the capacity to form tumors in xenografts, it is a clear evidence that deficiencies in DRC of lesions that cause transcriptional

mutagenesis can lead to the expression of mutated proteins that can potentially result in cellular transformation.

Altogether, the present work further extends the implications of transcriptional mutagenesis in three important ways: **(a)** by the elucidation of new *in vivo* transcriptional mutagenesis properties of common endogenously formed DNA lesions (hypoxanthine, uracil and abasic sites) **(b)** by the development of tools to study the transcriptional mutagenesis properties of virtually any lesion that can be site-specifically positioned into a plasmid reporter and **(c)** by providing methods to screen for individuals that could potentially be at higher disease risk based on their DRC of lesions that cause transcriptional mutagenesis.

HCR-Seq Considerations

In this next-generation sequencing era, RNA-Seq provides an extra layer of complexity and multiplexing for the study of lesions that cause transcriptional blockage and/or transcriptional mutagenesis. Because next-generation sequencing can be used to quantitate the expression levels of thousands of transcripts simultaneously, our assay has the potential to measure expression of dozens of reporters for multiple individuals in a single sequencing lane; this would make characterization of global DRC in large populations both efficient and affordable. Moreover, we have shown here that HCR-Seq has the potential to reveal new biological phenomena in the basic research setting. Because the plasmids are not replicated in the cell, and transcript sequence changes were observed at an elevated rate in repair deficient cells, these changes are likely to reflect transcriptional mutagenesis events due to unrepaired DNA lesions in the transcribed DNA strand. In particular the Samson lab previously showed that RNAPII can sometimes bypass, *in vivo* T^T (thymine dimers) [5]. Here we report bypass at εC (etheno-cytosines). The identities of the misincorporations were for the first time measured in an *in vivo* setting for both T^T (GA>AT(AU)>CG=AG) [5] and εC (A>T(U)>C). Coincidently the bypass frequency for both lesions was

almost 10 %. These frequencies are likely underestimates, given the expectation that repaired reporter plasmids may be transcribed at a higher rate, and because error-free bypass cannot be distinguished from transcripts arising from repaired plasmid.

Modeling cellular responses to DNA damaging agents based on DRC measurements.

Tumor resistance to chemotherapy and radiotherapy are unfortunately not rare events for cancer patients [20, 21]. The ability to determine ahead of time whether a particular treatment would be effective would improve therapy and avoid putting patients through unnecessary exposure to toxic agents. Increasing amount of evidence indicates that DRC can be an important factor in cancer treatment [22]. Consequently, if one were able to determine which agents a tumor would be susceptible to (by virtue of its impaired DRC), while minimizing toxicity for a patient's healthy tissues (by virtue of more proficient DRC compared to that of the tumor), therapy could be personalized in order increase efficacy and minimize side effects, decreasing the risk of secondary as a consequence of treatment [23-25].

To try to address some of these interactions, we built mathematical models by performing multiple linear regression (MLR) analysis with the goal of determining the contribution of our DRC measurements to cellular sensitivity to DNA damaging agents. One of the most striking results of our models was the pivotal role that 8oxoG repair seems to play in resistance to alkylation treatment with MMS and to a lesser degree to 5-FU. The models indicate that an increase in 8oxoG repair capacity contributes in resistance towards these agents, implying unexpectedly that an oxidative stress response might be the driving force of the treatment. Given the potential clinical implications that such mathematical models can have, exhaustive model validation as well as different modeling techniques should take place. Preliminary alternative modeling techniques such as PLSR

(Partial Least Square Regression), which is very similar to PCA (Principal Component Analysis) but considers the response variables in order to build principal component vectors, have resulted in comparable trends to the ones showed with MLR. Specifically with 8oxoG and abasic site repair appearing as the strongest contributors for the 1st and 2nd principal components (data not shown).

Leave-one-out cross-validation was used as a validation technique for these models; nevertheless corroboration of the model predictions by other means would further support the model outcome. Two key experiments that are currently ongoing and are anticipated to contribute in validation of the model are as follows: **(a)** overexpression of OGG1 in cells that are sensitive to MMS and report low 8oxoG repair activity, our expectation if the mathematical model is correct, is that this overexpression will render cells more resistant to MMS treatment; and **(b)** to use the multiple linear regression models built for the panel of 24 B-lymphoblastoid cells to predict sensitivity to the same DNA damaging agents in a new panel of cells lines for which we have never measured sensitivity before.

Even though it would be tempting to assume these models would apply for any cell type of interest, we must remain cautious when attempting to extrapolate these results and predictions to a context different than human B-lymphoblastoid cells lines. Similar analyses and validation in other cells lines will elucidate whether these predictions are conserved in a variety of cellular contexts. Moreover, models that integrate several different chemo and radio resistance mechanisms besides variation in DRC, such as changes in DNA damage signaling [26], damage detoxification mechanisms [27], and drug efflux pumps will likely help improve the accuracy and prediction capacity of the models with the ultimate goal of personalizing and enhancing cancer treatment.

PBMCs as surrogates for an individual's DRC

Peripheral Blood Mononuclear Cells (PBMCs) have been widely used to determine an individual's DRC. A longstanding question in the field is whether these cells do indeed represent the overall DRC of a person. The main reason for using blood cells for these an many other population studies has mainly to do with how easily accessible blood is and how non-invasive the procedure is.

There is unequivocal evidence that DRC can vary in different tissues [28-30]. For example MGMT activity varies almost a 100-fold within human tissues (**Figure 7.3**). However, determining if ranked inter-individual differences in DRC measured in PBMCs are preserved regardless of the tissue studied is still an open question. Nevertheless, experimental evidence shows that 8oxoG repair capacity is conserved between PBMCs and healthy lung tissue of 7 lung cancer patients (R^2 = 0.858) [31]. Moreover, our work shows that NER capacity is also maintained between B-lymphoblastoid cells and fibroblasts derived from the same people (R^2 = 0.94) [5]. These are of course only two positive relationships between two cell types in two different DNA repair pathways. Future efforts should focus on performing side-by-side comparison of inter-individual DRC in different tissues in all major repair pathways, as it is also possible that PBMCs could work as a surrogate for DRC in some pathways but not all of them.

A main caveat of the HCR method (in any of its formats, classic HCR, FM-HCR or HCR-Seq) is that the cells being tested need to be transfected and subsequently incubated for roughly a day in order to measure DRC. As such, any primary cells that cannot be isolated as single cells or that are not amenable for tissue culture or transfection are not compatible with this method. It is also important to point out that PBMCs used for the HCR assay require previous stimulation with a mitogenic agent, usually PHA for T-lymphocyte stimulation. In contrast, most other functional DRC assays such the comet assay or assays based on *in vitro* protein cell extracts generally use non-stimulated PBMCs, and

thus measure DRC of a mixture of T-lymphocytes, B-lymphocytes, NK cells and monocytes, many of which are quiescent.

A very interesting and feasible possibility for future DRC studies that would help overcome many of the drawbacks and caveats of using PBMCS is the use of iPSCs (induced pluripotent stem cells) (**Figure 7.4**). iPSCs would allow one to test in a minimally invasive manner the DRC of virtually any cell type that the iPSCs can be differentiated into. Moreover, by virtue of all the cell types being completely isogenic (as they are all differentiated from the same iPSCs, that are derived from one individual) any confounding factors that could arise from the use of samples from different donors are minimized. Importantly, iPSC can also act as ideal disease models to test the potential role of DNA repair capacity in cells that wouldn't otherwise be available from patients with a variety of diseases [32].

Figures

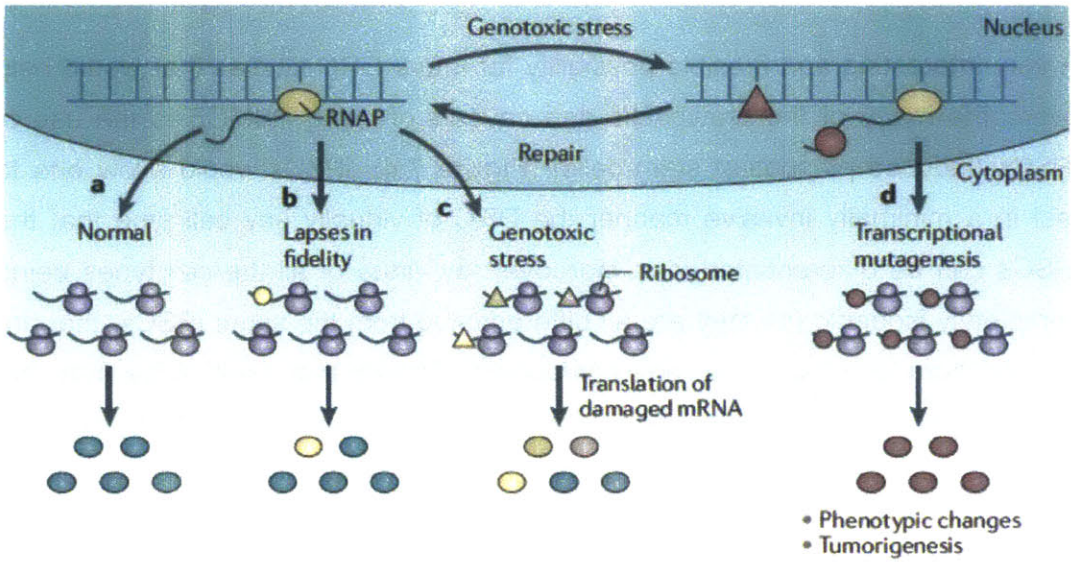


Figure 7.1 DNA replication-independent production of erroneous proteins (From Brégeon and Doetsch 2011).

Under normal conditions (a), transcription in the nucleus produces error-free mRNAs that are translated by ribosomes to normal proteins (blue ovals) in the cytoplasm. In some cases (b), lapses in RNA polymerase (RNAP) fidelity can generate aberrant transcripts (yellow circle) that are translated into erroneous proteins (yellow oval). This random, low-frequency production of erroneous proteins can also be caused by lapses in ribosome fidelity. When exposed to a genotoxic agent (c), RNA molecules in a cell may contain various lesions (triangles) that could induce the production of erroneous proteins during translation because of their potentially altered codon–anticodon pairing during tRNA selection. DNA is the other target for genotoxic stress (d). RNAP can bypass numerous unrepaired damaged deoxyribonucleotides on the transcribed strand of a gene (red triangle) that can result in misincorporation events in the transcript sequence (red circles) as long as the DNA damage is not removed by one of the DNA repair pathways. Transcriptional mutagenesis results in the

production of a mostly homogenous mutant transcript population, which in turn leads to the production of high levels of erroneous proteins that possess the same mutant sequence, and that could alter the phenotype of the cell [14].

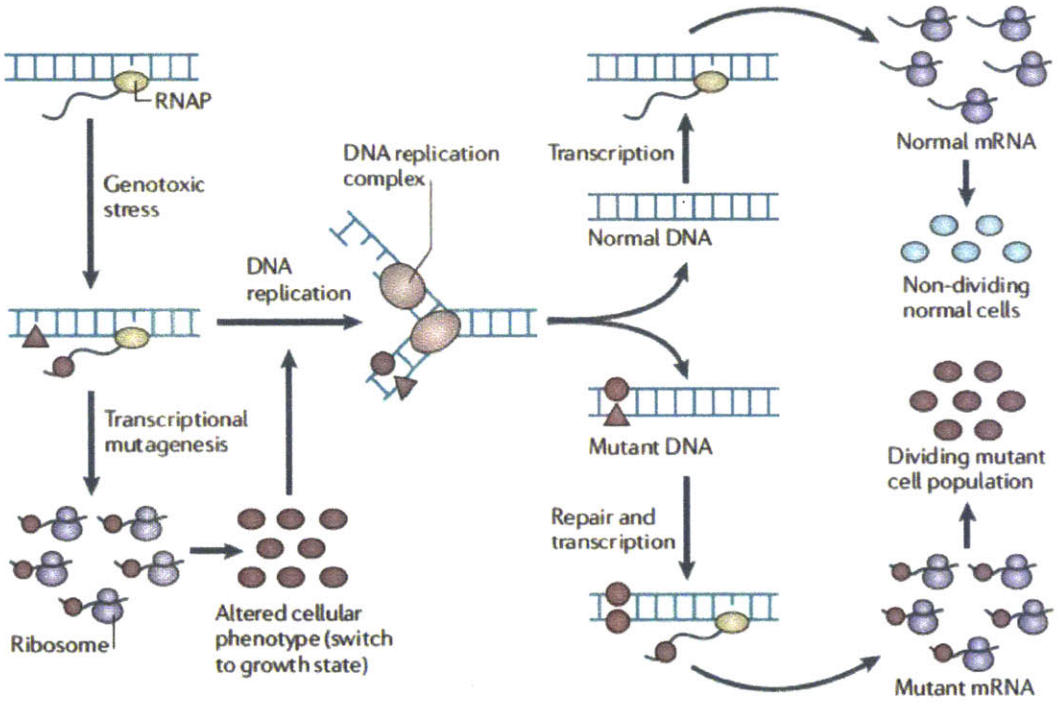


Figure 7.2 The potential role of transcriptional mutagenesis in tumour development (From Brégeon and Doetsch 2011).

Following genotoxic stress, a DNA lesion (red triangle) can appear on the transcribed strand of a gene, resulting in the production of high levels of erroneous protein by transcriptional mutagenesis The resulting mutant proteins (red ovals) may have the ability to alter the phenotype of the cell in such a way that a growth advantage is conferred, leading to initiation of DNA replication. If left unrepaired, the DNA lesion will subsequently be encountered by the replication machinery and will probably cause similar miscoding during DNA synthesis, which will result in the fixation of the mutation into the genome of one of the replicated progeny (red circle). Subsequent rounds of replication of the mutated chromosome could lead to tumour development. RNAP, RNA polymerase [14].

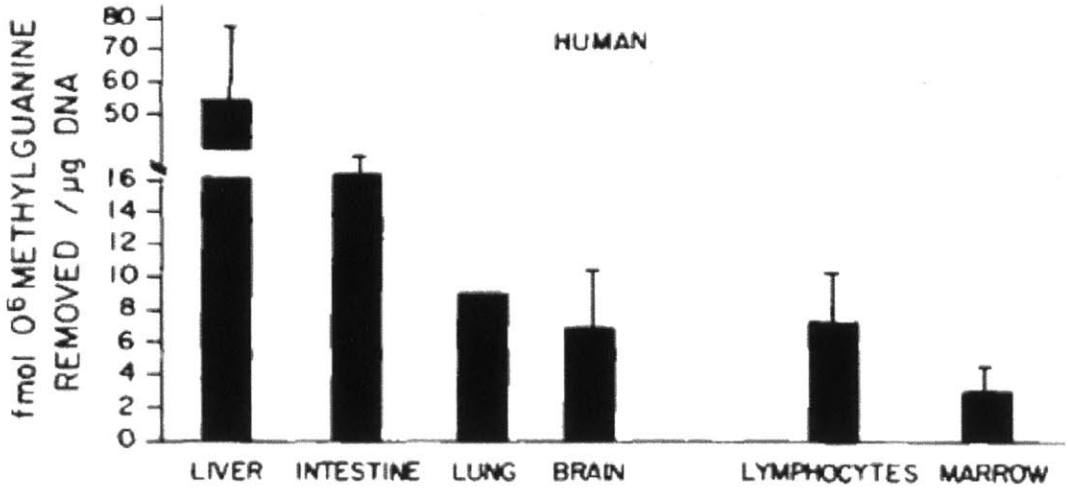


Figure 7.3 MGMT activity varies in human tissues.

From Gerson *et al* 1986 [30].

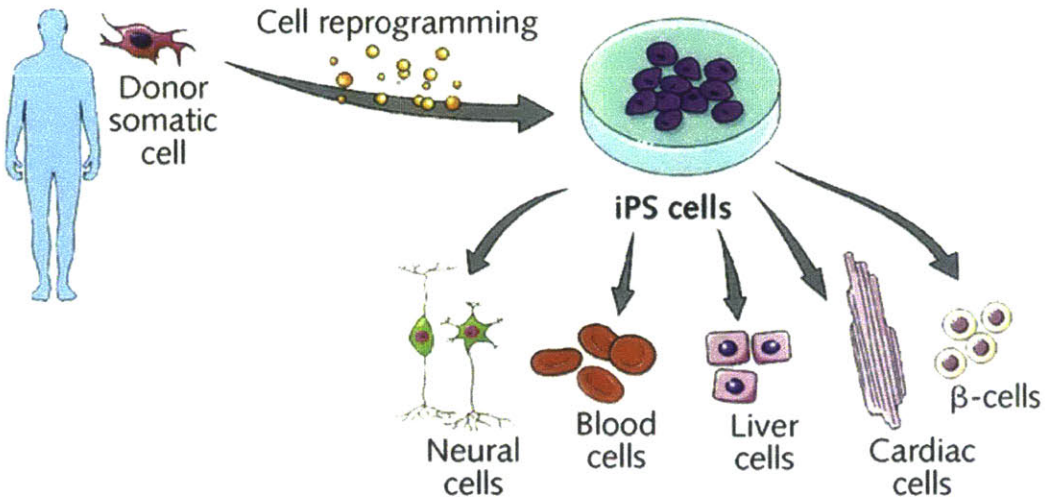


Figure 7.4 Derivation and use of iPS Cells.
Modified from Power and Rasko 2011 [32].

References

1. Imai, K., G. Slupphaug, W.-I. Lee, P. Revy, S. Nonoyama, N. Catalan, L. Yel, M. Forveille, B. Kavli, H.E. Krokan, H.D. Ochs, A. Fischer, and A. Durandy, *Human uracil-DNA glycosylase deficiency associated with profoundly impaired immunoglobulin class-switch recombination.* Nature immunology, 2003. **4**(10): p. 1023-1028.
2. Jones, S., P. Emmerson, J. Maynard, J.M. Best, S. Jordan, G.T. Williams, J.R. Sampson, and J.P. Cheadle, *Biallelic germline mutations in MYH predispose to multiple colorectal adenoma and somatic G: C→ T: A mutations.* Human Molecular Genetics, 2002. **11**(23): p. 2961-2967.
3. Shao, C., S. Xiong, G.-M. Li, L. Gu, G. Mao, W.R. Markesbery, and M.A. Lovell, *Altered 8-oxoguanine glycosylase in mild cognitive impairment and late-stage Alzheimer's disease brain.* Free Radical Biology and Medicine, 2008. **45**(6): p. 813-819.
4. Sevilya, Z., Y. Leitner-Dagan, M. Pinchev, R. Kremer, D. Elinger, H.S. Rennert, E. Schechtman, L.S. Freedman, G. Rennert, and T. Paz-Elizur, *Low integrated DNA repair score and lung cancer risk.* Cancer Prevention Research, 2013: p. canprevres. 0318.2013.
5. Nagel, Z.D., C.M. Margulies, I.A. Chaim, S.K. McRee, P. Mazzucato, A. Ahmad, R.P. Abo, V.L. Butty, A.L. Forget, and L.D. Samson, *Multiplexed DNA repair assays for multiple lesions and multiple doses via transcription inhibition and transcriptional mutagenesis.* Proceedings of the National Academy of Sciences, 2014. **111**(18): p. E1823-E1832.
6. Athas, W.F., M.A. Hedayati, G.M. Matanoski, E.R. Farmer, and L. Grossman, *Development and field-test validation of an assay for DNA repair in circulating human lymphocytes.* Cancer Res, 1991. **51**(21): p. 5786-93.
7. Lee, C.Y.I., J.C. Delaney, M. Kartalou, G.M. Lingaraju, A. Maor-Shoshani, J.M. Essigmann, and L.D. Samson, *Recognition and Processing of a New Repertoire of DNA Substrates by Human 3-Methyladenine DNA Glycosylase (AAG).* Biochemistry, 2009. **48**(9): p. 1850-1861.
8. Monden, Y., T. Arai, M. Asano, E. Ohtsuka, H. Aburatani, and S. Nishimura, *Human MMH (OGG1) type 1a protein is a major enzyme for repair of 8-hydroxyguanine lesions in human cells.* Biochemical and biophysical research communications, 1999. **258**(3): p. 605-610.
9. Dou, H., S. Mitra, and T.K. Hazra, *Repair of oxidized bases in DNA bubble structures by human DNA glycosylases NEIL1 and NEIL2.* Journal of Biological Chemistry, 2003. **278**(50): p. 49679-49684.
10. Bessho, T., R. Roy, K. Yamamoto, H. Kasai, S. Nishimura, K. Tano, and S. Mitra, *Repair of 8-hydroxyguanine in DNA by mammalian N-methylpurine-DNA glycosylase.* Proceedings of the National Academy of Sciences, 1993. **90**(19): p. 8901-8904.
11. Wilson, T.M., A. Vaisman, S.A. Martomo, P. Sullivan, L. Lan, F. Hanaoka, A. Yasui, R. Woodgate, and P.J. Gearhart, *MSH2–MSH6 stimulates DNA*

polymerase η, suggesting a role for A: T mutations in antibody genes. The Journal of experimental medicine, 2005. **201**(4): p. 637-645.
12. Hadi, M.Z. and D.M. Wilson, *Second human protein with homology to the Escherichia coli abasic endonuclease exonuclease III.* Environmental and molecular mutagenesis, 2000. **36**(4): p. 312-324.
13. Fortini, P., B. Pascucci, E. Parlanti, M. D'errico, V. Simonelli, and E. Dogliotti, *8-Oxoguanine DNA damage: at the crossroad of alternative repair pathways.* Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis, 2003. **531**(1): p. 127-139.
14. Brégeon, D. and P.W. Doetsch, *Transcriptional mutagenesis: causes and involvement in tumour development.* Nature Reviews Cancer, 2011. **11**(3): p. 218-227.
15. Dreij, K., J.A. Burns, A. Dimitri, L. Nirenstein, T. Noujnykh, and D.A. Scicchitano, *DNA damage and transcription elongation: Consequences and RNA integrity.* The Chemical Biology of DNA Damage, 2010: p. 399-437.
16. Ciccia, A. and S.J. Elledge, *The DNA damage response: making it safe to play with knives.* Molecular cell, 2010. **40**(2): p. 179-204.
17. Kim, Y.-J. and D.M. Wilson III, *Overview of base excision repair biochemistry.* Current molecular pharmacology, 2012. **5**(1): p. 3.
18. Csárdi, G., A. Franks, D.S. Choi, E.M. Airoldi, and D.A. Drummond, *Accounting for experimental noise reveals that mRNA levels, amplified by post-transcriptional processes, largely determine steady-state protein levels in yeast.* PLoS Genet, 2015. **11**(5): p. e1005206.
19. Saxowsky, T.T., K.L. Meadows, A. Klungland, and P.W. Doetsch, *8-Oxoguanine-mediated transcriptional mutagenesis causes Ras activation in mammalian cells.* Proceedings of the National Academy of Sciences, 2008. **105**(48): p. 18877-18882.
20. Coley, H.M., *Mechanisms and strategies to overcome chemotherapy resistance in metastatic breast cancer.* Cancer treatment reviews, 2008. **34**(4): p. 378-390.
21. Frosina, G., *DNA repair and resistance of gliomas to chemotherapy and radiotherapy.* Molecular Cancer Research, 2009. **7**(7): p. 989-999.
22. Jalal, S., J.N. Earley, and J.J. Turchi, *DNA repair: From genome maintenance to biomarker and therapeutic target.* Clinical Cancer Research, 2011.
23. Leone, G., M.T. Voso, S. Sica, R. Morosetti, and L. Pagano, *Therapy related leukemias: susceptibility, prevention and treatment.* Leukemia & lymphoma, 2001. **41**(3-4): p. 255-276.
24. Oeffinger, K.C. and S. Bhatia, *Second primary cancers in survivors of childhood cancer.* The Lancet, 2009. **374**(9700): p. 1484-1485.
25. Wood, M.E., V. Vogel, A. Ng, L. Foxhall, P. Goodwin, and L.B. Travis, *Second malignant neoplasms: assessment and strategies for risk reduction.* Journal of Clinical Oncology, 2012. **30**(30): p. 3734-3745.
26. Hanahan, D. and R.A. Weinberg, *Hallmarks of Cancer: The Next Generation.* Cell, 2011. **144**(5): p. 646-674.
27. Rocha, C.R., C.C. Garcia, D.B. Vieira, A. Quinet, L.C. de Andrade-Lima, V. Munford, J.E. Belizario, and C.F. Menck, *Glutathione depletion sensitizes*

cisplatin- and temozolomide-resistant glioma cells in vitro and in vivo. Cell Death Dis, 2014. **5**: p. e1505.
28. Myrnes, B., K.-E. Giercksky, and H. Krokan, *Interindividual variation in the activity of O6-methyl guanine-DNA methyltransferase and uracil-DNA glycosylase in human organs.* Carcinogenesis, 1983. **4**(12): p. 1565-1568.
29. Christmann, M., B. Verbeek, W.P. Roos, and B. Kaina, *O 6-Methylguanine-DNA methyltransferase (MGMT) in normal tissues and tumors: enzyme activity, promoter methylation and immunohistochemistry.* Biochimica et Biophysica Acta (BBA)-Reviews on Cancer, 2011. **1816**(2): p. 179-190.
30. Gerson, S.L., J.E. Trey, K. Miller, and N.A. Berger, *Comparison of O6-alkylguanine-DNA alkyltransferase activity based on cellular DNA content in human, rat and mouse tissues.* Carcinogenesis, 1986. **7**(5): p. 745-749.
31. Paz-Elizur, T., M. Krupsky, S. Blumenstein, D. Elinger, E. Schechtman, and Z. Livneh, *DNA repair activity for oxidative damage and risk of lung cancer.* Journal of the National Cancer Institute, 2003. **95**(17): p. 1312-1319.
32. Power, C. and J.E. Rasko, *Will cell reprogramming resolve the embryonic stem cell controversy? A narrative review.* Annals of internal medicine, 2011. **155**(2): p. 114-121.