

Copyright

by

Catherine Michelle Rondelli

2013

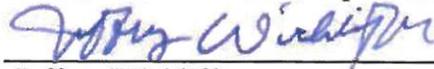
The Dissertation Committee for Catherine Michelle Rondelli Certifies that this is
the approved version of the following Dissertation:

**XPC Haplotypes Alter DNA Repair Capacity and Levels of Genetic
Damage**

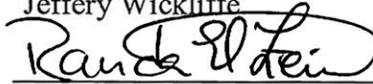
Committee:



Sherif Abdel-Rahman, Supervisor



Jeffery Wickliffe



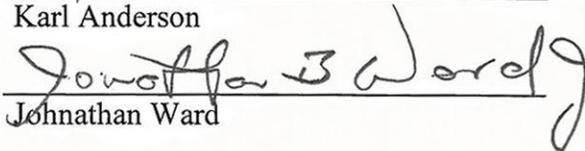
Randa El-Zein



José Barral



Karl Anderson



Johnathan Ward



Dean, Graduate School of Biomedical Sciences

***XPC* Haplotypes Alter DNA Repair Capacity and Levels of Genetic
Damage**

by

Catherine Michelle Rondelli, BS

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas Medical Branch

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas Medical Branch

July, 2013

Dedication

To all my family, friends, and mentors for their love, patience, and guidance. This could not have been done without you.

Robert Bell

Mike Rondelli

Theresa Rondelli

Mary Rondelli

Words cannot express my gratitude for your love and support.

Acknowledgements

Thanks to Dr. Carol Etzel and Dr. Courtney Cross for their biostatistics help and analysis, to Dr. Meixiang Xu for her molecular biology expertise, and Dr. Jeffery Witcliffe for his guidance in the development of the experiments presented. Thank you to Jordan Speidel, who has helped complete confirmatory experiments and the first sets of future experiments, and Dr. Ilona Nekhyava, who has helped edit papers generated by this work. Thank you to Dr. El-Zein, Syo Lopez, and Claudia Monroy, who helped create and unlock the secrets to confounding protocols. Thank you to the other lab members, Dr. Carla Kinslow and Mia Tolba for their insightful discussions. Thank you to Dr. Golda Leonard, Dr. Darren Boehning, and Lisa Davis who helped me through the requirements of both the graduate school and our program. Thank you to my committee members Dr. Sherif Abdel-Rahman, Dr. Jeffery Wickliffe, Dr. Jonathan Ward, Dr. Randa El-Zein, Dr. José Barrel, and Dr. Karl Anderson for their guidance and expertise. Thank you to Dr. William Ameredes for his guidance and training both in and out of the Toxicology Training grant, and extreme thanks to Dr. Sherif Abdel-Rahman for his mentorship, patience, and financial support.

This work was supported in part by NIEHS Predoctoral Toxicology Training Grant (T32-07454), NIH Ruth L Kirschstein NRSA NIEHS Predoctoral Fellowship (F31-ES019053), NIEHS Center award (ES006676), and NCI awards (P30-006676, CA093592).

***XPC* Haplotypes Alter DNA Repair Capacity and Levels of Genetic Damage**

Publication No. _____

Catherine Michelle Rondelli, PhD
The University of Texas Medical Branch, 2013

Supervisor: Sherif Abdel-Rahman

Xeroderma pigmentosum complementation group C (*XPC*) is the key recognition factor of DNA damage in global genome nucleotide excision repair (NER). The disease Xeroderma pigmentosum (XP) results from mutations leading to structural defects of the encoding gene and in some instances trace to changes in a single nucleotide. The *XPC* gene is highly polymorphic and while polymorphisms in general have no discernible phenotypic effects, some can alter the structure and function of the encoded protein. To date, the effect of single nucleotide polymorphisms (SNPs) in *XPC* have not been properly characterized. Documented associations exist between only a few *XPC* SNPs and cancer risk, leaving a majority of SNPs untested. ***My hypothesis is that specific XPC SNP combinations (haplotypes) alter DNA repair capacity and levels of genetic damage by altering transcriptional processes and/or protein function.*** I used bioinformatics to conduct a comprehensive haplotype analysis of the entire genomic sequence of *XPC* and characterize the effect of haplotypes on genetic damage in a population of smokers as an environmentally exposed population. All genomic region *XPC* polymorphisms with a minor allele frequency (MAF) ≥ 0.05 , from the HapMap CEPH population were analyzed

using PHASE, generating a series of likely phylogenetically clustered haplotypes. Cigarette smokers and matched non-smokers from a White, non-Hispanic population residing in the Houston-Galveston area were genotyped and recoded for these haplotype groups. Association between smoking status and DNA damage was determined using chromosomal aberrations as a biomarker. To characterize the biological effects of the *XPC* haplotypes, I determined how they affected DNA damage and repair capacity over time (i.e. the genotype/phenotype relationship) using representative cell lines. I evaluated the effect of these haplotypes on NER capacity using ELISA (Enzyme Linked Immunosorbent Assay) following exposure to ultraviolet (UV) radiation. I characterized the functional significance of *XPC* haplotypes by determining the effects of these haplotypes on transcriptional processing and stability using real-time analysis, and protein expression and stability with Western blot analysis. *I found that the haplotypes not only conferred differential repair capacity, but that they did so through uniquely different mechanisms.*

TABLE OF CONTENTS

List of Tables	ix
List of Figures	xiii
List of Abbreviations	xv
Chapter 1: INTRODUCTION.....	19
General Background	19
Smoking as a Genotoxicant	20
Single Nucleotide Polymorphisms and DNA Repair	21
Challenges in SNP Studies	25
Approaches for SNP Functionality Determination.....	27
Nucleotide Excision Repair Pathway	30
SNPs in <i>XPC</i> , Historically	35
Overview and Objectives.....	36
Chapter 2: MATERIALS AND METHODS	39
Overview.....	43
Study subjects and blood collection.....	43
Determination of background and mutagen-induced chromosomal aberrations	45
<i>XPC</i> haplotype-tagging SNPs and population genotyping	46
<i>XPC</i> haplotype inference and phylogenetic grouping	49
Establishment of cell culture for mechanistic studies.....	50
UV-B Exposure of cell culture for mechanistic studies	52
DNA isolation and quantification	54
ELISA standard preparation	55
Adduct determination by ELISA	56
Picogreen DNA damage Assay.....	57
RNA isolation	58
RNA Real-Time analysis	59
mFOLD RNA analysis.....	60
Protein isolation and analysis	61

<i>XPC</i> antibody preparation.....	61
Western Blot protein analysis	62
Statistical analysis.....	63
Chapter 3: RESULTS	67
Overview.....	67
Aim 1: Haplotype Determination and Associations	67
Aim 2: DRC Analysis.....	81
Aim 3: mRNA and Protein Analysis	98
Chapter 4: DISSCUSION AND CONCLUSIONS.....	108
Haplotype Analysis and CA Evaluation	108
Haplotype Affects on DRC.....	113
RNA and Protein Analysis.....	120
Conclusion	126
SUPPLEMENTAL DATA	128
REFERENCES	157
Curriculum Vita	184
Other Required Information.....	192

List of Tables

Table 1: Overview of SNP literature for DNA repair genes.....	22
Table: Chemicals	39
Table: Special equipment and software	42
Table 2: Coriell Institute Biorepository cell lines	51
Table 3: HapMap breakdown of <i>XPC</i> SNPs.....	68
Table 4: CEU inferred haplotypes for the <i>XPC</i> gene as determined by PHASE analysis.....	70
Table 5: The percent genetic divergence between the 6 clades	73
Table 6: CEU inferred haplotypes for the <i>XPC</i> gene as determined by PHASE analysis and clustered by haplotype.....	73
Table 7: Demographic breakdown of the UTMB experimental population	77
Table 8: Percent genetic divergence between haplotypes comprising each PGH	83
Table 9: Measure of central tendency for CPDs and 6-4PPs at intervals of repair after UV-B exposure	88
Table 10: Measure of significance for CPDs and 6-4PPs over the time course of repair after UV-B exposure	93

Table 11: Measure of significance for CPDs and 6,4-PPs over the time course of repair after UV-B exposure of the combined PGHs	94
Table 12: Measure of significance for total damage over the time course of repair after UV-B exposure of the combined PGHs	97
Table 13: Fold change over the time course of repair after UV-B exposure of the combined PGHs	99
Table 14: The mFOLD free energies for each virtual segment of each haplotype	102
Table 15: Amount of XPC protein over the time course of repair after UV-B exposure of the combined PGHs	106
Table 16: Statistical significance in the amount of XPC protein over the time course of repair after UV-B exposure of the combined PGHs.....	107
Table 17: Haplotype SNP (htSNP) analysis by tagger software.....	128
Table 18: Genotyping of the htSNPs alleles by study participant for the UTMB White non-Hispanic cohort of the experimental population	129
Table 19: Assignment of haplotype by study participant for the UTMB White non-Hispanic cohort of the experimental population.....	132
Table 20: Raw CA stats by haplotype for both baseline and 24 hours after in vitro NNK treatment of isolated primary lymphocytes for the UTMB White non-Hispanic cohort of the experimental population	135
Table 21: Percent genetic divergence between all 21 haplotypes as determined by PHASE analysis	140

Table 22: Raw growth counts for each of the haplotype lines over time	141
Table 23: Raw viability data for each of the haplotype lines over time	142
Table 24: Shapiro-Wilk normality data for each of the haplotypes over time by adduct	142
Table 25: Mann-Whitney test of significance data for each of the haplotypes over time by adduct.....	143
Table 26: Mann-Whitney test of significance data for the haplotypes at each time by adduct.....	146
Table 27: Mann-Whitney test of significance data of grouped sensitive verses insensitive haplotype by adduct.....	148
Table 28: Raw data for picogreen analysis of the haplotypes by time after UV-B exposure	149
Table 29: Shapiro-Wilk normality data for each of the haplotypes over time for fold induction of <i>XPC</i> mRNA by real time analysis	150
Table 30: Mann-Whitney test of significance data for each of the haplotypes over time for fold induction of <i>XPC</i> mRNA by real time analysis.....	150
Table 31: Mann-Whitney test of significance data for the haplotypes at each time for fold induction of <i>XPC</i> mRNA by real time analysis	152
Table 32: T-test analysis of significance data for each of the haplotypes over time for <i>XPC</i> protein by Western blot analysis	153

Table 33: T-test analysis of significance data for the haplotypes at each time for XPC protein by Western blot analysis.....	155
---	-----

List of Figures

Figure 1: DNA damage structures	32
Figure 2: Nucleotide Excision Repair.....	34
Figure 3: <i>XPC</i> gene.....	36
Figure 4: Endpoints after UV-induction	54
Figure 5: Phylogenetic analysis of the 21 haplotypes.....	72
Figure 6: CA frequency divided by clade and smoking status at baseline	79
Figure 7: CA frequency divided by clade and smoking status after mutagen exposure	80
Figure 8: Cell Viability after UV-B treatment.....	84
Figure 9: Cell growth after UV-B treatment.....	85
Figure 10: Test of antibody specificity	87
Figure 11: Amount of adduct remaining over time after UV-B exposure.....	92
Figure 13: Amount of DNA damage remaining over time after UV-B exposure	97
Figure 15: MFOLD structures for virtual segments	103
Figure 16: Arbitrary densitometry units verses β -actin over time after UV-B exposure	106

Figure 17: Comparison of CA data by clade137

Figure 18: Percent viability for each of the haplotype lines over time141

List of Abbreviations

%D	percent genetic divergence
6,4-PP	6,4-pyrimidine-pyrimidone photoproduct
BCA	bicinchoninic acid
BER	base excision repair
BPDE	benzo(a)pyrene diolepoxide
BSA	bovine serum albumin
CA	chromosome aberration
CENT2	centrin 2
CEPH	Centre d'Etude du Polymorphisme Humain
CEU	Utah population of CEPH collection
CPD	cyclopyrimidine dimer
CRISP	Computer Retrieval of Information on Scientific Projects
CSA	cockayne syndrome a
CSB	cockayne syndrome b
Ct	cycle threshold
DDB2	damage-specific DNA binding protein 2
DMSO	dimethylsulfoxide
DNA	dioxyribonucleic acid
DRC	DNA repair capacity
ECL	enhanced chemilumescence
EDTA	ethylenediaminetetraacetic acid
ELISA	enzyme linked immunosorbent assay

FBS	fetal bovine serum
FRET	Förster resonance energy transfer
HCl	hydrochloric acid
hHR23B	human radiation-sensitive(RAD)23 homolog
HR	homologous recombination
htSNP	haplotype tagging SNP
HWE	Hardy-Weinberg equilibrium
IQR	inner quartile range
IRB	internal review board
LD	linkage disequilibrium
LOD	logarithm of odds
MAF	minor allele frequency
MEGA	molecular evolutionary genetics analysis
miRNA	micro RNA
MMR	mismatch repair
MS	mutagen sensitivity
NEHJ	non-homologous end joining
NER	nucleotide excision repair
NIH RePORT	National Institutes of Health Research Portfolio Online Reporting Tools
NNK	4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone
PAGE	polyacrylamide gel electrophoresis
PAH	polyaromatic hydrocarbon

PBL	peripheral blood lymphocyte
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PGH	phylogenetically grouped haplotype
PHA	phytohemagglutinin
PMSF	phenylmethanesulfonylfluoride
PVDF	polyvinylidene fluoride
RIPA	radio-immunoprecipitation assay
RNA	ribonucleic acid
RPA	replication protein A
RPMI	Roswell Park Memorial Institute
RS	reference SNP
SDS	sodium dodecyl sulfate
SEM	standard error of mean
SNP	single nucleotide polymorphism
TBS	tris buffered saline
TBS-T	tris buffered saline with tween
TE	tris-EDTA
TFIIH	transcription factor II human/helicase
TMB	3,3',5,5'-tetramethylbenzidine
Tris	Tris(hydrozomethyl)aminomethane
UTMB	University of Texas Medical Branch
UV	ultraviolet radiation

XP	xeroderma pigmentosum
XPA	xeroderma pigmentosum complementation group A
XPC	xeroderma pigmentosum complementation group C
XPB	xeroderma pigmentosum complementation group D
XPG	xeroderma pigmentosum complementation group G
XRCC1	X-ray repair cross-complementing protein 1
XRCC3	X-ray repair cross-complementing protein 3

Chapter 1: INTRODUCTION

GENERAL BACKGROUND

Maintaining the integrity of cellular DNA is critical to the life cycle of the cell. As the blueprint for creating proteins and regulating their function, the information needs to remain readable and uncorrupted. Failure to do so can result in a number of potentially harmful changes such as mutations, dysregulation, or genomic instability. Mutations can cause misread or total loss of a cellular component such as a protein, or it can be misregulated such as misdirecting the protein to the wrong cellular compartment. Dysfunction can result from changes to regulatory sites such as promoter regions or regulatory components such as miRNAs or their target sequences. Genome instability results in poorly replicated DNA, leading to broken or mutated daughter strands that are passed on to new cells. These lesions have the potential to ultimately lead to the development of cancer (Jackson and Bartek, 2009; Nowsheen and Yang, 2012).

Cells are exposed to several thousand DNA damaging agents daily, from both endogenous and exogenous sources (Jackson and Bartek, 2009; Nowsheen and Yang, 2012). For example, endogenous byproducts of cellular respiration, such as hydrogen peroxide and superoxide, are highly reactive and can damage the DNA molecule directly resulting in damaged bases (e.g. 8-hydroxydeoxyguanosine) (Nohl et al., 2003; Nowsheen and Yang, 2012) or indirectly (e.g. via acrolein accumulation, which itself is a genotoxicant) (Wang et al., 2012). Exogenously, there are a wide range of chemicals and physical agents that can damage DNA through a number of mechanisms. For example, both car exhaust and smoking produce the aforementioned acrolein as well as other compounds such as polycyclic aromatic hydrocarbons (PAH) (Wang et al., 2012, 2007). Diet can also be a source of mutagens through consumption, not only of dietary contaminants such as bioaccumulated pesticides (Alavanja et al., 2012; Aylward et al.,

2013) and heavy metals (Tchounwou et al., 2012), but of natural byproducts such as PAHs found in charbroiled meats (Chien and Yeh, 2010; Hoelzl et al., 2008; Rothman et al., 1990). Organisms are regularly exposed to other naturally occurring sources of mutagenic agents such as UV radiation (Jackson and Bartek, 2009; Schuch and Menck, 2010). DNA damage needs to be faithfully repaired for normal healthy cellular processes.

SMOKING AS A GENOTOXICANT

Perhaps one of the best-studied genotoxicant is tobacco smoke. According to the National Cancer Institute (www.cancer.gov), tobacco smoking is the leading cause of preventable death in the United States, resulting in roughly 1 in 5 deaths each year. Smoking is associated with high cancer risks at a number of different organ sites (Boffetta et al., 2012; IARC, 1976; Peterson, 2010). This is due in part to the fact that tobacco smoke is comprised of more than 5000 compounds, over 60 of which are classified as carcinogenic (Klassen, 2001; Peterson, 2010). Upon inhalation into the lungs, down the bronchi and bronchioles and into the alveoli, these compounds are capable of diffusing or are transported through the cellular membrane to interact and modify various cellular components such as proteins and DNA. These compounds not only act on cells at the site of exposure such as the nose, throat, and lung, but are distributed via the circulatory system to distant sites such as liver, pancreas, and colon (Klassen, 2001). A number of epidemiological studies have shown that both immediate and peripheral exposure is closely correlated with an increase in the risk of cancer when exposed to tobacco smoke (Adlkofer, 2001; Caporaso and Landi, 1994; Gangwar et al., 2009; Gao et al., 2011; Henríquez-Hernández et al., 2009; Hirayama, 1981; Hsu et al., 2009; Iodice et al., 2008; Ladeiras-Lopes et al., 2008; Liang et al., 2012; Lodovici and Bigagli, 2009; Mohelnikova-duchonova et al., 2011; Morita et al., 2010; Mucha et al.,

2006; Pryor, 1997; Sasco et al., 2004; Stern et al., 2009; Thorgeirsson and Stefansson, 2010; Veglia et al., 2003; Yan et al., 2009; Zeegers et al., 2000).

It is notable that not all smokers develop cancer (Boffetta et al., 2012). In fact, while smoking closely correlates with cancer risk, the response to smoking is highly variable between individuals. In a practical sense, there are a number of different possible outcomes when a person smokes. This is not because the individuals responding using innately different pathways, but rather the components within the pathways are highly variable for metabolic production, detoxification, and the effect on the genetic level by the chemical compounds. There is wide variation in the genes encoding metabolic enzymes of tobacco carcinogens (Campayo et al., 2011; Lui et al., 2005; Russo et al., 2011; Ter-Minassian et al., 2012), as well as significant association between reduced DNA repair capacity (DRC) and an increased risk of tobacco-related cancers (Campayo et al., 2011; Shen et al., 2003; Zhu et al., 2007). This interindividual variability makes understanding complex disease processes such as cancer difficult. As more information about the various genetic components become available, accurate mechanistic understanding becomes more obtainable and outcome prediction viable.

SINGLE NUCLEOTIDE POLYMORPHISMS AND DNA REPAIR

Single Nucleotide Polymorphisms (SNPs) can heavily influence this individual variation. SNPs are differences in a single base of a known genetic sequence. They are inherited generationally and can be found in all DNA carrying cells, thus making them attractive biomarkers as they are readily definable, stable, and common (Blitzblau and Weidhaas, 2010; Calzone, 2012; Chen et al., 2012; Cordero and Ashley, 2012; Dandona et al., 2012; Geenen et al., 2012; Herazo-Maya and Kaminski, 2012; Johnson et al., 2012; Lam et al., 2010; Mir, 2009). Unlike mutations, which are extremely rare, SNPs typically exist at a frequency of at least 1% (Brookes, 1999; Panagiotou et al., 2010) in a given

population. Recent evidence points to SNPs having subtle but still defining functional effects (Gorlov et al., 2011). However, while it is possible to link certain SNPs to a certain phenotype (e.g. cancer risk), this does not elucidate the mechanisms by which such SNPs can impart that phenotype (Chung and Chanock, 2011; Cooper and Shendure, 2011; Geenen et al., 2012; Keller et al., 2010; Parliament and Murray, 2010; Simonelli et al., 2012; Weiss et al., 2012).

A number of studies link some DNA repair SNPs to cancer risk by associating them with genetic damage or cancer risk. Table 1 is a brief summary of some reported SNPs associated with cancer or genetic damage in epidemiological studies. The RS (reference SNP) number nomenclature is used for SNP identification where possible, in accordance with NCBI Entrez SNP (ncbi.nlm.nih.gov/snp), NIEHS SNP (egp.gs.washington.edu), and International HapMap Project (hapmap.ncbi.nlm.nih.gov) databases. SNPs that do not follow this nomenclature are italicized in the table.

Table 1: Overview of SNP literature for DNA repair genes

Pathway	Gene	Function	SNP (rs#)	association with cancer	Ref
BER	<i>ADPRT/PARP1</i>	recruits ligase complex	1136410	head and neck, lung, bladder	1, 2, 3, 32
	<i>MUTHY</i>	glycosylates oxidized purines	34612342	head and neck	28
	<i>OGG1</i>	glycosylates oxidized purines	125701	bladder	3
			1052133	head and neck	28
			1052134	liver	4
	<i>POLB</i>	gap filling polymerase	3136717	bladder	3
	<i>XRCCI</i>	ligase component	25487	colorectal, esophagus, lung	2, 5, 6, 7, 32
1799782			bladder (protective), mouth	8, 9	
Direct	<i>MGMT</i>	transfer methyl to alkylated DNA	1625649	lung	10
			12917	lung	32

			2308321	breast	11
			12268840	esophagus	6
NER	CSA	core recruitment factor	60217257	PBL chromatid breaks	12
			60223979	PBL chromatid breaks	12
			60236422	PBL chromatid breaks	12
			60283572	PBL chromatid breaks	12
			60289598	PBL chromatid breaks	12
	CSB/ERCC6	core recruitment factor	50348723	PBL chromatid breaks	12
			50383065	PBL chromatid breaks	12
			50401891	PBL chromatid breaks	12
			50411435	PBL chromatid breaks	12
			50427510	PBL chromatid breaks	12
			50432753	PBL chromatid breaks	12
	CSB/ERCC6	core recruitment factor	2228526, 3793784, 4253160, 12571445	lung (additive)	13
	DDB2	damage recognition and recruitment	830083	lung	14
	ERCC1	incision complex component	11615	esophagus	6
			IVS5+33A>C	bladder	15
	ERCC2	nicks duplex DNA, recognizes ssDNA	238406	bladder	15
	ERCC4/XPF	5' endonuclease	1800067	pancreas	16
	ERCC5/XPG	3' endonuclease	1047769	bladder	15
			2296148	prostate	17
	MMS19L	helicase	872106	pancreatic (protective)	18
			2211243	pancreatic (protective)	18
			2236575	pancreatic cancer	18
	Rad23B	damage recognition complex	IVS5-15A>G	bladder	15
	XPA	damage recognition and core recruitment	1800975	esophagus	19
	XPC	damage recognition and core recruitment	2228000	colorectal lung	20 32

			2228001	esophagus, bladder	19, 21
	<i>XPD</i>	TFIIH component	1052559	bladder	8
			1799793	bladder lung	8 & 22 32
			13181	head and neck	28
NHEJ/BER	<i>XRCC2</i>	nicks duplex DNA, recognizes ssDNA	2040639	mouth	9
			<i>C41657T</i>	esophagus	23
	<i>XRCC3</i>	nicks duplex DNA, recognizes ssDNA	861539	mouth	9
			3212024	lymphoma	24
			3212038	lymphoma	24
			3212090	lymphoma	24
	<i>XRCC4</i>	ligase/kinase complex	2075685	mouth	9
			6869366	bladder	25
DSB	<i>ATM</i>	phosphoinositide 3- kinase	189037 228597 228592 664677 609261 599558 609429 227062 664982	lung	29
NHEJ/HR	<i>RAG1</i>	5' endonuclease	2227973	bladder	22
Mismatch Repair	<i>MSH2</i>	damage recognition	3732183	lung	32
Cell cycle signaling	<i>CCND1</i>	G1 to S transition signal	9344 678653	lung	30
Inflammation	<i>IL10</i>	cytokine signal	1800871	multi-site	33
nucleotide pools	<i>MTH1/NUDT1</i>	triphosphate hydrolysis	4866	lung	26
	<i>MTHFR</i>	DNA methylation	1801133 1801131	prostate	31
post-repair	<i>Rad18</i>	ubiquitin conjugating enzyme	373572	colorectal	27

Table 1: Overview of SNP literature for DNA repair genes. “RS#” refers to the Reference SNP number as designated in the NCI dbSNP database. References: 1=(Li et al., 2007), 2=(Zhang et al., 2005), 3=(Figuroa et al., 2007), 4=(Peng et al., 2003), 5=(Stern et al., 2006), 6=(Doecke et al., 2008), 7=(Kiyohara et al., 2006), 8=(Andrew et al., 2006), 9=(Yen et al., 2008), 10=(Hu et al., 2007), 11=(J. Shen et al., 2005), 12=(Leng

et al., 2008), 13=(Ma et al., 2009), 14=(Hu et al., 2006), 15=(García-Closas et al., 2006), 16=(McWilliams et al., 2008), 17=(Hooker et al., 2008), 18=(McWilliams et al., 2009), 19=(Guo et al., 2008), 20=(Huang et al., 2006), 21=(Fontana et al., 2008), 22=(Wu et al., 2006), 23=(Wang et al., 2009), 24=(Smedby et al., 2006), 25=(Chang et al., 2009), 26=(Kohno et al., 2006), 27=(Kanzaki et al., 2007), 28=(Sliwinski et al., 2011), 29=(Lo et al., 2010), 30=(Hsia et al., 2011), 31=(Wu et al., 2010), 32=(Kim et al., 2010), 33=(Ding et al., 2013).

CHALLENGES IN SNP STUDIES

While the work summarized in table 1 appears impressive, it only depicts association between cancer or genetic damage with a few SNPs. However, each of the studied genes carries over 100 SNPs, a vast majority of which have never been studied. Additionally, each pathway listed consists of many proteins (Nelson and Cox, 2005), each encoded by a number of genes with their own SNP complements. Many of these SNPs do not segregate independently in the genome, but rather as combinations. These combinations form defined groups (Browning and Browning, 2011). It is well known that genetic variations in humans are not arrayed simply as independent SNPs but, rather, as various combinations of SNPs or “haplotypes” (Gabriel et al., 2002). This is because some of the individual SNPs, often those located in close proximity to one another, are correlated and exist in degrees of linkage disequilibrium (LD). This creates identifiable and unique haplotypes, comprising several SNPs (Gabriel et al., 2002; Huang et al., 2011). Haplotypes are indicative of normal gene structure and are representative of the actual biology, defined both within the gene and between genes of the same pathway. Studying haplotypes enables functional analysis of genetic variation as it exists in nature, as opposed to the artificial approach of studying isolated SNPs without regard to the rest of the gene (Gulcher, 2012; Mir, 2009). Doing so will likely resolve many discrepancies found between current epidemiological SNP studies, as full haplotype analysis may reveal as yet unexplored SNPs and SNP combinations which drive functional effects (Ding et al., 2013; Lee et al., 2005; H. Ma et al., 2012; J. Ma et al., 2012; Mahmoudi et

al., 2011; Pereira et al., 2010; Qiu et al., 2011; Sakoda et al., 2012; Sharma et al., 2011; Wang et al., 2007).

Additionally, different ethnic populations can have widely disparate frequencies for the same SNP (Fu et al., 2011; García-Martín, 2008; Nakai et al., 2007; Shriener et al., 2011; Tian et al., 2009; Woo et al., 2009; Zabaleta et al., 2008). For example, for the *XPC* gene, the HapMap project reports significantly different frequencies in the allele rs2228000 for Europeans (0.300), Japanese (0.456), and Sub-Saharan Africans (0.058). This variation in frequency influences the actual structure of the haplotypes within a population. In fact, this is exacerbated in mixed populations, where the variable structure complicates the matter further. The issues of admixtures (combinations from multiple ethnic groups) can greatly change the LD of SNPs (Fu et al., 2011; Leng et al., 2012; Pabalan et al., 2012; Schwartz et al., 2009; Seldin et al., 2011; Shriener et al., 2011).

As more information becomes available from the various builds (or versions of total compiled genetic frequency data for the human genome project) at public databases like dbSNP and HapMap, there is an increasing need to analyze this raw data in a coherent and meaningful manner (Hollox, 2012; Huang et al., 2011). In the last few years, many commercial enterprises have sprung up to support this research through development of specialty probes from established companies to new sequencing approaches and bioinformatics analysis programs (Ng and Kirkness, 2010). Additionally, the research enterprise is still developing data mining and analysis techniques for this data, which leaves the area open for new specialties (Cordero and Ashley, 2012; Hollox, 2012; Mir, 2009; Ng and Kirkness, 2010). Perusal of the NIH RePORT database (the replacement database to CRISP, which lists both current and previous funding awarded by the NIH) for grants awarded for SNP research in 2012 sits at 3355 grants funded, up from 1168 in 2009, 513 in 2007, 324 in 2004, and from 40 grants in 1999 (projectreporter.nih.gov/reporter.cfm). The data generated from SNP and haplotype results have many translational applications. Warfarin research, for example, has

documented several key SNPs that effect blood clotting potential and side effect. Indeed, various components of the bioactivation pathway have proven well-defined gene-drug interaction. Current research is attempting to use genetic information in initial dosing schemes, with the idea that patients should reach effective therapeutic dose faster and with fewer possible side effects that with traditional dosing schemes alone (Gulseth et al., 2009; Krynetskiy and McDonnell, 2007; Takeuchi et al., 2009; Thomas et al., 2004; Wadelius et al., 2009). This idea can be extrapolated to many pathways, including the DNA repair pathways affecting risk/response to environmental damage and responses to chemotherapeutics (e.g. cisplatin, bleomycin, oxaliplatin) (de Haas et al., 2008; Kim et al., 2009; Sun et al., 2009; Wu et al., 2011; Zhang et al., 2012; Zhu et al., 2010).

APPROACHES FOR SNP FUNCTIONALITY DETERMINATION

Researchers use several approaches including mutagen sensitivity (MS) to evaluate the effect of SNPs in DNA repair genes. First proposed by Hsu in 1989, this approach measures the accumulation of DNA damage in the form of chromosomal aberrations (CAs) after in vitro exposure of cells to DNA damaging agents (Hsu et al., 1989). Cytogenetic analysis of CAs in lymphocytes is an accepted indirect biomarker of cancer risk (Abdel-Rahman and El-Zein, 2011; Bonassi et al., 2000; Decordier et al., 2010; Hagmar et al., 1994). The mutagen sensitivity assay draws its credibility as an epidemiological model from the link between CAs and cancer, whereby the repair of DNA damage after exposure indirectly measures inherited susceptibility to cancer risk (Abdel-Rahman and El-Zein, 2011). For one such mutagen sensitivity assay, in brief, cells are grown in culture and exposed to a DNA damaging agent such as bleomycin, cisplatin, or tobacco carcinogens such as benzo(a)pyrene diolepoxide (BPDE) or 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone (NNK). After exposure, cells are spread onto slides and chromatid breaks are counted. More exposed cells are maintained in

culture and allowed to repair for a defined length of time before being spread and subsequently the chromatid breaks are counted. Comparing the amount of chromatid breaks before and after repair gives an indirect measure of the cells' repair capacity, which can act as an intermediate phenotype of cancer risk (Abdel-Rahman and El-Zein, 2011; Hsu et al., 1989; Maekawa et al., 2006; Spitz and Bondy, 1993; Spitz et al., 1995, 1989; Wu et al., 1995).

The MS assay has been used to evaluate the effect of SNPs on genetic damage (Abdel-Rahman and El-Zein, 2011; Decordier et al., 2010). For example, previous studies in this laboratory have used the MS approach to look at the effect of two nonsynonymous SNPs each in the *XRCC1* and *XPB* genes on repair of NNK damage in lymphocytes from healthy individuals (Abdel-Rahman and El-Zein, 2000; Affatato et al., 2004). Other laboratories have used the same assay to evaluate the effect of SNPs on DNA repair. Leng et al. used the MS assay to test the NER capacity with different genotypes in *CSA* and *CSB* genes and smoking status using BPDE treatment. By measuring the number of chromatid breaks in isolated lymphocytes, they were able to determine associations between the 37 different SNPs and repair capacity. Additionally, associations between smoking and repair capacity were determined for the various SNPs and SNP combinations. Leng's analysis showed 5 *CSA* SNPs and 6 *CSB* SNPs correlated with reduced repair capacity in smokers after BPDE treatment (Leng et al., 2008). Similar studies were used by Aka et al. and Angelini et al. for *XRCC3* with hydrogen peroxide and *XPB* SNPs with bleomycin, respectively (Aka et al., 2004; Angelini et al., 2008). The MS approach used in the initial studies of this project correlates the effect of multiple SNPs (haplotypes) on CA accumulation after mutagen treatment.

To determine haplotypes for a gene, the LDs between SNPs are calculated from a small population with known sequence information using computational methods. While there are a number of algorithms available for determining haplotypes, PHASE analysis is the preferred method for small genomic segments like a single gene (Browning and

Browning, 2011). In 2003, Stephens and Donnelly published the current PHASE program. PHASE is an iterative algorithm, generating a large number of probabilities that are mathematically scored for likelihood based on a set of known sequences, ultimately resulting in a list of haplotypes with estimated frequencies that are indicative of the whole population (Stephens and Donnelly, 2003). These unique haplotypes can be assigned to individual subjects of a population based on SNP sequencing.

A difficulty in haplotype functional analysis can be the potential number of real combinations that can exist in a population. As humans have a diploid genome, for n number of haplotypes there are 2^n possible haplotype combinations. Even as few as 20 haplotypes can result in over a billion combinations, which can be very limiting to analyze in a population. In an effort to increase statistical power, the relatedness of each haplotype can be determined by phylogenetic analysis. Sequence similarity between individual haplotypes can be calculated to create highly similar groupings of haplotypes (clades, also called phylogenetically grouped haplotypes or PGHs). Phylogenies constructed on the basis of sequence similarities between haplotypes provide an objective tool for grouping haplotypes that share genealogical similarities. This similarly, comparing haplotypes by group can highlight potential SNPs of interest, where a given clade phenotype is likely to be governed by SNPs that exist as all variant (or all ancestral) in the haplotypes within a given PGH. For example, a clade X may consist of haplotypes that contain variants at position 5, 7, 14, and 20. Another clade Y may have variants at positions 3, 5, 18, 19, and 20. If clade X responds twice as well to a treatment than clade Y, it is likely that this difference is being driven by the SNPs at positions 3, 7, 15, 18, and 19, with 3 and 7 enhancing the effect and 15, 18, and 19 lowering the effect in comparison. Such clade assessments are useful for breaking down functional analysis results and providing more detailed associations.

Again, like haplotype analysis, there are several algorithms available for computing sequence similarity, the standard since the 1990s remains the MEGA analysis

program. The molecular evolutionary distance is computed using assumptions encoded in the MEGA method to determine the total divergence between haplotypes. This divergence is determined for each SNP pair individually based on “nucleotide substitution parameters” (the mathematical predetermined likelihood of changing one nucleotide to another) and added together for the total haplotype (Tamura et al., 2007). The shorter the distance between two haplotypes, the greater their similarity and, hypothetically, the greater the probability of sharing similar mechanistic traits. This phylogenetic analysis increases the statistical power by reducing the total number of comparisons made, while still capturing the unique sequence characteristics that constitute related haplotypes (Bardel et al., 2009; Rzhetsky and Nei, 1992; Tamura et al., 2007; Yang, 1997). These groups were used to evaluate the hypothesis of a haplotype based differences in response to environmental mutagens.

NUCLEOTIDE EXCISION REPAIR PATHWAY

DNA repair is critical for day to day cellular maintenance. Many endogenous and exogenous compounds are capable of entering the nucleus and causing DNA damage throughout the genome, resulting in a variety of aberrations including adducts (polycyclic aromatic hydrocarbons), methylation (arsenic), alkylations (nitrosamines), etc (Cornetta et al., 2006; Klassen, 2001; Pryor, 1997; Veglia et al., 2003). To maintain genetic fidelity during cellular replication, this damage must be fixed. To do so, cells have developed a number of repair mechanisms that target different forms of damage. The most common of these are Nucleotide Excision Repair (NER), Base Excision Repair (BER), Non-Homologous End Joining (NEHJ), Homologous Recombination (HR), and Direct Repair (Klassen, 2001).

BER occurs through removal of a damaged base itself by cleavage of the glycoyl bond in the nucleotide, resulting in an abasic site. These abasic sights are recognized,

removed, and repaired using the complimentary sequence in the same manner as NER. Direct Repair is a repair that occurs at the site of damage without cleave or gap filing. A common example of this is by methyltransferases detecting simple alkylation damage and transferring a methyl group to the residue, which subsequent reactions use directly to regenerate the correct base at the transfer site. Less common, but still important repair mechanisms are NHEJ and HR. NEHJ allows for repair of double stand breaks, such as those generated after replication when damage has not been repaired. Conversely, while HR also occurs at double strand breaks, it has a recognizable strand inversion step that can result in a crossing-over event at homologous sites in a pair of chromosomes. Massive changes in double helical structure, such as adducts and dimers are corrected by the NER. In this repair pathway, several bases surrounding the aberration site are removed by exonucleases and the gap filled in by a repair specific polymerase using the complementary sequence in the same manner as replication. This pathway is important in maintaining long term chromosomal stability and genetic fidelity, as even untranscribed sequences are repaired prior to replication of cellular division (Klassen, 2001; Nelson and Cox, 2005).

Figure 1: DNA damage structures

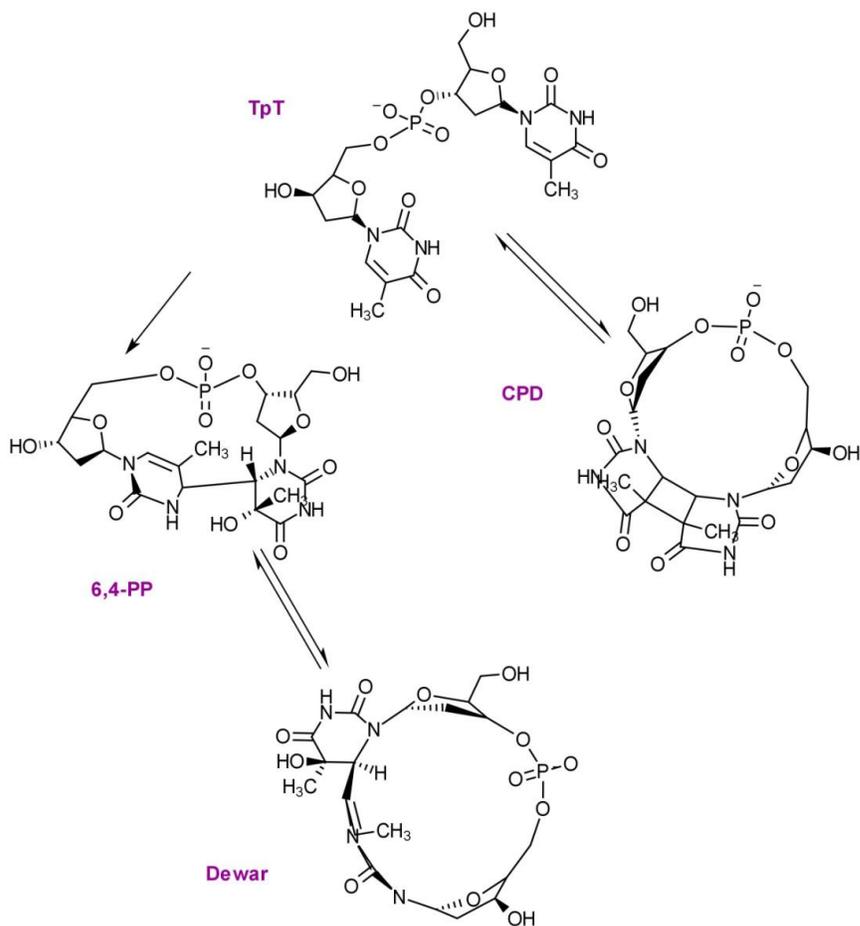


Figure 1: DNA damage structures. UV-induced DNA dimers presented as: DNA containing thymine (TpT), cyclopyrimidine dimers (CPD), and pyrimidine-6-4-pyrimidone photoproducts (6,4-PP). Image created in ChemDoodle (Copyright © 2008-2013 iChemLabs, LLC).

Each of these pathways is a multistep process involving a number of proteins and enzymes whose actions must occur in a coordinated fashion, and often in large conglomerate complexes that have well defined interactions. Incomplete activity, binding, or interaction can result in loss of repair function, which in turn can lead to accumulation of DNA damage. This damage can lead to larger defects in cellular processes and loss of genetic fidelity in successive divisions, both of which can lead to initiation and development of cancer (Bonassi et al., 2000).

NER is the most flexible of these repair pathways, capable of recognition of any perturbation of the classical helical DNA structure regardless of transcriptional status before it can be passed onto daughter chromosomes as breaks or mutations (Bunick et al., 2006; Friedberg, 2001; Sugasawa et al., 1998; Wang, 2008). This critical pathway has been shown to be involved in cancer risk when the DNA repair capacity is reduced (Cheng et al., 1998; Cornetta et al., 2006; Decordier et al., 2010; Hsu et al., 1991; Langie et al., 2006; Ming et al., 2012; Slyskova et al., 2011; Spitz and Bondy, 1993; Wei et al., 1996; Zelle and Lohman, 1979). In NER DNA damage is recognized by the damage recognition complex, which contains the *xeroderma pigmentosum* complementation group C (XPC) protein complexed with hHR23B, CENT2, and several other proteins, with XPC being the critical direct DNA binding protein (Araki et al., 2001; Bunick et al., 2006; Craig et al., 2006; Dantas et al., 2012; El-Mahdy et al., 2006; Lujsterburg et al., 2012; Maillard et al., 2007; Nishi et al., 2005; Park and Choi, 2006; Renaud et al., 2011; Sugasawa et al., 1998; Trego and Turchi, 2006; You et al., 2003). The recognition of DNA damage recruits the next set of repair components, XPA, XPG, RPA, and TFIIH to the site of damage, displacing the XPC complex (Bunick et al., 2006; Park and Choi, 2006; Sugasawa, 2011; You et al., 2003). TFIIH unwinds the DNA at the site of recruitment, ligases XPG and ERCC1 nick the strand about 25 nucleotides apart, allowing the repair polymerase complex to fill the gap, and ligases are recruited to seal the nicks (Nelson and Cox, 2005).

Figure 2: Nucleotide Excision Repair

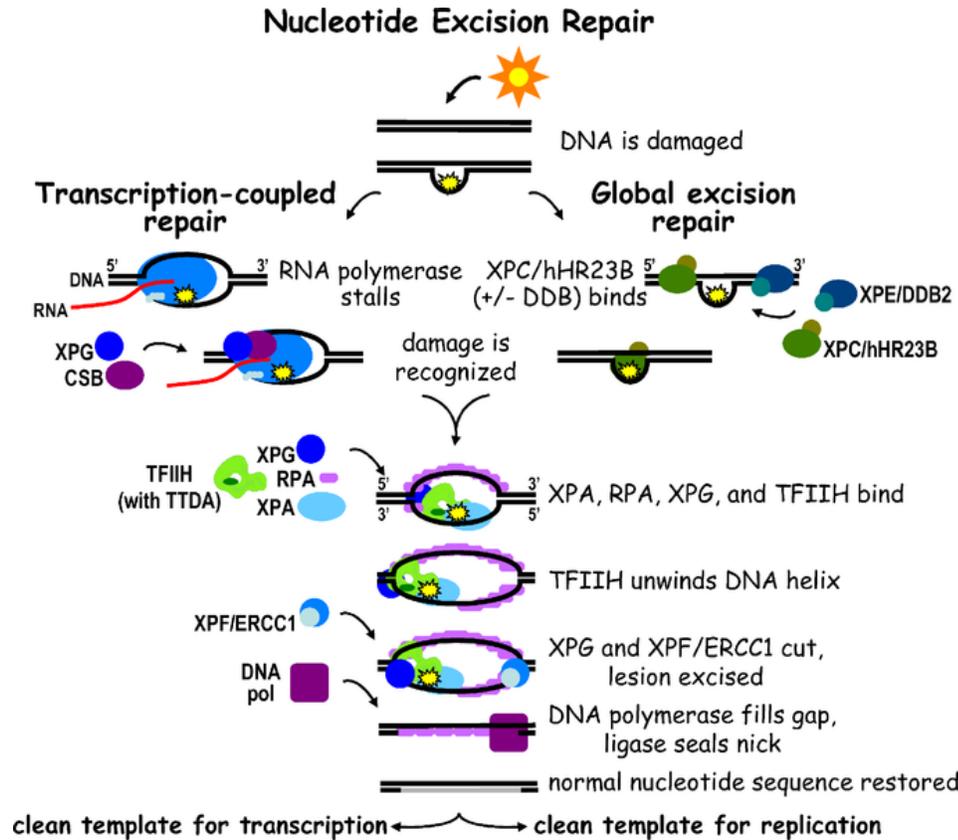


Figure 2: Nucleotide Excision Repair. Shown is a diagram of the nucleotide excision repair pathway. XPC is the DNA binding protein recognizing DNA damage in the initial step of the Global Genome Repair arm of the NER pathway. Image obtained from WikiCommons under the Creative Commons attribute license (Fuss and Cooper, 2006).

The focus of this dissertation is on XPC. As the recognition factor, XPC has significant influence on the repair capacity of the entire pathway. Loss of XPC function is rare, and results in *xeroderma pigmentosum* type C disease (XP-C, OMIM 278720), which is characterized by increased hypersensitivity to UV-induced genetic damage and skin cancer predisposition (Friedberg, 2001; Klassen, 2001). Supporting XPC function are studies involving *XPC*-null mice lacking the XPC protein. These deficient animals have been shown to have significantly higher sensitivity to DNA damaging compounds

(Wickliffe et al., 2006). The much more subtle changes in XPC function encoded by SNPs in the *XPC* gene are significantly more common, though less understood.

SNPs IN *XPC*, HISTORICALLY

XPC is a 940 amino acid protein (Genbank Accession Number AC090645) encoded by a highly polymorphic 33kB gene region on 3p25, with over 90 SNPs currently listed in the dbSNP and HapMap databases. However, only a few of these polymorphisms have been studied thus far. Of those, only 4 have been studied as potential risk modifiers for cancer susceptibility: exon 16 (K939Q; rs2228001), exon 8 (A499V; rs2228000), intron 11-5 splice site C/A (rs3729587), and intron 9 poly-AT insertion. As stated earlier, there are many conflicting results from SNP epidemiological studies, and XPC is no exception to this. An et al. reported an association between the rs2228000 SNP and head and neck cancer risk, while Guo et al. did not find association in esophageal squamous cell carcinoma (An et al., 2007; Guo et al., 2008). Hansen et al. found a significant association between colorectal cancer risk and rs2228001 SNP, as did Zhu et al. for DNA damage accumulation after BPDE exposure (Hansen et al., 2007; Zhu et al., 2007), however, Vodicka et al. reported no effect for the same SNP in relation to accumulation of DNA damage after α -radiation (Vodicka et al., 2004). Similarly, associations between SNPs in XPC and DNA repair capacity (DRC) are also controversial. Cornetta et al. reported that the rs2228001 SNP was associated with increased repair capacity in the heterozygous form, while both Slyskova et al. and Vodicka et al. showed no change in DRC with this polymorphism (Cornetta et al., 2006; Slyskova et al., 2011; Vodicka et al., 2004). Langie et al. showed that both rs2228000 and rs2228001 had no significant effect on DRC while Zhu did (Langie et al., 2010; Zhu et al., 2008). Additionally, Qiao et al. found a significant effect on DRC with the intron 9

insertion showing a decrease in DRC, yet Slyskova et al. did not (Qiao et al., 2002; Slyskova et al., 2011).

These clear discrepancies are not surprising, given that each of these studies looked at only individual SNPs. Without LD information for each of the SNPs, and all the others not evaluated, it is plausible that these differences may be due to unevaluated SNPs. These unevaluated SNPs may exist in variable LD with the SNPs that were evaluated, or other unevaluated functional SNPs. Admixtures, sampling inconsistency, or incomplete LD may not have captured SNPs with functional effects (Abdel-Rahman and El-Zein, 2011; Keller et al., 2010). To avoid similar problems, I chose to evaluate the haplotype effect of all reported common *XPC* SNPs comprehensively.

Figure 3: *XPC* gene



Figure 3: *XPC* gene. Shown is a diagram of the total *XPC* genomic region under study. The UTR regions are colored in blue, the introns in peach, and the exons in purple. The downward hash-marks are the locations for each of the SNPs with a minor allele frequency (MAF) of 0.05 in the as reported in the HapMap population. (Additional details are given in Material and methods.)

OVERVIEW AND OBJECTIVES

The objective of the study was to address a number of significant gaps in knowledge regarding the effect of *XPC* SNPs, in their biological context of whole gene haplotypes, on genetic damage as well as their potential role in disease risk. This is the first comprehensive analysis of the entire *XPC* genomic sequence and mechanistic evaluation of the effect of *XPC* haplotypes on genetic damage. In addition, this work is the first time the effects of *XPC* haplotypes on transcription or translation were evaluated.

In order to understand the effect of *XPC* haplotypes on genetic damage, a comprehensive haplotype map encompassing all common *XPC* SNPs reported to exist with a minor allele frequency of ≤ 0.05 was created using bioinformatics. The HapMap CEU population was used as a reference population of known genotypes. Using this information, the first hypothesis tested was if certain *XPC* haplotypes have phenotypic or functional effects, there would be a correlation between these haplotypes and genetic damage in individuals exposed to environmental mutagens. To test this hypothesis, an exploratory haplotype-phenotype study was performed using lymphocytes from a study population composed of healthy individuals. Results from these studies guided additional investigations to understand the mechanistic effects of *XPC* haplotypes on genetic damage.

The mechanistic relationships were studied *in vitro* using cell lines representative of each PGH, and evaluated the haplotype effect on DRC. The working hypothesis was that *XPC* haplotypes influence levels of accumulated DNA damage by affecting DNA repair capacity. To test this hypothesis, I exposed human lymphoblastoid cell lines representing the different PGHs to low dose UV-B radiation, inducing the formation of pyrimidine-6-4-purimidone photoproducts (6-4PPs) and cyclopyrimidine dimers (CPDs). The rate of removal of these DNA damage adducts represents the DRC of the cells. The haplotype effect was determined for both rate of repair and adduct preference.

The effects of the haplotypes on gene and protein expression were evaluated to test the working hypothesis that *XPC* haplotypes influence DNA repair capacity through transcriptional and/or translational differences. To test this hypothesis, I again exposed human lymphoblastoid cell lines representing different PGHs to low dose UV-B radiation, inducing the formation of UV adducts. I evaluated the changes in amount of *XPC* mRNA over time as a measure of the induction of the gene. Changes in the rate of nascent *XPC* production indicate sensitivity to DNA damage at a transcriptional level. Additionally, I evaluated the changes in amount of XPC protein over time as another

measure of induction and, therefore, sensitivity to DNA damage at a translational level. Collectively, these results provide a mechanistic link between *XPC* haplotypes and differences in DNA repair. From a public health prospective, this study provides a mechanistic link between whole-gene variation and cancer susceptibility.

Chapter 2: MATERIALS AND METHODS

Table: Chemicals

Chemical	Company	Location	Product #
3,3',5,5'-Tetramethylbenzidine (TMB)	Sigma-Aldrich	St. Louis, MO	T0440
4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone (NNK: CAS#64091-91-4, National Cancer Institute)	Midwest Carcinogen Repository	Kansas City, MO	E0698
6,4-PP anti-mouse (64M-2)	CosmoBio	Tokyo, Japan	CAC-NM-DND-002
7.5% SDS-PAGE Tris-HCl precast Ready Gels	Bio-Rad Laboratories	Hercules, CA	161-1100
BCA Assay kit	Pierce Thermo Fisher Scientific	Waltham, MA	23225
Bovine Serum Albumin (BSA)	Sigma-Aldrich	St. Louis, MO	A7906
BSA protein standard	Pierce Thermo Fisher Scientific	Waltham, MA	23209
Colcemid	Gibco-Invitrogen	Carlsbad, California	15212-012
Commassie Blue R250	Pierce Thermo Fisher Scientific	Waltham, MA	20279
CPD anti-mouse (TDM-2)	CosmoBio	Tokyo, Japan	CAC-NM-DND-001
dimethylsulfoxide (DMSO)	Sigma-Aldrich	St. Louis, MO	472301
DNase/RNase-free water	Invitrogen	Carlsbad, CA	AM9937
DNase-I	Qiagen	Venlo, Netherlands	79254
ECL DualVue Molecular Weight Marker	GE Healthcare Life Sciences	Little Chalfont, England	RPN810
ECLplus	Pierce Thermo Fisher Scientific	Waltham, MA	32132
ELISA wash buffer	Cayman Chemical	Ann Arbor, MI	400062

Ethanol	via UTMB pharmacy	N/A	N/A
Ethylenediaminetetraacetic acid (EDTA)	Sigma-Aldrich	St. Louis, MO	E9884
Giemsa	Sigma-Aldrich	St. Louis, MO	G9641
Glacial acetic acid	Sigma-Aldrich	St. Louis, MO	A6283
glutamine (L) 2 μ M	Invitrogen	Carlsbad, CA	205030-081
Glycine	Sigma-Aldrich	St. Louis, MO	G8898
hydrochloric acid (HCl)	Sigma-Aldrich	St. Louis, MO	320331
hypotonic solution (0.075 M potassium chloride: KCl)	Purgene	Minneapolis, MN	158902
Isopropanol	Sigma-Aldrich	St. Louis, MO	109827
Lammelli loading dye 2x	Bio-Rad Laboratories	Hercules, CA	161-0737
Methanol	Sigma-Aldrich	St. Louis, MO	32213
Penicillin/Streptomycin	Invitrogen	Carlsbad, CA	205030-081
phosphate buffered saline (PBS)	Sigma-Aldrich	St. Louis, MO	P3813
phytohemagglutinin (PHA, reagent grade)	Remel	Lenexa, KS	3085271
Picogreen	Life Technologies	Carlsbad, CA	P11495
PMSF	Sigma-Aldrich	St. Louis, MO	P7626
premium fetal bovine serum (FBS)	Atlanta Biologics	Lawrenceville, GA	S11150
Protamine sulfate	Sigma-Aldrich	St. Louis, MO	P4020
protease inhibitor complete cocktail tablets	Roche	Penzberg, Germany	11836153001
PVDF membrane	Bio-Rad Laboratories	Hercules, CA	162-0177
Qiagen QiaAmp DNA isolation mini-kit	Qiagen	Venlo, Netherlands	51306
rabbit anti-mouse HRP-conjugated	Zymo Research	Irvine, CA	N/A
rabbit anti-mouse HRP-conjugated	Invitrogen	Carlsbad, CA	61-6520

Rainbow Molecular Weight Marker	GE Healthcare Life Sciences	Little Chalfont, England	RPN800E
RIPA buffer 10x	Cell Signaling	Danvers, MA	9806
RNA Direct-zol mini-prep kit	Zymo Research	Irvine, CA	R2050
RPMI 1640	Gibco	Carlsbad, California	11875
Sodium Chloride (NaCl)	Sigma-Aldrich	St. Louis, MO	S9888
Sodium dodecyl sulfaste (SDS)	Sigma-Aldrich	St. Louis, MO	L3771
Sodium hydroxide (NaOH)	Sigma-Aldrich	St. Louis, MO	S5881
TaqMan® genotyping primers	Applied Biosystems	Foster City, CA	various by rsNumber
TaqMan® high capacity RNA to cDNA kit	Applied Biosystems	Foster City, CA	4387406
TaqMan® β-actin specific VIC fluorophore (ATCB)	Applied Biosystems	Foster City, CA	4326315E
TaqMan® universal master mix	Applied Biosystems	Foster City, CA	4304437
TaqMan® XPC specific FAM fluorophore (Hs01104206_m1)	Applied Biosystems	Foster City, CA	4331182
TE buffer RNase/DNase free 20x	Invitrogen	Carlsbad, CA	T11493
TRI reagent	Invitrogen	Carlsbad, CA	AM9738
Tris(hydrozomethyl)aminomethane (Tris)	Sigma-Aldrich	St. Louis, MO	T4661
Trypan Blue	Sigma-Aldrich	St. Louis, MO	T8154
Tween-20	Sigma-Aldrich	St. Louis, MO	P1379
Urea	Sigma-Aldrich	St. Louis, MO	U6504
XPC Primary Antibody	UTMB Protein Chemistry Core in the Biomolecular Resource Facility	N/A	N/A
β-mercapitoethanol	Bio-Rad Laboratories	Hercules, CA	161-0710

Table: Special equipment and software

Special Equipment/Software	Company	Location
AlphaImager 2200	Alpha Innotech Corporaion	San Leandro, CA
CellBIND® Low Profile Corning flasks (13700433)	Fisher Scientific	Hampton, NH
Chromo4 computerized real-time PCR detection system	Bio-Rad Laboratories	Hercules, CA
Haploview software (version 4.1)	MIT/Harvard Broad Institute	Cambridge, MA
MEGA 4 (http://www.megasoftware.net/)	The Biodesign Institute	Tempe, AZ
MJ Research DNA Engine thermocycler	Bio-Rad Laboratories	Hercules, CA
NCSS/PASS software Dawson Edition	NCSS LLC	Kaysville, UT
Nikon 400 light microscope	Nikon Instruments	Melville, NY
PHASE ver. 2.1 software (www.stat.washington.edu/stephens/phase.html)	University of Chicago	Chicago, IL
SigmaPlot	Systat Software Inc	Chicago, IL
SoftMax Pro software	Molecular Devices	Sunnyvale, CA
SpectraMax190 plate reader	Molecular Devices	Sunnyvale, CA
S-Plus	TIBCO Software Inc	Palo Alto, CA
SPSS18 Statistics	IBM Corporation	Endicott, NY
Tagger software (www.broad.mit.edu/mpg/tagger)	MIT/Harvard Broad Institute	Cambridge, MA
Tecan GENios Pro plate reader	Tecan Group Ltd	San Jose, CA
UV Simulator	Oriel Instruments	Stratford, CT
UV-B specific meter	National Biological Corporation	Beachwood, OH

UV-transparent 96 well Corning plates (07200848)	Fisher Scientific	Hampton, NH
Wester mini-blot	Bio-Rad Laboratories	Hercules, CA

OVERVIEW

To evaluate the effect of all the common *XPC* SNPs on DRC and levels of genetic damage comprehensively, I determined all common *XPC* SNPs reported to exist with a minor allele frequency of ≥ 0.05 and constructed unique and biologically viable haplotypes. Phylogenetic analysis of these haplotypes created a small subset of groups consisting of highly related haplotypes. I then used these groups to evaluate the effect of *XPC* haplotypes on the accumulation of genetic damage after environmental exposure to mutagenic agents, as well as the haplotype effect on accumulation of genetic damage and DRC in cell culture after acute exposure to DNA damaging agents. I further investigated the effect of *XPC* haplotypes on NER DRC mechanistically by changes of the various PGHs on transcriptional and translational endpoints.

STUDY SUBJECTS AND BLOOD COLLECTION

The study involved a subset of blood samples obtained from a larger cohort of subjects who were recruited without regard to age, sex, or ethnicity from the smoking and non-smoking staff and student population of University of Texas Medical Branch (UTMB). This cohort was comprised of individuals who had responded to posted notices and advertisements requesting volunteers for studies aimed at understanding the functional and biological significance of sequence variability in DNA repair genes. The study protocol was approved by the Institutional Review Board (IRB 04-131), and all study subjects signed a written consent form that described the purpose of the study. This large cohort was recruited, characterized, and studied by other members of our laboratory

(Hill et al., 2005). Individuals were defined as non-smokers if they had smoked less than 100 cigarettes during their lifetime or as smokers if they had smoked at least five cigarettes per day for at least one year prior to enrollment in the study. Participants filled out a questionnaire that provided demographic, occupational, general health information, and smoking habits (number of cigarettes per day, number of years smoked, preferred brand, duration of smoking, former tobacco use, and use of other tobacco products). To control for possible confounders, exclusion criteria included a recent acute viral or bacterial infection, a major chronic illness, a recent blood transfusion, treatment with mutagenic agents, excessive alcohol consumption, defined as more than a 10 gram serving per day (as determined by nationwide standard practices), and employment involving exposure to potentially mutagenic agents (e.g. the petrochemical industry).

For the purposes of this study, I further limited the data to the ethnic population of white-non-Hispanics as self-reported on the questionnaire. This population was chosen as the focus because of the availability and completeness of the online data (HapMap online database at www.hapmap.org data release 22 phase II NCBI assembly B36 dbSNP b126), and to limit the impact of possible admixtures (Abdel-Rahman and El-Zein, 2011; Fu et al., 2011; García-Martín, 2008; Keller et al., 2010; Leng et al., 2012; Nakai et al., 2007; Pabalan et al., 2012; Schwartz et al., 2009; Seldin et al., 2011; Shriner et al., 2011; Tian et al., 2009; Woo et al., 2009; Zabaleta et al., 2008). The CEU population (Utah residents from the CEPH population with ancestry from northern and western Europe) accurately represents white non-Hispanics in HapMap. As such, my study included 123 individuals. (Further information is available in the Results chapter of this dissertation.)

DETERMINATION OF BACKGROUND AND MUTAGEN-INDUCED CHROMOSOMAL ABERRATIONS

Data for background and mutagen-induced chromosomal aberrations (CAs) were generated through a different study (Hill et al., 2005) and was readily available for this project. In these studies, 10mL of blood was obtained from each study participant and 1mL was used to establish cytogenetic cultures of peripheral blood lymphocytes (PBLs) according to standard procedures (Abdel-Rahman and El-Zein, 2000; Affatato et al., 2004; Evans et al., 1975). Genotyping analysis used isolated DNA extracted from a second 5mL sample.

Each subject had two cultures established; the first provided a baseline *in vivo* CA frequency and the second used for the determination of mutagen-induced CA levels. For this assay, PHA stimulated cells PBLs were resuspended in serum-free media supplemented with 0.24mM of the tobacco specific mutagen, 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone (NNK), for 1 hour. Following NNK treatment, the PBLs were washed, transferred, and resuspended in the original growth medium until harvested 24 hours after NNK treatment. (Mutagen concentration and harvest times were previously established to produce measurable genetic damage with low toxicity (Abdel-Rahman and El-Zein, 2000; Affatato et al., 2004).) Cells were arrested in metaphase prior to harvest with 0.1 µg/ml colcemid treatment for 1-hour. PBLs were resuspended in hypotonic solution, fixed with Carnoy's fixative (3 parts methanol to 1 part acetic acid), and stored at 4-0°C. Fixed cells were spread on duplicate slides, coded, and stained with Giemsa. One hundred metaphase cells per coded slide were scored for CAs in a blinded manner, using mixed batches representing smokers and non-smokers on a Nikon 400 light microscope, according to standard procedures (ISCN, 1985). Chromatid breaks were scored as one break, chromosome breaks as two, and total aberrations were defined as breaks per 100 cells. For quality control, 20% of the slides were randomly selected for

blind rescoring, with agreement measured using Cohen's kappa statistical test. The results for both baseline and NNK-induced CAs were used for statistical comparisons.

For DNA isolation, red blood cells were lysed and the white blood cells separated from the rest of the blood product by centrifugation. The white blood cells were lysed by proteinase K digestion and the proteins removed by salting out. The released DNA was cleaned using RNaseA digestion followed by alcohol precipitation. The precipitated DNA was washed in ethanol solution and briefly air dried before redissolving in hydration solution (TE buffer). The isolated DNA was stored at -80°C until genotyping analysis.

***XPC* HAPLOTYPE-TAGGING SNPs AND POPULATION GENOTYPING**

Using genotype data from the International HapMap Project database for the CEU population (CEPH population Utah residents with ancestry from northern and western Europe; data release #24 (October 2008) National Center for Biotechnology Information Build 36 assembly (dbSNP b126)), I analyzed the full genomic region plus an additional 2kb both 5' and 3' to cover any potential UTRs (chromosome 3, bases 14159650 to 14197142 on the minus strand). This population contains family trio genotypes, meaning that the CEU population contains individuals that are mother, father, and offspring families. Genotypes for all individuals were screened using Haploview to ensure that only SNPs with a minimum allele frequency (MAF) of 0.05 or greater were used for haplotype inference. This analysis revealed 35 SNPs, which were analyzed for this study.

To determine the genetic profile of the individuals, I used TaqMan[®] based PCR assay. This assay uses Förster (or Fluorescence) resonance energy transfer (FRET) technology to detect individual SNPs using PCR amplification. In brief, the assay contains a set of short oligonucleotides (oligo) with both a fluorescent probe and the matching fluorophore's quencher attached. When the oligo is intact, any excitation energy

that is absorbed by the fluorophore (probe) will pass it to a low energy dark molecule (quencher) in close proximity instead of being emitted as a photon, thereby suppressing the fluorescence. These oligos contain sequences specific for annealing to the region of the gene containing the SNP site, one fluorophore for the variant and another for the ancestral allele. Primers specific for the genetic region are also included. During the annealing phase of the PCR reaction, one of the probe oligos will set down along with the primer pairs. The forward primer is extended by the TaqMan AmpliTaq Gold DNA polymerase, which contains 5' exonuclease activity. During the extension phase, the primer is extended by the polymerase and when the polymerase reaches the annealed SNP oligo, the exonuclease portion of the polymerase cleaves the oligo to release the fluorophore. Meanwhile, the quencher remains stuck to the DNA as it contains a minor groove binding protein. Consequently, when the fluorophore is excited the energy is emitted as a photon, which is detected at the emission wavelength. The emissions are read as an indicator of the presence or absence of the variant SNP (or both, in the case of a heterozygote).

This genotyping can be very labor intensive and costly. While it would be optimal to sequence the entire gene to determine unequivocally the SNP complement, that is not practical for many labs. Even rigorous MAF cutoffs can still result in a large number of SNPs needing to be genotyped for each individual, resulting in many of the same problems. To simplify the processes, researchers use haplotype tagging SNPs (htSNPs, also called tag SNPs) to reduce the number of genotyping reactions required for haplotype analysis. The use of htSNPs is based on linkage disequilibrium analysis, where a program such as Haploview's (Barrett et al., 2005) Tagger (de Bakker et al., 2005) compares the genetic information at a series of given SNPs for a small population (or reference panel) and determines which SNPs exist in [near] full disequilibrium. As a consequence, one SNP (designated the htSNP) can be used to predict the genotype of another SNP when in disequilibrium, reducing the number of genotyping reactions needed to obtain all the SNPs (de Bakker et al., 2005; Gabriel et al., 2002). I used Tagger

software (broadinstitute.org/mpg/tagger) to identify tagging SNPs (htSNPs), e.g. those that are capable of determining all 35 SNPs without the need to genotype each SNP individually (de Bakker et al., 2005). Genotypes for the 30 family trios at all 35 SNP positions was run in Tagger to identify haplotype tagging htSNPs for assay design. I used an aggressive multi-marker approach (up to 6 markers) with a conservatively set the r^2 threshold to ≥ 0.8 (mean value 0.971) and coupled with a logarithm of odds score (LOD; for estimating a recombination-frequency heterogeneity) threshold of 2 (de Bakker et al., 2005; Goode et al., 2007; Nam et al., 2007).

Subsequently, I identified 11 htSNPs for the 35 evaluated SNPs in *XPC*. Using, custom-designed real-time polymerase chain reaction (PCR)-based assays and the TaqMan[®] chemistry, assays were developed based on the sequences listed in the NCBI dbSNP database. For each reference SNP (rs) number, either allele-specific probes labeled with a FAM or a VIC fluorophore and an appropriate quencher were developed. The PCR consisted of TaqMan[®] universal master mix, template DNA isolated from the study subjects (or water as a no template control), and TaqMan[®] target-assay mix in a total reaction volume of 12 μ l at concentrations recommended by Applied Biosystems. Thermal cycling was carried out in our laboratory on a MJ Research DNA Engine thermocycler equipped with a Chromo4 real-time PCR detection system under recommended conditions (50°C, 2 min; 95°C, 10 min; and 40 cycles at 95°C for 15 sec and 58-61°C for 1 min). Designation of referent and polymorphic forms was determined by the FAM to VIC ratio. For quality control, I ran all PCR reactions in duplicate, alongside both no-template negative controls and positive controls for each possible genotypic combination (when possible). Samples were blind coded for smoking status, and samples from smokers and non-smokers were run together in mixed batches, with 10% of the samples randomly selected and re-genotyped. Genotypes for all htSNPs were analyzed for deviations from Hardy-Weinberg equilibrium (HWE) on a locus-by-locus basis.

***XPC* HAPLOTYPE INFERENCE AND PHYLOGENETIC GROUPING**

Haplotype analysis analyzes all the SNPs comprehensively, using linkage to determine the number of real permutations possible for all the SNPs in aggregate, based on the reference set. PHASE is a software package from the Stephens lab has been considered a gold standard for haplotype analysis, and is still considered optimal for small genomic regions such as a single gene (Browning and Browning, 2011; Stephens and Donnelly, 2003). Very briefly, PHASE uses Bayesian inference to score the likelihood of random genotype combinations based on a known reference population. This likelihood is calculated over multiple iterations and the final output is the combined sets of real combinations possible based on the reference.

I inferred the haplotypes using the CEU population of the HapMap database. I used Bayesian statistics implemented in PHASE with the algorithm designed for family trios. The number of iterations was increased to 10,000, the thinning interval was increased to 10, and the burn-in was increased to 200 to improve the accuracy of the inferred haplotypes. The real power of haplotype analysis is that it can extract the real haplotypes that are present (or calculated to be present) in a population of similar ethnicity (Browning and Browning, 2011; Srkar-Roy et al., 2011; Stephens and Donnelly, 2003). Family trios offer an advantage in this analysis, as the child is a valid recombination of the parents, thus improving the accuracy of the algorithm. The PHASE analysis generated 21 unique real haplotypes out of a possible 34,359,738,368 (2^{35}) theoretical haplotypes.

To reduce the number of statistical comparisons (2,097,152 theoretical combinations, or 2^{21}) for the haplotype diplotypes, I used a phylogenetic grouping approach to cluster evolutionarily related haplotypes together. Grouping of haplotypes, based on genealogical or phenotypic relationships, previously has been used successfully

by others (Maekawa et al., 2006; Rieder et al., 2005; Veenstra et al., 2005). Genetic distances were computed among haplotypes using the maximum likelihood composite model implemented in MEGA 4, then used to phylogenetically group the haplotypes using the neighbor-joining method. Phylogenetically related haplotypes were given group designations (phylogenetically grouped haplotype: PGH, groups A-F) for further statistical comparisons and analysis. This analysis revealed 6 distinct clades (groups). Much like haplotyping, grouping shared strong genealogically similar haplotypes substantially increases the statistical power of analyses by reducing the number of comparisons, in this instance dropping the potential number of haplotype combinations from 2 million to 64 (2^6). Genotyped individuals from the UTMB experimental population were then coded into haplotypes and, subsequently into clades. To ensure accuracy of reported results, I excluded from the analysis any individuals lacking defined genotype data for more than one SNP, as well as any individuals lacking identification of a single SNP that prevented the accurate assignment of full haplotypes.

ESTABLISHMENT OF CELL CULTURE FOR MECHANISTIC STUDIES

Due to the sheer number of cells required for mechanistic studies, it was impractical to use primary lymphocytes collected from human subjects. Therefore it was imperative to use cell lines representing the haplotypes evaluated. I used Epstein bar transformed human lymphoblast cells from the Coriell Institute Biorepository. These cells have the advantage of being originally obtained and sequenced for the HapMap project and, therefore, not only have known genotypes, but are the same genotypes which were used in the PHASE and haplotype grouping studies described previously. I chose cell lines based on haplotype grouping (PGH) determined in PHASE and Mega analysis as stated earlier. I used a recessive model for the determination of the effects of haplotypes on DRC, where I chose only cells that are homozygous for a given PGH.

Since the repository does not contain samples homozygous for either PGH-B or PGH-C, these PGHs were not evaluated in this study. Table 2 shows the cell lines used in our study. As a negative control, I used a cell line from a confirmed Xeroderma pigmentosum patient, denoted as XP-C. XP patients are characterized by reduced DNA repair due to mutation in XPC gene leading to a defective XPC protein.

Table 2: Coriell Institute Biorepository cell lines

Lines	PGH
GM12812	AA
GM10257	DD
GM12144	EE
GM11882	FF
GM02246	XP

Table 2: Coriell Institute Biorepository cell lines. The cell lines used for the functional analysis of the homozygous haplotypes. PGH lists the clade, Lines lists the designated Coriell cell line by the catalogue “GM” number.

Cell lines were shipped from the Repository as live cultures. After overnight acclimation in the incubator (humidified 37°C with 5% CO₂ in dark conditions), I cultured the cells under conditions optimized for doubling using CellBIND[®] Corning 100cm² low profile tissue culture flasks with growth medium (RPMI 1640 supplemented with Pen/Strep, with 15% heat inactivated premium fetal bovine serum and 2μM L-glutamine). The FBS was heat inactivated at 65°C with rocking for 30 minutes, then rapidly chilled on ice and prealiquoted for refreezing. FBS was refrozen no more than once after heat inactivation. Cells were disaggregated by aggressive pipetting before passage. Used media was removed by centrifugation at approximately 400g for 10 minutes at room temperature then decanted. Fresh culture media was added to cell pellet and disaggregated by aggressive pipetting before cell density and viability was determined using the Trypan blue exclusion assay. Cultures were maintained for 72 hours

(the optimized doubling time) at cells densities between 1.11×10^5 to 3.33×10^5 cells per mL (5-15 million cells per flask), with the higher end of the range preferred.

Low passage cells (passage number <15) were frozen for long term storage. I collected each culture as above, but the cell pellet was instead resuspended in cold 100% FBS at a 30 million per mL and 0.5mL per tube was transferred to individual cryotubes using a wide mouthed pipette. The remainder (0.5mL per tube) of the freezing solution was added on top of the cold cell mixture, which consisted of 90% FBS and 10% dimethylsulfoxide (DMSO), and then gently mixed by inversion. The cryotube was placed in a Nalgene “Mr. Frosty” freezing container containing isopropanol and allowed to controlled freeze at a rate of 1°C per minute overnight (16 hours) at -80°C. The next morning I transferred the cells to long term storage in liquid nitrogen.

Cells were cultured up to 50 passages or until viability fell below 80% by Trypan Blue exclusion assay, at which point I started fresh cultures from the frozen cryopreserved cells. Once cells were pulled from long term storage, they were thawed to slush conditions and carefully transferred to a large (50mL) sterile polypropylene centrifuge tube with 50x volume thawing media (RPMI 1640 with 20% FBS) using a wide mouth pipette and then gently mixed with soft pipetting. This tube was incubated overnight under culture condition with the cap cracked to allow air exchange and cells to reacclimatize for 16 hours. The medium containing the diluted DMSO was removed.

UV-B EXPOSURE OF CELL CULTURE FOR MECHANISTIC STUDIES

The mutagen sensitivity (MS) assay has been used with a number of compounds including the NNK. However, this is an indirect measure of DNA repair and does not differentiate between the different repair pathways. While there are NNK derived DNA-adducts that are repaired by NER (Brown et al., 2008; Peterson, 2010), there are a number of forms of damage and, therefore, repair processes (Affatato et al., 2004; Brown

and Massey, 2009; Lacoste et al., 2007). Additionally, there are metabolic activation and/or detoxification processes which also can confound the results (Smith et al., 1999) making mechanistic endpoints difficult to quantify. To avoid this, I chose instead to use ultraviolet (UV) radiation as my DNA damaging agent. Not only does UV-B (290-320nm) produce DNA adducts that are preferential NER substrates (Trego and Turchi, 2006), but also lack any biotransformation requirements to either produce the damage, damaging agent, or removal of the agent. This represents the cleanest system possible to study NER activity.

To prepare cultures for exposure, I pooled cultured flasks at a cell density of 15 million cells per flask (3.33×10^6 cells per mL) and then allowed the cells to recover for 16 hours under standard culture conditions. Immediately prior to UV exposure, I gently resuspended the cells by pipetting the media within the flask. One flask was designated for each individual time point per experiment, and cells were maintained in dark conditions to minimize the amount of spontaneous dimer photoreversion after exposure. Flasks containing the newly resuspended cells were placed under the exposure eye of an Oriel Instruments UV simulator and were irradiated at $35\text{mJ}/\text{cm}^2$ UV-B over 2 minutes. I monitored the dose intensity using a UV-B specific meter before and after exposures, and adjusted the intensity to account wavelength absorption for the culture medium and 1 side of the polystyrene flask.

UV-induced dimers were allowed to repair for designated times as depicted in Figure 4 under standard culture conditions. I collected DNA from exposed cells to measure the amount of UV-derived DNA adducts (CPDs and 6,4-PPs) and measure DRC. I also collected RNA for real time analysis to determine the amount of *XPC* transcriptional induction as well as protein to determine the total translational response to the DNA damage.

Figure 4: Endpoints after UV-induction

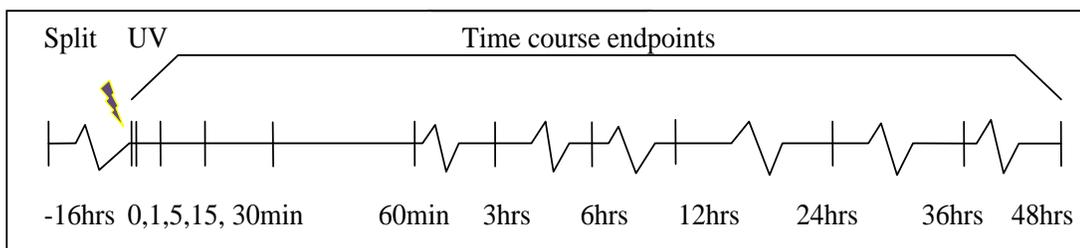


Figure 4: Endpoints after UV-induction. A diagram depicting the time points after exposure for the cell-based experiments. Cells were split 16 hours prior to UV-B (lightning bolt) exposure. UV was administered for 2 minutes, and time points ranging from 0 minutes post-exposure to 48 hours were used for endpoint analysis.

At each designated time point, the corresponding flask was removed from the incubator and rapidly chilled by immersion in wet ice for 15 minutes under dark conditions. I then transferred the chilled media suspension to a pre-chilled centrifuge tube and collected by centrifugation at 400g for 10 minutes at 4°C under dark conditions. The culture media was removed by decanting and the pellet allowed to drip dry for 15 minutes inverted at 4°C under dark conditions for DNA and protein samples, while the cell pellet was immediately snap frozen in liquid nitrogen for RNA samples. The pellets were stored at conditions optimized for each cell product. Cells were stored at -20°C under dark conditions for DNA isolation for adduct determination, at -20°C for protein isolation for XPC quantification, and at -80°C for RNA isolation for mRNA quantification. I exposed each cell line to UV in 3 separate experiments for all time points for each endpoint (DNA, RNA, or protein).

DNA ISOLATION AND QUANTIFICATION

I maintained dark conditions of the samples for all DNA work, and all samples were maintained on ice throughout all isolation and quantification experiments. Thirty minutes prior to processing, cell pellets were thawed on ice, and then the DNA was extracted using the QiaAmp DNA isolation mini-kit. In brief, I resuspended the cells in

phosphate buffered saline (PBS) and added kit supplied lysis buffer. The mixture was incubated for 15 minutes at 65°C with gentle rocking to lyse the cells, and protein was precipitated from the nucleic acids using the kit buffer followed by centrifugation as per manufacturer's instructions. The various cellular debris were collected into the pellet and the supernatant containing the DNA was transferred onto the membrane containing column, using centrifugation as per manufacturer's instructions. The DNA loaded membrane was washed with a series of ethanol containing buffers and dried using centrifugation as per manufacturer's instructions. To elute the purified DNA from the membrane, I centrifuged as per manufacturer's instructions using the kit specific elution buffer through the membrane, then transferred the eluent to a prelabelled screw topped microfuge tube with O-ring closure and stored the purified DNA at -20°C in a dark box until adduct determination. All procedures were conducted under dark conditions to minimize the amount of spontaneous dimer photoreversion post-exposure and for consistency between assay batches.

Prior to adduct analysis, I determined the concentration ($\mu\text{g}/\mu\text{L}$) of each DNA sample directly from the 280nm absorbance measurement and the purity of the sample from the 260/280nm absorbance ratio. Briefly, an aliquot of purified DNA was serially diluted with DNase free 20mM Tris-EDTA (TE) buffer. The dilution series was plated in duplicate into UV-transparent 96 well plates and read on SpectraMax190 plate reader with SoftMax Pro.

ELISA STANDARD PREPARATION

I maintained dark conditions of the samples for all DNA work, and standards were maintained on ice until plating. Standards of the UV-specific DNA adducts cyclopurimidine dimers (CPDs) and pyrimidine-6-4-pyrimidone photoproducts (6,4-PPs) were synthesized at the University of Texas Medical Branch Molecular Biosynthesis

Core Facility. In brief, short 10-mer oligos containing the sequence 5'-CGTATTATGC-3' were synthesized and exposed to irradiation to produce oligos containing a single dimer between positions 5 and 6. The CPDs and 6,4-PPs were separated from each other and the parent oligo using a Beckman-Coulter HPLC system with a Phenomenex Jupiter C18 reverse phase chromatography column. Purified dimers were dried completely and stored in dark containers as a powder at -20°C until reconstitution into 20nM in DNase-free TE buffer at a concentration of 10µM oligo. The stock solution was aliquoted and stored at -20°C. Stocks were thawed only once for dilution into standard curves. Concentrations ranging from 0 to 20nM were made by diluting the stock oligo solution with more DNase-free TE buffer. The concentration range was calculated based upon the linear range of the ELISA assay as determined during antibody specificity testing.

ADDUCT DETERMINATION BY ELISA

Amounts of UV-induced 6,4-PPs and CPDs were determined using an enzyme linked immunosorbent assay (ELISA), modified from CosmoBio's protocol. I maintained dark conditions of the samples for all DNA work, and all samples and standards were maintained on ice until plating. I precoated high binding affinity flat bottom ELISA plates with 0.003% protamine sulfate solution (50µL per well dried completely at 37°C), which were stored at room temperature until the start of the ELISA assay. Samples and pre-aliquoted standards were thawed on ice and the samples were diluted with TE buffer to 50µg per assay well. I prepared enough standards and samples (50µL per assay well) for plating in duplicate for each adduct to be tested, and boiled each for 15 minutes immediately followed by rapid chilling with wet ice for 5 minutes to denature the DNA and maintain the open conformation. Standards were loaded on the same plate as the test duplicate samples for every experiment, and every experiment was repeated to limit

variability. The loaded plates were once again dried completely at 37°C to adhere the DNA.

Once dried, I washed the DNA 5 times with 3x sample volume of ELISA wash buffer and blocked in 2x sample volume of 2%FBS for 30 minutes at 37°C with the plates covered with microplate film. After another 5 washes, the plates were incubated at 2x sample volume of 1:5,000 dilution of primary monoclonal mouse antibody for either 6,4-PP or CPD (64M-2 or TDM-2, respectively) at 37°C covered in microplate film for 30 minutes. After another 5 washes, the plates were then incubated at 2x sample volume of 1:10,000 dilution of rabbit anti-mouse secondary antibody at 37°C covered in microplate film for another 30 minutes. After yet another 5 washes, the plates were incubated with 3,3',5,5'-tetramethylbenzidine (TMB) at 37°C for 15 minutes and development was rapidly stopped with the addition of 2 N hydrochloric acid (HCl). The TMB and HCl were mixed by gently swirling the plate until all the blue was converted to a uniform (per well) yellow. To quantitate the amount of adduct present in sample, the plates were read at 420 nm on a SpectraMax plate reader with SoftMax software. I plotted each standard curve and extrapolated the sample concentrations for each corresponding plate.

PICOGREEN DNA DAMAGE ASSAY

The picogreen assay measures total DNA damage. As such, it is a fluorescence assay which uses a picogreen fluorescent “dye” that preferentially interacts with double stranded DNA. The positively charged nitrogen of the picogreen compound interacts with DNA phosphate group (negative) of the DNA minor groove. The dye intercalates into the DNA double strand uniformly. Using a method for alkaline denaturation of the DNA helical strand followed by neutralization renaturation allows relative measurement of the amount of DNA damage by fluorescence emission.

Cells were exposed using the protocol listed above with the following changes: at each time point an aliquot of cell culture was pipette into pre-chilled 15mL conical tubes with additional media to dilute the cells to 3×10^5 cells/mL, and the remaining cells were returned to the incubator until the next time point. In a 96 well microtiter assay plate, 2x sample volume TE buffer with 10% DMSO at pH7.4 was pre-plated into each well and the chilled and diluted cells were added. To each well, 2x sample volume of lysis buffer (9M urea, 0.1%SDS 0.2M EDTA at pH10.0) was added and mixed by careful pipetting. The plate was covered and incubated in the dark for 40 minutes at 4°C, then 2x sample volume denaturant solution (0.025M sodium hydroxide at pH 12.4) was added and mixed by careful pipetting. The plate was covered and incubated in the dark for 20 minutes at 37°C. The plate was immediately read on a Tecan spectrophotometer at ~480nm (excitation) (filter 485/20) and~ 520nm (emission) (filter 535/25).

Data was normalized to a blank for each plate and each well normalized to μ g protein as measured by Bradford assay. In brief, Bradford reagent was mixed 1:25 sample to 1x Bradford reagent and the absorbance read at 595nm on the Tecan spectrophotometer. Protein concentrations were extrapolated using a curve of absorbance from the BSA standards on the same plate. DNA damage was calculated as treated (sample) divided by untreated (control).

RNA ISOLATION

Approximately one hour prior to processing, snap frozen UV-treated cell pellets were thawed on ice, and then the RNA was extracted using the RNA Direct-zol mini-prep kit. In brief, I resuspended the cells by vortexing the pellets in TRI reagent for a final concentration of no more than 1×10^7 cells per mL. Cell debris were removed through centrifugation at 12,000g for 1 minute and the supernatant transferred to clean tubes in aliquotes of 350 μ L. Equal volumes of 100% ethanol were added to each sample and

vortexed, and the RNA was loaded onto the Zymo-Spin IIC by centrifugation at 12,000g for 1min. After discarding the eluent, the RNA was washed with the kit wash buffer by recentrifugation, then the sample was treated for contaminating DNA with DNase-I diluted in RNA wash buffer for 30 minutes at room temperature. The DNase solution was cleared by centrifugation and all protein removed with a wash series from the kit (1 pre-wash followed by 2 RNA wash sets) by centrifugation. The purified RNA was eluted from the column in DNase/RNase –free water using centrifugation at 12,000g for 30 seconds. The RNA was aliquoted and stored at -80°C until real time analysis. No sample was refrozen after freezing. Initial isolation samples were tested for concentration and purity between samples and sample isolation batches by Agilent assay at the UTMB Molecular Genomics Core. Subsequent tests used 260nm absorbance measurements for concentration and 260/280/230 nm ratio measurements to determine purity as needed.

RNA REAL-TIME ANALYSIS

Immediately before testing, I thawed the isolated RNA and all assay components on ice, and all reaction master mixes were made fresh immediately prior to use. I used a 2-step PCR process to quantitate the amount of *XPC* specific mRNA. To generate cDNA template from the RNA, the PCR mix contained TaqMan[®] high capacity RNA to cDNA buffer, TaqMan[®] high capacity reverse transcriptase enzyme, purified RNA template samples (or water as a no template control), and water to a total reaction volume of 20µL as per manufacturer's instructions. Thermal cycling was carried out in our laboratory on a MJ Research DNA Engine thermocycler equipped with a Chromo4 real-time PCR detection system under recommended conditions (25°C, 5 min; 42°C, 30 min; and 85°C, 5 min). Once cDNA was generated, it could be used immediately or frozen at -20°C before the second strand reaction.

I ran the second step of the 2-step real time analysis on the cDNAs generated by the high-capacity kit using PCR-based assay kits using the TaqMan[®] chemistry with primers specific for *XPC* mRNA (FAM fluorophore and appropriate quencher) or β -actin control (VIC fluorophore and appropriate quencher). The PCR consisted of TaqMan[®] universal master mix, template cDNA (or water as no template control), nuclease-free water, and both sample and control TaqMan[®] target-assay mixes in a total reaction volume of 20 μ l at concentrations recommended by Applied Biosystems. Thermal cycling was carried out in our laboratory on a MJ Research DNA Engine thermocycler equipped with a Chromo4 real-time PCR detection system under recommended conditions (50°C, 2 min; 95°C, 10 min; and 40 cycles at 95°C for 15 sec and 60°C for 1 min). Cycle threshold (Ct) was set to the lowest linear range for all samples per plate for both FAM and VIC together. I analyzed data using the comparative $\Delta\Delta$ Ct method with β -actin as the endogenous control per well and the no treatment sample as the referent. Each time point for each cell line was analyzed for fold-change ($2^{-\Delta\Delta C_t}$) compare to the no treatment control respective for the given experiment.

mFOLD RNA ANALYSIS

Folding bioinformatics was run using mFOLD (Zuker, 2003) online folding prediction server. Each haplotype virtual pre-mRNA was created by hand using a backbone sequence downloaded from NCBI Entrez database and the SNP positional information from the dbSNP database. Each sequence was divided uniformly into five sections for folding analysis due to the length limitations of the folding software. Current limitations in software coding disallowed full pre-mRNA folding analysis by any online server or desktop standalone application. For comparison, the lowest theoretical folding energies for each of the individual sequences were compared as a measure of the folding change.

PROTEIN ISOLATION AND ANALYSIS

Frozen UV-treated cell pellets were slowly thawed on ice, and cells were resuspended in fresh RIPA buffer plus protease inhibitors (1x protease inhibitor cocktail plus 1mM PMSF) with pipette mixing. I then sonicated each sample on ice for 1 minute in short bursts, taking care to minimize foaming as much as possible. Samples were then allowed to stand on ice 30 minutes to further reduce any foaming before centrifugation at 14,100g for 30 minutes at 4°C. Supernatant was stored at -20°C for further analysis. I measured the concentration of total protein in these samples using the bicinchoninic acid (BCA) assay. Briefly, samples are serially diluted with RIPA buffer plus protease inhibitors and pipetted (25µL) in duplicate into clear 96 well assay plates alongside bovine serum albumin (BSA) standards. A 10 point standard BSA curve ranging from 5µg to 2mg were pre-made by serial dilution and stored at -20°C (the 11th point as 0mg, or 100% diluent). An assay no protein blank of RIPA buffers plus protease inhibitors was loaded onto each plate as well. BCA working reagent (200µL) was mixed into each sample and incubated for 30 minutes at 37°C, and absorbance was read on a Tecan GENios Pro plate reader. I determined the total protein concentration of the samples by extrapolating from the linear regression analysis of the BSA standard curve in Excel and subtracting the no protein blank.

***XPC* ANTIBODY PREPERATION**

Due to inconsistency between preparations in commercial antibodies available for XPC, I chose to develop an in-house antibody with the help of the UTMB Protein Chemistry Core in the Biomolecular Resource Facility. In brief, a short – (910)DEEKQKLKGGPKKTKREKKA(929) – 19 amino acid peptide in the C-terminal

domain of XPC was synthesized and sent to a secondary facility where monoclonal antibodies against the single epitope were raised in a research rabbit, and the XPC antibody was affinity purified from the rabbit blood. The antibody was aliquoted and stored at -20°C for short term use, while the remaining aliquots were stored long term at -80°C. Western blot antibody concentrations were determined first by using standard dot blot procedures then confirmed with test samples using the western blot protocol described below. For the Dot Blot tests, the XPC antibody, and later protein samples, was spotted as drops on PVDF membrane and developed with ECLplus reagent and visualized using an Alpha Imager 2200. Both the primary and secondary antibody concentrations and the total protein concentrations were determined using side-by-side series dilution comparisons optimized for signal vs background intensities.

WESTERN BLOT PROTEIN ANALYSIS

I thawed the isolated protein, sample buffers, and molecular weight markers on ice and warmed the SDS-PAGE precast gel to room temperature. Aliquotes of 5µg of total protein were diluted with fresh RIPA buffer plus protease inhibitors to 10µL total volume, then the samples and markers (volumes as per manufacturer's instructions) were each mixed with equal volume of Lammelli loading dye and the samples only were boiled for 10 minutes. Samples were then chilled for 10 minutes and both the samples and the molecular weight markers were loaded into sample wells of the SDS-PAGE precast gel in the BioRad mini-Protean Western blot system. I loaded the gel and ran it at 90V for approximately 110 minutes, or until the blue dye front reached the bottom of the gel. The separated protein bands were then transferred to PVDF membranes using the BioRad mini-Protean transfer apparatus at 100V for 1 hour at 4°C with constant stirring. The transferred membrane was blocked with 5%BSA in TBS-T for 3 hours at room temperature with constant rocking and the transfer was confirmed with Commassie Blue

protein staining of the SDS-PAGE gel (destaining revealed no bands remained). Blots were then incubated with 1:5,000 rabbit anti-XPC antibody and 1:10,000 mouse anti- β -actin antibody in fresh blocking solution approximately 16 hours at 4°C with constant rocking. The blots were then washed in tris buffered saline with tween (TBS-T) for 15 minutes and two 5 minute washes at 4°C with constant rocking. The blots were then incubated with 1:7,500 anti-rabbit and anti-mouse HRP conjugated antibodies in TBS-T for 1 hour at 4°C with constant rocking. The blots were then washed with TBS-T for 15 minutes, then for twice for 5 minutes each, then in TBS twice for 5 minutes each. Finally, the blot was developed in the ECLplus (now known as ECL2) reagent system by incubating the washed blot for 5 minutes at room temperature in ECL working reagent as per manufacturer's instructions. I then visualized the membrane on the Alpha imager imaging system. XPC was visible at the expected 125kDa and the β -actin was visible at the expected 42kDa. The band intensity of both were measured with the ImageQuant software, and I analyzed the data as relative XPC concentrations using β -actin as the endogenous normalization control per lane and the no treatment sample as the referent. Each time point for each cell line was analyzed for fold-change compare to the no treatment control respective for the given experiment.

STATISTICAL ANALYSIS

In the prior study completed by other members of the laboratory (Hill et al., 2005), MS was evaluated using Cohen's kappa statistical test. A statistically significant value of $P < 0.001$ was obtained for both baseline and mutagen-induced CA, indicating that the agreement between the original and rescored data was not attributable to random chance.

Genotypes generated for all *XPC* htSNPs were analyzed for deviations from Hardy-Weinberg equilibrium (HWE) on a locus-by-locus basis using two methods

implemented in Linkage Disequilibrium Analyzer: one a standard 2-sided Pearson chi-squared test and the second a Monte Carlo permutation-based test. The chi-squared test is rapid and computationally simple, while the Monte Carlo permutation-based exact test estimates deviations from HWE. Any SNP failing these tests was excluded from any further analysis.

Each sample used in this study was coded for the presence (+) or absence (-) of each PGH. Using NCSS/PASS Dawson Edition software, I calculated the means and standard error of the mean (SEMs) for continuous variables and frequencies for categorical variables to characterize the study population. CA frequencies for each PGH (present vs. absent) were compared using preliminary Student's two-sample T-tests in combination with a permutation test with 1000 replicates to calculate empirical P-values, respectively for each PGH comparison, to account for multiple comparisons. Permutations test corrections are robust and benefit from having a empirical P-value that is constructed directly from experimental data (Cheverud, 2001). Comparisons were then repeated with stratification by smoking status (non-smokers vs. smokers). Graphical data is presented as means \pm SEM. A general linear statistical model with final parameters estimated from the comparison analysis was then fit to evaluate differences in CA frequency involving interactions between each PGH, genetic damage, and smoking, additionally adjusted for age and gender. Error-bar plots (depicting mean and 95% confidence interval limits corrected for the numbers of comparisons) to graphically visualize statistically significant interactions using SigmaPlot. $P \leq 0.05$ was defined as statistically significant.

For adduct analysis, I used SPSS18 to test for outliers and normality. I analyzed the amount of adduct for possible outliers within each time point for each cell line by inner quartile analysis. The inner quartile range (IQR) was calculated as $IQR=Q3-Q1$, with the first quartile (Q1) define as the 25th percentile and the third quartile (Q3) defined as the 75th percentile. As such, outliers were determined for each time point for each cell

line and were defined as any value that was outside the respective bounds [Q1-1.5 IQR, Q3+1.5 IQR]. Any data point that was defined as an outlier by this test was omitted from further analysis. Additionally, dataset normality was determined using the Shapiro Wilk test, and any failures were analyzed using non-parametric tests.

I used the amount of dimers remaining after the given repair time as a direct measure of the DRC, where comparisons between time points within a single cell line show the relative repair of the cell line, while between cell lines within a single time point show relative repair between haplotypes. I used linear regression within a cell line to determine the rate of repair over time, and used the comparison of regression lines for the impact of haplotype on rate. Again using SPSS18, I ran point-by-point comparisons using the Independent Sample Kruskal-Wallis test, while comparisons between cell lines or time points were determined using the Mann-Whitney U test. I ran the linear regression in NCSS using full and reduced models of the median values while the log-likelihood analysis was completed in S-PLUS. For all test, p values ≤ 0.05 were considered significant.

Similarly, I used the picogreen analysis to determine the total amount of DNA damage remaining after the given repair time as a direct measure of the total DRC. The relative repair of each cell line and between cell lines were calculated by the same method as for the adduct analysis, and linear regression for the amount of repair of over time. For all test, p values ≤ 0.05 were considered significant.

I used real-time threshold comparative analysis to determine the changes in transcription by measuring specific *XPC* mRNA against the control gene *β -actin*. The fold-changes for each sample were calculated independently and I used SPSS18 to test for outliers and normality. Specifically, I used the inner quartile analysis to test for outliers and the Shapiro Wilk test to test for normality. Any data point that was defined as an outlier was omitted from further analysis and any dataset failing the normality test was subsequently analyzed using non-parametric tests. Comparisons between time points

within a single cell line show the relative change in XPC expression over time, while between cell lines within a single time point show the relative influence of haplotype on XPC expression. Using SPSS18, I ran point-by-point comparisons using the Independent Sample Kruskal-Wallis test, while comparisons between cell lines or time points were determined using the Mann-Whitney U test. For all test, p values ≤ 0.05 were considered significant.

I used relative comparative analysis to determine the changes in translation by measuring XPC intensity against the no treatment sample for each experiment. Each XPC densitometry measurement was normalized independently to each sample's respective β -actin intensity. I used SPSS18 again to test for outliers (using the inner quartile analysis) and normality (with the Shapiro Wilk test). Any data point that was defined as an outlier was omitted from further analysis. Graphical data is presented as means \pm SEM. Comparisons between time points within a single cell line show the relative change in XPC translation over time, while between cell lines within a single time point show the relative influence of haplotype on XPC protein pools. Using SPSS18, I ran point-by-point comparisons using the T-test, while comparisons between cell lines or time points were determined using the ANOVA test. For all test, p values ≤ 0.05 were considered significant.

Chapter 3: RESULTS

OVERVIEW

The overarching goal of the study was to determine the biological significance of *XPC* haplotypes on the repair of DNA damage. I evaluated the effect of *XPC* haplotypes on DNA repair capacity (DRC) and levels of genetic damage, and the underlying mechanisms involved. To achieve this goal, the naturally occurring haplotypes were first established, their biological effect evaluated using a biomarker of cancer susceptibility on a population level, and the impact of the haplotypes on DRC determined as a loss of DNA damage over time after exposure. These haplotypes were analyzed for functional changes by the differences between haplotypes in the amount of mRNA and protein over time after exposure.

AIM 1: HAPLOTYPE DETERMINATION AND ASSOCIATIONS

The International HapMap Project is an open database available for download (www.hapmap.org) that contains bulk sequencing variation data for SNPs carried by the participating individuals, subdivided by population. The project officially began in 2002 and has since offered three unique builds that has integrated data sets from 2005, 2007, and 2009; the current data release for HapMap3 (phases I+II+III merged) is #28. The choice of using a reference population to run the haplotyping analysis stemmed from logistical concerns, as the type of next-generation/deep sequencing that would be required for this type of study was (and largely remains) cost prohibitive. Additionally, the HapMap data is subdivided into distinct ethnic populations. This is necessary as SNP frequencies can vary greatly between ethnicities (Fu et al., 2011). In an attempt to avoid confounding the results by admixtures, I focused on the HapMap population that most closely resembled the population of individuals previously tested in our laboratory, who

self-identified primarily as White non-Hispanic (Hill et al., 2005). Therefore, the CEU (Centre d'Etude du Polymorphisme Humain (CEPH) collection of Utah residents with ancestry from northern and western Europe) population, was used as the reference population as it represents a White non-Hispanic population. The CEU data was obtained from HapMap2 Data Release #22 (April 2007 release of phase II data) which includes NCBI B36 assembly of dbSNP b126. I focused on the data from chromosome 3 at base 14159650 to 14197142 on the minus strand. Of the 90+ SNPs documented in this region, there were 35 common SNPs in the CEU population, defined as minor allele frequency (MAF) of ≥ 0.05 (Table 3).

Tagger software (broadinstitute.org/mpg/tagger) was used to analyze the extent of linkage disequilibrium (LD) between the 35 SNPs for haplotype tagging SNP (htSNP) determination (de Bakker et al., 2005). The aggressive multi-marker approach (up to 6 markers) was coupled with conservative thresholds (r^2 threshold mean value of 0.971 and LOD of 2) (de Bakker et al., 2005; Goode et al., 2007; Nam et al., 2007), resulting in 11 htSNP (designated as 1-11 in the htSNP column of Table 3, and by RS# in Table 17-A and B of supplementary data). The htSNPs that were chosen to represent each set of SNPs in the LD group (1-11) are bold in Table 3. These bolded SNPs were used to design the genotyping reactions.

Table 3: HapMap breakdown of *XPC* SNPs

RS#	alleles	ancestral	htSNP	site
8516	C/T	T	9	3' UTR
10468	C/T	T	9	3' UTR
1126547	C/G	G	1	3' UTR
2470352	-/A/T*	A	2	3' UTR
2229090	C/G	C	9	3' UTR
2228001	A/C	C	3	K939Q
2733532	C/T	T	3	intron 15
2733533	A/C	C	11	intron 15
2733534	C/G	G	11	intron 15

2279017	G/T	G	3	intron 12
2470353	A/C/G**	G	11	intron 12
2607734	A/G	G	3	intron 11
2607736	A/G	G	3	intron 11
2607737	C/T	T	11	intron 11
3731149	A/C	A	8	intron 10
3731146	G/T	T	8	intron 10
9653966	G/T	T	4	intron 10
1124303	G/T	T	5	intron 10
3731143	C/T	T	6	intron 10
2228000	C/T	C	9	V499R
2227999	A/G	A	6	H492R
3731127	C/T	C	7	intron 8
3731125	A/G	A	4	intron 7
3731124	A/C	A	8	intron 7
13099160	A/G	A	7	intron 7
1106087	G/T	G	9	intron 5
3731108	C/T	C	8	intron 5
3731106	A/G	A	8	intron 5
3729587	C/G	G	8	intron 5
3731093	C/T	T	4	intron 3
2733537	A/G	A	10	intron 3
3731081	G/T	G	8	intron 3
3731068	A/C	C	8	intron 2
1350344	A/G	G	11	intron 1
2607775	C/G	C	11	5' UTR

Table 3: HapMap breakdown of XPC SNPs. “RS#” is the designated Reference SNP number as assigned by dbSNP NCI database, “alleles” are the respective Ancestral and Variant bases at the SNP, “ancestral” is the designated ancestral form of the SNP as assigned by the dbSNP NCI database, “htSNP” is the respective haplotype tagging SNP as determined from linkage analysis, and “site” is the relative genetic region of the gene. The bolded SNPs were the alleles chosen for Applied Biosystem assay development for htSNP genotyping. (*Null allele of rs2470352 is not expected to exist in the sampled population. It has not been reported for CEU population.)(**The A allele of 2470353 is not expected to exist in the sampled population. It has not been reported for the CEU population.)

Similarly, the CEU data also analyzed by PHASE analysis using the inference calculations corrected for trio data, as the CEU population contains designated mother, father, and child genotype data. The real power of haplotype analysis is that it can extract

the real haplotypes that are present (or estimated to be present) in a population of similar ethnicity (Browning and Browning, 2011; Srkar-Roy et al., 2011; Stephens and Donnelly, 2003). Using the 35 SNPs, theoretically 2^{35} (34,359,738,368) combinations could be visualized. Using the CEU population data for PHASE inference analysis resulted in only 21 unique real haplotypes. These are listed in Table 4. The 21 haplotypes are shown as rows, with columns as the allele (ancestral as black or variant as red) for each of the 35 positions.

Table 4: CEU inferred haplotypes for the *XPC* gene as determined by PHASE analysis

hap	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	
1	T	T	C	A	C	A	C	G	G	G	G	G	C	A	T	G	T	T	C	G	C	G	A	A	G	C	A	C	C	G	G	C	G	C	G	C
2	T	T	C	A	C	A	C	G	G	G	G	G	C	A	T	G	T	T	C	G	T	G	A	G	G	C	A	C	C	G	G	C	G	C	G	C
3	T	T	C	A	C	A	C	G	C	G	G	T	C	G	T	T	T	C	G	C	A	C	A	G	T	G	G	T	A	T	C	A	G			
4	T	T	C	A	C	A	C	G	C	G	G	T	C	G	T	T	T	C	G	C	A	C	A	G	T	G	G	T	A	T	A	A	G			
5	T	T	C	A	C	A	C	G	C	G	G	T	C	G	T	T	T	T	G	C	A	C	A	G	T	G	G	T	A	T	A	A	G			
6	T	T	C	A	C	A	C	G	C	G	G	T	C	G	T	G	T	C	G	C	A	C	A	G	T	G	G	T	A	T	A	A	G			
7	T	T	C	A	C	C	A	C	G	C	G	G	T	C	G	T	T	T	C	G	C	A	C	A	G	T	G	G	T	A	T	C	A	G		
8	T	T	C	A	C	T	C	G	T	G	A	G	C	A	T	T	T	T	C	G	C	A	A	A	G	C	A	C	T	A	G	C	G	C		
9	T	T	C	A	C	T	C	G	T	G	A	A	C	A	T	T	T	T	C	G	C	A	A	A	G	C	A	C	T	A	G	C	G	C		
10	T	T	C	A	G	A	C	G	G	G	G	G	C	A	T	T	T	T	T	G	C	A	A	A	T	C	A	C	T	G	G	C	G	C		
11	T	T	C	T	C	A	C	G	G	G	G	G	C	A	T	G	T	T	C	G	T	G	A	G	G	C	A	C	C	G	G	C	G	C		
12	T	T	C	T	C	A	C	G	C	G	G	T	C	G	T	G	T	C	G	C	A	C	A	G	T	G	G	T	A	T	A	A	G			
13	T	T	G	A	C	C	C	G	T	G	A	A	C	A	T	T	T	T	C	G	C	A	A	A	G	C	A	C	T	A	G	C	G	C		
14	T	T	G	A	C	T	C	G	T	G	A	A	C	A	T	T	T	T	C	G	C	A	A	A	G	C	A	C	T	A	G	C	G	C		
15	T	C	C	A	C	A	C	G	C	G	G	T	A	G	T	T	T	C	G	C	A	A	A	G	C	G	G	T	A	G	C	A	G			
16	C	C	C	A	C	A	C	G	C	G	G	T	A	T	T	C	T	A	C	A	A	A	T	C	A	C	T	G	G	C	A	G				
17	C	C	C	A	G	A	C	G	C	G	G	T	A	T	T	T	T	T	G	C	A	A	A	T	C	A	C	T	G	G	C	A	G			
18	C	C	C	A	G	A	C	G	C	G	G	T	A	T	T	C	T	A	C	A	A	A	T	C	A	C	T	G	G	C	A	G				
19	C	C	C	T	G	A	C	G	C	G	G	G	G	C	A	T	T	T	T	G	C	A	A	A	T	C	A	C	T	G	G	C	G	C		
20	C	C	C	T	G	A	C	G	C	G	G	T	A	T	T	T	T	T	T	G	C	A	A	A	T	C	A	C	T	G	G	C	A	G		
21	C	C	C	T	G	A	C	G	C	G	G	T	A	T	T	C	T	A	C	A	A	A	T	C	A	C	T	G	G	C	A	G				

Table 4: CEU inferred haplotypes for the *XPC* gene as determined by PHASE analysis. Haplotypes were coded into numbers 1-21 and are listed in rows. The columns are the individual 35 SNPs, with ancestral alleles presented as black text and variant alleles presented as red.

However, humans are diplotype, and the theoretical number of diplotypes possible given the 21 haplotypes is 2^{21} (2,097,152). Again, the potential number of combinations is problematic due to the large number. However, as SNPs are in LD with each other due to recombination events, the haplotypes themselves share similarities. As such, these haplotypes can be analyzed for their evolutionary relatedness and theoretically could be grouped together using a phylogenetic approach. Mega works by grouping haplotypes based on the likelihood estimates of genetic distance. In other words, the program groups haplotypes together based on how many and what kinds of changes are required to change from one haplotype to another. Using Mega 4 phylogenetic analysis, I was able to calculate theoretical genetic distances between the haplotypes and group them as per Figure 5. This distance is indicative of the number and complexity of the change. For example, haplotype 1 has “fewer” changes than 8 to become haplotype 19, although both less than what would be required for haplotype 14. As such, there is the least genetic distance between 1 and 19, more between 9 and 19, and even more than 14 and 19.

This analysis revealed 6 distinct clades A through F (Phylogenetically Grouped Haplotype: PGH). The percent genetic divergence within PGH clades (Table 5A) and between PGH clades (Table 5B) are presented below. The percent genetic divergence within groups was 8.6% or less (Table 5A), while between the PGHs ranged from 18.6% to 57.3% (Table 5B and Table 21 of supplemental data). Objectively, there is no defining limits for clade definition, yet the grouping of haplotypes, based on genealogical or phenotypic relationships, previously has been used successfully by others (Maekawa et al., 2006; Rieder et al., 2005; Veenstra et al., 2005). The natural groups as depicted in the tree of Figure 5 were additionally supported by bootstrapping analysis ($\geq 90\%$). Much like haplotyping, grouping shared strong genealogically similar haplotypes substantially increases the statistical power of analyses by reducing the number of comparisons, in this instance dropping the theoretical number of haplotype diplotype combinations down to 2^6 (64), thereby increasing the likelihood of finding these groups in a smaller population.

Table 6 shows the haplotypes in detail arranged by group. The red and black clusters show the extent of sequence similarity between the groups.

Figure 5: Phylogenetic analysis of the 21 haplotypes

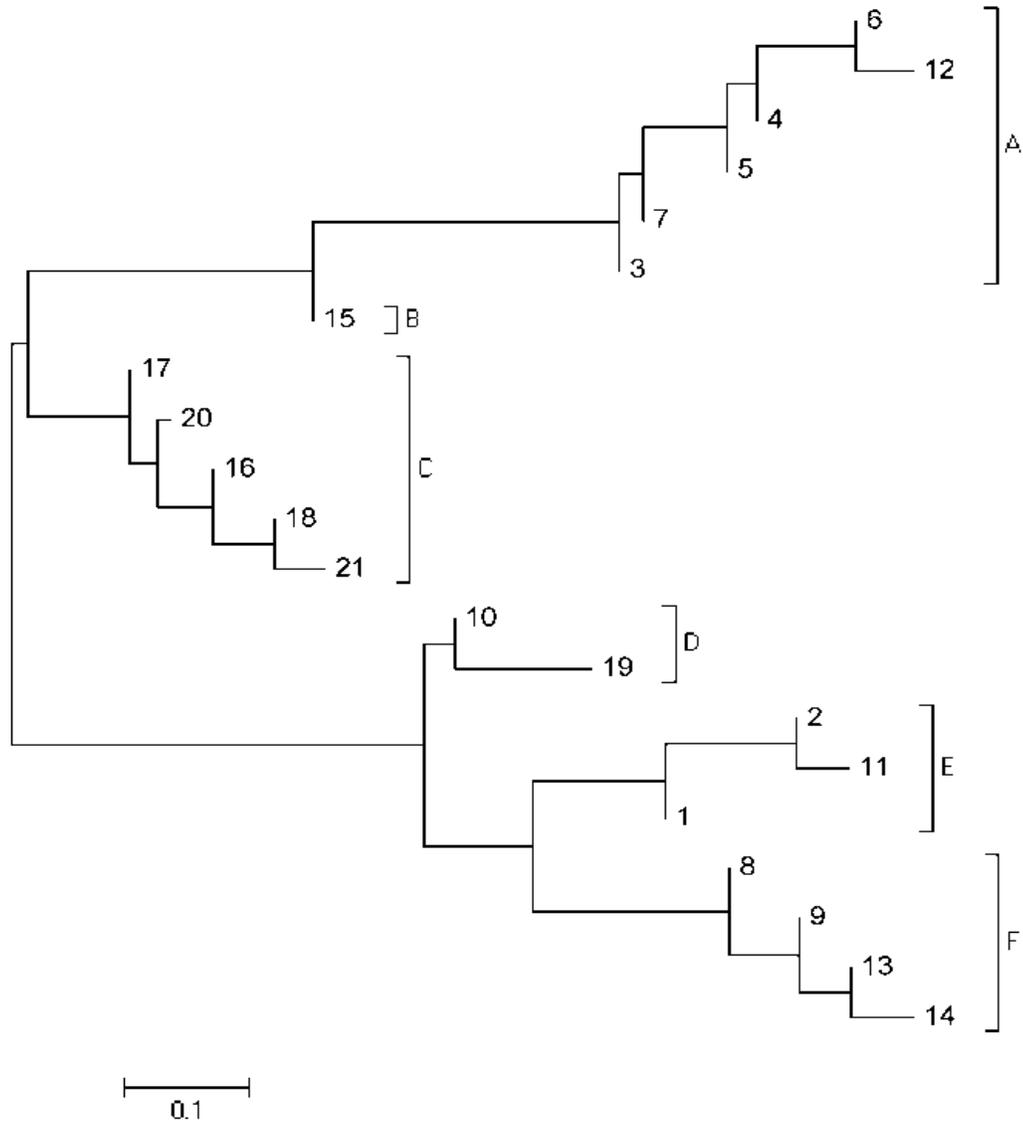


Figure 5: Phylogenetic analysis of the 21 haplotypes. The haplotypes listed by number correspond to the designations listed in Table 4. The length of the line represents the genetic distance between haplotypes

Table 5: The percent genetic divergence between the 6 clades

A

Grp	distance
A	5.9
B	n/c
C	6.3
D	8.6
E	5.7
F	4.8

B

	A	B	C	D	E	F
A		18.6	44.1	55.7	57.3	54.8
B	18.6		26.9	41.4	44.8	42.9
C	44.1	26.9		25.4	48	52.6
D	55.7	41.4	25.4		25.2	30
E	57.3	44.8	48	25.2		30.5
F	54.8	42.9	52.6	30	30.5	

Table 5: The percent genetic divergence between the 6 clades. A: The percent genetic divergence within each PGH from figure 5 as computed by Mega analysis. **B:** The percent genetic divergence between each PGH from figure 5 as computed by Mega analysis.

Table 6: CEU inferred haplotypes for the *XPC* gene as determined by PHASE analysis and clustered by haplotype

RS#	Group	Freq	8516	10468	1126547	2470352
Ancestral			T	T	G	A
Variant			C	C	C	T
haplotype (PGH)	%	3' UTR	3' UTR	3' UTR	3' UTR	3' UTR
3	A		T	T	C	A
4	A		T	T	C	A
5	A	20.83	T	T	C	A
6	A		T	T	C	A
7	A		T	T	C	A
12	A		T	T	C	T
15	B	1.67	T	C	C	A
16	C		C	C	C	A
17	C		C	C	C	A
18	C	26.66	C	C	C	A
20	C		C	C	C	T
21	C		C	C	C	T
10	D	2.50	T	T	C	A
19	D		C	C	C	T
2	E		T	T	C	A
11	E	8.33	T	T	C	T
1	E		T	T	C	A
8	F		T	T	C	A
9	F	40.00	T	T	C	A
13	F		T	T	G	A
14	F		T	T	G	A

Table 7: Demographic breakdown of the UTMB experimental population

Characteristic	N(%)
Total N	99 (100)
Sex	
Male	21 (21.2)
Female	78 (78.8)
Smoking Status	
Non Smoker	49 (49.5)
Smoker	50 (50.5)
Haplotype Frequency (%)	
A	26
B	3
C	20
D	4
E	7
F	40
	Mean (SEM)
Age	39 (1.30)
Cigarettes/day	17.6 (1.47)
Years smoked	19.9 (1.81)
Pack-years	18.2 (2.67)
CA (baseline)	0.79 (0.10)
CA (MS)	5.24 (0.29)

Table 7: Demographic breakdown of the UTMB experimental population. Breakdown by sex, age, smoking use, haplotype, and cytogenetics. The haplotypes are presented as a percentage of the total number of chromosomes (2 for each individual therefore percentage is out of 198 chromosomes). Cytogenetic data is presented as the mean for the total population (all 99 individuals) with chromosomal aberrations (CA) at baseline and after mutagen sensitivity (MS) analysis.

I performed correlation analysis on the CA data using a dominant “haplotype group copy”(HGC) model whereby either one or two HGCs were counted as one (presence; +) and zero HGCs were counted as none (absence; -). CA data was broken down into baseline (no 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone (NNK) treatment) or mutagen induced (72hours after NNK treatment). After adjusting for age

and sex, a general linear model for the effect of either smoking or PGHs alone did not show any correlation on CA frequencies. However, there was a statistically significant interaction between haplotype and smoking for group C ($p=0.046$), with smokers carrying PGH-C have increased CA frequencies at baseline. The lowest baseline level of CAs were seen in non-smokers lacking PGH-C (mean \pm SEM = 0.53 ± 0.192), which was significantly lower than smokers carrying PGH-C (mean \pm SEM = 1.21 ± 0.29). Comparison of smoking status in the group C positive individuals showed a 3.5 times higher CA frequency in smokers compared to non-smokers. The breakdown of CA data is listed in table 20 of the supplementary data, as well as graphically represented in Figure 17 panel A (baseline) and panel B (mutagen induced) also in the supplementary data. For clarity, I am presenting only the data that showed statistical significance. Figure 6 shows the CA data at baseline for individuals with or without PGH-C. Individuals who self-report smoking cigarettes (designated as “smokers”) are in red, while non-smokers are in blue.

Baseline levels of CA are indicative of the long term DNA damage, but do not reflect the short-term response to an acute exposure. For acute response, the mutagen sensitivity assay was used. In this assay, isolated cells were exposed to NNK for 1 hour and then allowed to repair for 72 after removal of the mutagen. Figure 7 shows the CA data after mutagen treatment for individuals with or without PGH-D (panel A) and PGH-F (panel B). Individuals who self-report smoking cigarettes (designated as “smokers”) are in red, while non-smokers are in blue. There was a statistically significant interaction (P values of ≥ 0.05) between haplotype and smoking for both groups D ($p=0.023$) and F ($p=0.031$), with smokers who had either group carrying a higher CA load after exposure. Comparison of smoking status in PGH-D individuals showed that non-smokers had significantly lower CA frequencies after MS exposure (mean \pm SEM = 3.75 ± 0.85) than smokers (mean \pm SEM = 8.75 ± 2.43). PGH-D non-smokers had 2.3 times lower CA

Figure 6: CA frequency divided by clade and smoking status at baseline

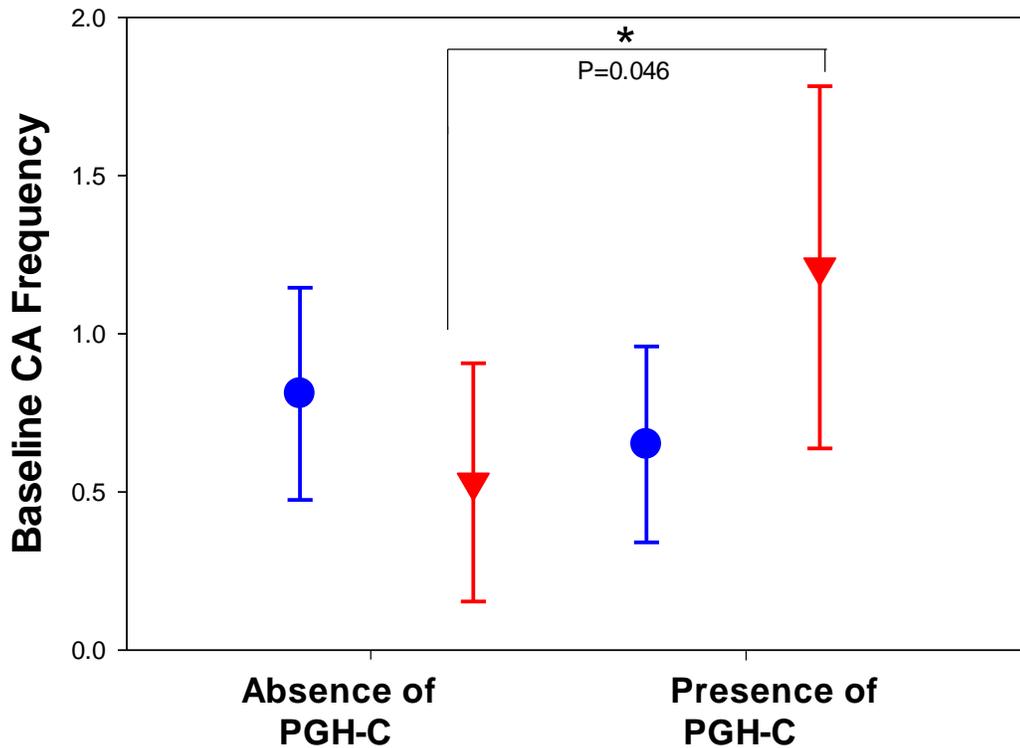


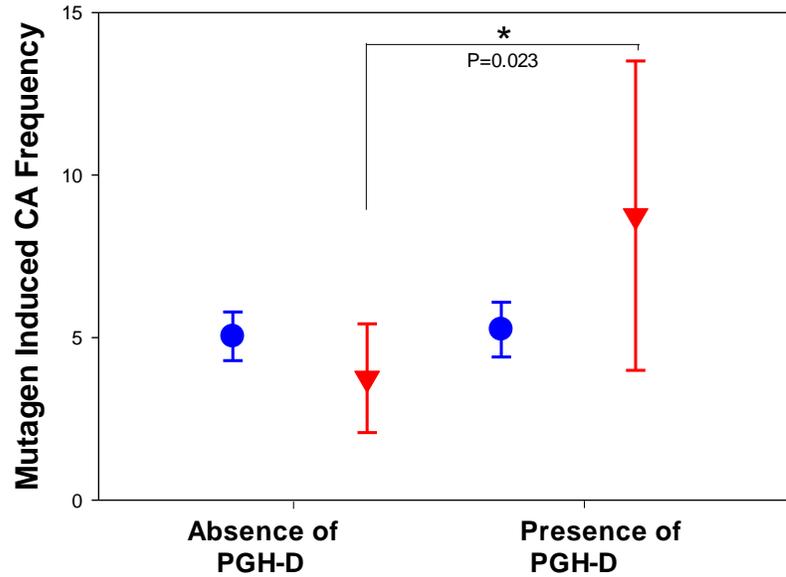
Figure 6: CA frequency divided by clade and smoking status at baseline. Smokers are in red, non-smokers in blue. Phylogenetically grouped haplotype (PGH)-C showed a statistically significant interaction between haplotype and smoking, with smokers who have PGH-C carrying a higher CA load at baseline. Smokers with C N=20, without C N=69; Non-Smokers with C N= 16, without C N=68. Statistical significance was defined as P values of ≤ 0.05 (*).

frequencies than smokers in the MS experiment. A similar comparison for group F individuals showed that smokers had a 1.3 times higher frequency of mutagen-induced CAs (mean \pm SEM = 6.03 ± 0.51) compared to non-smokers (mean \pm SEM = 4.63 ± 0.47).

To summarize, the data from Aim 1 revealed the presence of 21 haplotypes that covered 35 the common SNPs ($MAF \geq 0.05$) spanning the *XPC* genomic region. Furthermore, these haplotypes were clustered together by their relatedness (as defined by percent sequence divergence) into 6 distinct clades (PGHs A-F). Using CAs as a

Figure 7: CA frequency divided by clade and smoking status after mutagen exposure

A



B

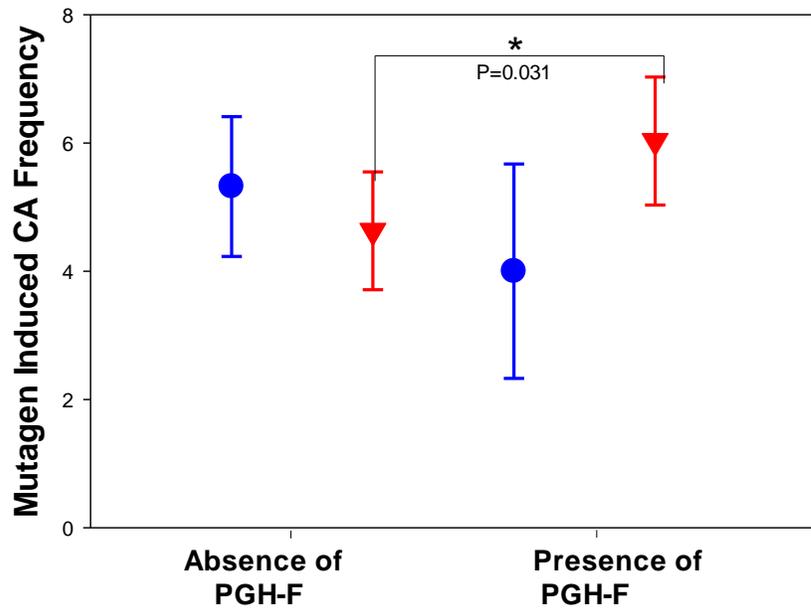


Figure 7: CA frequency divided by clade and smoking status after mutagen exposure. Smokers are in red, non-smokers in blue. **A:** PGH-D showed a statistically significant interaction between haplotype and smoking, with smokers who have PGH-D carrying a higher CA load after mutagen exposure. Smokers with D N=3, without D N=86; Non-Smokers with D N= 4, without C N=80. **B:** PGH-F showed a statistically significant interaction between haplotype and smoking, with smokers who have PGH-F carrying a higher CA load at baseline. Smokers with F N=39, without F N=50; Non-Smokers with F N= 26, without F N=58. Statistical significance was defined as P values of ≤ 0.05 (*).

biomarker, the data indicate a significant interaction between PGH-C and cigarette smoking, while a significant interaction between PGHs D and F and cigarette smoking was observed after mutagen exposure. The results suggest smokers (and possibly other individuals undergoing chronic DNA insult) with PGH-C may have a higher risk of developing genomic instability. Similarly, the results suggest that individuals with PGHs D or F may also be predisposed to developing cancer in response to an acute genotoxic dose, as reduced DRC predisposes cells to genomic instability. Together, the results indicate a clear gene-environment interaction that can have implications in disease predisposition and clinical health concerns.

AIM 2: DRC ANALYSIS

To determine the underlying mechanisms that drive this differential correlation between mutagen-induced CAs and PGHs, I measured the DRC. A concern for using cigarette smoking as a model of genotoxicity is the complexity of the mixture. Cigarette smoke contains more than 5000 compounds, over 60 of which are classified as carcinogenic (Klassen, 2001; Peterson, 2010). These mutagenic agents are aerosolized and inhaled as a complex mixture, which has the potential to activate a number of DNA repair pathways, among many others (Klassen, 2001). A good example of this is the aforementioned NNK. While there are NNK derived DNA-adducts that are repaired by NER (Brown et al., 2008; Peterson, 2010), the repair of NNK derived damage is not NER specific, and repair is complicated by overlap of alternate forms of damage and, therefore, repair processes such as BER or specific glycosylases (Affatato et al., 2004; Brown and Massey, 2009; Lacoste et al., 2007). Additionally, there are various biotransformation enzymes that act on NNK to produce additional metabolites or reduce the active intermediates (Smith et al., 1999). To avoid confounding reactions of these

other pathways, I focused on the NER pathway by using a mutagen that is repaired predominantly by NER. Ultraviolet radiation, specifically in the region of 290-320nm (UV-B), produces adducts that are preferentially repaired by NER (Trego and Turchi, 2006). Additionally, UV-B has the added benefit of lacking any biotransformation requirements to either produce the damage, damaging agent, or removal of the agent. This represents the cleanest system possible to study NER activity.

Furthermore, the large number of cells required for mechanistic studies made it impractical to use primary lymphocytes collected from human subjects. To overcome this limitation, I used lymphoblast cells derived from members of the CEU populations (previously used as the reference population in Aim 1) that were transformed with Epstein bar and cultivated by the Coriell Insititue Biorepositiory. The cell lines were chosen to represent the homozygous forms of each PGH, to avoid confounding influences of clade mixtures. This limited the analysis to 4 cell lines, designated AA, DD, EE, and FF, as there were no homozygous forms for clades B or C present in the CEU population due to the rarity of these haplotype groups. Table 8 panel A shows the percent genetic divergence (%D) between the haplotypes within a designated PGH, while panel B shows all the percent genetic divergences between all of the haplotypes in these cell lines. The percent genetic divergence between PGHs ranged from 25.7% (DD haplotype 19 and EE haplotype 1) up to 62.9% (AA haplotype 12 and EE haplotype 2). For comparison, the full list of percent genetic divergences is given in Table 21 in the supplemental data section.

Exposure conditions were determined using a dose response study whereby cells for PGH AA and XP were exposed to UV-B for a total dose of 10-50mJ/cm², and cut offs of 90% viability (dashed black line) versus no treatment at 48 hours and 85% viability (solid black line) versus no treatment at 72 hours was used to determine the optimal damage causing dose that did not induce cell death. Figure 8 shows the relative viability

Table 8: Percent genetic divergence between haplotypes comprising each PGH

A						B						
GM	Hap1	PGH1	Hap2	PGH2	%D	%D	7	12	19	1	2	8
AA	7	A	12	A	11.4	7		11.4	60.0	51.4	57.1	45.7
DD	19	D	19	D	na	12	11.4		60.0	57.1	62.9	57.1
EE	1	E	2	E	5.7	19	60	60		25.7	31.4	31.4
FF	8	F	8	F	na	1	51.4	57.1	25.7		5.7	22.9
						2	57.1	62.9	31.4	5.7		28.6
						8	45.7	57.1	31.4	22.9	28.9	

Table 8: Percent genetic divergence between haplotypes comprising each PGH represented by the tested cell lines. Numbers correspond to the haplotypes as listed in table 6 figure 5. **A:** The percent genetic divergence between the haplotypes within a given PGH listed by haplotype and clade for each cell line tested. **B:** The percent genetic divergence between all haplotypes for each cell line tested.

over 24, 48, and 72 hours after exposure for both cell lines. The green bars represent the PGH AA cell line. As a control, the cut off for the XP cell line (represented as the purple bars) viability was 75% compared to no treatment control at 48 hours (dashed blue line) and 70% compared to no treatment control at 72 hours (solid blue line). Using these criteria, the optimal dose was estimated to be 35mJ/cm² total exposure (red line). An additional cell line was obtained from Coriell (GM02246) that was derived from a clinically diagnosed xeroderma pigmentosum group C patient. The patient has been characterized as having a cDNA dinucleotide frameshift deletion in the *XPC* gene, creating a premature stop codon, though the deletion has not been determined as homozygous or hemizygous, which has been published as having repressed (but not completely eradicated) repair as determined by UV-induced unscheduled synthesis (Khan et al., 2006). Repair in these cells is expected to be weak and likely to be from overlapping functions of other pathways, the XP cells were only used as a negative control in the dose response viability studies to determine the optimal UV-B exposure and not analyzed further.

Figure 8: Cell Viability after UV-B treatment

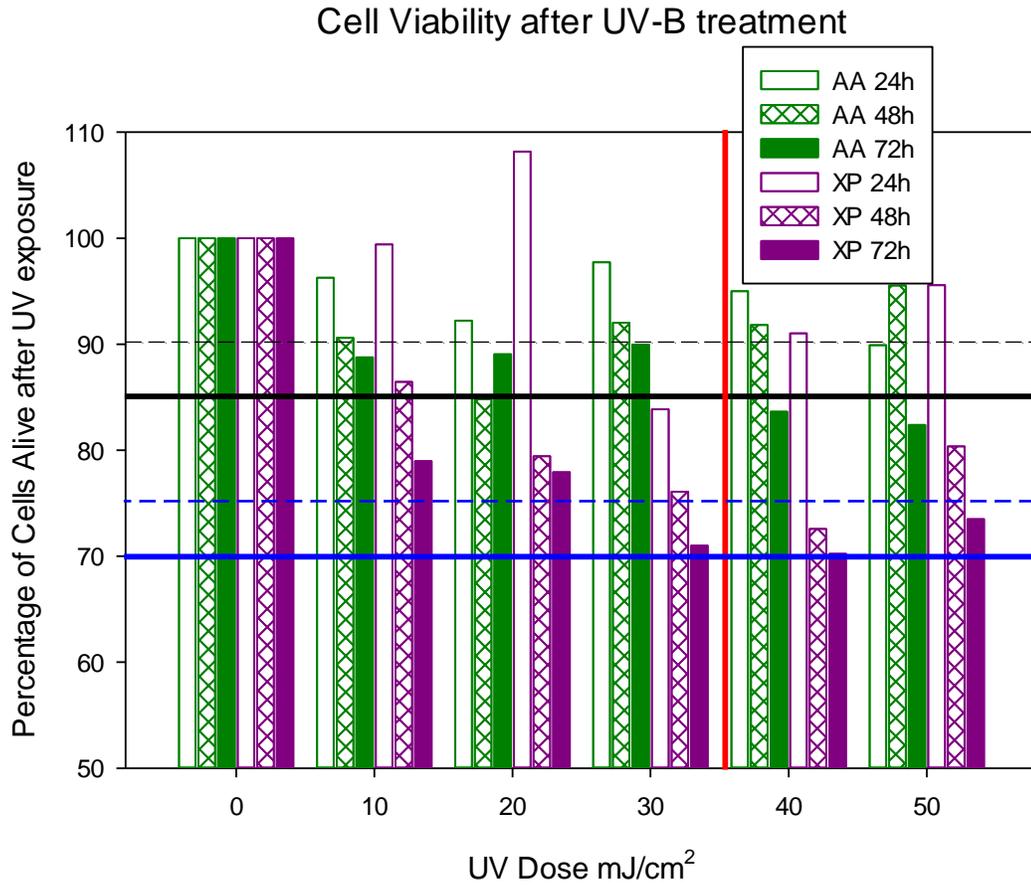


Figure 8: Cell Viability after UV-B treatment. A dose response study for lines AA (green) and XP (purple) at 24 hours (outline), 48 hours (hashed), and 72 hours (solid) after UV-B treatment. Cut off viability criteria was set to AA at 48h of 90% (dashed black), AA at 72h of 85% (solid black), XP at 48h of 75% (dashed blue) and XP at 72 h of 70% (solid blue). The final exposure concentration was chosen to be 35mJ/cm² (red line).

The dose was confirmed for each of the 4 haplotype cell lines by measuring doubling time as seen in Figure 9 (corresponding data table is found in the supplemental data section, Table 22). The solid color bars represent the exposed cells, while the hashed bars represent the untreated cells. The solid line at 100 represents the initial baseline amount of cells at seeding (cell split at 16 hours prior to exposure). The dashed line is the doubling level. Untreated (hashed bars) cell lines were expected to have doubled by the 48h time point, while some of the treated (solid bars) showed delayed growth (AA and

EE) and did not pass the doubling line until the 72h time point. This delay was not unexpected, and may be explained given XPC's role in cell cycle control (Melis et al., 2011; Ray et al., 2013). Concurrently, the viability of these exposed cells were confirmed to lose <30% viability as compared to unexposed cells of the same cell line (NT – no treatment). (See Figure 18 and Table 23 of the supplemental data section.)

Figure 9: Cell growth after UV-B treatment

Growth vs Initial Seeding for 35mJ/cm² by PGH

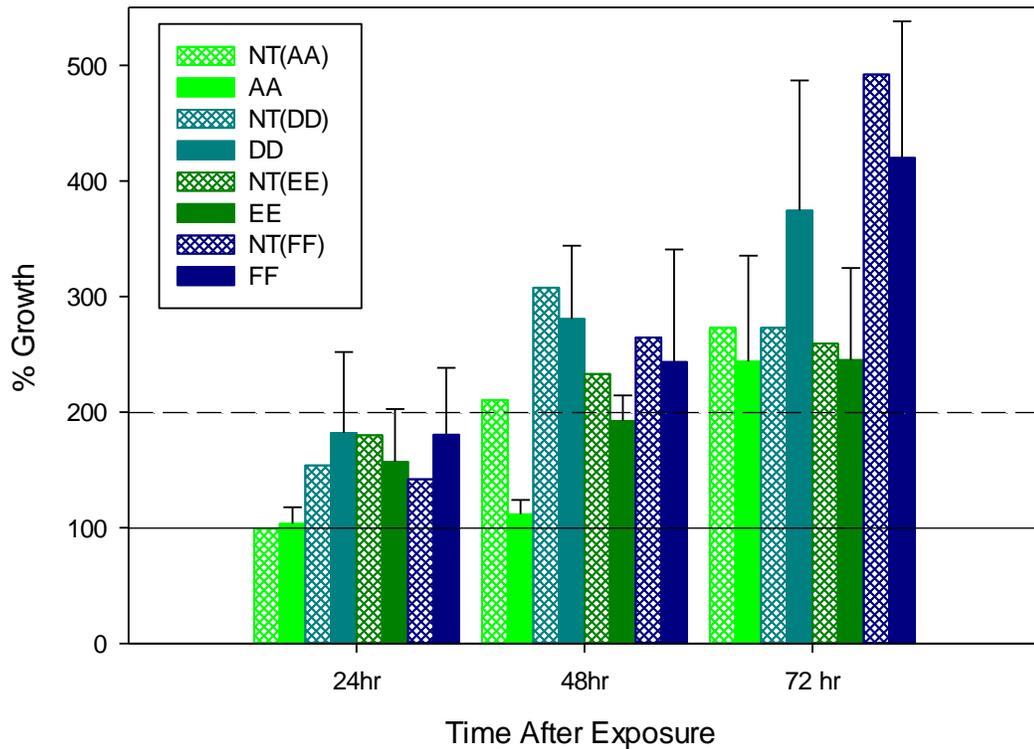


Figure 9: Cell growth after UV-B treatment. Test of exposure concentration (35m/cm²) was tested for all 4 PGHs at 24h, 48h, and 72h post UV-B exposure. PGH AA in bright green, DD in light blue, DD in dark green, and FF in dark blue. The hashed bars are untreated cells for the respective cell line, for comparison. The solid line at 100 represents the initial baseline amount of cells at seeding (cell split at 16 hours prior to exposure). The dashed line is the doubling level. Untreated (hashed bars) cell lines were expected to have doubled by the 48h time point, while treated (solid bars) showed

delayed growth in some lines (AA and EE) and did not pass the doubling line until the 72h time point.

The amount of the two major UV-B DNA damage adducts – cyclopyrimidine dimers (CPDs) and 6-4 Photoproducts (6,4-PPs) – were determined by ELISA analysis using antibodies specific for each of the respective adduct dimers. Standard curves were generated from the sequence 5'-CGTATTATGC-3' and used for antibody specificity testing. Figure 10 shows the specificity of the antibody against the expected oligo (CPDs for panel A and 6,4-PPs for panel B), the alternative oligo (6,4-PPs for panel A and CPDs for panel B), and the parent oligo (no adduct). The values for the standards were chosen to be in the linear range and to have little to no cross reactivity. Due to the variability of the biological samples, the data was analyzed using non-parametric methods. (The Shapiro-Wilk normality test values are listed in Table 24 of the supplemental data section.) As the medians are a better representation of the central tendency for non-normal data, Table 9 shows the median values per time point after UV-B exposure for each of the respective haplotype groups, with the error as the approximate 95% CI using the formulas stated in the statistics section of the methods chapter. Panel A of Figure 10 shows the amount of CPD adduct as pmol dimer per 50µg DNA above the baseline (no UV-B treatment) while panel B shows the 6,4-PP adduct as pmol dimer per 50µg DNA above the baseline.

The haplotypes were tested for significant differences between each time point both within the haplotypes (supplemental tables 25, panels A-H, significant p-values ($p \leq 0.05$) in italics) and between the haplotypes (supplemental tables 26, panels A and B, significant p-values ($p \leq 0.05$) in italics) using the Mann-Whitney test (non-parametric T-test). For clarity, I am presenting the data as median only graphs. Figure 11 panel A shows the amount of remaining CPD adduct in pmol/50µg DNA above baseline no treatment for each haplotype as a function of time in minutes after UV-B exposure, while

panel B shows the amount of remaining 6,4-PP adduct in pmol/50 μ g DNA above baseline no treatment for each haplotype as a function of time after UV-B exposure. The vertical line emphasizes the 15 minute time point that was used as the starting point for the linear regression analysis.

Figure 10: Test of antibody specificity

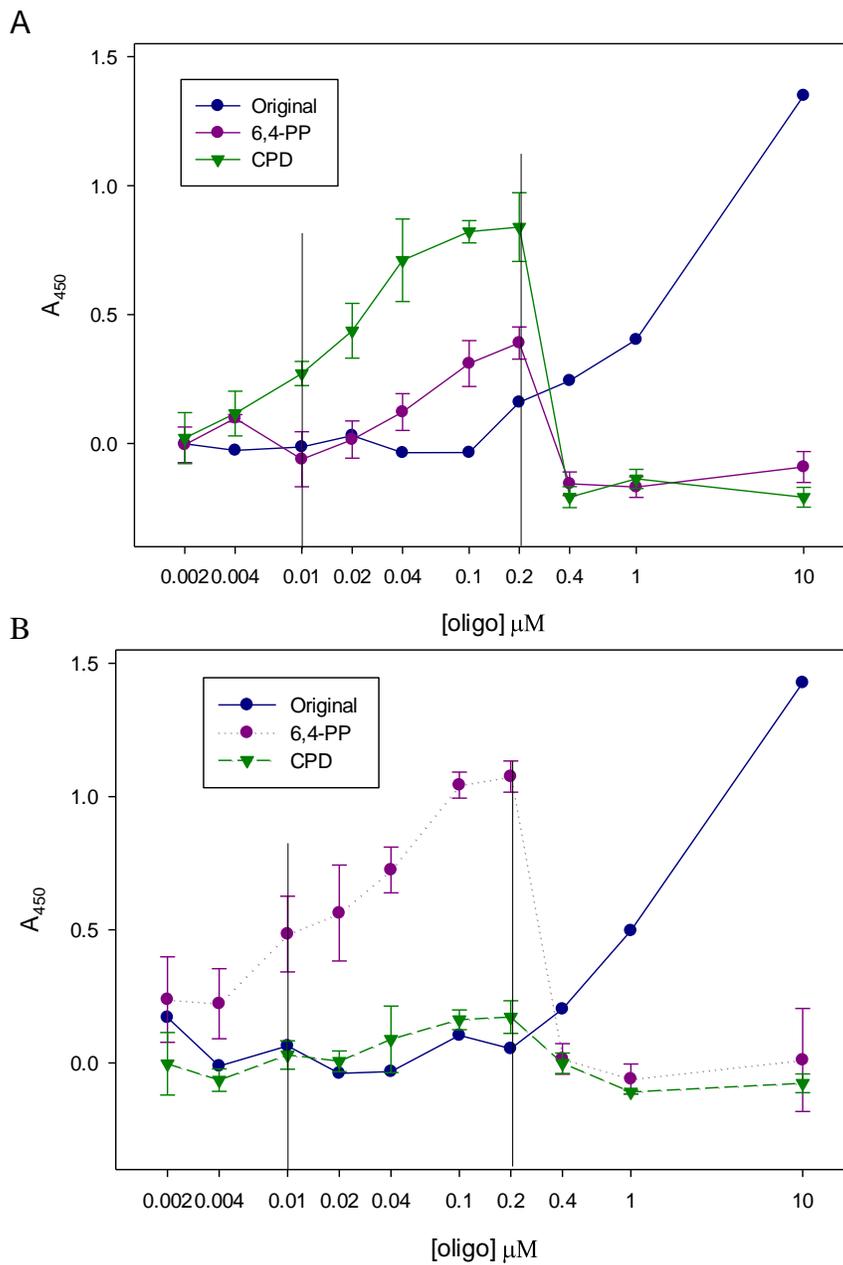


Figure 10: Test of antibody specificity. **A:** Test of CPD specific antibody against isolated CPD adduct oligo in green, against 6,4-PP adduct specific oligo in purple, and against parent/no adduct oligo in blue. **B:** Test of 6-4PP specific antibody against isolated CPD adduct oligo in green, against 6,4-PP adduct specific oligo in purple, and against parent/no adduct oligo in blue. Drop lines indicate the values chosen for the standards with criteria determined to be in the linear range and have little to no cross reactivity.

Table 9: Measure of central tendency for CPDs and 6-4PPs at intervals of repair after UV-B exposure

A

CPD adduct remaining (pmol/50ugDNA) above baseline (no treatment)				
Time Point min	AA	DD	EE	
0min	0	1.581 (0.246-5.896)	9.446 (3.732-14.302)	-3.4035 (-7.535-10.6)
1min	1	3.398 (-0.625-12.543)	6.999 (2.427-9.986)	0.5675 (-1.135-7.259)
5min	5	-0.568 (-2.894-6.222)	5.221 (-2.557-10.193)	-0.3995 (-0.719-6.803)
15min	15	0.9795 (-4.808-2.92)	7.4575 (-6.396-10.149)	0.298 (0.021-9.642)
30min	30	4.975 (2.381-7.963)	9.958 (5.103-13.279)	0.7685 (0.399-10.916)
1hr	60	4.206 (2.117-6.599)	15.414 (5.202-22.865)	2.0745 (-0.295-4.11)
3hr	180	1.884 (-2.426-7.349)	7.8615 (-13.065-12.18)	-0.6895 (-1.039-10.957)
6hr	360	3.391 (-7.481-7.558)	4.7935 (-16.149-12.348)	1.124 (0.593-6.18)
12hr	720	2.957 (-5.314-8.229)	7.942 (1.272-11.524)	1.459 (0.186-3.908)
24hr	1440	3.2235 (-1.811-14.898)	5.2368 (-10.337-9.125)	1.367 (-0.09-4.93)

FP	AAEE	DDFF
9.7985 (-2.167- 116.199)	1.131 (-3.729-5.896)	9.455 (2.632-17.434)
3.103 (-20.641-21.205)	0.78 (-0.625-7.259)	6.786 (-0.323-9.986)
14.232 (3.048-93.503)	-0.3995 (-0.854-6.115)	9.4795 (1.059-12.938)
15.2045 (-1.508- 169.929)	0.3445 (0.024-2.648)	8.394 (-1.231-16.153)
16.734 (6.164-127.291)	2.959 (0.518-7.963)	11.955 (7.334-16.042)
39.202 (7-124.284)	3.4335 (1.799-4.67)	15.414 (8.299-23.544)
13.351 (-1.21-195.291)	1.2185 (-0.821-3.788)	11.278 (-1.21-13.748)
11.2675 (-7.575-18.303)	1.4235 (0.593-5.242)	5.757 (-7.575-12.348)
22.1945 (1.806- 166.496)	1.5185 (0.077-5.407)	9.8075 (1.806-25.863)
48.6065 (2.282- 137.844)	2.3235 (-0.082-4.93)	7.866 (-0.469-21.991)

6,4-PP adduct remaining (pmol/50ugDNA) above baseline (no treatment)				
Time Point min	AA	DD	EE	FF
0min	-1.7565 (-10.244-0.905)	30.9825 (11.096-69.791)	8.5355 (-8.35-18.4)	6.781 (-18.208-30.655)
1min	4.68 (-15.988-28.607)	22.949 (-0.781-60.414)	4.371 (-16.835-22.155)	1.361 (-6.43-14.36)
5min	-4.438 (-49.326-7.842)	8.605 (0.768-28.068)	3.156 (1.3-20.672)	5.282 (0.839-8.936)
15min	-4.162 (-27.976-6.754)	11.955 (-2.184-23.63)	7.0265 (-19.728-16.129)	5.3565 (3.916-21.683)
30min	5.959 (-10.433-17.48)	18.2995 (8.218-45.397)	5.063 (-2.22-21.797)	11.6915 (-2.434-23.967)
1hr	-2.383 (2.718-14.245)	-0.6505 (-18.882-86.404)	-6.467 (-23.743-0.807)	5.2255 (-1.377-19.013)
3hr	-5.78 (-41.607--1.312)	-10.1025 (-20.918--2.866)	-5.814 (-23.743-9.59)	6.744 (-15.629-12.777)
6hr	-9.302 (-16.418-6.668)	-8.649 (-19.615--3.18)	-4.2685 (-26.73-0.6)	1.417 (-4.245-10.318)
12hr	-2.867 (-14324-16.89)	-7.037 (-17.84-5.483)	-7.0035 (-24.656--0.311)	1.3855 (-6.401-19.182)
24hr	-2.363 (-20.211-3.221)	-10.3705 (-19.359--6.655)	-8.747 (-19.428--1.185)	-0.724 (-2.949-8.662)

	AAFF	DDEE
	-1.518 (-10.244-9.237)	14.295 (7.108-27.515)
	1.922 (-6.43-14.36)	10.995 (1.945-23.818)
	1.156 (-4.753-6.761)	7.333 (1.3-20.672)
	4.177 (-3.612-6.754)	8.6205 (-2.184-16.4)
	7.9505 (-2.434-16.055)	9.192 (3.695-22.171)
	3.162 (-5.15-10.917)	-1.1385 (-17.239-12.776)
	-2.1005 (-15.629-7.377)	-9.124 (-18.349--2.59)
	-0.008 (-8.098-2.844)	-8.519 (-19.485--2.59)
	-0.12 (-6.401-2.482)	-7.0035 (15.341--0.311)
	-1.3635 (-3.261-2.559)	-9.4975 (-15.558--6.93)

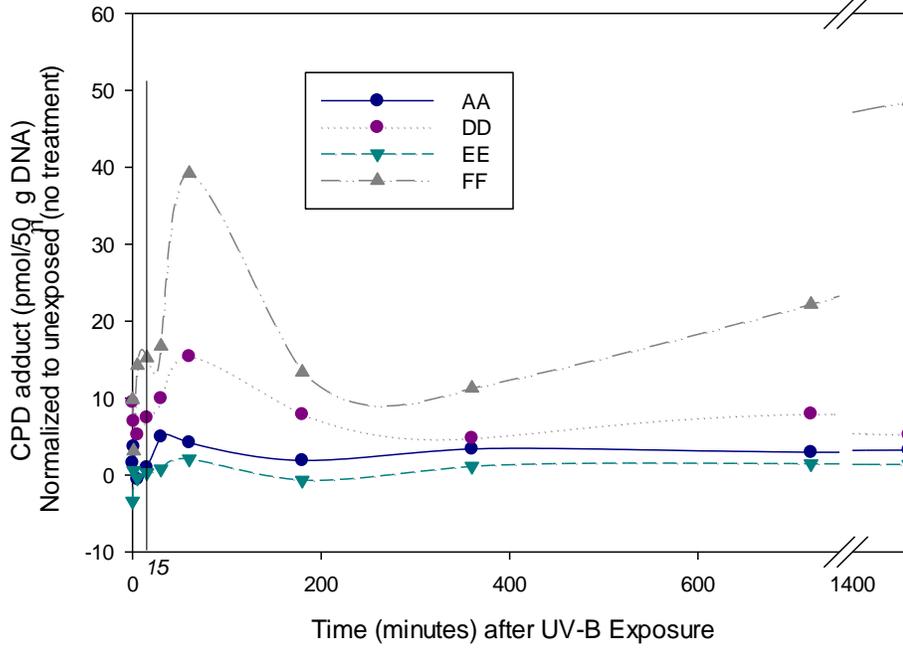
Table 9: Measure of central tendency for CPDs and 6-4PPs at intervals of repair after UV-B exposure. A: Median (~95%CI) CPD in pmol/50µg DNA above baseline. **B:** Median (~95%CI) 6,4-PP in pmol/50µg DNA above baseline.

The DNA repair capacity (DRC) was determined for each haplotype by the loss of adducts over time using liner regression analysis starting from 15 minutes. The slope of the line was taken as a measure of the DRC. The relationship between PGHs and DRC by adduct is presented in Table 10 panel A. The haplotypes were tested for significance of adduct remaining over time within each line by the Kruskal-Wallis test (non-parametric ANOVA). Similarly, Table 10 panel B presents the p-values by haplotype across times, with significance at $p \leq 0.05$. All haplotypes showed statistical significance in the level of DNA damage over time for CPD adducts, while only DD and EE showed statistically significant differences in the level of DNA damage over time for the 6,4-PP adducts (significant p-values ($p \leq 0.05$) in italics).

Figure 11: Amount of adduct remaining over time after UV-B exposure

A

Amount of CPDs over time after UV-B exposure at 35mJ/cm²



B

Amount of 6,4-PPs over time after UV-B exposure at 35mJ/cm²

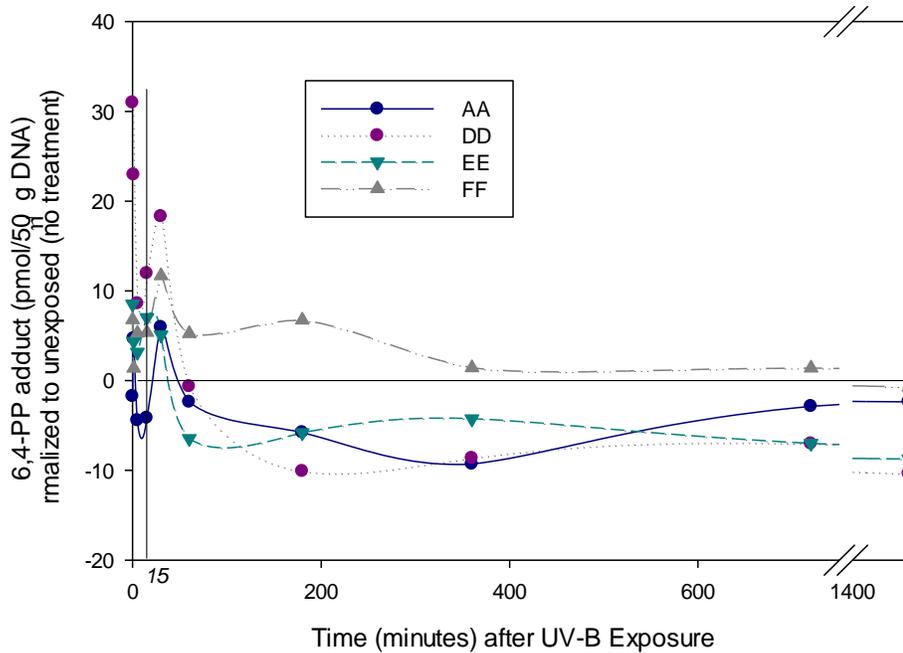


Figure 11: Amount of adduct remaining over time after UV-B exposure. The dark blue line represents remaining adduct in line AA, the purple line the amount of adduct remaining in DD, the light blue line the amount of adduct remaining in EE, and the grey line the amount of adduct remaining in FF. Time after exposure is presented in minutes. **A:** The amount of CPD remaining in pmol/50µg DNA after normalization to no UV baseline. **B:** The amount of 6,4-PP remaining in pmol/50µg DNA after normalization to no UV baseline.

Table 10: Measure of significance for CPDs and 6-4PPs over the time course of repair after UV-B exposure

A			B		
PGH\Adduct	CPD	6,4-PP	PGH\Adduct	CPD	6,4-PP
AA	-0.0019	-0.0099	AA	<i>0.025</i>	<i>0.170</i>
DD	0.0014	-0.0145	DD	<i>0.002</i>	<i>0.000</i>
EE	-0.0038	-0.0135	EE	<i>0.031</i>	<i>0.006</i>
FF	0.0296	-0.0052	FF	<i>0.038</i>	<i>0.346</i>

Table 10: Measure of significance for CPDs and 6-4PPs over the time course of repair after UV-B exposure. **A:** Linear regression slopes for each haplotype by adduct. **B:** Kruskal-Wallis p-values, with significance at $p \leq 0.05$, of each haplotype over time for each adduct. All haplotypes showed statistical significance for CPDs, while only DD and EE showed statistically significant differences over time for the 6,4-PPs (significant p-values in italics).

Taken together, the data shows that the relative rate of repair for **CPDs**, in descending order, are

$$\text{PGH E} > \text{PGH A} > \text{PGH D} > \text{PGH F.}$$

Similarly, the relative rate of repair for **6,4-PPs** are (also in descending order)

$$\text{PGH D} > \text{PGH E} > \text{PGH A} > \text{PGH F.}$$

Comparison of the calculated DRC rates by the likelihood ratio test resulted in a p-values of 3.2×10^{-5} for CPDs and $< 1 \times 10^{-5}$ for 6,4-PPs. This indicates that the trend across time points is significantly different between the haplotype clades (PGHs). It is interesting to note that the orders listed above are different depending on the specific adduct being repaired, indicating a haplotype preference for the adducts being repaired.

Consistent with the mutagen sensitivity experiments of aim 1, the PGHs that showed significant differences were D and F. The repair order for CPDs reflect this relationship. However, the 6,4PPs do not. Consequently, I hypothesized that “sensitive” (haplotypes that show good repair) and “insensitive” (haplotypes that show poorer repair) haplotypes have clear differences in specificity for the type of adduct, as opposed to a weak spectrum, and consequently would show clear differences in the rate of repair when combined into a sensitivity-based grouping. Table 11 shows the DRC rates (as determined by the loss of adduct over time) of these sensitivity-based groupings, which were designated as sensitive or insensitive by the individual DRC rates for each adduct. Sensitive haplotypes for CPDs were clades A and E, while haplotypes sensitive for 6,4-PPs were clades D and E. Consequently, the insensitive haplotypes for CPDs were clades D and F, while insensitive haplotypes for 6,4-PPs were clades A and F. Again using the likelihood ratio test to show differences between those relative rates of repair, the p-values for the sensitive compared to the insensitive haplotype groups were $\leq 1 \times 10^{-4}$ for both CPDs and 6,4-PPs. For comparison, graphical representations have been presented in figure 12 below, and p-values are presented in Table 27 of the supplementary data section.

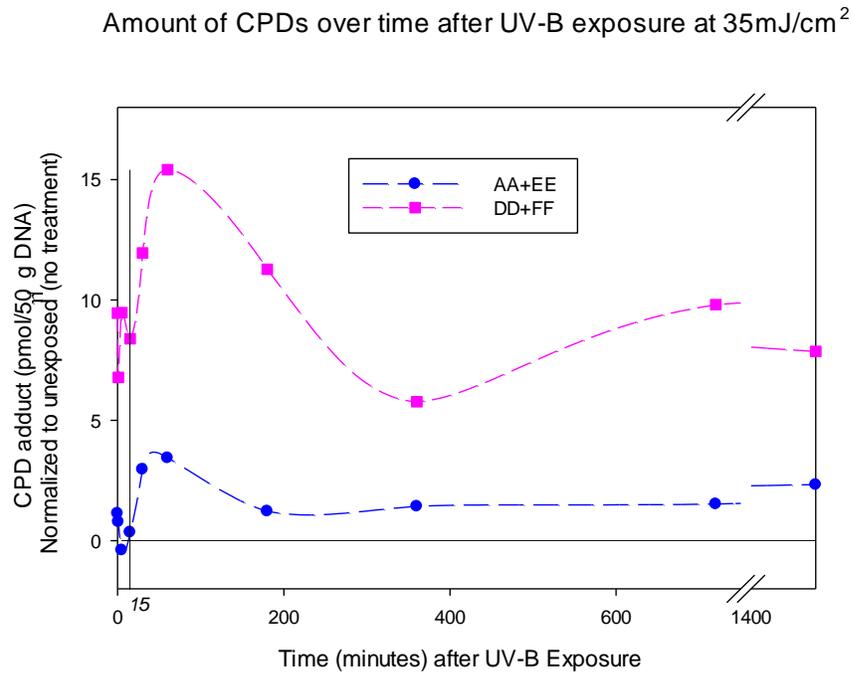
Table 11: Measure of significance for CPDs and 6,4-PPs over the time course of repair after UV-B exposure of the combined PGHs

A			B		
sensitive	line	slope	sensitive	line	pvalue
AA+EE	CPD	-0.0009	AA+EE	CPD	0.004
DD+EE	6,4-PP	-0.0018	DD+EE	6,4-PP	0
insensitive	line	slope	insensitive	line	pvalue
DD+FF	CPD	0.0058	DD+FF	CPD	0
AA+FF	6,4-PP	-0.0123	AA+FF	6,4-PP	0.213

Table 11: Measure of significance for CPDs and 6,4-PPs over the time course of repair after UV-B exposure of the combined PGHs. Sensitive (upper level) and insensitive (lower level) cell lines were determined from regression analysis of the

individual haplotypes as shown in Table 10. **A:** Linear regression slopes for each haplotype combination by adduct. **B:** Kruskal-Wallis p-values, with significance at $p \leq 0.05$, for each combined haplotype over time by adduct.

Figure 12: Amount of adduct remaining over time after UV-B exposure
A



B

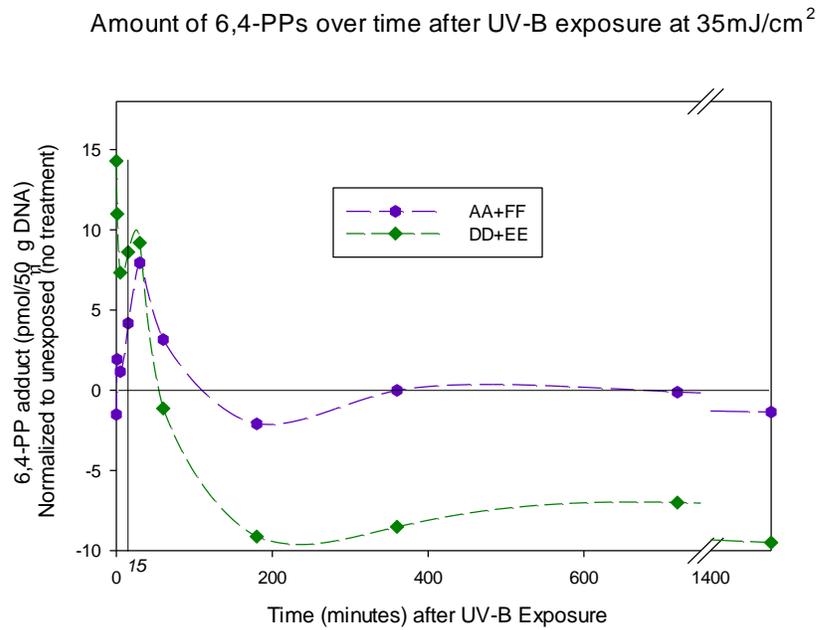


Figure 12: Amount of adduct remaining over time after UV-B exposure. Time after exposure is presented in minutes. **A:** The amount of CPD remaining in pmol/50 μ g DNA above baseline no treatment. The blue line represents remaining adduct in the sensitive lines (AA+EE), while the pink line represents remaining adduct in the insensitive lines (DD+FF). **B:** The amount of 6,4-PP remaining in pmol/50 μ g DNA above baseline no treatment. The green line represents remaining adduct in the sensitive lines (DD+EE), while the purple line represents the remaining adduct in the insensitive lines (AA+FF).

The adduct data shows clear differences between types of adducts, but the question as to what this means to total damage was not directly addressed. I hypothesized that the haplotypes, regardless of the form of damage, would show clear differences in repair rates of total DNA damage. To confirm this, I ran a series of picogreen experiments to measure the overall (non-specific) amount of DNA damage that is repaired over time. Individual time points are presented graphically in figure 13 and in table 28 of the supplemental results section. Table 12 shows the linear regression results by haplotype along with the respective p-values. Since this is a general measure of total DNA damage, other possible sources of DNA damage may be measured such as UV-induced ROS DNA damage (Cadet et al., 2005; D'Errico et al., 2006; Melis et al., 2013; Nagira et al., 2002; Rastogi et al., 2010) or alternative adducts such as the Dewar form (Cadet et al., 2005; Sinha and Häder, 2002).

In summary, the data from Aim 2 indicate that DRCs are significantly different for each haplotype over time after exposure. DRC differences calculated for total (using a picogreen-based assay) and adduct specific repair (using the ELISA method) show similar haplotype responses to the epidemiological study. Furthermore, the repair efficiency for the haplotypes is dependent on the type of adduct repaired. Overall, the data indicates haplotypes influence DRC.

Figure 13: Amount of DNA damage remaining over time after UV-B exposure

Remaining DNA damage by picogreen assay

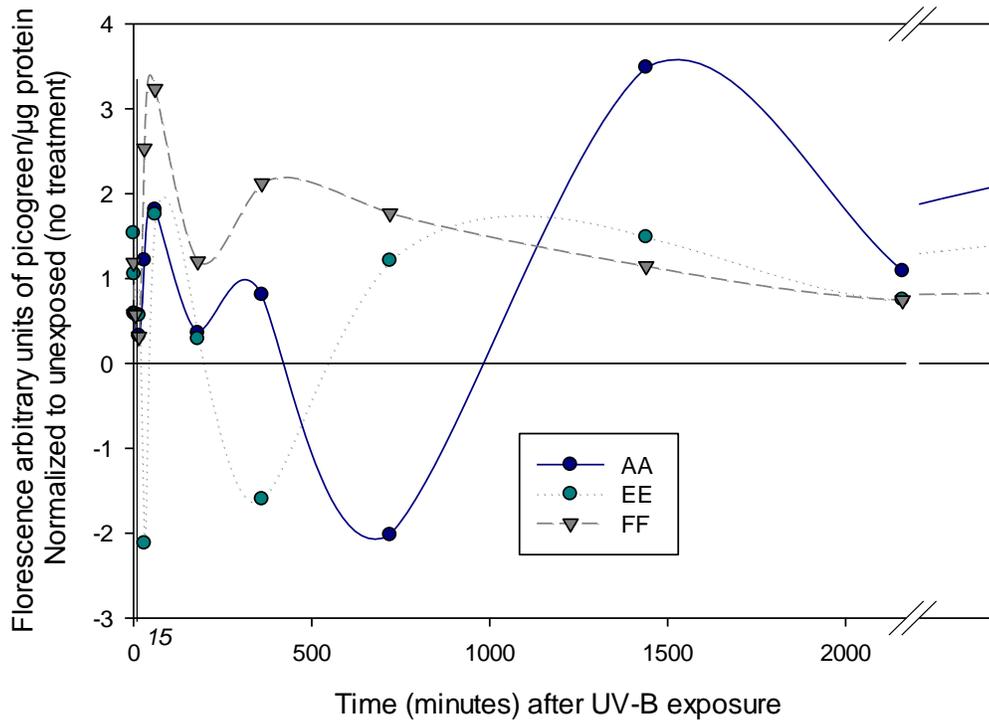


Figure 13: Amount of DNA damage remaining over time after UV-B exposure. The dark blue line represents remaining adduct in cell line AA, the light blue line the amount of adduct remaining in EE, and the grey line the amount of adduct remaining in FF. Time after exposure is presented in minutes. Data was normalized to μg of protein using Bradford analysis.

Table 12: Measure of significance for total damage over the time course of repair after UV-B exposure of the combined PGHs

F	slope	p-value
AA	-1.09E-03	0.003
EE	1.11E-03	0.002
FF	-1.47E-04	0.672

Table 12: Measure of significance for total damage over the time course of repair after UV-B exposure of the combined PGHs. Slope is the linear regression of the amount of picogreen signal lost over time, normalized to the amount of signal in the untreated no exposure of the same haplotype.

AIM 3: MRNA AND PROTEIN ANALYSIS

The differences in DRC rates and the relative preference for different types of adducts depicted in results from the damage studies indicate differences in the underlying mechanisms by which XPC exerts its effect. For example, a rapid change in repair may be due to an induction of either mRNA or protein production. Therefore, to further elucidate the mechanisms driving these DRC differences, I measured the levels of *XPC* mRNA and XPC protein by relative real-time analysis and western blot analysis, respectively. Additional studies such as immunocytochemistry, microRNA, immunoprecipitation, and pull down analysis are currently in progress in our laboratory and will not be presented in this dissertation.

XPC mRNA was measured using Applied Biosystem's real time assay (Hs01104206_m1), which recognizes a bridging region that spans ~90 nucleotides over the junction of exons 15 and 16. Thus, the assay measures mature transcripts only, as the intron 15 region splicing removes ~1.1kB and conditions are not optimized for amplifying a product the size of an unspliced region, nor does the reaction test for other common splicing variants at other introns (Khan et al., 2002). Given these caveats, the amount of *XPC* mRNA should be a measure of the functional RNA pool available for translation into theoretically functional protein. The amount of *XPC* transcript was normalized with a control transcript (β -actin). Table 29 of the supplemental data section shows the values of the Shapiro-Wilk normality test. Due to the variability inherent with biological samples, the data was analyzed using non-parametric methods. As the medians are a better representation of the central tendency for non-normal data, Table 13 shows the median values per time point after UV-B exposure for each of the respective

haplotype groups, with the error as the approximate 95% CI using the formulas stated in the statistics section of the methods chapter.

Table 13: Fold change over the time course of repair after UV-B exposure of the combined PGHs

		Fold Change from no treatment after UV-B exposure			
Time	min	AA	DD	EE	FF
0min	0	0.83 (0.35-1.55)	4.03 (0.10-36.25)	0.10 (0.00-2.23)	1.53 (0.49-4.63)
1min	1	7.14 (0.26-103.25)	3.25 (0.63-5.94)	1.09 (0.03-1.91)	2.55 (0.58-6.87)
5min	5	18.41 (1.04-65.34)	5.45 (1.31-83.29)	0.38 (0.02-1.87)	3.61 (0.41-99.73)
15min	15	3.01 (0.08-22.01)	0.87 (0.33-33.20)	1.00 (0.01-3.94)	3.70 (1.43-9.65)
30min	30	1.33 (0.28-199.47)	2.77 (0.34-12.21)	2.36 (0.05-7.26)	1.90 (0.14-6.63)
1hr	60	1.04 (0.10-3.94)	4.02 (1.21-9.65)	1.01 (0.22-3.92)	8.46 (3.61-22.94)
3hr	180	6.63 (3.89-58.08)	0.55 (0.37-17.27)	0.79 (0.02-6.54)	2.15 (0.65-5.17)
6hr	360	1.03 (0.30-14.22)	5.66 (1.26-19.70)	1.13 (0.00-7.52)	3.41 (0.95-16.45)
12hr	720	1.02 (0.29-3.01)	2.81 (0.31-5.31)	0.93 (0.03-4.69)	9.79 (0.14-11.71)
24hr	1440	0.95 (0.43-5.39)	4.76 (NA)	1.66 (0.00-4.11)	8.17 (0.66-143.01)
36hr	2160	0.45 (0.02-1.28)	0.44 (0.43-0.79)	0.38 (0.00-5.31)	1.28 (0.55-18.00)
48hr	2880	1.21 (0.47-65.80)	NA	1.75 (0.00-5.21)	8.88 (0.57-88.03)

Table 13: Fold change over the time course of repair after UV-B exposure of the combined PGHs. Median (~95%CI) fold change from baseline for *XPC* mRNA by real time analysis.

The haplotypes were tested for significance between each time point both within the haplotypes (supplemental tables 30, panels A-D, significant p-values ($p \leq 0.05$) in italics) and between the haplotypes (supplemental table 31, significant p-values ($p \leq 0.05$) in italics) using the Mann-Whitney test (non-parametric T-test). For clarity, I present the data as median only graphs. Figure 14 shows the fold change from baseline of the *XPC* mRNA as a function of time in minutes after UV-B exposure.

Figure 14: Fold change in *XPC* mRNA above baseline over time after UV-B exposure

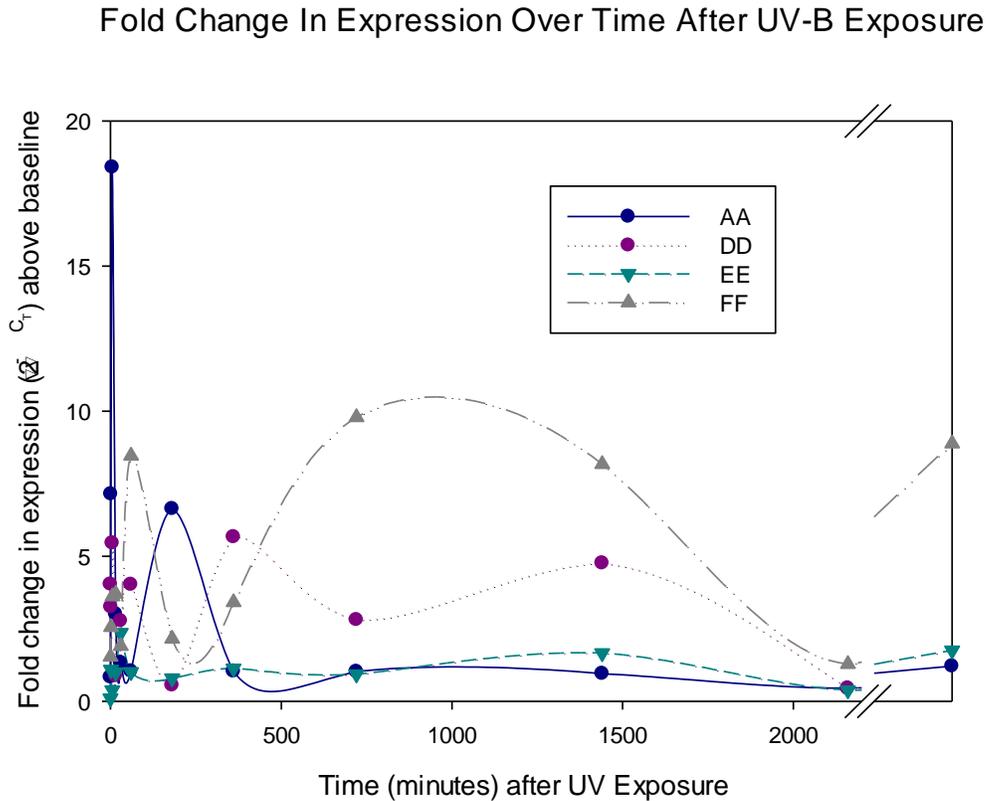


Figure 14: Fold change in *XPC* mRNA above baseline over time after UV-B exposure. The dark blue line represents remaining adduct in line AA, the purple line the amount of adduct remaining in DD, the light blue line the amount of adduct remaining in EE, and the grey line the amount of adduct remaining in FF. Time after exposure is presented in minutes.

The data shows clear and statistically significant (Tables 30 and 31 of the supplemental data section) changes in the amount of *XPC* mRNA over time in response to UV-B exposure that is dependent on the haplotype. Cells from the haplotype AA group responded very quickly, returning to baseline rapidly, while cells from the FF group were slow to respond but did reach statistical significance at later time points. Additionally, cells from the DD group showed a response that was slower than the AA cells but more rapid than the FF cells, for both longer than the AA cells but shorter than the FF cells. Additionally, the DD cells *XPC* mRNA fold change peaked at approximately 5.5 fold

median value, while the cells from FF peaked at about 10.0 fold and the cells from AA peaked as high as approximately 18.5 fold. These changes were not observed with EE cells, where the mRNA barely peaked at any point above 2 fold. As such the relative **rate** of response (that is, rapidity of the fold change in mRNA) can be ranked, in descending order as

PGH A > PGH D > PGH F > PGH E.

Additionally, the **intensity** (maximal fold change) can be ranked in descending order as

PGH A > PGH F > PGH D > PGH E.

It is interesting to note that these rankings do *not* directly follow either patterns for the CA or DRC data. These differences appear to indicate that the haplotypes have very different mechanisms driving the clear differences in repair response.

These differences in both rate and intensity may indicate mechanisms that are driven by changes in mRNA context, such as changes in folding, splicing, transport, or degradation (Berman et al., 2004; Gorlov et al., 2011; Komar, 2007; Park et al., 2010; Qiao et al., 2011a; Rouzaud et al., 2010; Rukov and Shomron, 2011; Zhang et al., 2009). For the purposes of this dissertation, I used computational methods to focus on the initial step of pre-mRNA folding, which is the folding of the nascent mRNA strand immediately after transcription but before processing. The web-based program mFOLD from the University of Albany is the gold standard among the many folding sites online. Due to the nucleotide length limit imposed on submissions, I virtually “cut” each of the haplotypes uniformly to yield five shorter sequences per haplotype, designated a-e. While this is less than the ideal folding parameters, it is an acceptable approximation on localized effects of partial *XPC* haplotypes on pre-mRNA folding structures. Table 14 lists the mFOLD calculated lowest folding free energies, given as kcal/mol, by haplotype of each segment.

Table 14: The mFOLD free energies for each virtual segment of each haplotype

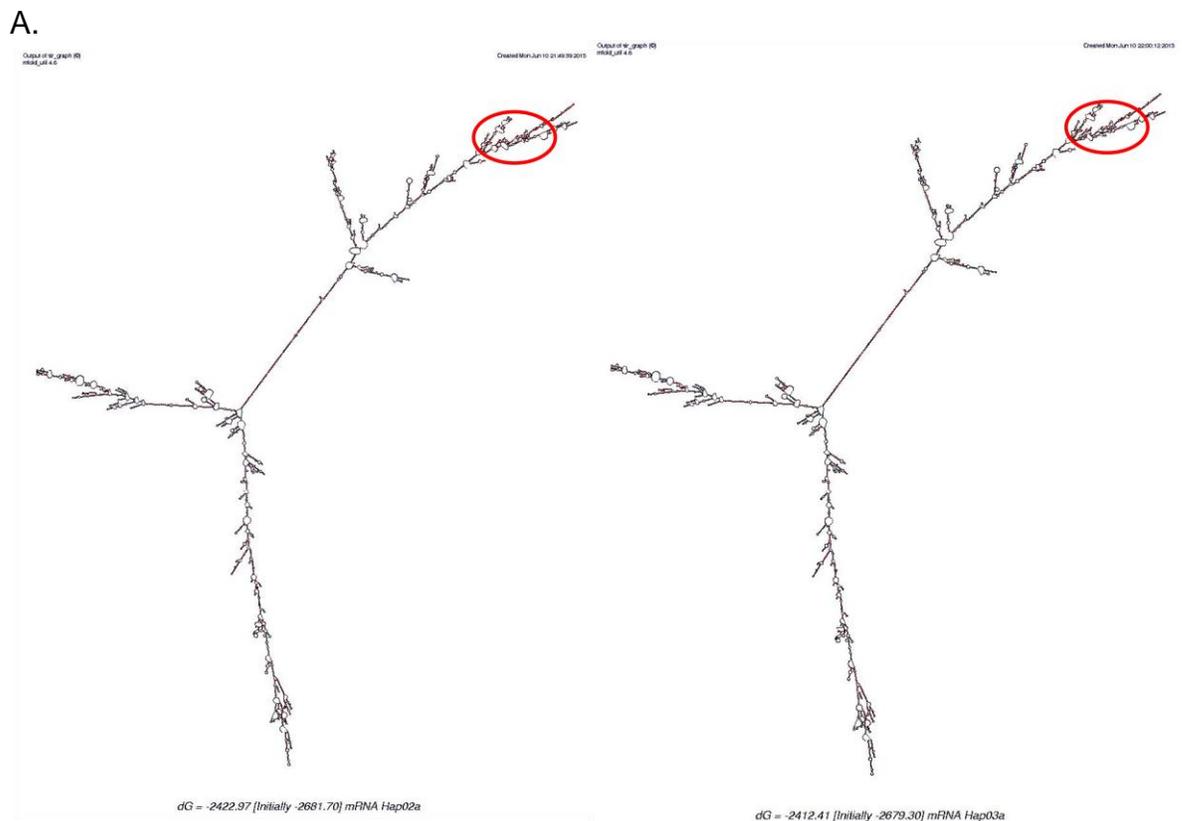
Hap	PGH	<i>a</i>	<i>b</i>	<i>c</i>	<i>d</i>	<i>e</i>
3	A	-2412.41	-2073.50	-2304.45	-2592.75	-2692.42
4	A	-2412.41	-2074.92	-2304.45	-2597.35	-2692.42
5	A	-2412.41	-2074.92	-2304.45	-2592.75	-2692.42
6	A	-2412.41	-2074.92	-2304.45	-2598.84	-2692.42
7	A	-2412.41	-2073.50	-2304.45	-2592.75	-2692.24
12	A	-2412.41	-2074.92	-2304.45	-2598.84	-2690.84
15	B	-2412.41	-2076.46	-2306.85	-2596.65	-2692.42
16	C	-2412.41	-2076.46	-2298.37	-2597.15	-2692.42
17	C	-2412.41	-2076.46	-2298.37	-2597.35	-2695.59
18	C	-2412.41	-2076.46	-2298.37	-2597.15	-2695.59
20	C	-2412.41	-2076.46	-2298.37	-2597.15	-2694.01
21	C	-2412.41	-2076.46	-2298.37	-2597.15	-2694.01
10	D	-2422.97	-2076.46	-2298.37	-2597.35	-2700.79
19	D	-2422.97	-2076.46	-2393.28	-2484.70	-2699.21
2	E	-2422.97	-2078.90	-2307.22	-2599.85	-2701.62
11	E	-2422.97	-2078.90	-2307.22	-2599.85	-2693.72
1	E	-2429.57	-2078.90	-2305.43	-2600.78	-2697.62
8	F	-2422.97	-2076.46	-2305.43	-2581.63	-2694.04
9	F	-2422.97	-2076.46	-2305.43	-2583.32	-2694.04
13	F	-2422.97	-2076.46	-2305.43	-2583.32	-2697.54
14	F	-2422.97	-2076.46	-2305.43	-2583.32	-2698.04

Table 14: The mFOLD free energies for each virtual segment of each haplotype. Haplotypes are organized by PGH, and each segment listed as *a-e*.

Comparing the range of free energies within the segments shows that greater differences in free energies can indicate greater differences in predicted structures at those locations. Comparing across the haplotypes, the data shows that segment *a* (bases 1-8000; includes two SNPs, rs2607775 - rs1350344) has a range of 10.56 kcal/mol, *b* (bases 8001-16200; includes four SNPs rs3729587 – rs3731125) a range of only 5.4 kcal/mol, *c* (bases 16201-24300; includes seven SNPs rs3731127 – rs2607736) a range of 94.91 kcal/mol, *d* (bases 24301-32640; includes ten SNPs rs3731127 – rs2607736) the

largest range of 116.08 kcal/mol, and *e* (bases 32641-40526; includes twelve SNPs rs2607734 – rs8516) a range of 9.95 kcal/mol. It is interesting to note that the difference in free energy is not dependent on the number of SNPs. As expected, the highest free energies show the largest changes in structure (Figure 15). For clarity, I am presenting only the two extremes of the differences in folding, the remainder of the structures can be found in the Supplemental Data 2 section.

Figure 15: MFOLD structures for virtual segments



B.

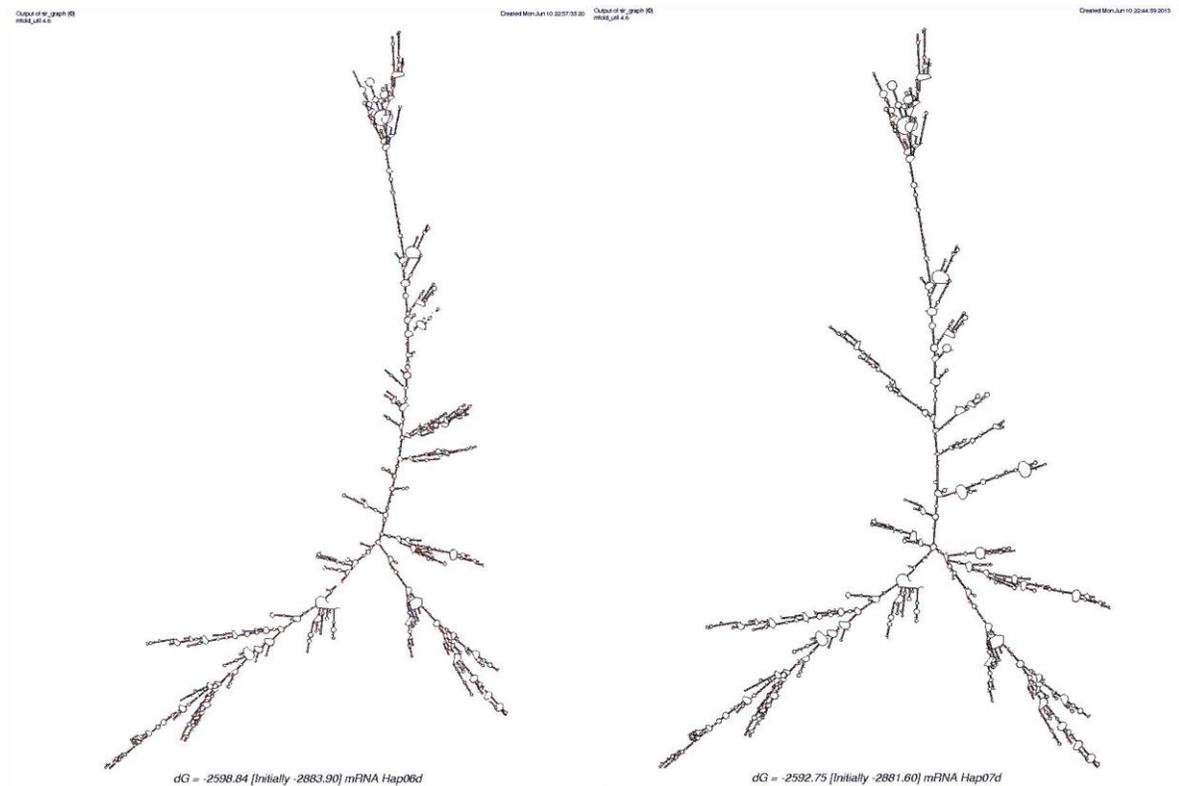


Figure 15: MFOLD structures for virtual segments. Representative structures show low free energy differences are indicative of small changes in mRNA folding while larger differences in free energy are indicative of larger changes in folding structures. A: Small changes separate the a segments of haplotypes 2 and 3 are circled in red. B: Large changes separate the d segments of haplotypes 6 and 7.

Given the clear differences in *XPC* mRNA, the next logical step was to determine the amount of XPC protein present in the cells at these various time points. Unfortunately, the available commercial antibodies available for XPC proved to have limited specificity and reproducibility, so in collaboration with the Biothysis and Biomarker Core Laboratory in the Biomolecular Resource Facility, we generated an in-house antibody from a synthesized polypeptide. As part of XPC's function is to not only recognize damage on the DNA strand, but to also recruit other components of the NER machinery to the site, I chose to focus on the TFIID binding region as a partial indicator of XPC functionality. The alternative, which was the DNA binding region, showed

extensive sequence homology using NCBI's BLAST analysis (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>) (Altschul et al., 1997). The short C-terminal polypeptide (910)DEEKQKLKGGPKKTKREKKA(929) BLAST results returned smaller E-values (that is, the calculated false positive value, or p-value) and higher similarity scores for XPC specific hits. The polyclonal antibody generated in rabbit and serum isolated was used for western blot analysis. Additional studies, such as localization by immunocytochemistry and confocal microscopy or binding assay analysis by co-immunoprecipitation and gel-shift analysis are ongoing but are beyond the scope of this dissertation.

XPC protein was measured using standard western blot analysis of total protein preparations from cells. As there was no enrichment for XPC, there was no differentiation between cytosolic or nuclear XPC, nor mature, immature, or degraded XPC proteins. Thus, the amount of XPC measurable by this analysis is the total protein pool and any changes to that total would represent either an increase in translation and therefore nascent proteins (if increased) or an increased degradation (if decreased). The XPC data was normalized against β -actin for each sample. The data was analyzed using parametric tests, as the samples passed the Shapiro-Wilk test of normality. Mean data with standard error of mean is presented in Table 15.

Comparison of the data as a function of time is shown graphically in Figure 16 below. ANOVA analysis is presented in Table 16, with panel A showing the difference between time points for each haplotype group and panel B showing the difference between cell lines at each time point. Interestingly there was only significant difference over the course of the repair in haplotype EE, while there was no statistically significance difference over time for haplotypes AA, DD, or FF. (The corresponding t-tests p-values are given in table 32, panels A-D of the supplemental data section.) However, comparison between the lines at the individual time points revealed significant differences between the haplotypes at the early time points – that is from baseline (no

treatment) out to 6 hours after UV-B exposure. (The corresponding t-test p-values are given in table 33 of the supplemental data section.)

Table 15: Amount of XPC protein over the time course of repair after UV-B exposure of the combined PGHs

time	AA	SEM	DD	SEM	EE	SEM	FF	SEM
NT	1.550	0.292	2.099	0.586	0.609	0.197	0.764	0.157
0 min	1.339	0.296	1.712	0.077	0.547	0.182	0.739	0.128
1 hr	1.213	0.193	1.588	0.184	0.514	0.153	0.670	0.113
6 hr	1.138	0.198	1.296	0.034	0.531	0.153	0.676	0.139
12 hr	0.996	0.095	1.084	0.093	0.545	0.156	0.847	0.226
24 hr	1.190	0.091	1.112	0.092	0.597	0.190	0.943	0.335
36 hr	0.833	0.050	1.490	0.188	0.546	0.202	0.827	0.211
48 hr	1.215	0.060	1.402	0.129	0.629	0.285	0.843	0.181

Table 15: Amount of XPC protein over the time course of repair after UV-B exposure of the combined PGHs. Mean and standard deviations of XPC specific protein from western blot analysis by haplotype. Time is presented in hours.

Figure 16: Arbitrary densitometry units versus β -actin over time after UV-B exposure

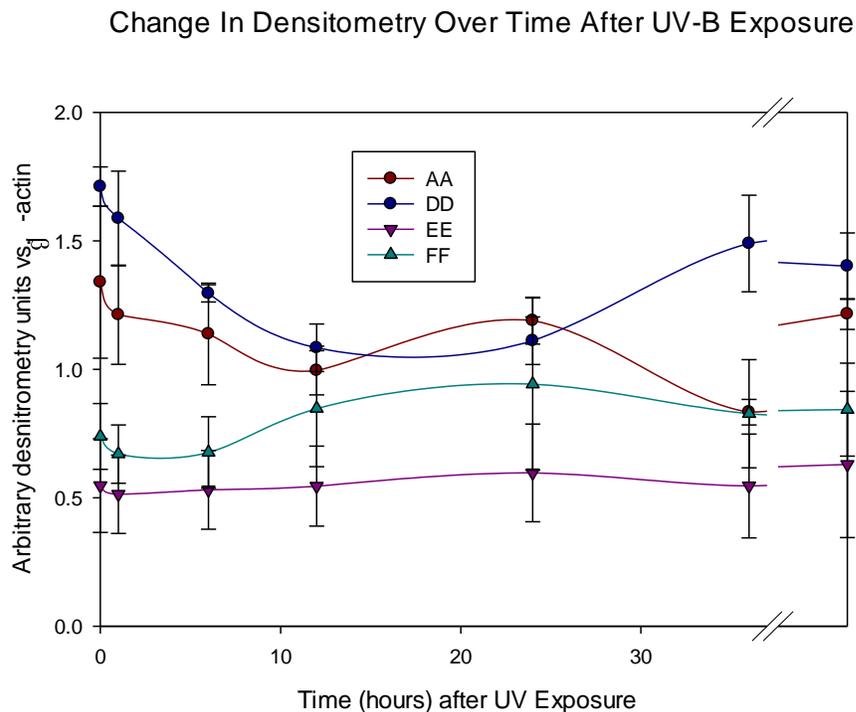


Figure 16: arbitrary densitometry units verses β -actin over time after UV-B exposure. The dark blue line represents remaining adduct in line AA, the purple line the amount of adduct remaining in DD, the light blue line the amount of adduct remaining in EE, and the grey line the amount of adduct remaining in FF. Time after exposure is presented in minutes.

Table 16: Statistical significance in the amount of XPC protein over the time course of repair after UV-B exposure of the combined PGHs

A		B	
Line	Densitometry	Time (hr)	Densitometry
AA	0.293	NT	<i>0.032</i>
DD	0.181	0	<i>0.023</i>
EE	<i>0.045</i>	1	<i>0.010</i>
FF	0.223	6	<i>0.038</i>
		12	0.200
		24	0.305
		36	0.052
		48	0.105

Table 16: Statistical significance in the amount of XPC protein over the time course of repair after UV-B exposure of the combined PGHs. ANOVA analysis of the densitometry analysis presented in Figure 16. Time is presented in hours, significant p-values are italicized. A: Statistical significance of each haplotype group over the course of the experiment. B: Statistical significance at each time point between the haplotypes.

It is clear that further analysis is needed to understand the underlying mechanisms of these correlations. While there are clear differences in the DRC both dependent on the haplotype and the type of damage, there is no clear mechanism for this, as the XPC specific mRNA levels determined from real time analysis do not directly correlate, though they are statistically significant. Oddly, there is a lack of change for the amount of protein over the course of the experiment, which does not correlate with the changes in the mRNA levels. Further confounding the situation is the weak but statically significant difference that exists between the amount of XPC protein and the different haplotypes.

Chapter 4: DISSCUSSION AND CONCLUSIONS

Since SNPs exist in combination, evaluation of their joint effects is more biologically relevant than individual SNP assessment. Full gene haplotype analysis is a new, largely unexplored comprehensive evaluation that has only recently become possible due to freely available genotype information from programs such as the Human Genome Project, HapMap, and the 1000 Genomes Project (Browning and Browning, 2011; Mir, 2009; Ng and Kirkness, 2010). Given this, I wanted to evaluate the effect of haplotypes of the common ($MAF \geq 0.05$) *XPC* SNPs.

HAPLOTYPE ANALYSIS AND CA EVALUATION

To determine the relationship between haplotypes present in the entire genomic region of the *XPC* gene and genetic damage, I evaluated a small subset of individuals from a larger cohort of individuals (Hill et al., 2005) for haplotype effects on background levels of CA as well as mutagen-induced CA levels. CAs, were used as an endpoint for this analysis as they are a substantially validated (by prospective studies) as human cancer risk biomarker (Bonassi et al., 2000, 1995; Hagmar et al., 1998, 1994). Mutagen sensitivity uses CA frequency as an indirect measure of the repair capacity and an indirect measure of cancer risk (Abdel-Rahman and El-Zein, 2011). While there have been studies that use the MS assay to evaluate the effect of a few isolated SNPs on genetic damage (Aka et al., 2004; Angelini et al., 2008; Leng et al., 2008), this is the first study to comprehensively evaluate whole-gene haplotypes in this context. Haplotype association, with both background and mutagen-induced CAs, shows a statistically significant interaction with smoking status. Specifically, the PGH-C clade is associated with increased CA frequency, which supports other association studies linking *XPC* SNPs to smoking associated cancer risk (An et al., 2007; Guo et al., 2008; Hansen et al., 2007).

This suggests that individuals with PGH-C are at greater risk of developing genomic instability from damage caused by tobacco-smoke carcinogens, likely via decreased repair capacity of genetic damage. The fact that this association is lost after MS analysis indicates that the baseline mechanistic difference is being driven by a component that changes in response to an acute challenge. For example, this mechanism may be abnormally low levels of a repair signal or component, such as XPC protein, which can be induced by a significant bolus of a toxic agent causing a high level of DNA damage.

Prior published association studies have addressed only a small number *XPC* polymorphisms, with inconsistent results. Positive association with cancer risk was reported for SNP rs2228000 (A499V) in some studies (An et al., 2007; D'Amelio et al., 2012; de Verdier et al., 2010; He et al., 2013; Sak et al., 2006; M. Shen et al., 2005; Song et al., 2013; Stern et al., 2009) but not others (Doherty et al., 2011; Guo et al., 2008; Ji et al., 2012; Liu et al., 2013; Roberts et al., 2011; Wang et al., 2010; L.-E. Wang et al., 2013; Weiss et al., 2006). Two studies showed a statistically significant increase in survival of cancer patients following chemotherapy with rs2228000 (Alvarez-larra et al., 2010; Dong et al., 2012), which would indicate a possibly decreased repair capacity as well. Conversely, another study shows a statistically significant decrease in survival after chemotherapy treatment in acute-myeloid-leukemia patients carrying the variant SNP (Strom et al., 2010), while a third study showed no effect (Fleming et al., 2012). Interestingly, a study in a Polish population showed a *decrease* in cancer risk for individuals carrying this SNP (Paszowska-Szczur et al., 2013). An association of intron 12 SNP rs2279017 with bladder cancer risk was reported in one study, yet these results still remain unconfirmed (Sak et al., 2006). Reported associations between the exon 16 variant rs2228001 (K939Q) and cancer risk were shown for esophageal, colorectal, lung, and liver cancers (Campayo et al., 2011; Guo et al., 2008; Hansen et al., 2007; Long et al., 2010b; Vogel et al., 2005). The rs2228001 was also associated with increased survival of patients in one hepatocellular carcinoma study (Long et al., 2010b). However,

others reported no association for rs2228001 and cancer risk (An et al., 2007; Engin et al., 2011; Long et al., 2010a; Millikan et al., 2006; Zhu et al., 2008). Interestingly, one paper reported an increased risk of cancer recurrence and eventual mortality in ovarian cancer patients after chemotherapy with the variant form of rs2228001. This suggests the poor chemotherapeutic response of the patients with the variant form have less chemotherapy induced DNA damage due to an *increase* in repair capacity (Kang et al., 2013).

Such inconsistencies between *XPC* polymorphisms and cancer risk have been reported for other genes, with a number of possible explanations ranging from differences of study design (such as population sizes and ethnicities) to linkage disequilibrium (LD) between other SNPs (Au et al., 2004; Kang et al., 2013; Manuguerra et al., 2006; Vogel et al., 2005). It is likely that a combination of these factors play a role in the discrepancies seen in the literature – variable degrees of linkage disequilibrium exist both between ethnic populations and between SNPs within a population (Fu et al., 2011; Gazdar and Boffetta, 2010; Tian et al., 2009; Zabaleta et al., 2008). Incomplete LD, coupled with possible admixture, can confound association studies such that it is possible that polymorphism(s) linked with the SNP of interest could drive the effect under study. Differences in the ability to capture all of the variants involved can result in the differences between the investigations. Conceivably, all the evaluated SNPs could have little to no biological effect individually but could present a phenotypic effect in the context of a haplotype of combined polymorphisms as either a direct contributor or coincidental member. Such a hypothesis has been shown with polymorphisms of the *NEIL2* gene, where the rs56037884 SNP had no independent effect on expression of the gene but showed a statistically significant reduction of 69% when found in conjunction with rs8191518 (Kinslow et al., 2008).

Haplotype analysis can be coupled with phylogenetic analysis to highlight similarities and provide a robustness of analysis that can be limited when studying

individual haplotypes. Phylogenetic analysis compares the relatedness of multiple haplotypes, measuring the mathematical divergence as shared genealogical similarities (Tamura et al., 2007). Grouping haplotypes by phylogenetic analysis is an objective tool for capturing unique sequence characteristics that may underlie shared mechanistic traits (Bardel et al., 2009; Rzhetsky and Nei, 1992; Tamura et al., 2007; Yang, 1997). For the current *XPC* haplotype analysis, linkage differences resulted in the distinct clades (A-F). The percent divergence (%D) between the haplotypes within each PGH ranged from 4.8-8.6 (table 5A), although the differences for all haplotypes ranged from 2.9-62.9 %D (table 21). The %D between each PGH ranged from 25.4-57.3 (table 5B). This variability is due to incomplete penetrance of the linkage (<100%) between several of the SNPs. The relatively low divergence within the clades generates confidence that the aggregate (PGH) is a good representative of the individual haplotypes that comprise it.

PGH-C data indicates an interaction between genetic damage and smoking. This is consistent with the reported associations of increased cancer risk for the rs2228000 variant, which is uniformly present in the clade (see table 6) (An et al., 2007; D'Amelio et al., 2012; de Verdier et al., 2010; He et al., 2013; Sak et al., 2006; M. Shen et al., 2005; Song et al., 2013; Stern et al., 2009). Subsequently, the idea that the clade has lower DNA repair capacity (DRC) resulting in higher background CAs is also supported by the report that the same rs2228000 SNP is associated with decreased DRC (Slyskova et al., 2012; Zhu et al., 2008). Somewhat expectedly, rs2228000 showed no association with DRC in another study (Slyskova et al., 2011). Whether the effect seen with PGH-C is due to this SNP (alone or in combination) remains to be determined by future experiments. A likely explanation for this haplotype-smoking interaction is that the reduced repair capacity differences are exacerbated in the presence of exposure, and is consistent with data from other polymorphisms in other DNA repair genes (Abdel-Rahman and El-Zein, 2000; Affatato et al., 2004).

The data from the mutagen sensitivity assay showed increased mutagen induced CAs with PGHs D and F in smokers, suggesting that smokers with these haplotypes have reduced DRC and therefore are likely to be predisposed to an increased risk of developing cancer, given the well establish correlation between decreased DRC and cancer risk (An et al., 2007; Cheng et al., 1998; Spitz et al., 1995; Wang et al., 2007). A possible explanation for the MS data lies in the underlying mechanistic differences inherent in the effect of the haplotypes. For example, individuals carrying PGH-C may have a high risk of genomic instability with low chronic DNA insults due to poor recognition of damage with subsequently poor induction of DNA repair, yet may have efficient repair after an acute insult results in significant levels of DNA damage to activate repair pathways. Conversely, individuals carrying either clades D or F may have efficient recognition and repair with lower and/or chronic exposures, but a high acute exposure may show an increased risk of genomic instability due to slow induction or poor recruitment, resulting in inefficient repair of the significant DNA damage.

While the mechanisms by which certain haplotypes affect DRC are not fully understood, there are good indicators for some of the potential effects on protein structure and/or function. The rs2228000 variant found in all haplotypes of PHGs C and D (and haplotype 5 of PGH-A) is located at the 5' end of the hHR23B binding region. Any alteration in this domain could potentially change the function of XPC by modifying it's interaction with its binding partner. Alternatively, rs2228000 is found in consistently in conjunction with rs8516, rs10468, and rs1126547 in clade C and in half of clade D, while rs2229090 is uniformly found in clade D and in all but one member of clade C. All four of these SNPs are in the 3'UTR region of *XPC*, which can affect mRNA half-life and folding stability or translocation rates (George Priya Doss et al., 2008). Other SNPs in LD with these include rs1350344 (intron 1) in clades A, B, and C; rs2733537 (intron 3) in D, E, and F; rs1106087 in C and D and rs3729587 in C, D, E, and, F (both in intron 5); rs2607737 in D, E, and F and rs2607734 in F (both in intron 11); and rs2470353 (intron

12) in A, B, and C. Intronic SNPs like these seven can effect translation via exon skipping and alternative splicing, aberrant mRNA folding or stability, or modified epigenetics (e.g. microRNA interaction or generation) (Blankenburg et al., 2005; Cheng et al., 2006; Duan et al., 2007; Kinslow et al., 2008; Law et al., 2007; Lin et al., 2006; Martin et al., 2012; Mittal et al., 2012; Qiao et al., 2011a; Song and Chen, 2011; Zhou et al., 2011). Finally, SNPs in the 5' region (such as rs2607775 found in clades A, B, and C) can affect promoter modulation and subsequent gene expression (Bai et al., 2007; Cheng et al., 2006). It is likely that a combination of these effects are dictated by the haplotypes, and those combines effects account for the differences in repair as determined by CA levels.

To summarize, the data indicates haplotypes found in PGHs C, D, and F have decreased repair efficiency as determined by background CA and by the MS assay. Although the sample size of the population is small (99 individuals after haplotyping) and larger studies are warranted to confirm these findings, mechanistic investigation into the underlying effects of haplotypes on DRC (and therefore disease risk) is necessary and ongoing.

HAPLOTYPE AFFECTS ON DRC

Given the correlation between mutagen-induced CAs and specific *XPC* haplotypes, I needed to measure directly DRC without confounding factors such as complex mixtures, effects of metabolism and/or clearance, or heterozygous haplotype effects. While a dominant model was useful for correlation studies, a mixture of haplotypes (especially from different clades) would complicate mechanistic analysis by diluting the haplotype affects. This necessitated using a recessive genetic model for cell-based mechanistic studies. All cells were homozygous for a given PGH. Due to the lack of homozygous PGHs B or C available from the HapMap biorepository (Coriell), the

analysis was restricted to clades A, D, E and F. Future experiments will include heterozygous forms for the purpose of analyzing both gene-dose effects and evaluating the effects from clades C and B.

To simplify the exposure model, I used a low dose UV-B radiation model to create adducts that are primarily repaired by the NER pathway. UV-B has a number of advantages over 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone (NNK) used previously in the MS assay to generate DNA damage. While NNK is a powerful mutagen (Hecht et al., 1983; Lacoste et al., 2007; Peterson, 2010), it produces multiple forms of DNA damage that are not primarily repaired by NER (Peterson, 2010; Rastogi et al., 2010). Indeed, the NNK derivatives repaired by NER are the pyridyloxobutyl DNA adducts formed on guanine (N^7 , N^2 and O^6), cytosine (O^2), and thymine (O^2) (Brown et al., 2008; Peterson, 2010). UV-B radiation induces the formation of cyclopyrimidine (CPD) and 6,4 pyrimidine-pyrimidone (6,4-PP) dimers without bioactivation or metabolism, do not produce any toxic metabolites, and are predominantly repaired by the NER pathway (Jhappan et al., 2003; Ona et al., 2009; Pfeifer, 1997). The advantage of using UV is it avoids any confounding effects from differences in phase I or phase II enzymes, from damage accumulation by secondary metabolites, and from other metabolic polymorphisms. Furthermore, as helix distorting bulky adducts, CPDs and 6,4-PPs are considered classical NER substrates and are primarily repaired by the NER pathway (Rastogi et al., 2010).

Cell viability studies were used to determine the optimal dose of UV-B irradiation that resulted in genetic damage but did not significantly affect cell viability (defined as >80% viability by trypan blue exclusion assay). Interestingly, the 35mJ/cm² dose chosen for this study represents the physiological exposure UV-B dose of natural sunlight (Diffey, 1991). Using the recessive genetic model, cell lines homozygous for the designated PGH were evaluated for persistent levels of CPDs or 6,4-PPs over time. Clades D and F, who showed increased levels of genetic damage induced by tobacco

mutagens, were expected to show increased levels of both CPDs and 6,4-PPs. Linear regression analysis of the picogreen data supported this as a measure of overall genetic damage and its repair over time, although it should be noted that the method is a measure of all DNA damage. UV radiation is known to not only generate bulky helix distorting dimers, but also other forms of damage such as oxidation products, which have many other (often overlapping) repair pathways (Melis et al., 2013). It is for this reason that the ELISA data is considered a better measure for the mechanisms of UV-induced DNA damage repair. However, looking at specific adduct quantification by ELISA analysis showed that this clade difference (poor D and F repair versus good A and E repair) was only partially the case.

Consistent with previous reports, overall, 6,4-PPs were repaired faster than CPDs (Nakagawa et al., 1998), likely due to the requirement of DDB2 protein colocalization and subsequent ubiquitination-directed removal of DDB2 that greatly enhances CPD removal but is less sensitive for 6,4-PP dimer repair (Ford, 2005). However, *XPC* haplotypes affected not only the rate of repair but that rate was differentially dependent on which adduct is being repaired. In fact, all haplotypes showed statistical significance *within* the PGH over time for the amount of CPD adduct while only DD and EE showed statistical significance for 6,4-PP adducts. Linear regression analysis for PGHs DD and FF showed preferential repair for 6,4-PPs compared to CPDs, as opposed to PGHs AA and EE which did not have a clear preference for either adduct. There was an increased haplotype effect (differences *between* PGHs) in the early time points for CPDs as seen by the Mann-Whitney tests, with statistical significance as early as 15 minutes. Contrast this to the 6,4-PPs' haplotype effect, which became pronounced at 180 minutes (3 hours) and this statistical significance between haplotypes remained for the rest of the experiment. In short, ordering the relative rates of repair illustrates the haplotype effect. In descending order of DRC by clade: **E > A > D > F** for CPDs while **D > E > A > F** for 6,4-PPs.

There are a number of possible mechanisms that can drive such a preferential difference in repair of the type of damage. By dividing the caldes into either sensitive or insensitive groups, I attempted to clarify what SNPs overlapped in the haplotypes which could be working together to produce the gradient of effects seen in the ELISA analysis. “Sensitive” (good repair) grouping for CPDs consists of PGHs AA and EE and for 6,4-PP PGHs DD and EE. “Insensitive” (poor repair) grouping for CPDs consisted of PGHs DD and FF and for 6,4-PPs AA and FF. Differences between sensitive and insensitive groups showed statistical significance for CPDs at earlier time points as compared to 6,4-PPs, which were later and continued to the end of the experiment. This was expected based on the individual clade analysis, so I attempted to determine what, if any, clear SNP combinations that existed unique to the haplotype members of the PGH combined groupings.

Looking at the clustering of SNPs for the CPD sensitive haplotypes, EE and AA share rs1126547 in the 3' UTR, [rs2228001 (all but haplotype #7 of PGH A) at K939Q], rs2733532 in intron 15, and rs2227999 at H492R. Additionally, they uniformly lack variants at rs8516, rs10468, and rs2229090 in the 3'UTR; rs2279017 in intron12; and rs1106087 in intron 5. Therefore, it appears that three (rs1126547, rs2733532, and rs2227999) variants **and** five (rs8516, rs10468, rs2229090, rs2279017, and rs1106087) ancestral SNPs are necessary for effective CPD repair.

For 6,4-PP sensitive haplotypes DD and EE, they share variants at rs1126547 in the 3'UTR, rs2228001 at K939Q, rs2733532 in intron 15, rs2607737 in intron 11, rs2227999 at H492R, rs3729587 in intron 5, and rs2733537 in intron 3. Additionally, they uniformly lack variants at rs8516 and rs10468 in the 3'UTR; rs2733533 and rs2733534 in intron 15; rs2279017 and rs2470353 in intron 12; rs2607734 and rs2607736 in intron 11; rs3731149, rs3731146, rs1124303 and rs3731143 in intron 10; rs3731124 in intron 7; rs3731108 and rs3731106 in intron 5; rs3731081 in intron 3; rs3431068 in

intron 2; rs1350344 in intron 1; and rs260775 in the 5'UTR. Therefore, it appears that seven variants **and** 19 ancestral SNPs are necessary for effective 6,4-PP repair.

Thus, identifying the overlapping aspects, group E+A do not show any unique variants that do not also exist in the E+D group. The E+D group, however, share three additional intronic variants (rs2607737, rs3729587, and rs2733537) that are good potential candidates for the driving mechanisms for the damage preference of 6,4-PPs over CPDs. Additionally, there are 18 ancestral forms of the SNPs that are present in the sensitive groups – two for the E+A CPD sensitive grouping and 16 for the D+E 6,4-PP sensitive grouping. It is likely the lack of SNP variants at these locations also contribute to the overall sensitive phenotype.

Interestingly, the CPD insensitive grouping of D+F share only three variants (rs2607737 in intron 11, rs2227999 at H492R, and rs3729587 in intron 5), yet have a larger ancestral requirement than the relative sensitive (E+A) grouping, with 19 of the 35 in agreement for the ancestral form between PGHs D and F. These include rs2733533 and rs2733534 in intron 15; rs2470353 in intron 12; rs3731149, rs3731146, rs9653966, rs1124303, and rs3731143 in intron 10; rs3731127 in intron 8; rs3731125, rs3731124, and rs13099160 in intron 7; rs3731108 and rs3731106 in intron 5; rs3731093 and rs3731081 in intron 3; rs3731068 in intron 2, rs1350344 in intron 1; and rs2607775 in the 5'UTR. Therefore, it appears that three variants **and** 19 ancestral SNPs are necessary for less effective CPD repair.

The 6,4-PP insensitive grouping of A+F have only one SNP in total agreement between the two, located at rs2227999, which encodes for H492R. It also has ancestral agreement in ten positions (rs8516, rs10468, and rs2229090 in the 3'UTR; rs9653966 and rs3731143 in intron 10; rs3731127 in intron 8; rs3731125, rs3731124, and rs13099160 in intron 7; and rs3731093 in intron 3). Therefore, it appears that only one variant **and** ten ancestral SNPs are necessary for less effective CPD repair.

Similarly identifying the unique SNPs for the insensitive grouping shows that there are no variants that are unique to the insensitivity of A+F group, while the D+F group shares two intronic SNP variants at rs2607736 and rs3729587. Again, it seems like these are good possible candidates for acting as a driving mechanism for less efficient DNA repair, but the ancestral forms of the SNPs are likely to play a role as well. Comparing the overlapping ancestral SNP constituents of the both groups show D+F have 12 of the 10 intronic ancestral SNPs unique to grouping while A+F have three unique ancestral SNPs exclusively in the 3'UTR. These clustering analysis help to focus on potential critical areas of the *XPC* gene that are likely driving these repair phenotypes; in these cases, later introns and the 3'UTR.

At this time, there are no studies investigating the effect of haplotypes on any mechanism, but there is some limited information available on individual *XPC* SNPs. One study (Qiao et al., 2011b) attempted to probe the mechanism for the variant A499V (rs2228000) allele's effect on bladder cancer risk. *In silico* analysis predicted this SNP to have no functional effect alone, but the SNP is found in high LD with two 3'UTR SNPs (rs2470352 and rs2470458). The authors used a plasmid-based assay to characterize the 3'UTR region and measure mRNA stability and both mRNA and protein expression. They found that each of the variants could independently reduce expression of both mRNA and protein, concluding that the 3'UTR SNPs drove the rs2228000 association with bladder cancer risk. Similarly, other studies by the same group linked four other rare SNPs (rs2470353, rs121965090, rs121965091, and rs121965092) with rs2228000 as well (Qiao et al., 2011a). Interestingly, rs2470353 is found consistently in PGH A, B, and C of the CEU population, but was not well linked to rs2228000 for clade A or B. This association is maintained with PGH C, but rs2228000 is part of all clade D haplotypes *without* the rs2470353 SNP. The discrepancy is likely due to the differences in population, where the bladder cancer association study samples were obtained from

Leeds Teaching Hospitals in Leeds, England as opposed to the Utah population with northern and western European ancestry.

Despite this lack of predicted effect for rs2228000, it is one of the many SNPs who have the potential to play a part of the haplotype effect. The SNP is located in the 5' end of the hHR23B binding domain of XPC and could effect this protein-protein interaction, possibly explaining at least some of the insensitive effect of PGH DD on CPD repair. Whether this same interaction could account for the increased repair sensitivity seen for the 6,4-PPs is still uncertain. Unfortunately, the applicability of *in silico* analysis in predicting the effects of SNPs on protein function is very limited. Computationally, these algorithms rely on evolutionary conservation scores, which themselves are mostly based on rare Mendelian disease mutation studies. This limits the application of characteristics more suited to analyzing pathologic as opposed to physiologic mechanisms (Nakken et al., 2007). Consequently, the impact of SNPs like rs2228000 likely have physiologic meaning despite the lack of predictive effects, particularly in conjunction with others of the same haplotype.

This haplotype-adduct interaction is still poorly understood. The data suggest the intrinsic structural differences between CPDs and 6,4-PPs can interact with the haplotype differences, exacerbating them. It is known that NER capacity can be affected by the structure of the adducts being repaired (Gunz et al., 1996). Additionally, the intrinsic differences in *XPC* expression, translation, or post-translational modifications can impact XPC interaction with not only the adducted DNA itself, but also other NER proteins. XPC binds not only to damaged DNA, but also hHR23B, CENT2, and TFIIH – all of which map to the C-terminus (residues 492-940) end of the protein – and XPA – which has been mapped to the N-terminal side of protein (residues 156-325) (Bunick et al., 2006; Popescu et al., 2003; Uchida et al., 2002). The striking differences seen in the data and the potential positional effects are all consistent with the central hypothesis of *XPC*

haplotypes influence levels of accumulated DNA damage via differences in the DRCs by unique intrinsic mechanisms.

In summary, lymphoblastoid cell lines representing different *XPC* haplotypes were exposed to UV-B radiation and the DNA repair capacity was analyzed by measuring the amount of DNA damage repair over time. The rate of removal of the UV dimers was indicative of the intrinsic DRC. These differences in DRC were dependent both the haplotype and on the type of damage to be repaired, indicating multiple mechanisms at work to create the differences seen in both preference and rate of repair. Further detailed mechanistic studies investigating into the underlying differences intrinsic to the haplotypes is warranted.

RNA AND PROTEIN ANALYSIS

Not only do *XPC* haplotypes exert differential correlation with CAs induced by tobacco mutagens, but also direct damage repair preference and timing for UV radiation induced DNA dimers (CPDs and 6,4-PPs). Together these data indicate that unique mechanisms encoded in the *XPC* haplotype underlie these differences in repair. To determine if these mechanisms are dependent on either transcriptional or translational changes, I needed to measure *XPC* specific mRNA and XPC protein over time after UV-induction. The goal was to correlate differences in mRNA or protein with the differences in adduct accumulation.

As SNPs in LD may contribute to the observed phenotypic effect of altered DRC, the changes in mRNA after UV exposure were a starting point, as a number of effects can be at play in a given haplotype. For example, SNPs in the 3'UTR can effect mRNA stability and half-life as well as folding directed rates of translation (George Priya Doss et al., 2008), while 5'UTR SNPs can modulate promoter activity and gene expression (Kinslow et al., 2008; Musunuru et al., 2010; Rouzaud et al., 2010). Perhaps some of the

most versatile are the intronic SNPs. When located at or near exon boundaries, SNPs can direct alternative splicing or aberrant mRNA folding events (Duan et al., 2007; Law et al., 2007). These intronic SNPs can also direct epigenetic mechanisms, which has been reported for other genes. Such epigenetic mechanisms can include protein regulation through changes in mRNA folding (that is, stabilizing mRNA in optimal or sub-optimal pre-translation configurations) (Martin et al., 2012), or through miRNA changes either by creating (or losing) binding sites for other existing miRNAs or by generating (or losing) novel miRNAs derived from the spliced introns (Cortez et al., 2010; Lin et al., 2006; Rukov and Shomron, 2011). An exhaustive comparison of each of these possibilities is beyond the scope of this dissertation.

Comparative real time analysis of *XPC* mRNA at each time point after UV-exposure shows the *XPC* mRNA dynamic changes resulting from a rapid insult and the response to acute DNA damage. Each haplotype responded to the UV exposure in a different manner. The PGH AA cells, which were able to repair DNA damage from tobacco mutagens, sensitive to CPDs yet less sensitive to 6,4-PPs, showed a rapid induction of *XPC* specific mRNA, increasing over 15 fold within the initial 10 minutes after UV exposure, then quickly returning to baseline. This was in contrast to the effect observed with the PGH EE cells, which were sensitive to all forms of damage (tobacco mutagens and UV induced CPDs and 6,4-PPs) yet did not increase the amount of mRNA above baseline after UV treatment. Between these two responses were the PGH DD and PGH FF cells. The PGH DD cells (poor repairers of DNA damage from tobacco mutagens, sensitive to 6,4-PPs but not CPDs) responded to UV induced damage with increased mRNA production transiently, increasing just over 5 fold rapidly (within the first 10 minutes) and maintained that induction for several hours, returning to baseline after 12 hours. Interestingly, the worst repair group (FF, which was insensitive to all forms damage be it tobacco mutagens or UV adducts) showed slow but steady increase in

XPC specific mRNA over the initial 30 minutes up to nearly 10 fold induction, which was maintained over the course of the experiment.

Differences in the rate of expression change as well as the intensity of the change may yield clues to the underlying mechanism at work. The rapid yet short lived burst of intense *XPC* mRNA induction found in AA cells could be an indicator of a rapid response to damage but perhaps less efficient actual repair as indicated by the less sensitive 6,4-PP DRC. The lack of *XPC* mRNA induction in the EE cells may reflect an opposite situation, with an effective repair scheme that is not as sensitive to the acute damage at this dosage. Although the FF cells do eventually respond to the insult, the actual repair in these cells is less, despite persistent mRNA expression. This seems to indicate that the mRNA response is not the driving force for the DRC for this haplotype. The DD cells fall mid-way between all the other cell types in terms of sensitivity, response, and DRC, indicating that the mechanism is likely a blend of the others.

Haplotypes can effect mRNA folding in a number of ways, including stabilizing alternative conformations that may or may not be conducive for stability or translation (Martin et al., 2012). Bioinformatics also suggests that there are small but clear differences in the potential folding states of the pre-mRNA (unmodified by splicing, etc) of these haplotypes. The analysis is somewhat limited due to the nature of the folding algorithms available, so the sequences were compared as five sections of near uniform length. It is expected that the differences would be even more pronounced if compared as a single molecule, but such a comparison is currently beyond the computational abilities of the available programs and systems at this time (unpublished communication). Of particular note are the numerous small stem loop structures that are predicted to vary between haplotypes. This is of particular interest due to the origin of intronic miRNA biogenesis. This could indicate changes in the miRNA intracellular pools, if new stem loops from these introns are processed accordingly (Lin et al., 2006; Z. Wang et al., 2013).

MicroRNAs can be generated from intronic mRNA with stem-loop structures, as pre-miRNAs, which are then processed into functional miRNAs in the cytoplasm. The problem in predicting if an intron can generate a miRNA is that it is currently beyond the scope of bioinformatics – the generated miRNA is only 18-25bases long, with the “seed” region of the guiding sequence only 2-8 bases long, and this region does not require perfect complementarity (Lin et al., 2006; Song and Chen, 2011). In light of this imperfect matching, of which no thermodynamic guiding parameters are currently identified, SNPs can only increase the number of potential miRNAs and their targets (Song and Chen, 2011). Others have shown miRNA SNPs can modify transcription of targeted genes and, ultimately, alter disease risk (Glinskii et al., 2011). For example, one experiment showed UV induced DNA damage (the UV type is not reported) triggers rapid damage induced both intercellular stress response and miRNA upregulation, whereby subsequent miR-16 knockdown resulted in S-phase accumulation of cells. Consequently, the authors suggest loss of miR-16 drives a blunted DNA damage response, halting the cell cycle (Adimoolam and Ford, 2002).

Western blot analysis determined if the significance in fold differences seen in the real time experiments were reflected in the protein levels. Interestingly, the amount of protein did not significantly change within a haplotype over time, however there were difference between the haplotypes at nearly every given time point. Haplotypes AA and DD had statistically significant increases for XPC specific protein over haplotypes EE and FF. Haplotype EE is particularly striking, given that these cells had the least amount of XPC. While this does fit with the unchanging *XPC* mRNA, it does not explain the enhanced repair capacity seen in the other experiments. One possible explanation for this would be significant differences in the protein folding that is directed by the mRNA allows the protein enhanced binding efficiency or stability, or perhaps increased localization to the nucleus leading to efficient repair with less protein. These same

explanations could account for the relative DRC insensitivity of the FF haplotypes. These experiments are a future direction of this research.

Although the XPC protein has been reported at ~105kDa based on the amino acid content, the functional protein has been reported at 132kDa or higher. After UV irradiation, the size of the XPC protein has been reported at a number of varying sizes from 150kDa and larger. This shift in mobility has been attributed to post-translational modification in response to damage, most especially to rapid (five minutes to one hour) ubiquitination (Sugasawa et al., 2005). XPC requires ubiquitination (and possibly sumoylation) for activity, as well as ubiquitination of other members of the repair complex (hHR23B) and cascade (DDB2), which was shown by proteasome inhibition and p53 dysregulation (reducing the amount DDB2) (Ford, 2005). A UV-C based experiment using fibroblasts and higher levels of radiation reported only a modest 4.6 fold maximal induction of XPC which was dependent on p53 activation, meanwhile p53-independent induction of XPC was much less (Adimoolam and Ford, 2002). It is possible that the UV-B dose used in my experiment is too low to induce p53, which in turn abrogates the XPC induction. Other experiments have shown that a lack of p53 induction is also a limiting factor for colocalization partners of XPC, such as DDB2, which has been shown to greatly enhance CPD repair (Ford, 2005). It should be noted, however, that p53 is not only involved in NER, but other repair pathways (e.g. BER, MMR, HJ, NHEJ...) and cellular processes (G1/S checkpoint control, apoptotic signaling, etc.) (Sengupta and Harris, 2005). Perhaps crosstalk between other systems can account for the variable repair seen in the AA haplotype cells, especially when coupled with the rapid, intense, and short lived *XPC* mRNA induction that does not correlate with any protein changes, yet still allows for repair of tobacco mutagens and is sensitive for CPD DRC (but not 6,4-PP). Coupled with the miRNA signaling, crosstalk mechanisms could easily play a role in the differences in repair seen for the different haplotypes.

Functional XPC can respond very rapidly, as was shown in a functional overexpression mutation study. In fact, localization to the sites of damage has been reported to occur within 15 seconds of irradiation, with a steady-state of <1 minute (Clement et al., 2011) confirming the speed of XPC's localization reported by others (Ford, 2005). This rapid response poses an interesting conundrum, as the locally high dose was administered with extreme precision (confocal based multiphoton fiber laser emission). It is possible that the physiological dose administered in the experiments presented in this dissertation are low enough to be within the capacity for the XPC and, lacking other intercellular signals to induce XPC, is the reason for a lack of detectable protein induction despite significantly increased mRNA in most of the haplotypes. Also, it appears that this signaling is more important for repair of CPDs as opposed to 6,4-PPs, which may be a mechanism for the apparent insensitivity of DD despite high sensitivity and repair of 6,4-PP. This may, in turn, account for the sustained increase in *XPC* mRNA induction over the longer time frame. The persistent damage may signal persistent *XPC* mRNA.

In summary, the same lymphoblastoid cell lines used to determine DRC of UV induced adducts were analyzed for induction of *XPC* specific gene products (mRNA and protein) over time. The haplotypes showed statistically significant differences in mRNA induction patterns, yet no significant changes in protein levels were observed. There were some statistically significant differences in protein levels between the haplotypes, which may provide some clues to the underlying mechanistic differences seen in the repair, however these differences may not be biologically significant. These clues provide a basis for further detailed analysis.

CONCLUSION

Haplotype analysis revealed clearly defined clusters of haplotypes present in a white non-Hispanic population of European ancestry. These haplotypes show differences in the repair response of tobacco mutagens. Mechanistic analysis shows these differences are persistent with other forms of damage, specifically UV-induced damage, but the relative DRC is dependent on the type of damage analyzed. Further analysis of *XPC* mRNA for these haplotypes shows unique patterns in mRNA induction that do not directly correlate to changes in XPC specific protein. These response patterns may contain further clues to specific mechanisms of action encoded within the haplotype. In short, this analysis shows that coding and non-coding SNPs of the *XPC* gene in LD with each other forming specific haplotypes act collectively to influence repair capacity by different mechanisms.

Future analysis will build on these studies to elucidate the individual driving mechanisms for the haplotypes. Clade A's results show variable DRC sensitivity, rapid and intense yet transient mRNA induction, and slightly higher protein levels, suggesting a possible mechanism based on a short term mRNA induction with long term effect. Such a mechanism might involve miRNA, as an example. All of the PGHs can be broken down in a similar manner. Clade E's results show variable DRC sensitivity, low but significant mRNA induction that is moderately sustained, and a slightly higher XPC protein content. A mechanism governing this would drive a weak mRNA change with no protein effect other than some inherent insensitivity, such as a change in XPC localization or binding partner interaction. Clade E results show high DRC sensitivity that has no corresponding mRNA induction and lower protein content. Lacking mRNA changes and having only a possible weak protein effect, the possible mechanism for this PGH may include conformational changes in XPC that enhance the DNA-protein interaction. Clade F, by comparison, shows low DRC sensitivity with a slow but moderately intense mRNA

induction that is sustained yet shows only low XPC protein content, suggesting a possible mechanism based on a sustained mRNA induction signal without protein changes such as lowered stability or higher turn-over of the mRNA signal. These various mechanisms are going to be probed using splicing and stability analysis, immunocytochemistry localization and colocalization of protein binding partners and damaged DNA, immunoprecipitation and coimmunoprecipitation, and miRNA analysis. Additionally, future analysis should include analysis of the individual haplotypes and unique SNP contributions by whole gene transgenetic analysis and DNA mutation analysis. From here, biomarker evaluation, prognostic analysis, and individualized treatment are ultimate goals.

The future of this work should have a number of possible implications to human health. With the current push towards individualized medicine, comprehensive biomarkers of risk are needed. Additionally, after disease development, treatment options can be evaluated in light of expected phenotypes. For example, an individual with a strong repair haplotype (such as one from clade E) is less likely to respond to a chemotherapeutic like cisplatin, whose adduct is likely to be effectively repaired via the NER pathway and thus mitigate the treatment. Understanding the mechanisms that drive each haplotype should allow for the development of more effective treatments or those with fewer side-effects. These are just some of the long reaching applications for this work.

SUPPLEMENTAL DATA

The following figures and tables are additional information for this dissertation. They have been included for the purposes of additional information and clarification for the reader if desired. Summaries and key data are included in the main body of the text.

Table 17: Haplotype SNP (htSNP) analysis by tagger software

A

#captured 35 of 35 alleles at $r^2 \geq 0.8$
 #captured 100 percent of alleles with mean r^2 of 0.971
 #using 11 Tag SNPs in 11 tests.

Allele	Best Test	r^2 w/test
rs8516	rs1106087	0.921
rs10468	rs1106087	0.845
rs1126547	rs1126547	1
rs2470352	rs2470352	1
rs2229090	rs1106087	0.951
rs2228001	rs2228001	1
rs2733532	rs2228001	0.901
rs2733533	rs2607775	1
rs2733534	rs2607775	1
rs2279017	rs2228001	0.966
rs2470353	rs2607775	1
rs2607734	rs2228001	0.966
rs2607736	rs2228001	0.958
rs2607737	rs2607775	1
rs3731149	rs3731124	1
rs3731146	rs3731124	0.903
rs9653966	rs9653966	1
rs1124303	rs1124303	1
rs3731143	rs2227999	1
rs2228000	rs1106087	0.961
rs2227999	rs2227999	1
rs3731127	rs3731127	1
rs3731125	rs9653966	1
rs3731124	rs3731124	1
rs13099160	rs3731127	1
rs1106087	rs1106087	1

rs3731108	rs3731124	1
rs3731106	rs3731124	0.906
rs3729587	rs3731124	0.9
rs3731093	rs9653966	1
rs2733537	rs2733537	1
rs3731081	rs3731124	1
rs3731068	rs3731124	0.806
rs1350344	rs2607775	1
rs2607775	rs2607775	1

B

Test	Alleles Captured							
rs3731124	rs3731124	rs3729587	rs3731149	rs3731146	rs3731108	rs3731081	rs3731106	rs3731068
rs2607775	rs2470353	rs2733533	rs2607775	rs2607737	rs1350344	rs2733534		
rs2228001	rs2279017	rs2733532	rs2607736	rs2228001	rs2607734			
rs1106087	rs1106087	rs10468	rs2229090	rs8516	rs2228000			
rs9653966	rs3731093	rs9653966	rs3731125					
rs3731127	rs13099160	rs3731127						
rs2227999	rs3731143	rs2227999						
rs1126547	rs1126547							
rs2733537	rs2733537							
rs2470352	rs2470352							
rs1124303	rs1124303							

Table 17: Haplotype SNP (htSNP) analysis by tagger software. A: SNPs paired for tagging my rs number and R^2 value of correlation. Threshold analysis of allele capture was set to $R^2=0.8$. 100% of alleles were captured with a mean $R^2=0.971$. B: The list of the SNPs captured by the htSNP alleles as determined by tagger analysis.

Table 18: Genotyping of the htSNPs alleles by study participant for the UTMB White non-Hispanic cohort of the experimental population

ID	SNP3	SNP4	SNP6	SNP17	SNP18	SNP21	SNP22	SNP24	SNP26	SNP31	SNP35
S126	CC	AT	AC	TT	GT	GG	CC	AA	GT	AG	CG
S132	CG	AT	AC	TT	TT	GG	CC	AA	GT	AG	CG
S136	CC	AA	CC	TT	TT	GG	CC	AA	GG	AA	CC
S140	CC	AA	AC	GT	TT	GG	CC	AA	GG	AG	CC
S142	CC	TT	AA	TT	TT	GG	CC	AA	TT	GG	GG
S144	CC	AT	AA	GT	TT	GG	CC	AA	GT	GG	CG
S147	CC	AA	AC	TT	GT	GG	CC	AC	GG	AA	CG
S148	CC	AA	AC	TT	TT	GG	CC	AC	GG	AA	CG
S156	CG	AA	CC	TT	TT	GG	CC	AA	GG	AA	CC
S157	CG	TT	AA	TT	TT	AG	CC	AA	TT	GG	GG

S161	CC	AA	AC	GT	TT	GG	CT	AA	GG	AG	CC
S180	CC	TT	AC	TT	TT	GG	CC	AA	GT	AG	CG
S187	CC	AA	CC	TT	TT	GG	CC	AA	GG	AA	CC
S190	CC	TT	AC	TT	GT	AG	CC	AA	GT	AG	n
S197	CC	AT	AC	GT	TT	GG	CC	AA	GG	AA	CC
S200	CC	AA	AC	TT	TT	GG	CC	AC	GG	AA	CG
S212	CC	AT	AC	TT	n	n	n	n	GT	AG	CG
S213	CC	AT	AA	TT	GT	GG	CC	CC	GG	AA	GG
S215	CC	TT	AA	TT	TT	GG	CC	AC	GT	AA	GG
S216	GG	AT	AC	TT	TT	GG	CC	AC	GG	AA	CG
S218	CG	AA	AA	GT	TT	GG	CT	AC	GG	AG	CG
S220	CC	AA	CC	TT	TT	GG	CC	AA	GG	AA	CC
S221	CC	AT	AA	TT	GT	GG	CC	AC	GG	AA	GG
S224	CC	AT	AC	TT	TT	GG	CC	AC	GG	AG	CG
S225	CG	AT	AA	TT	TT	AG	CC	AA	TT	GG	CG
S227	CC	AA	CC	TT	TT	GG	CC	AA	GG	AA	CC
S228	CC	TT	AC	TT	TT	AG	CC	AA	GT	AG	CG
S230	CC	AT	AC	TT	TT	GG	CC	AA	GT	AG	CG
S234	CC	AT	AC	TT	TT	GG	CC	AC	GG	AA	CG
S235	CC	AT	AC	TT	TT	AG	CC	AA	GG	AG	CG
S236	CC	AT	AC	TT	TT	GG	CC	AC	GG	AG	CG
S244	CC	TT	AA	TT	GT	GG	CC	AA	GT	AG	CG
S246	CC	AT	CC	TT	GT	GG	CC	AA	GG	AA	CC
S247	CC	TT	AC	TT	TT	GG	CC	AA	GT	AG	CG
S263	CC	AA	AC	GT	TT	GG	CT	AA	GG	AG	CC
S264	CC	AA	AA	TT	GT	GG	CC	AC	GG	AA	GG
S268	CC	AT	CC	TT	TT	GG	CC	AA	GG	AA	CC
S269	CG	AA	AA	GG	TT	GG	CT	AA	GT	AA	CC
S320	CC	AT	AC	TT	TT	AG	CC	AA	GT	AG	CG
S322	CC	AT	AC	TT	TT	GG	CC	AA	GT	AG	CG
S323	n	AT	AA	GT	TT	GG	CT	AA	GT	GG	CG
S328	CC	AT	AC	GT	TT	GG	CT	AA	GG	AG	CC
S329	CC	AA	AC	GT	TT	GG	CT	AA	GG	AG	CC
S366	CC	TT	AC	TT	TT	GG	CC	AA	GT	AG	CG
S371	CC	AT	AA	TT	TT	AG	CC	AC	GT	AG	GG
S372	CC	AT	AC	TT	TT	GG	CC	AC	GG	AA	CG
S375	CC	AA	AC	TT	TT	GG	CC	AC	GG	AG	n
S376	CC	AA	AA	GT	TT	GG	CT	AA	GT	GG	CC
S388	CC	AA	AC	TT	TT	GG	CC	AC	GG	AA	CG
S391	CC	AA	AA	GT	TT	GG	CC	AA	GT	GG	CC
S393	CC	AT	AC	TT	n	GG	CC	AC	GG	AA	CG
S394	CC	AT	AC	GT	TT	GG	CC	AC	GG	AG	CC
S400	CC	AT	AC	TT	n	GG	CC	AC	GG	AA	CG

S404	CG	AT	AC	TT	TT	GG	CC	AC	GG	AA	CG
S408	CC	AA	AA	TT	GT	GG	CC	CC	GG	AA	GG
S409	CG	AA	AA	TT	GT	n	CC	CC	GG	AA	GG
S412	CG	AT	AA	TT	GT	GG	CC	CC	GG	AA	GG
S420	CC	AA	AC	TT	TT	GG	CT	AC	GG	AA	CG
S424	CC	AA	CC	TT	TT	GG	CC	AA	GG	AA	CC
S426	n	AA	AA	GT	TT	GG	CT	AC	GG	AG	CG
S428	CC	AA	CC	TT	TT	GG	CC	AA	GG	AA	CC
S431	CC	AA	AC	TT	TT	GG	CC	AA	GT	AG	CC
S436	CC	AA	AA	TT	GT	GG	CC	CC	GG	AA	GG
S439	CG	AA	AA	TT	TT	GG	CC	AC	GT	AG	CG
S442	CC	TT	AC	TT	n	GG	CC	AA	GT	AG	CG
S446	CC	n	AC	GT	GT	AG	CT	n	GG	AG	CG
S467	CC	AT	AC	TT	n	GG	CC	AA	GT	AG	CG
S469	CG	TT	AA	TT	TT	GG	CC	AA	TT	GG	GG
S472	CC	AT	CC	TT	TT	GG	CC	AA	GG	AA	CC
S494	CG	AA	AC	TT	TT	GG	CC	AA	GG	AA	CG
S496	CC	AA	AA	TT	TT	GG	CC	AC	GG	AA	GG
S497	CC	AT	AC	TT	TT	GG	CC	AC	GG	AA	CG
S525	CC	TT	AA	TT	GT	GG	CC	AC	GT	AG	GG
S526	GG	TT	AC	TT	TT	GG	CC	AA	GT	AG	CG
S527	CC	AA	AA	TT	TT	GG	CC	AC	GG	AG	CG
S538	CC	TT	AA	TT	TT	GG	CC	AA	TT	GG	GG
S541	CG	AT	AC	TT	TT	GG	CC	AC	GG	AA	CG
S542	CC	AT	AC	GT	n	GG	CT	AA	n	AG	CC
S544	CG	AT	AC	TT	n	GG	CC	AC	GG	AA	CG
S550	CG	AA	AA	TT	TT	GG	CC	AC	GG	AA	GG
S556	CC	AA	AA	GT	TT	GG	CT	AC	GG	AG	CG
S560	CC	AT	AC	TT	TT	AG	CC	AA	GT	AG	CG
S563	CC	AT	AA	GT	TT	GG	CC	AA	GT	GG	CG
S565	CG	AT	AA	TT	GT	GG	CC	CC	GG	AA	GG
S580	CC	AA	AC	TT	TT	GG	CC	AC	GG	AA	CG
S582	CC	TT	AA	TT	GT	GG	CC	AC	GT	AG	GG
S583	CC	AT	AC	TT	TT	GG	CC	AA	GG	AA	CG
S585	CC	AA	AA	TT	TT	GG	CC	CC	GG	AA	GG
S586	CC	AA	AC	GT	TT	GG	CT	AA	GG	AG	CC
S589	CC	AT	AA	TT	GT	GG	CC	AC	GT	AG	GG
S590	CC	TT	AC	TT	GT	GG	CC	AA	GT	AG	CG
S591	CC	AA	AC	GT	TT	GG	CT	AA	GG	AG	CC
S593	CC	AA	AC	TT	TT	GG	CC	AA	GT	AG	CC
S607	CC	AT	AA	GT	TT	GG	CC	AA	GT	GG	CG
S610	CC	AT	AC	TT	TT	GG	CC	AA	GT	AG	CG
S612	CC	AT	CC	TT	GT	GG	CC	AA	GG	AA	CC

S614	CC	AA	CC	TT	TT	GG	CC	AA	GG	AA	CC
S635	CC	AT	AA	TT	GT	GG	CC	AC	GT	AG	CG
S636	CC	AA	AC	TT	n	n	n	n	n	AA	n
S654	CC	AA	CC	TT	TT	GG	CC	AA	GG	AA	CC
S655	CC	AT	AA	GT	GT	GG	CT	AC	GG	AG	CG
S658	CC	AT	AC	TT	TT	GG	CC	AA	GT	AG	CG
S661	CC	TT	AA	TT	GT	AG	CC	AC	GT	AG	GG
S664	GG	n	AA	TT	n	n	CC	AA	GT	GG	n
S666	CG	n	AA	n	n	n	CC	AA	TT	GG	n
S667	CC	TT	AA	GT	n	GG	CT	AA	GT	GG	CG
S672	CC	TT	AA	TT	n	GG	CC	AC	GT	AG	GG
S677	CG	AA	AA	GT	GT	GG	CC	AC	GG	AG	CG
S682	GG	AT	AC	TT	GT	AG	CT	AA	GT	AG	CG
S687	CC	AA	AA	GT	n	n	CT	AC	GG	AG	n
S688	CC	AA	AC	TT	n	GG	CC	AC	GG	AA	CG
S692	CC	AT	AC	TT	TT	GG	CC	AA	GT	AG	CG
S700	CC	AT	AC	TT	GT	GG	CC	AC	GG	AA	CG
S701	CC	TT	AC	TT	GT	GG	CC	AA	GG	AG	CG
S704	CC	AT	AC	GT	TT	GG	CT	AA	GG	AG	CC
S705	CC	AA	AC	TT	TT	GG	CC	AC	GG	AA	CG
S710	CG	AT	AA	TT	TT	AG	CC	AC	GT	AG	GG
S715	CG	AT	AA	TT	TT	GG	CC	AC	GT	AG	GG
S726	CC	AT	AC	TT	TT	GG	CC	AA	GG	AA	CC
S728	CC	AT	AC	TT	TT	GG	CC	AA	GT	AG	CG
S733	CC	AT	CC	TT	n	GG	CC	AA	GG	AA	CC
S754	CG	TT	AC	TT	TT	GG	CC	AA	GT	AG	CG
S785	CG	AA	AC	TT	GT	GG	CC	AC	GG	AA	CG

Table 18: Genotyping of the htSNPs alleles by study participant for the UTMB White non-Hispanic cohort of the experimental population. Each participant was designated by an S number as listed in column one corresponding with a larger population as previously described. SNPs are listed by their position number in the larger haplotype of 35. Alleles are listed by nucleotide base, with A= adenine, C= cytosine, G= guanine, and T= thymine. An “n” denotes a non-callable genotype by Taqman analysis.

Table 19: Assignment of haplotype by study participant for the UTMB White non-Hispanic cohort of the experimental population

ID	Sex	Age	Smoking Status	Baseline	NNK	Haps
S-126	M	55	Smoker	0	5	
S-132	M	73	Smoker	0	4	C,F
S-136	F	67	Smoker	0	5	F,F
S-140	F	42	non	2	7	F,F

S-142	F	49	Smoker	1	3	C,C
S-144	M	40	Smoker	0	11	C,F
S-147	M	73	non	1	0	A,F
S-148	F	61	Smoker	1	3	A,F
S-156	F	35	non	3	3	F,F
S-157	M	35	non	0	7	C,C
S-161	F	68	Smoker	0	2	E,F
S-180	F	51	non	2	5	C,F
S-187	F	65	non	1	7	F,F
S-190	M	51	non	1	9	
S-197	F	46	Smoker	0	10	F,F
S-200	M	72	Smoker	0	3	A,F
S-213	F	28	Smoker	1	3	A,A
S-215	F	25	non	1	5	
S-216	M	53	Smoker	0	4	A,F
S-218	M	42	Smoker	0	0	A,E
S-220	F	40	Smoker	0	3	F,F
S-221	M	53	non	0	9	A,B
S-224	F	48	Smoker	2	3	
S-225	F	23	non	0	3	C,D
S-227	F	40	Smoker	0	5	F,F
S-228	F	28	Smoker	2	5	C,F
S-230	F	23	non	0	6	C,F
S-234	M	47	Smoker	0	10	A,F
S-235	F	36	non	2	5	
S-236	F	27	Smoker	0	6	
S-244	F	30	non	2	6	A,D
S-246	M	33	Smoker	2	3	
S-247	F	28	Smoker	1	5	C,F
S-263	F	23	non	1	4	E,F
S-264	F	24	non	1	2	A,B
S-268	M	54	non	0	1	F,F
S-269	F	70	non	0	7	
S-320	F	39	non	0	7	C,F
S-322	F	24	smoker	0	12	C,F
S-323	F	37	smoker	1	2	C,E
S-328	F	27	smoker	0	8	E,F
S-329	F	44	smoker	0	8	E,F
S-366	F	29	Smoker	0	7	C,F
S-371	F	44	non	0	5	A,C
S-372	F	48	smoker	1	3	A,F
S-375	F	41	non	1	5	
S-376	F	42	non	0	2	D,E

S-388	F	24	smoker	1	4	A,F
S-391	F	42	smoker	2	7	D,F
S-393	F	34	non	1	7	A,F
S-394	F	26	non	1	8	
S-400	F	49	non	1	6	A,F
S-404	F	25	smoker	1	4	A,F
S-408	F	24	smoker	0	5	A,A
S-409	M	42	non	1	3	A,A
S-412	F	48	non	0	7	A,A
S-420	F	58	non	2	16	
S-424	F	36	smoker	2	9	F,F
S-426	F	23	non	1	10	A,E
S-428	F	44	smoker	2	9	F,F
S-431	F	36	smoker	3	16	D,F
S-436	F	44	non	1	4	A,A
S-439	F	44	smoker	2	6	A,D
S-442	M	34	smoker	1	4	C,F
S-446	F	40	non	2	4	
S-467	M	42	smoker	2	8	C,F
S-469	F	41	non	1	4	C,C
S-472	F	27	non	0	0	F,F
S-494	F	37	non	0	7	B,F
S-496	F	34	non	0	3	A,B
S-497	F	28	non	0	2	A,F
S-525	F	49	smoker	3	11	A,C
S-526	F	26	smoker	2	3	C,F
S-527	M	32	non	0	10	
S-538	F	34	smoker	1	4	C,C
S-541	F	49	non	4	4	A,F
S-544	F	51	smoker	2	10	A,F
S-550	F	39	smoker	0	1	A,B
S-556	F	41	non	0	1	A,E
S-560	M	23	non	2	6	C,F
S-563	F	30	smoker	5	6	C,F
S-565	F	47	smoker	1	3	A,A
S-580	F	58	non	0	7	A,F
S-582	F	50	non	1	8	A,C
S-583	F	45	smoker	0	4	B,F
S-585	F	35	non	0	8	A,A
S-586	M	34	smoker	1	7	E,F
S-589	M	38	non	1	6	A,C
S-590	F	47	smoker	1	8	
S-591	F	23	non	2	8	E,F

S-593	F	38	smoker	0	6	D,F
S-607	F	24	non	0	2	C,F
S-610	F	30	smoker	0	5	C,F
S-612	F	46	non	1	6	
S-614	M	23	smoker	0	4	F,F
S-635	F	25	non	1	4	A,D
S-654	F	24	non	0	2	F,F
S-655	M	47	non		4	A,E
S-658	M	37	non		6	C,F
S-661	F	26	non	0	4	A,C
S-667	F	28	smoker	1	3	C,E
S-672	F	47	non	0	9	A,C
S-677	M	69	non	1	5	A,F
S-682	F	21	smoker	0	5	
S-688	F	21	non	0	2	A,F
S-692	F	23	smoker	0	2	C,F
S-700	F	29	smoker	0	4	A,F
S-701	F	29	smoker	0	1	
S-704	F	28	non	0	3	E,F
S-705	F	24	non	1	4	A,F
S-710	M	33	non	0	8	A,C
S-715	F	39	smoker	1	7	A,C
S-726	F	20	non	1	3	
S-728	F	37	non	1	7	C,F
S-733	F	46	non	1	7	F,F
S-754	F	28	smoker	2	6	C,F
S-785	F	58	smoker	0	3	A,F

Table 19: Assignment of haplotype by study participant for the UTMB White non-Hispanic cohort of the experimental population. Each individual is listed by S number, and the corresponding haplotype assignment is given in the Haps column. For individuals with no haplotype listed, they were dropped from the study. Demographic breakdown is listed for age, gender, and smoking status. Chromosomal aberration (CA) data is given as baseline (no *in vitro* exposure) and NNK (72 hours after NNK exposure) of isolated primary lymphocytes.

Table 20: Raw CA stats by haplotype for both baseline and 24 hours after *in vitro* NNK treatment of isolated primary lymphocytes for the UTMB White non-Hispanic cohort of the experimental population

A

BASELINE	
----------	--

smoking	hap	presence	mean	SEM	CI 95%
no	A	y	0.74	0.174	0.341
yes	A	y	0.78	0.207	0.406
no	A	n	0.7	0.206	0.404
yes	A	n	0.91	0.208	0.408
no	B	y	0.25	0.25	0.49
yes	B	y	0	0	0
no	B	n	0.77	0.141	0.276
yes	B	n	0.9	0.155	0.304
no	C	y	0.43	0.173	0.339
yes	C	y	1.21	0.292	0.572
no	C	n	0.85	0.169	0.331
yes	C	n	0.65	0.158	0.31
no	D	y	0.75	0.479	0.939
yes	D	y	1.75	0.629	1.233
no	D	n	0.72	0.139	0.272
yes	D	n	0.78	0.152	0.298
no	E	y	0.86	0.34	0.666
yes	E	y	0.43	0.202	0.396
no	E	n	0.7	0.144	0.282
yes	E	n	0.93	0.171	0.335
no	F	y	0.88	0.211	0.414
yes	F	y	0.82	0.184	0.361
no	F	n	0.55	0.143	0.28
yes	F	n	1	0.246	0.482

B

NNK					
smoking	hap	presence	mean	SEM	CI 95%
no	A	y	4.46	0.516	1.011
yes	A	y	3.89	0.62	1.215
no	A	n	4	0.512	1.004
yes	A	n	5.12	0.56	1.098
no	B	y	5	1.826	3.579
yes	B	y	2.5	1.5	2.94
no	B	n	4.2	0.37	0.725
yes	B	n	4.77	0.438	0.858
no	C	y	5.4	0.533	1.045
yes	C	y	4.47	0.735	1.441
no	C	n	3.76	0.449	0.88

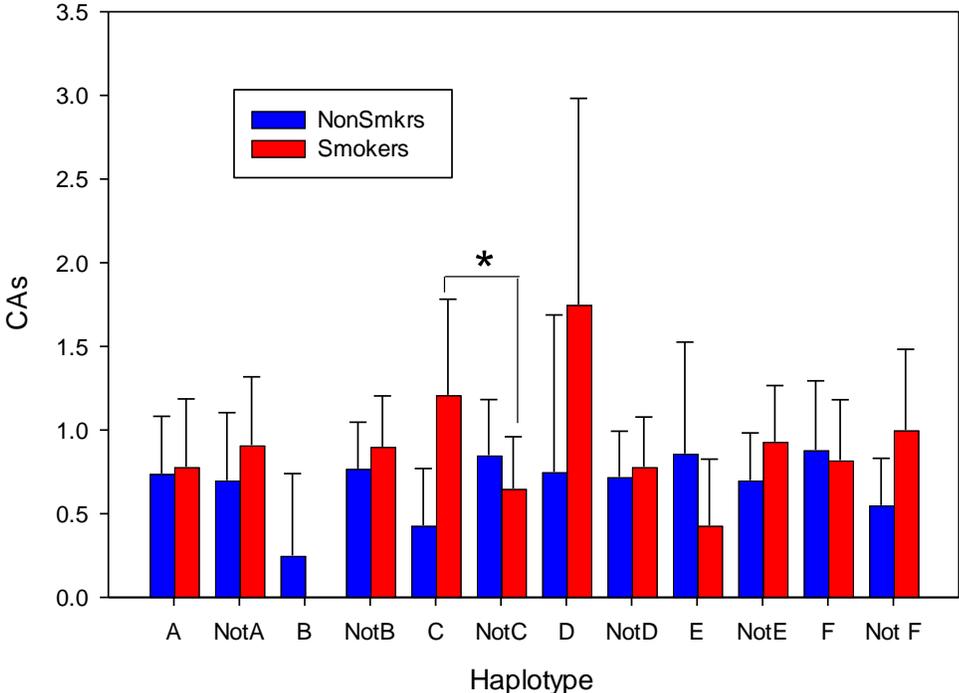
yes	C	n	4.81	0.53	1.039
no	D	y	3	0.408	0.8
yes	D	y	7	2.041	4
no	D	n	4.38	0.392	0.768
yes	D	n	4.48	0.425	0.833
no	E	y	4	0.886	1.737
yes	E	y	3.86	1.28	2.509
no	E	n	4.32	0.405	0.794
yes	E	n	4.81	0.454	0.89
no	F	y	3.77	0.478	0.937
yes	F	y	5.21	0.493	0.966
no	F	n	4.83	0.547	1.072
yes	F	n	3	0.674	1.321

Table 20: Raw CA stats by haplotype for both baseline and 24 hours after in vitro NNK treatment of isolated primary lymphocytes for the UTMB White non-Hispanic cohort of the experimental population. The presence or absence of haplotypes were determined using a dominant model. Smoking status was designated as either non-smoker or cigarette smoker. Data is compiled as mean with standard error of mean and the 95% confidence interval (CI) is listed for each combination. A. Baseline CAs. B. CAs after NNK treatment.

Figure 17: Comparison of CA data by clade

A

CAs at Baseline by Haplotype



B

CAs after NNK by Haplotype

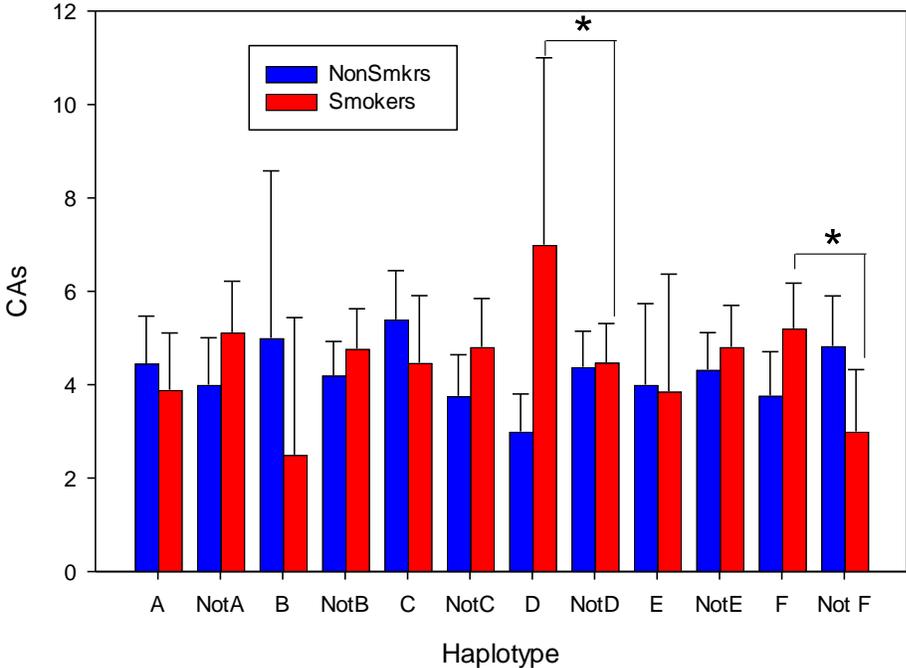


Figure 17: Comparison of CA data by clade. Individual haplotype designations are presented as presence (X) or absence (NotX). Cigarette smokers are presented in red bars, while non-smokers are presented in blue. Statistical significance (*) was determined as a p-value ≤ 0.05 for associations between smoking status and haplotype. Data values are presented in table 20. PGHs A: smokers N=18, non-smokers N=27; notA: smokers N=71, non-smokers 57; B: smokers N=2, non-smokers N=4; notB: smokers N=87, non-smokers N=80; C: smokers N=20, non-smokers N=16; notC: smokers N=69, non-smokers N=68; D: smokers N=3, non-smokers N=4; not D: smokers N=86, non-smokers N=80; E: smokers N=7, non-smokers N=7; notE: smokers N=82, not-smokers N=77; F: smokers N=39, non-smokers N=26; notF: smokers N=50, non-smokers N=58. **A:** CA data at baseline by smoking status for each haplotype designation. Haplotype group C was the only haplotype determined to show statistical significance at baseline, with presense of C showing an interaction with smoking for increased CAs. **B:** CA data at 72 hours after NNK exposure by smoking status for each haplotype designation. Haplotypes D and F were the only haplotypes determined to show statistical significance at 72 hours after NNK exposure, with presense of D or F showing an interaction with smoking for increased CAs.

Table 21: Percent genetic divergence between all 21 haplotypes as determined by PHASE analysis

Hap#	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
1		5.7	48.6	51.4	54.3	54.3	51.4	22.9	25.7	17.1	8.6	57.1	25.7	28.6	40.0	42.9	40.0	45.7	25.7	42.9	48.6
2	5.7		54.3	57.1	60.0	60.0	57.1	28.6	31.4	22.9	2.9	62.9	31.4	45.7	48.6	45.7	51.4	51.4	31.4	48.6	54.3
3	48.6	54.3		2.9	5.7	5.7	2.9	48.6	51.4	48.6	8.6	51.4	54.3	14.3	40.0	37.1	42.9	57.1	57.1	40.0	45.7
4	51.4	57.1	2.9		2.9	2.9	2.9	51.4	54.3	51.4	60.0	5.7	54.3	17.1	42.9	40.0	45.7	40.0	45.7	60.0	42.9
5	54.3	60.0	5.7	2.9		5.7	8.6	54.3	57.1	48.6	8.6	57.1	60.0	20.0	40.0	37.1	42.9	57.1	57.1	40.0	45.7
6	54.3	60.0	5.7	2.9	5.7		8.6	54.3	57.1	54.3	2.9	57.1	60.0	20.0	45.7	42.9	48.6	62.9	62.9	45.7	51.4
7	51.4	57.1	2.9	5.7	8.6	8.6		45.7	48.6	51.4	60.0	11.4	48.6	17.1	42.9	40.0	45.7	40.0	60.0	42.9	48.6
8	22.9	28.6	48.6	51.4	54.3	54.3	45.7		2.9	22.9	31.4	57.1	8.6	5.7	40.0	48.6	45.7	51.4	31.4	48.6	54.3
9	25.7	31.4	51.4	54.3	57.1	57.1	48.6	2.9		25.7	34.3	60.0	5.7	2.9	42.9	51.4	48.6	54.3	34.3	51.4	57.1
10	17.1	22.9	48.6	51.4	48.6	54.3	51.4	22.9	25.7		25.7	57.1	25.7	28.6	40.0	31.4	22.9	28.6	8.6	25.7	31.4
11	8.6	2.9	57.1	60.0	62.9	62.9	60.0	31.4	34.3	25.7		60.0	34.3	37.1	48.6	51.4	48.6	54.3	28.6	45.7	51.4
12	57.1	62.9	8.6	5.7	8.6	2.9	11.4	57.1	60.0	57.1	60.0		60.0	62.9	22.9	48.6	45.7	51.4	60.0	42.9	48.6
13	25.7	31.4	51.4	54.3	57.1	57.1	48.6	8.6	5.7	25.7	34.3	60.0		2.9	42.9	51.4	48.6	54.3	34.3	51.4	57.1
14	28.6	34.3	54.3	57.1	60.0	60.0	51.4	5.7	2.9	28.6	37.1	62.9	2.9		45.7	54.3	51.4	57.1	37.1	54.3	60.0
15	40.0	45.7	14.3	17.1	20.0	20.0	17.1	40.0	42.9	40.0	48.6	22.9	42.9	45.7		25.7	22.9	28.6	42.9	25.7	31.4
16	42.9	48.6	40.0	42.9	40.0	45.7	42.9	48.6	51.4	51.4	48.6	45.7	51.4	54.3	25.7		8.6	2.9	28.6	11.4	5.7
17	40.0	45.7	37.1	40.0	37.1	42.9	40.0	45.7	48.6	22.9	48.6	45.7	48.6	51.4	22.9	8.6		5.7	20.0	2.9	8.6
18	45.7	51.4	42.9	45.7	42.9	48.6	45.7	51.4	54.3	28.6	54.3	51.4	54.3	57.1	28.6	2.9	5.7		25.7	8.6	2.9
19	25.7	31.4	57.1	60.0	57.1	62.9	60.0	31.4	34.3	8.6	28.6	60.0	34.3	37.1	42.9	28.6	20.0	25.7		17.1	22.9
20	42.9	48.6	40.0	42.9	40.0	45.7	42.9	48.6	51.4	25.7	45.7	42.9	51.4	54.3	25.7	11.4	2.9	8.6	17.1		5.7
21	48.6	54.3	45.7	48.6	45.7	51.4	48.6	54.3	57.1	31.4	51.4	48.6	57.1	60.0	31.4	5.7	8.6	2.9	22.9	5.7	

Table 21: Percent genetic divergence between all 21 haplotypes as determined by PHASE analysis. Numbers correspond to the haplotypes as listed in table 6 figure 5. Values for the haplotypes chosen for each cell culture and mechanistic analysis are highlighted for emphasis. The percent genetic divergence between PGHs ranged from

25.7% (DD haplotype 19 and EE haplotype 1) up to 62.9% (AA haplotype 12 and EE haplotype 2).

Table 22: Raw growth counts for each of the haplotype lines over time

Growth Counts	NT(AA)	AA	SD	NT(DD)	DD	SD	NT(EE)	EE	SD	NT(FF)	FF	SD
24hr	99.26	103.40	14.33	153.85	182.05	70.12	180.00	157.06	45.64	141.91	180.50	57.89
48hr	210.48	111.67	12.41	307.69	280.77	63.32	232.94	192.35	22.04	264.71	243.53	97.19
72 hr	272.98	244.03	91.46	273.08	374.36	112.62	259.41	245.29	79.47	492.16	420.00	117.95

Table 22: Raw growth counts for each of the haplotype lines over time. Raw data for the total number of cells present by cell line after 35mJ/cm² UV-B exposure at 24, 48, and 72 hours as determined from trypan blue exclusion. NT designates the amount of cells in a matched unexposed culture from the same cell line/haplotype. All haplotypes showed at least the expected doubling of total cells within 72 hours post-exposure.

Figure 18: Percent viability for each of the haplotype lines over time

Percent Viability vs No Treatment for 35mJ/cm² by PGH

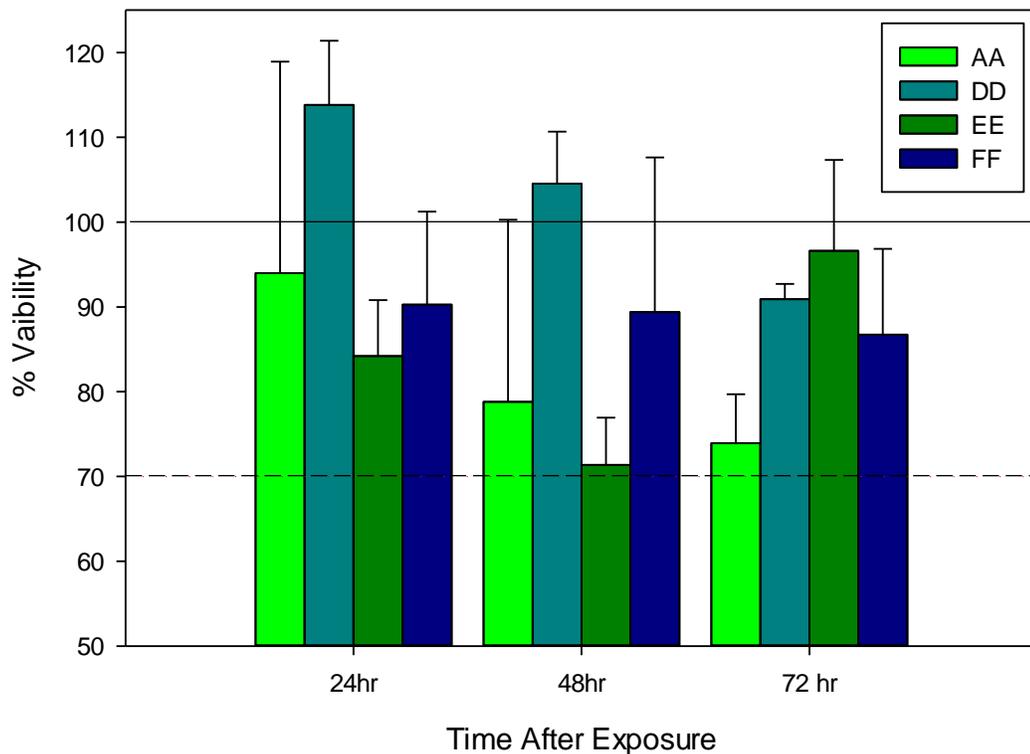


Figure 18: Percent viability for each of the haplotype lines over time. Comparison of the percent viability of cells present by cell line after 35mJ/cm² UV-B exposure at 24, 48, and 72 hours. An ideal target of 100% is shown with a solid line, while a minimal viability cutoff is represented with a dashed line. Viability was calculated as alive versus dead cells as determined by trypan blue exclusion analysis, and each time point was normalized to the respective viability of the untreated culture at the corresponding time point. Bright green represents haplotype AA, light blue DD, dark green EE, and blue FF. All haplotypes were above the 70% threshold, with many peaking over the 100% expected values.

Table 23: Raw viability data for each of the haplotype lines over time

viability	AA	SD	DD	SD	EE	SD	FF	SD
24hr	93.99	24.94	113.82	7.56	84.20	6.61	90.27	10.96
48hr	78.78	21.50	104.54	6.14	71.36	5.57	89.41	18.22
72 hr	73.90	5.78	90.91	1.81	96.62	10.71	86.75	10.11

Table 23: Raw viability data for each of the haplotype lines over time. For comparison, the percent viability of cells present by cell line after 35mJ/cm² UV-B exposure at 24, 48, and 72 hours.

Table 24: Shapiro-Wilk normality data for each of the haplotypes over time by adduct

A

minPE	AA	p	DD	p	EE	p	FF	p	AA+EE	p	DD+FF	p
0	0.957	0.786	0.970	0.916	0.779	0.005	0.870	0.100	0.913	0.074	0.701	0.000
1	0.920	0.428	0.952	0.692	0.795	0.008	0.930	0.483	0.876	0.018	0.912	0.080
5	0.839	0.074	0.882	0.094	0.741	0.004	0.807	0.025	0.820	0.004	0.629	0.000
15	0.865	0.134	0.931	0.392	0.682	0.001	0.810	0.012	0.793	0.001	0.663	0.000
30	0.955	0.750	0.954	0.696	0.705	0.001	0.710	0.001	0.855	0.008	0.521	0.000
60	0.986	0.988	0.942	0.545	0.889	0.114	0.831	0.021	0.950	0.313	0.660	0.000
180	0.929	0.508	0.848	0.056	0.673	0.000	0.865	0.057	0.812	0.001	0.746	0.000
360	0.926	0.517	0.840	0.031	0.736	0.002	0.696	0.001	0.918	0.104	0.629	0.000
720	0.953	0.746	0.870	0.101	0.856	0.043	0.786	0.007	0.955	0.442	0.610	0.000
1440	0.885	0.208	0.927	0.354	0.852	0.039	0.834	0.024	0.845	0.004	0.720	0.000

B

minPE	AA	p	DD	p	EE	p	FF	p	AA+FF	p	DD+EE	p
0	0.846	0.087	0.859	0.047	0.945	0.570	0.923	0.385	0.920	0.130	0.870	0.005
1	0.934	0.553	0.917	0.334	0.955	0.710	0.937	0.547	0.966	0.747	0.930	0.141
5	0.832	0.062	0.854	0.041	0.886	0.180	0.949	0.679	0.704	0.000	0.851	0.004
15	0.887	0.219	0.983	0.992	0.910	0.213	0.840	0.028	0.919	0.096	0.965	0.548
30	0.927	0.451	0.917	0.292	0.930	0.449	0.971	0.898	0.952	0.431	0.916	0.073
60	0.957	0.747	0.794	0.008	0.960	0.781	0.899	0.154	0.976	0.835	0.751	0.000
180	0.761	0.011	0.913	0.300	0.959	0.768	0.921	0.295	0.944	0.281	0.976	0.851
360	0.866	0.171	0.941	0.530	0.935	0.431	0.925	0.401	0.948	0.430	0.960	0.473
720	0.926	0.480	0.961	0.800	0.955	0.716	0.789	0.007	0.854	0.006	0.981	0.927
1440	0.888	0.226	0.972	0.931	0.968	0.887	0.879	0.085	0.938	0.222	0.970	0.662

Table 24: Shapiro-Wilk normality data for each of the haplotypes over time by adduct. Time is presented in minutes after UV-B exposure (minPE). Failure of the test is indicated by statistical significance, or p-value ≤ 0.05 . **A:** Normality tests by haplotype for CPD ELISA analysis. **B:** Normality tests by haplotype for 6,4-PP ELISA analysis.

Table 25: Mann-Whitney test of significance data for each of the haplotypes over time by adduct

A

AA(CPD)	0	1	5	15	30	60	180	360	720	1440
0		0.844	0.623	0.976	0.188	0.251	0.984	0.757	0.713	0.915
1	0.844		0.953	0.680	0.543	0.507	0.616	0.442	1.000	0.794
5	0.623	0.953		0.208	<i>0.013</i>	<i>0.031</i>	0.345	0.522	0.188	0.256
15	0.976	0.680	0.208		<i>0.018</i>	<i>0.023</i>	0.557	0.481	0.188	0.201
30	0.188	0.543	<i>0.013</i>	<i>0.018</i>		0.903	0.467	0.267	0.620	0.493
60	0.251	0.507	<i>0.031</i>	<i>0.023</i>	0.903		0.567	0.419	0.795	0.585
180	0.984	0.616	0.345	0.557	0.467	0.567		0.966	0.517	0.591
360	0.757	0.442	0.522	0.481	0.267	0.419	0.966		0.479	0.549
720	0.713	1.000	0.188	0.188	0.620	0.795	0.517	0.479		0.585
1440	0.915	0.794	0.256	0.201	0.493	0.585	0.591	0.549	0.585	

B

AA (6,4-PP)	0	1	5	15	30	60	180	360	720	1440
0		<i>0.011</i>	<i>0.008</i>	<i>0.020</i>	<i>0.001</i>	0.106	0.931	0.716	0.425	0.853
1	<i>0.011</i>		0.854	0.792	0.223	0.826	0.240	0.405	0.753	0.312
5	<i>0.008</i>	0.854		<i>0.032</i>	<i>0.002</i>	0.085	0.358	0.275	0.177	0.467
15	<i>0.020</i>	0.792	<i>0.032</i>		<i>0.002</i>	0.188	0.896	0.685	0.367	0.901
30	<i>0.001</i>	0.223	<i>0.002</i>	<i>0.002</i>		0.660	0.117	0.117	0.365	0.102
60	0.106	0.826	0.085	0.188	0.660		0.605	0.970	0.839	0.649
180	0.931	0.240	0.358	0.896	0.117	0.605		0.132	<i>0.037</i>	0.184
360	0.716	0.405	0.275	0.685	0.117	0.970	0.132		0.111	0.386
720	0.425	0.753	0.177	0.367	0.365	0.839	<i>0.037</i>	0.110		0.820
1440	0.853	0.312	0.467	0.901	0.102	0.649	0.184	0.386	0.820	

C

DD(CPD)	0	1	5	15	30	60	180	360	720	1440
0		<i>0.010</i>	<i>0.025</i>	<i>0.017</i>	0.103	0.173	0.410	<i>0.004</i>	<i>0.046</i>	<i>0.032</i>
1	<i>0.010</i>		0.244	0.222	0.853	0.635	0.346	0.087	0.324	0.191
5	<i>0.025</i>	0.244		0.934	0.311	0.476	0.894	0.543	0.888	0.935
15	<i>0.017</i>	0.222	0.934		0.654	0.837	0.790	0.344	0.852	0.671
30	0.103	0.853	0.311	0.654		0.215	0.074	<i>0.007</i>	0.057	<i>0.041</i>

60	0.173	0.635	0.476	0.837	0.215		0.005	0.000	0.007	0.004
180	0.410	0.346	0.894	0.790	0.074	0.005		0.417	0.888	0.756
360	0.004	0.087	0.543	0.344	0.007	0.000	0.417		0.767	0.870
720	0.046	0.324	0.888	0.852	0.057	0.007	0.888	0.767		0.266
1440	0.032	0.191	0.935	0.671	0.041	0.004	0.756	0.870	0.266	

D

DD (6,4-PP)	0	1	5	15	30	60	180	360	720	1440
0		0.003	0.000	0.001	0.021	0.001	0.000	0.000	0.000	0.000
1	0.003		0.011	0.031	0.204	0.026	0.001	0.005	0.012	0.001
5	0.000	0.011		0.187	0.962	0.096	0.001	0.004	0.017	0.001
15	0.001	0.031	0.187		0.974	0.129	0.005	0.016	0.050	0.004
30	0.021	0.204	0.962	0.974		0.005	0.000	0.000	0.001	0.000
60	0.001	0.026	0.096	0.129	0.005		0.164	0.521	0.777	0.266
180	0.000	0.001	0.001	0.005	0.000	0.164		0.029	0.005	0.032
360	0.000	0.005	0.004	0.016	0.000	0.521	0.029		0.010	0.075
720	0.000	0.012	0.017	0.050	0.001	0.777	0.005	0.010		0.536
1440	0.000	0.001	0.001	0.004	0.000	0.266	0.032	0.075	0.536	

E

EE(CPD)	0	1	5	15	30	60	180	360	720	1440
0		0.214	0.038	0.110	0.016	0.040	0.189	0.407	0.139	0.180
1	0.214		0.397	0.561	0.024	0.076	0.606	0.813	0.326	0.433
5	0.038	0.397		0.231	0.012	0.040	0.740	0.735	0.213	0.360
15	0.110	0.561	0.231		0.044	0.189	0.803	0.973	0.533	0.935
30	0.160	0.024	0.012	0.044		0.241	0.765	0.826	0.708	0.987
60	0.040	0.076	0.040	0.189	0.241		0.947	0.987	0.436	0.612
180	0.189	0.606	0.740	0.803	0.765	9.470		0.946	0.391	0.423
360	0.407	0.813	0.735	0.973	0.826	0.987	0.946		0.585	0.857
720	0.139	0.326	0.213	0.533	0.708	0.391	0.391	0.585		0.612
1440	0.180	0.433	0.360	0.935	0.987	0.423	0.423	0.857	0.612	

F

EE (6,4-PP)	0	1	5	15	30	60	180	360	720	1440
0		0.813	0.238	0.552	0.499	0.311	0.024	0.046	0.148	0.021
1	0.813		0.738	0.761	0.108	0.696	0.158	0.157	0.319	0.075
5	0.238	0.738		0.894	0.217	0.309	0.033	0.014	0.047	0.004
15	0.552	0.761	0.894		0.098	0.912	0.203	0.447	0.715	0.266
30	0.499	0.108	0.217	0.098		0.538	0.041	0.065	0.171	0.024
60	0.311	0.696	0.309	0.912	0.538		0.582	0.223	0.107	0.472
180	0.024	0.158	0.033	0.203	0.041	0.582		0.335	0.189	0.567

360	<i>0.046</i>	0.157	<i>0.015</i>	0.447	0.065	0.223	0.335		0.130	0.452
720	0.148	0.319	<i>0.048</i>	0.715	0.171	0.107	0.189	0.130		0.327
1440	0.021	0.075	<i>0.004</i>	0.266	<i>0.024</i>	0.472	0.567	0.452	0.327	

G

FF(CPD)	0	1	5	15	30	60	180	360	720	1440
0		0.129	0.237	0.202	0.471	0.493	0.252	0.084	0.291	0.267
1	0.129		0.530	0.680	0.543	0.495	0.953	0.992	0.659	0.692
5	0.237	0.530		<i>0.024</i>	<i>0.060</i>	<i>0.082</i>	<i>0.029</i>	<i>0.007</i>	<i>0.037</i>	<i>0.038</i>
15	0.202	0.680	<i>0.024</i>		0.114	0.121	<i>0.044</i>	<i>0.015</i>	0.064	0.056
30	0.471	0.543	<i>0.060</i>	0.144		<i>0.008</i>	<i>0.002</i>	<i>0.000</i>	<i>0.003</i>	<i>0.002</i>
60	0.493	0.495	<i>0.082</i>	0.121	<i>0.008</i>		<i>0.005</i>	<i>0.001</i>	<i>0.007</i>	<i>0.007</i>
180	0.252	0.953	<i>0.029</i>	<i>0.044</i>	<i>0.002</i>	<i>0.005</i>		<i>0.050</i>	0.156	0.120
360	0.084	0.992	<i>0.007</i>	<i>0.015</i>	<i>0.000</i>	<i>0.001</i>	<i>0.050</i>		0.522	0.434
720	0.291	0.659	<i>0.037</i>	0.064	<i>0.003</i>	<i>0.007</i>	0.156	0.520		0.017
1440	0.267	0.692	<i>0.038</i>	0.056	<i>0.002</i>	<i>0.007</i>	0.120	0.434	0.017	

H

FF (6,4-PP)	0	1	5	15	30	60	180	360	720	1440
0		0.359	0.931	0.722	0.132	0.927	0.356	0.587	1.000	0.444
1	0.359		0.870	0.573	0.064	0.837	0.099	0.156	0.378	<i>0.039</i>
5	0.931	0.870		0.909	0.133	0.483	0.080	<i>0.032</i>	0.103	<i>0.009</i>
15	0.722	0.573	0.909		0.440	0.145	<i>0.008</i>	<i>0.004</i>	<i>0.015</i>	<i>0.001</i>
30	0.132	0.064	0.133	0.440		0.164	<i>0.009</i>	<i>0.013</i>	0.067	<i>0.006</i>
60	0.927	0.837	0.483	0.145	0.164		<i>0.026</i>	<i>0.033</i>	0.126	<i>0.011</i>
180	0.356	0.099	0.080	<i>0.008</i>	<i>0.009</i>	<i>0.026</i>		0.187	0.444	0.132
360	0.587	0.156	<i>0.032</i>	<i>0.004</i>	<i>0.013</i>	<i>0.033</i>	0.187		0.471	<i>0.042</i>
720	1.000	0.378	0.103	<i>0.015</i>	0.067	0.126	0.444	0.471		<i>0.027</i>
1440	0.444	<i>0.039</i>	<i>0.009</i>	<i>0.001</i>	<i>0.006</i>	<i>0.011</i>	0.132	<i>0.042</i>	<i>0.027</i>	

Table 25: Mann-Whitney test of significance data for each of the haplotypes over time by adduct. Time is presented in minutes after UV-B exposure. Statistical significance is presented as a p-value ≤ 0.05 , listed in italics. **A:** Mann-Whitney test by time point for CPD ELISA analysis of haplotype AA. **B:** Mann-Whitney test by time point for 6,4-PP ELISA analysis of haplotype AA. **C:** Mann-Whitney test by time point for CPD ELISA analysis of haplotype DD. **D:** Mann-Whitney test by time point for 6,4-PP ELISA analysis of haplotype DD. **E:** Mann-Whitney test by time point for CPD ELISA analysis of haplotype EE. **F:** Mann-Whitney test by time point for 6,4-PP ELISA analysis of haplotype EE. **G:** Mann-Whitney test by time point for CPD ELISA analysis of haplotype FF. **F:** Mann-Whitney test by time point for 6,4-PP ELISA analysis of haplotype FF.

Table 26: Mann-Whitney test of significance data for the haplotypes at each time by adduct

A

CPD				
0 min	AA	DD	EE	FF
AA		0.172	0.260	0.092
DD	0.172		0.000	0.769
EE	0.260	0.000		0.002
FF	0.092	0.769	0.002	
1 min	AA	DD	EE	FF
AA		0.569	0.183	0.158
DD	0.569		0.005	0.282
EE	0.183	0.005		0.022
FF	0.158	0.282	0.022	
5 min	AA	DD	EE	FF
AA		0.020	0.138	0.015
DD	0.020		0.418	0.042
EE	0.138	0.418		0.007
FF	0.015	0.042	0.007	
15 min	AA	DD	EE	FF
AA		0.036	0.829	0.031
DD	0.036		0.259	0.082
EE	0.829	0.259		0.065
FF	0.031	0.082	0.065	
30 min	AA	DD	EE	FF
AA		0.596	0.010	0.261
DD	0.596		0.000	0.590
EE	0.010	0.000		0.111
FF	0.261	0.590	0.111	
60 min	AA	DD	EE	FF
AA		0.297	0.027	0.172
DD	0.297		0.000	0.612
EE	0.027	0.000		0.023
FF	0.172	0.612	0.023	
180 min	AA	DD	EE	FF
AA		0.105	0.319	0.780
DD	0.105		0.196	0.079
EE	0.319	0.196		0.013
FF	0.780	0.079	0.013	
360 min	AA	DD	EE	FF

AA		0.080	0.914	<i>0.048</i>
DD	0.080		0.452	<i>0.041</i>
EE	0.914	0.452		0.074
FF	<i>0.048</i>	<i>0.041</i>	0.074	
720 min	AA	DD	EE	FF
AA		0.100	0.962	<i>0.026</i>
DD	0.100		<i>0.010</i>	0.265
EE	0.962	<i>0.010</i>		0.051
FF	<i>0.026</i>	0.265	0.051	
1440 min	AA	DD	EE	FF
AA		0.624	0.175	0.208
DD	0.624		<i>0.000</i>	<i>0.040</i>
EE	0.175	<i>0.000</i>		<i>0.020</i>
FF	0.208	<i>0.040</i>	<i>0.020</i>	

B

6,4-PP				
0 min	AA	DD	EE	FF
AA		0.146	0.315	<i>0.006</i>
DD	0.146		<i>0.000</i>	<i>0.000</i>
EE	0.315	<i>0.000</i>		0.276
FF	<i>0.006</i>	<i>0.000</i>	0.276	
1 min	AA	DD	EE	FF
AA		0.881	0.372	1.000
DD	0.881		<i>0.001</i>	<i>0.006</i>
EE	0.372	<i>0.001</i>		0.887
FF	1.000	<i>0.006</i>	0.887	
5 min	AA	DD	EE	FF
AA		<i>0.043</i>	0.147	<i>0.032</i>
DD	<i>0.043</i>		<i>0.003</i>	0.055
EE	0.147	<i>0.003</i>		0.363
FF	<i>0.032</i>	0.055	0.363	
15 min	AA	DD	EE	FF
AA		0.117	0.362	<i>0.017</i>
DD	0.117		<i>0.009</i>	0.079
EE	0.362	<i>0.009</i>		0.862
FF	<i>0.017</i>	0.079	0.862	
30 min	AA	DD	EE	FF
AA		0.921	0.214	0.777
DD	0.921		<i>0.000</i>	<i>0.000</i>
EE	0.214	<i>0.000</i>		0.758
FF	0.777	<i>0.000</i>	0.758	
60 min	AA	DD	EE	FF

AA		0.337	0.934	0.275
DD	0.337		0.410	0.797
EE	0.934	0.410		<i>0.004</i>
FF	0.275	0.797	<i>0.004</i>	
180 min	AA	DD	EE	FF
AA		<i>0.006</i>	<i>0.021</i>	<i>0.000</i>
DD	<i>0.006</i>		<i>0.010</i>	<i>0.000</i>
EE	<i>0.021</i>	<i>0.010</i>		<i>0.031</i>
FF	<i>0.000</i>	<i>0.000</i>	<i>0.031</i>	
360 min	AA	DD	EE	FF
AA		<i>0.044</i>	0.173	<i>0.005</i>
DD	<i>0.044</i>		<i>0.020</i>	<i>0.000</i>
EE	0.173	<i>0.020</i>		<i>0.007</i>
FF	<i>0.005</i>	<i>0.000</i>	<i>0.007</i>	
720 min	AA	DD	EE	FF
AA		0.195	0.523	<i>0.035</i>
DD	0.195		0.223	<i>0.007</i>
EE	0.523	0.223		<i>0.002</i>
FF	<i>0.035</i>	<i>0.007</i>	<i>0.002</i>	
1440 min	AA	DD	EE	FF
AA		0.085	0.260	<i>0.006</i>
DD	0.085		<i>0.000</i>	<i>0.000</i>
EE	0.260	<i>0.000</i>		<i>0.001</i>
FF	<i>0.006</i>	<i>0.000</i>	<i>0.001</i>	

Table 26: Mann-Whitney test of significance data for the haplotypes at each time by adduct. Time is presented in minutes after UV-B exposure. Statistical significance is presented as a p-value ≤ 0.05 , listed in italics. **A:** Mann-Whitney test by haplotype for CPD ELISA analysis by each time point. **B:** Mann-Whitney test by haplotype for 6,4-PP ELISA analysis by each time point.

Table 27: Mann-Whitney test of significance data of grouped sensitive verses insensitive haplotype by adduct

Grouping by Regression Analysis Data			
	sensitive	insensitive	0min
CPD	AA+EE	DD+FF	<i>0.002</i>
6,4-PP	DD+EE	AA+FF	<i>0.001</i>
	sensitive	insensitive	1min
CPD	AA+EE	DD+FF	0.285
6,4-PP	DD+EE	AA+FF	0.151
	sensitive	insensitive	5min
CPD	AA+EE	DD+FF	<i>0.017</i>

6,4-PP	DD+EE	AA+FF	<i>0.021</i>
	sensitive	insensitive	15min
CPD	AA+EE	DD+FF	0.073
6,4-PP	DD+EE	AA+FF	0.248
	sensitive	insensitive	30min
CPD	AA+EE	DD+FF	<i>0.000</i>
6,4-PP	DD+EE	AA+FF	0.231
	sensitive	insensitive	60min
CPD	AA+EE	DD+FF	<i>0.000</i>
6,4-PP	DD+EE	AA+FF	0.281
	sensitive	insensitive	180min
CPD	AA+EE	DD+FF	0.085
6,4-PP	DD+EE	AA+FF	0.205
	sensitive	insensitive	360min
CPD	AA+EE	DD+FF	0.351
6,4-PP	DD+EE	AA+FF	<i>0.038</i>
	sensitive	insensitive	720min
CPD	AA+EE	DD+FF	<i>0.004</i>
6,4-PP	DD+EE	AA+FF	<i>0.019</i>
	sensitive	insensitive	1440min
CPD	AA+EE	DD+FF	0.109
6,4-PP	DD+EE	AA+FF	<i>0.002</i>

Table 27: Mann-Whitney test of significance data of grouped sensitive verses insensitive haplotype by adduct. Time is presented in minutes after UV-B exposure. Statistical significance is presented as a p-value ≤ 0.05 , listed in italics.

Table 28: Raw data for picogreen analysis of the haplotypes by time after UV-B exposure

Time	AA	EE	FF
0	n.d.	2.869	0.891
1	0.170	0.636	0.201
5	n.d.	0.789	0.753
15	0.698	0.769	0.712
30	0.977	3.031	1.475
60	1.421	1.363	1.583
180	0.329	-0.309	1.238
360	0.992	1.848	1.256
720	0.406	1.109	0.604
1440	0.572	0.516	0.776
2160	-5.881	6.563	0.300
2880	0.480	2.884	0.715

Table 28: Raw data for picogreen analysis of the haplotypes by time after UV-B exposure. Time is presented in minutes after UV-B exposure. Arbitrary fluorescence data has been normalized to amount of protein in μg using Bradford analysis and matched untreated (no UV-B exposure) controls. Fluorescence data not detected is designated as n.d.

Table 29: Shapiro-Wilk normality data for each of the haplotypes over time for fold induction of *XPC* mRNA by real time analysis

PEmin	AA	p	DD	p	EE	p	FF	p
0	0.974	0.921	0.984	0.923	0.733	0.014	0.847	0.186
1	0.925	0.546	1.000	0.995	0.829	0.105	0.797	0.056
5	0.866	0.209	0.694	0.010	0.860	0.189	0.940	0.666
15	0.903	0.390	0.869	0.261	0.886	0.298	0.915	0.472
30	0.670	0.005	0.861	0.230	0.880	0.271	0.884	0.290
60	0.867	0.214	0.854	0.171	0.825	0.098	0.866	0.250
180	0.871	0.272	0.817	0.111	0.804	0.063	0.864	0.205
360	0.751	0.021	0.818	0.113	0.826	0.099	0.918	0.517
720	0.882	0.317	0.914	0.506	0.837	0.124	0.781	0.039
1440	0.792	0.049	NA	NA	0.892	0.330	0.789	0.066
2160	0.950	0.740	0.787	0.084	0.796	0.075	0.762	0.038
2880	0.843	0.205	NA	NA	0.823	0.093	0.805	0.089

Table 29: Shapiro-Wilk normality data for each of the haplotypes over time for fold induction of *XPC* mRNA by real time analysis. Time is presented in minutes after UV-B exposure (minPE). Failure of the test is indicated by statistical significance, or p-value ≤ 0.05 . Undetermined values are presented as NA (not available).

Table 30: Mann-Whitney test of significance data for each of the haplotypes over time for fold induction of *XPC* mRNA by real time analysis

A

AA	0	1	5	15	30	60	180	360	720	1440	2160	2880
0		0.230	0.184	0.125	0.311	0.084	0.427	0.064	0.440	0.058	0.917	0.301
1	0.230		0.398	0.190	0.148	0.327	0.123	0.470	0.192	0.568	0.062	0.273
5	0.184	0.398		0.042	0.023	0.050	0.015	0.078	0.020	0.097	0.004	0.128
15	0.125	0.190	0.042		0.068	1.000	0.484	0.892	0.629	0.836	0.254	0.831
30	0.311	0.148	0.023	0.068		0.622	0.979	0.482	0.851	0.419	0.603	0.786
60	0.084	0.327	0.050	1.000	0.622		0.513	0.161	0.515	0.213	0.876	0.503
180	0.427	0.123	0.015	0.484	0.979	0.513		0.079	0.600	0.119	0.003	0.083
360	0.064	0.470	0.078	0.892	0.482	0.161	0.079		0.735	0.468	0.378	1.000
720	0.440	0.192	0.020	0.629	0.851	0.515	0.600	0.735		0.094	0.954	0.277
1440	0.058	0.568	0.097	0.836	0.419	0.213	0.119	0.468	0.094		0.484	0.715
2160	0.917	0.062	0.004	0.254	0.603	0.876	0.003	0.378	0.954	0.484		0.114
2880	0.301	0.273	0.128	0.831	0.786	0.503	0.083	1.000	0.277	0.715	0.114	

B

DD	0	1	5	15	30	60	180	360	720	1440	2160	2880
0		0.975	0.669	0.608	0.542	0.860	0.517	0.955	0.583	0.844	0.237	0.699
1	0.975		0.667	0.738	0.567	0.917	0.678	0.894	0.642	0.938	0.237	0.784
5	0.669	0.667		0.171	0.127	0.250	0.068	0.254	0.106	0.264	0.023	0.278
15	0.608	0.738	0.171		0.788	0.312	0.640	0.269	0.707	0.326	0.644	0.634
30	0.542	0.567	0.127	0.788		0.517	0.243	0.482	0.197	0.644	0.073	0.415
60	0.860	0.917	0.250	0.312	0.517		0.053	0.416	0.111	0.351	0.020	0.330
180	0.517	0.678	0.068	0.640	0.243	0.053		0.021	0.188	0.021	0.525	0.221
360	0.955	0.894	0.254	0.269	0.482	0.416	0.021		0.041	0.273	0.024	0.308
720	0.583	0.642	0.106	0.707	0.197	0.111	0.188	0.041		0.793	0.358	0.877
1440	0.844	0.938	0.264	0.326	0.644	0.351	0.021	0.273	0.793		0.212	0.509
2160	0.237	0.237	0.023	0.644	0.073	0.020	0.525	0.024	0.358	0.212		0.235
2880	0.699	0.784	0.278	0.634	0.415	0.330	0.221	0.308	0.877	0.509	0.235	

C

EE	0	1	5	15	30	60	180	360	720	1440	2160	2880
0		0.014	0.018	0.024	0.032	0.013	0.036	0.019	0.053	0.026	0.147	0.128
1	0.014		0.176	0.136	0.192	0.069	0.243	0.114	0.412	0.108	0.795	0.273
5	0.018	0.176		0.027	0.043	0.018	0.040	0.015	0.060	0.023	0.233	0.114
15	0.024	0.136	0.027		0.311	0.129	0.455	0.125	0.372	0.195	0.959	0.429
30	0.032	0.192	0.043	0.311		0.744	0.779	0.512	0.923	0.468	0.378	0.903
60	0.013	0.069	0.018	0.129	0.744		0.401	0.190	0.530	0.133	0.795	0.465
180	0.036	0.243	0.040	0.455	0.779	0.401		0.136	0.412	0.178	0.756	0.465
360	0.019	0.114	0.015	0.125	0.512	0.190	0.136		0.664	0.378	0.133	0.938
720	0.053	0.412	0.060	0.372	0.923	0.530	0.412	0.664		0.133	0.938	0.394
1440	0.026	0.108	0.023	0.195	0.468	0.133	0.178	0.378	0.133		0.856	0.563
2160	0.147	0.795	0.233	0.959	0.378	0.795	0.756	0.133	0.938	0.856		0.174
2880	0.128	0.273	0.114	0.429	0.903	0.465	0.465	0.938	0.394	0.563	0.174	

D

FF	0	1	5	15	30	60	180	360	720	1440	2160	2880
0		0.522	0.914	0.841	0.914	0.517	0.815	0.547	0.747	0.525	0.326	0.812
1	0.522		0.595	0.557	0.288	0.744	0.316	0.892	0.334	0.836	0.066	0.503
5	0.914	0.595		0.564	0.485	0.795	0.364	0.940	0.420	0.862	0.166	0.684
15	0.841	0.557	0.564		0.134	0.363	0.500	0.343	0.082	0.351	0.011	0.273
30	0.914	0.288	0.485	0.134		0.641	0.762	0.651	0.595	0.717	0.254	0.951
60	0.517	0.744	0.795	0.363	0.641		0.010	0.042	0.018	0.119	0.003	0.077
180	0.815	0.316	0.364	0.500	0.762	0.010		0.557	0.866	0.622	0.162	1.000
360	0.547	0.892	0.940	0.343	0.651	0.042	0.557		0.283	0.729	0.040	0.455
720	0.747	0.334	0.420	0.082	0.595	0.018	0.866	0.283		0.351	0.049	0.260
1440	0.525	0.836	0.862	0.351	0.717	0.119	0.622	0.729	0.351		0.028	0.197

2160	0.326	0.066	0.166	0.011	0.254	<i>0.003</i>	0.162	<i>0.040</i>	<i>0.049</i>	<i>0.028</i>	0.946
2880	0.812	0.503	0.684	0.273	0.951	0.077	1.000	0.455	0.260	0.197	0.946

Table 30: Mann-Whitney test of significance data for each of the haplotypes over time for fold induction of XPC mRNA by real time analysis. Time is presented in minutes after UV-B exposure. Statistical significance is presented as a p-value ≤ 0.05 , listed in italics. A: Mann-Whitney test by time point for real time analysis of haplotype AA. B: Mann-Whitney test by time point for real time analysis of haplotype DD. C: Mann-Whitney test by time point for real time analysis of haplotype EE. D: Mann-Whitney test by time point for real time analysis of haplotype FF.

Table 31: Mann-Whitney test of significance data for the haplotypes at each time for fold induction of XPC mRNA by real time analysis

		fold change			
0 min	AA	DD	EE	FF	
AA		0.201	0.337	0.201	
DD	0.201		0.055	0.327	
EE	0.337	0.055		0.144	
FF	0.201	0.327	0.144		
1 min	AA	DD	EE	FF	
AA		0.302	0.078	0.631	
DD	0.302		0.197	0.796	
EE	0.078	0.197		0.055	
FF	0.631	0.796	0.055		
5 min	AA	DD	EE	FF	
AA		0.831	<i>0.006</i>	0.273	
DD	0.831		<i>0.019</i>	0.462	
EE	<i>0.006</i>	<i>0.019</i>		<i>0.045</i>	
FF	0.273	0.462	<i>0.045</i>		
15 min	AA	DD	EE	FF	
AA		0.465	0.262	0.749	
DD	0.465		0.855	0.068	
EE	0.262	0.855		<i>0.037</i>	
FF	0.749	0.068	<i>0.037</i>		
30 min	AA	DD	EE	FF	
AA		0.465	0.855	0.715	
DD	0.465		0.465	0.361	
EE	0.855	0.465		1.000	
FF	0.715	0.361	1.000		
60 min	AA	DD	EE	FF	
AA		0.055	0.873	<i>0.018</i>	
DD	0.055		0.078	0.273	

EE	0.876	0.078		<i>0.018</i>
FF	<i>0.018</i>	0.273	<i>0.018</i>	
180 min	AA	DD	EE	FF
AA		<i>0.009</i>	<i>0.018</i>	<i>0.018</i>
DD	<i>0.009</i>		0.855	<i>0.045</i>
EE	<i>0.018</i>	0.855		0.337
FF	<i>0.018</i>	<i>0.045</i>	0.337	
360 min	AA	DD	EE	FF
AA		0.144	0.631	0.465
DD	0.144		0.144	0.465
EE	0.631	0.144		0.361
FF	0.465	0.465	0.361	
720 min	AA	DD	EE	FF
AA		0.327	1.000	0.201
DD	0.327		0.286	0.286
EE	1.000	0.286		0.078
FF	0.201	0.286	0.078	
1440 min	AA	DD	EE	FF
AA		0.317	0.873	0.144
DD	0.317		0.134	0.770
EE	0.873	0.134		0.201
FF	0.144	0.770	0.201	
2160 min	AA	DD	EE	FF
AA		0.796	1.000	0.055
DD	0.796		0.655	0.053
EE	1.000	0.655		0.175
FF	0.055	0.053	0.175	
2880 min	AA	DD	EE	FF
AA		NA	1.000	0.142
DD	NA		NA	NA
EE	1.000	NA		0.100
FF	0.142	NA	0.100	

Table 31: Mann-Whitney test of significance data for the haplotypes at each time for fold induction of XPC mRNA by real time analysis. Time is presented in minutes after UV-B exposure. Statistical significance is presented as a p-value ≤ 0.05 , listed in italics.

Table 32: T-test analysis of significance data for each of the haplotypes over time for XPC protein by Western blot analysis

A

AA	NT	0	1	6	12	24	36	48
NT		0.639	0.390	0.308	0.145	0.304	0.073	0.324

0	0.639		0.738	0.602	0.330	0.654	0.167	0.702
1	0.390	0.738		0.800	0.369	0.919	0.130	0.991
6	0.308	0.602	0.800		0.551	0.824	0.210	0.727
12	0.145	0.330	0.369	0.551		0.213	0.205	0.121
24	0.304	0.654	0.919	0.824	0.213		<i>0.026</i>	0.826
36	0.073	0.167	0.130	0.210	0.205	<i>0.026</i>		<i>0.008</i>
48	0.324	0.702	0.991	0.727	0.121	0.826	<i>0.008</i>	

B

DD	NT	0	1	6	12	24	36	48
NT		0.579	0.492	0.304	0.229	0.238	0.426	0.365
0	0.579		0.597	<i>0.038</i>	<i>0.035</i>	<i>0.038</i>	0.389	0.203
1	0.492	0.597		0.260	0.135	0.147	0.746	0.496
6	0.304	<i>0.038</i>	0.260		0.165	0.202	0.417	0.510
12	0.229	<i>0.035</i>	0.135	0.165		0.853	0.193	0.183
24	0.238	<i>0.038</i>	0.147	0.202	0.853		0.213	0.209
36	0.426	0.389	0.746	0.417	0.193	0.213		0.737
48	0.365	0.203	0.496	0.510	0.183	0.209	0.737	

C

EE	NT	0	1	6	12	24	36	48
NT		0.830	0.722	0.770	0.289	0.967	0.835	0.955
0	0.830		0.894	0.948	0.993	0.860	0.996	0.822
1	0.722	0.894		0.941	0.892	0.750	0.905	0.738
6	0.770	0.948	0.941		0.951	0.800	0.955	0.775
12	0.289	0.993	0.892	0.951		0.844	0.997	0.807
24	0.967	0.860	0.750	0.800	0.844		0.864	0.928
36	0.835	0.996	0.905	0.955	0.997	0.864		0.823
48	0.955	0.822	0.738	0.775	0.807	0.928	0.823	

D

FF	NT	0	1	6	12	24	36	48
NT		0.908	0.654	0.699	0.778	0.654	0.822	0.757
0	0.908		0.709	0.758	0.698	0.600	0.739	0.662
1	0.654	0.709		0.973	0.523	0.484	0.548	0.463
6	0.699	0.758	0.973		0.555	0.504	0.583	0.505
12	0.778	0.698	0.523	0.555		0.824	0.951	0.990
24	0.654	0.600	0.484	0.504	0.824		0.785	0.807
36	0.822	0.739	0.548	0.583	0.951	0.785		0.956
48	0.757	0.662	0.463	0.505	0.990	0.807	0.956	

Table 32: T-test analysis of significance data for each of the haplotypes over time for XPC protein by Western blot analysis. Time is presented in hours after UV-B exposure. Statistical significance is presented as a p-value ≤ 0.05 , listed in italics. **A:** T-

test analysis by time point for western analysis of haplotype AA. **B**: T-test analysis by time point for western analysis of haplotype DD. **C**: T-test analysis by time point for western analysis of haplotype EE. **D**: T-test analysis by time point for western analysis of haplotype FF.

Table 33: T-test analysis of significance data for the haplotypes at each time for XPC protein by Western blot analysis

Densitometry				
NT	AA	DD	EE	FF
AA		0.411	0.056	0.077
DD	0.411		0.060	0.069
EE	0.056	0.060		0.572
FF	0.077	0.069	0.572	
0 hr	AA	DD	EE	FF
AA		0.406	0.085	0.136
DD	0.406		0.017	0.011
EE	0.085	0.017		0.439
FF	0.136	0.011	0.439	
1 hr	AA	DD	EE	FF
AA		0.280	0.047	0.072
DD	0.280		0.021	0.020
EE	0.047	0.021		0.457
FF	0.072	0.020	0.457	
6 hr	AA	DD	EE	FF
AA		0.582	0.072	0.129
DD	0.582		0.031	0.042
EE	0.072	0.031		0.520
FF	0.129	0.042	0.520	
12 hr	AA	DD	EE	FF
AA		0.572	0.069	0.577
DD	0.572		0.085	0.487
EE	0.069	0.085		0.333
FF	0.577	0.487	0.333	
24 hr	AA	DD	EE	FF
AA		0.606	0.048	0.516
DD	0.606		0.136	0.725
EE	0.048	0.136		0.420
FF	0.516	0.725	0.420	0.420
36 hr	AA	DD	EE	FF
AA		0.024	0.239	0.978
DD	0.024		0.050	0.119

EE	0.239	<i>0.050</i>		0.390
FF	0.978	0.119	0.390	
48 hr	AA	DD	EE	FF
	AA	0.227	0.114	0.123
	DD	0.227	0.135	0.114
	EE	0.114	0.135	0.561
	FF	0.123	0.114	0.561

Table 33: T-test analysis of significance data for the haplotypes at each time for XPC protein by Western blot analysis. Time is presented in hours after UV-B exposure. Statistical significance is presented as a p-value ≤ 0.05 , listed in italics.

REFERENCES

- Abdel-Rahman, S.Z., El-Zein, R.A., 2000. The 399Gln polymorphism in the DNA repair gene *XRCC1* modulates the genotoxic response induced in human lymphocytes by the tobacco-specific nitrosamine NNK. *Cancer Letters* 159, 63–71.
- Abdel-Rahman, S.Z., El-Zein, R.A., 2011. Evaluating the effects of genetic variants of DNA repair genes using cytogenetic mutagen sensitivity approaches. *Biomarkers* 16, 393–404.
- Adimoolam, S., Ford, J.M., 2002. p53 and DNA damage-inducible expression of the xeroderma pigmentosum group C gene. *Proceedings of the National Academy of Sciences of the United States of America* 99, 12985–90.
- Adlkofer, F., 2001. Lung cancer due to passive smoking--a review. *International Archives of Occupational and Environmental Health* 74, 231–41.
- Affatato, A.A., Wolfe, K.J., Lopez, M.S., Hallberg, C., Ammenheuser, M.M., Abdel-Rahman, S.Z., 2004. Effect of *XPD/ERCC2* polymorphisms on chromosome aberration frequencies in smokers and on sensitivity to the mutagenic tobacco-specific nitrosamine NNK. *Environmental and Molecular Mutagenesis* 44, 65–73.
- Aka, P., Mateuca, R., Buchet, J.-P., Thierens, H., Kirsch-Volders, M., 2004. Are genetic polymorphisms in *OGGI*, *XRCC1* and *XRCC3* genes predictive for the DNA strand break repair phenotype and genotoxicity in workers exposed to low dose ionising radiations? *Mutation Research* 556, 169–81.
- Alavanja, M.C.R., Bonner, M.R., Alvanja, M.C., 2012. Occupational pesticide exposures and cancer risk: a review. *Journal of Toxicology and Environmental Health Part B, Critical Reviews* 15, 238–263.
- Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* 25, 3389–402.
- Alvarez-larra, A., Guillem, V.M., Cervantes, F., Sureda, A., Maffioli, M., Camo, M., Maruga, I., 2010. *XPC* genetic polymorphisms correlate with the response to imatinib treatment in patients with chronic phase chronic myeloid leukemia. *American Journal of Hematology* 85, 482–6.
- An, J., Liu, Z., Hu, Z., Li, G., Wang, L.-E., Sturgis, E.M., El-Naggar, A.K., Spitz, M.R., Wei, Q., 2007. Potentially functional single nucleotide polymorphisms in the core nucleotide excision repair genes and risk of squamous cell carcinoma of the head and neck. *Cancer Epidemiology, Biomarkers & Prevention* 16, 1633–8.

- Andrew, A.S., Nelson, H.H., Kelsey, K.T., Moore, J.H., Meng, A.C., Casella, D.P., Tosteson, T.D., Schned, A.R., Karagas, M.R., 2006. Concordance of multiple analytical approaches demonstrates a complex relationship between DNA repair gene SNPs, smoking and bladder cancer susceptibility. *Carcinogenesis* 27, 1030–7.
- Angelini, S., Kumar, R., Carbone, F., Bermejo, J.L., Maffei, F., Cantelli-Forti, G., Hemminki, K., Hrelia, P., 2008. Inherited susceptibility to bleomycin-induced micronuclei: correlating polymorphisms in *GSTT1*, *GSTM1* and DNA repair genes with mutagen sensitivity. *Mutation Research* 638, 90–7.
- Araki, M., Masutani, C., Takemura, M., Uchida, A., Sugasawa, K., Kondoh, J., Ohkuma, Y., Hanaoka, F., 2001. Centrosome protein centrin 2/caltractin 1 is part of the xeroderma pigmentosum group C complex that initiates global genome nucleotide excision repair. *The Journal of Biological Chemistry* 276, 18665–72.
- Au, W.W., Navasumrit, P., Ruchirawat, M., 2004. Use of biomarkers to characterize functions of polymorphic DNA repair genotypes. *International Journal of Hygiene and Environmental Health* 207, 301–13.
- Aylward, L.L., Kirman, C.R., Schoeny, R., Portier, C.J., Hays, S.M., 2013. Evaluation of biomonitoring data from the CDC National Exposure Report in a risk assessment context: perspectives across chemicals. *Environmental Health Perspectives* 121, 287–94.
- Bai, Y., Xu, L., Yang, X., Hu, Z., Yuan, J., Wang, F., Shao, M., Yuan, W., Qian, J., Ma, H., Wang, Y., Liu, H., Chen, W., Yang, L., Jing, G., Huo, X., Chen, F., Liu, Y., Jin, L., Wei, Q., Huang, W., Shen, H., Lu, D., Wu, T., 2007. Sequence variations in DNA repair gene *XPC* is associated with lung cancer risk in a Chinese population: a case-control study. *BMC Cancer* 7, 81.
- Bardel, C., Danjean, V., Morange, P., Génin, E., Darlu, P., 2009. On the use of phylogeny-based tests to detect association between quantitative traits and haplotypes. *Genetic Epidemiology* 33, 729–39.
- Barrett, J.C., Fry, B., Maller, J., Daly, M.J., 2005. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 21, 263–5.
- Berman, D.M., Wang, Y., Liu, Z., Dong, Q., Burke, L.-A., Liotta, L.A., Fisher, R., Wu, X., 2004. A functional polymorphism in *RGS6* modulates the risk of bladder cancer. *Cancer Research* 64, 6820–6.
- Blankenburg, S., König, I.R., Moessner, R., Laspe, P., Thoms, K.-M., Krueger, U., Khan, S.G., Westphal, G., Berking, C., Volkenandt, M., Reich, K., Neumann, C., Ziegler, A., Kraemer, K.H., Emmert, S., 2005. Assessment of 3 *xeroderma pigmentosum group C* gene polymorphisms and risk of cutaneous melanoma: a case-control study. *Carcinogenesis* 26, 1085–90.

- Blitzblau, R.C., Weidhaas, J.B., 2010. MicroRNA binding-site polymorphisms as potential biomarkers of cancer risk. *Molecular Diagnosis & Therapy* 14, 335–42.
- Boffetta, P., Winn, D.M., Ioannidis, J.P., Thomas, D.C., Little, J., Smith, G.D., Cogliano, V.J., Hecht, S.S., Seminara, D., Vineis, P., Khoury, M.J., 2012. Recommendations and proposed guidelines for assessing the cumulative evidence on joint effects of genes and environments on cancer occurrence in humans. *International Journal of Epidemiology* 41, 686–704.
- Bonassi, S., Abbondandolo, A., Camurri, L., Dal Prá, L., De Ferrari, M., Degrassi, F., Forni, A., Lamberti, L., Lando, C., Padovani, P., Sbrana, I., Vecchio, D., Putoni, R., 1995. Are chromosome aberrations in circulating lymphocytes predictive of future cancer onset in humans? Preliminary results of an Italian cohort study. *Cancer Genetics and Cytogenetics* 79, 133–5.
- Bonassi, S., Hagmar, L., Montagud, A., 2000. Chromosomal aberrations in lymphocytes predict human cancer independently of exposure to carcinogens. *Cancer Research* 60, 1619–25.
- Brookes, A.J., 1999. The essence of SNPs. *Gene* 234, 177–86.
- Brown, P.J., Bedard, L.L., Massey, T.E., 2008. Repair of 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone-induced DNA pyridyloxobutylation by nucleotide excision repair. *Cancer Letters* 260, 48–55.
- Brown, P.J., Massey, T.E., 2009. In vivo treatment with 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone (NNK) induces organ-specific alterations in in vitro repair of DNA pyridyloxobutylation. *Mutation Research* 663, 15–21.
- Browning, S.R., Browning, B.L., 2011. Haplotype phasing: existing methods and new developments. *Nature Reviews, Genetics* 12, 703–14.
- Bunick, C.G., Miller, M.R., Fuller, B.E., Fanning, E., Chazin, W.J., 2006. Biochemical and structural domain analysis of xeroderma pigmentosum complementation group C protein. *Biochemistry* 45, 14965–79.
- Cadet, J., Sage, E., Douki, T., 2005. Ultraviolet radiation-mediated damage to cellular DNA. *Mutation Research* 571, 3–17.
- Calzone, K.A., 2012. Genetic biomarkers of cancer risk. *Seminars in Oncology Nursing* 28, 122–8.
- Campayo, M., Viñolas, N., Navarro, A., Carcereny, E., Casas, F., Gel, B., Diaz, T., Gimferrer, J.M., Marrades, R.M., Ramirez, J., Monzo, M., 2011. Single nucleotide polymorphisms in tobacco metabolism and DNA repair genes and prognosis in resected non-small-cell lung cancer. *The Journal of Surgical Research* 167, e5–12.

- Caporaso, N.E., Landi, M.T., 1994. Molecular epidemiology: a new perspective for the study of toxic exposures in man. A consideration of the influence of genetic susceptibility factors on risk in different lung cancer histologies. *La Medicina del Lavoro* 85, 68–77.
- Chang, C.-H., Wang, R.-F., Tsai, R.-Y., Wu, H.-C., Wang, C.-H., Tsai, C.-W., Chang, C.-L., Tsou, Y.-A., Liu, C.-S., Bau, D.-T., 2009. Significant association of *XPD* codon 312 single nucleotide polymorphism with bladder cancer susceptibility in Taiwan. *Anticancer Research* 29, 3903–7.
- Chen, H.-Y., Yu, S.-L., Li, K.-C., Yang, P.-C., 2012. Biomarkers and transcriptome profiling of lung cancer. *Respirology* 17, 620–6.
- Cheng, A., Mao, Y., Cui, R., 2006. [The effect of gene polymorphism in promoter and intron 1 on human ApoA I expression]. *Chinese Journal of Medical Genetics* 23, 610–3.
- Cheng, L., Eicher, S.A., Guo, Z., Hong, W.K., Spitz, M.R., Wei, Q., 1998. Reduced DNA repair capacity in head and neck cancer patients. *Cancer Epidemiology Biomarkers & Prevention* 7, 465–8.
- Cheverud, J.M., 2001. A simple correction for multiple comparisons in interval mapping genome scans. *Heredity* 87, 52–8.
- Chien, Y.-C., Yeh, C.-T., 2010. Excretion characteristics of urinary 8-hydroxydeoxyguanosine after dietary exposure to polycyclic aromatic hydrocarbons. *Environmental and Molecular Mutagenesis* 51, 243–50.
- Chung, C.C., Chanock, S.J., 2011. Current status of genome-wide association studies in cancer. *Human Genetics* 130, 59–78.
- Clement, F.C., Kaczmarek, N., Mathieu, N., Tomas, M., Leitenstorfer, A., Ferrando-may, E., Naegeli, H., 2011. Dissection of the Xeroderma Pigmentosum Group C protein function by Site-directed mutagenesis. *Antioxidants & Redox Signaling* 14, 2479–90.
- Cooper, G.M., Shendure, J., 2011. Needles in stacks of needles: finding disease-causal variants in a wealth of genomic data. *Nature Reviews, Genetics* 12, 628–40.
- Cordero, P., Ashley, E.A., 2012. Whole-genome sequencing in personalized therapeutics. *Clinical Pharmacology and Therapeutics* 91, 1001–9.
- Cornetta, T., Festa, F., Testa, A., Cozzi, R., 2006. DNA damage repair and genetic polymorphisms: assessment of individual sensitivity and repair capacity. *International Journal of Radiation Oncology Biology Physics* 66, 537–45.

- Cortez, M.A., Ivan, C., Zhou, P., Wu, X., Ivan, M., Calin, G.A., 2010. MicroRNAs in cancer: from bench to bedside. *Advances in Cancer Research* 108, 113–57.
- Craig, T.A., Benson, L.M., Bergen, H.R., Venyaminov, S.Y., Salisbury, J.L., Ryan, Z.C., Thompson, J.R., Sperry, J., Gross, M.L., Kumar, R., 2006. Metal-binding properties of human centrin-2 determined by micro-electrospray ionization mass spectrometry and UV spectroscopy. *Journal of the American Society for Mass Spectrometry* 17, 1158–71.
- D'Amelio, A.M., Monroy, C., El-Zein, R., Etzel, C.J., 2012. Using haplotype analysis to elucidate significant associations between genes and Hodgkin lymphoma. *Leukemia Research* 36, 1359–64.
- D'Errico, M., Parlanti, E., Teson, M., de Jesus, B.M.B., Degan, P., Calcagnile, A., Jaruga, P., Bjørås, M., Crescenzi, M., Pedrini, A.M., Egly, J.-M., Zambruno, G., Stefanini, M., Dizdaroglu, M., Dogliotti, E., 2006. New functions of XPC in the protection of human skin cells from oxidative damage. *The EMBO Journal* 25, 4305–15.
- Dandona, S., Stewart, A.F.R., Roberts, R., 2012. Genomics: is it ready for primetime? *The Medical Clinics of North America* 96, 113–22.
- Dantas, T.J., Daly, O.M., Morrison, C.G., 2012. Such small hands: the roles of centrins/caltractins in the centriole and in genome maintenance. *Cellular and Molecular Life Sciences* 69, 2979–97.
- De Bakker, P.I.W., Yelensky, R., Pe'er, I., Gabriel, S.B., Daly, M.J., Altshuler, D., 2005. Efficiency and power in genetic association studies. *Nature Genetics* 37, 1217–23.
- De Haas, E.C., Zwart, N., Meijer, C., Nuver, J., Boezen, H.M., Suurmeijer, A.J.H., Hoekstra, H.J., van der Steege, G., Sleijfer, D.T., Gietema, J.A., 2008. Variation in bleomycin hydrolase gene is associated with reduced survival after chemotherapy for testicular germ cell cancer. *Journal of Clinical Oncology* 26, 1817–23.
- De Verdier, P.J., Sanyal, S., Bermejo, J.L., Steineck, G., Hemminki, K., Kumar, R., 2010. Genotypes, haplotypes and diplotypes of three *XPC* polymorphisms in urinary-bladder cancer patients. *Mutation Research* 694, 39–44.
- Decordier, I., Loock, K. Vande, Kirsch-Volders, M., 2010. Phenotyping for DNA repair capacity. *Mutation Research* 705, 107–29.
- Diffey, B.L., 1991. Solar ultraviolet radiation effects on biological systems. *Physics in Medicine and Biology* 36, 299–328.

- Ding, Q., Fan, B., Fan, Z., Ding, L., Li, F., Tu, W., Jin, X., Shi, Y., Wang, J., 2013. Interleukin-10-819C>T polymorphism contributed to cancer risk: Evidence from 29 studies. *Cytokine* 61, 139–45.
- Doecke, J., Zhao, Z.Z., Pandeya, N., Sadeghi, S., Stark, M., Green, A.C., Hayward, N.K., Webb, P.M., Whiteman, D.C., 2008. Polymorphisms in *MGMT* and DNA repair genes and the risk of esophageal adenocarcinoma. *Journal International du Cancer* 123, 174–80.
- Doherty, J.A., Weiss, N.S., Fish, S., Fan, W., Loomis, M.M., Sakoda, L.C., Rossing, M.A., Zhao, L.P., Chen, C., 2011. Polymorphisms in nucleotide excision repair genes and endometrial cancer risk. *Cancer Epidemiology Biomarkers & Prevention* 20, 1873–82.
- Dong, J., Hu, Z., Shu, Y., Pan, S., Chen, W., Wang, Y., Hu, L., Jiang, Y., Dai, J., Ma, H., Jin, G., Shen, H., 2012. Potentially functional polymorphisms in DNA repair genes and non-small-cell lung cancer survival: a pathway-based analysis. *Molecular Carcinogenesis* 51, 546–52.
- Duan, Z., Zhu, P., Dong, H., Gu, W., Yang, C., Liu, Q., Wang, Z., Jiang, J., 2007. Functional significance of the *TLR4*/11367 polymorphism identified in Chinese Han population. *Shock* 28, 160–4.
- El-Mahdy, M.A., Zhu, Q., Wang, Q., Wani, G., Praetorius-Ibba, M., Wani, A.A., 2006. Cullin 4A-mediated proteolysis of DDB2 protein at DNA damage sites regulates in vivo lesion recognition by XPC. *Journal of Biological Chemistry* 281, 13404–11.
- Engin, A.B., Karahalil, B., Engin, A., Karakaya, A.E., 2011. DNA repair enzyme polymorphisms and oxidative stress in a Turkish population with gastric carcinoma. *Molecular Biology Reports* 38, 5379–86.
- Evans, H.J., O’Riordan, M.L., Evans, J., 1975. Human peripheral blood lymphocytes for the analysis of chromosome aberrations in mutagen tests. *Mutation Research* 31, 135–48.
- Figueroa, J.D., Malats, N., Real, F.X., Silverman, D., Kogevinas, M., Chanock, S., Welch, R., Dosemeci, M., Tardón, A., Serra, C., Carrato, A., García-Closas, R., Castaño-Vinyals, G., Rothman, N., García-Closas, M., 2007. Genetic variation in the base excision repair pathway and bladder cancer risk. *Human Genetics* 121, 233–42.
- Fleming, N.D., Agadjanian, H., Nassanian, H., Miller, C.W., Orsulic, S., Karlan, B.Y., Walsh, C.S., 2012. *Xeroderma pigmentosum complementation group C* single-nucleotide polymorphisms in the nucleotide excision repair pathway correlate with prolonged progression-free survival in advanced ovarian cancer. *Cancer* 118, 689–97.

- Fontana, L., Bosviel, R., Delort, L., Guy, L., Chalabi, N., Kwiatkowski, F., Satih, S., Rabiau, N., Bioteux, J.-P., Chamoux, A., Bignon, Y.-J., Bernard-Gallon, D.J., 2008. DNA repair gene *ERCC2*, *XPC*, *XRCC1*, *XRCC3* polymorphisms and associations with bladder cancer risk in a French cohort. *Anticancer Research* 1856, 1853–6.
- Ford, J.M., 2005. Regulation of DNA damage recognition and nucleotide excision repair: another role for p53. *Mutation Research* 577, 195–202.
- Friedberg, E.C., 2001. How nucleotide excision repair protects against cancer. *Nature Reviews, Cancer* 1, 22–33.
- Fu, J., Festen, E.A.M., Wijmenga, C., 2011. Multi-ethnic studies in complex traits. *Human Molecular Genetics* 20, R206–13.
- Fuss, J.O., Cooper, P.K., 2006. DNA repair: dynamic defenders against cancer and aging. *PLoS Biology* 4, e203.
- Gabriel, S.B., Schaffner, S.F., Nguyen, H., Moore, J.M., Roy, J., Blumenstiel, B., Higgins, J., DeFelice, M., Lochner, A., Faggart, M., Liu-Cordero, S.N., Rotimi, C., Adeyemo, A., Cooper, R., Ward, R., Lander, E.S., Daly, M.J., Altshuler, D., 2002. The structure of haplotype blocks in the human genome. *Science* 296, 2225–9.
- Gangwar, R., Ahirwar, D., Mandhani, A., Mittal, R.D., 2009. Do DNA repair genes *OGG1*, *XRCC3* and *XRCC7* have an impact on susceptibility to bladder cancer in the North Indian population? *Mutation Research* 680, 56–63.
- Gao, Y., Hayes, R.B., Huang, W.-Y., Caporaso, N.E., Burdette, L., Yeager, M., Chanock, S.J., Berndt, S.I., 2011. DNA repair gene polymorphisms and tobacco smoking in the risk for colorectal adenomas. *Carcinogenesis* 32, 882–7.
- García-Closas, M., Malats, N., Real, F.X., Welch, R., Kogevinas, M., Chatterjee, N., Pfeiffer, R., Silverman, D., Dosemeci, M., Tardón, A., Serra, C., Carrato, A., García-Closas, R., Castaño-Vinyals, G., Chanock, S., Yeager, M., Rothman, N., 2006. Genetic variation in the nucleotide excision repair pathway and bladder cancer risk. *Cancer Epidemiology Biomarkers & Prevention* 15, 536–42.
- García-Martín, E., 2008. Interethnic and intraethnic variability of NAT2 single nucleotide polymorphisms. *Current Drug Metabolism* 9, 487–97.
- Gazdar, A.F., Boffetta, P., 2010. A risky business--identifying susceptibility loci for lung cancer. *Journal of the National Cancer Institute* 102, 920–3.
- Geenen, S., Taylor, P.N., Snoep, J.L., Wilson, I.D., Kenna, J.G., Westerhoff, H. V., 2012. Systems biology tools for toxicology. *Archives of Toxicology* 86, 1251–71.

- George Priya Doss, C., Sudandiradoss, C., Rajasekaran, R., Choudhury, P., Sinha, P., Hota, P., Batra, U.P., Rao, S., 2008. Applications of computational algorithm tools to identify functional SNPs. *Functional & Integrative Genomics* 8, 309–16.
- Glinskii, A.B., Ma, S., Ma, J., Grant, D., Lim, C.-U., Guest, I., Sell, S., Buttyan, R., Glinsky, G. V, 2011. Networks of intergenic long-range enhancers and snpRNAs drive castration-resistant phenotype of prostate cancer and contribute to pathogenesis of multiple common human disorders. *Cell Cycle* 10, 3571–97.
- Goode, E.L., Fridley, B.L., Sun, Z., Atkinson, E.J., Nord, A.S., McDonnell, S.K., Jarvik, G.P., Andrade, M. De, Slager, S.L., 2007. Comparison of tagging single-nucleotide polymorphism methods in association analyses. *BMC Proceedings* 1, S6.
- Gorlov, I.P., Gorlova, O.Y., Frazier, M.L., Spitz, M.R., Amos, C.I., 2011. Evolutionary evidence of the effect of rare variants on disease etiology. *Clinical Genetics* 79, 199–206.
- Gulcher, J., 2012. Microsatellite markers for linkage and association studies. *Cold Spring Harbor Protocols* 2012, 425–32.
- Gulseth, M.P., Grice, G.R., Dager, W.E., 2009. Pharmacogenomics of warfarin: uncovering a piece of the warfarin mystery. *American Journal of Health-System Pharmacy* 66, 123–33.
- Gunz, D., Hess, M.T., Naegeli, H., 1996. Recognition of DNA Adducts by Human Nucleotide Excision Repair. *Journal of Biological Chemistry* 271, 25089–98.
- Guo, W., Zhou, R.M., Wan, L.L., Wang, N., Li, Y., Zhang, X.J., Dong, X.J., 2008. Polymorphisms of the DNA repair gene *xeroderma pigmentosum groups A and C* and risk of esophageal squamous cell carcinoma in a population of high incidence region of North China. *Journal of Cancer Research & Clinical Oncology* 134, 263–70.
- Hagmar, L., Bonassi, S., Strömberg, U., Brøgger, A., Knudsen, L.E., Norppa, H., Reuterwall, C., and the European Study on Cytogenetic Biomarkers and Health, 1998. Chromosomal aberrations in lymphocytes predict human cancer: a report from the European Study Group on Cytogenetic Biomarkers and Health (ESCH). *Cancer Research* 58, 4117–21.
- Hagmar, L., Brøgger, A., Hansteen, I.-L., Heim, S., Högstedt, B., Knudsen, L., Lambert, B., Linnainmaa, K., Mitelman, F., Nordenson, I., Reuterwall, C., Salomaa, S., Skerving, S., Sorsa, M., 1994. Cancer risk in humans predicted by increased levels of chromosomal aberrations in lymphocytes: Nordic study group on the health risk of chromosome damage. *Cancer Research* 54, 2919–22.

- Hansen, R.D., Sørensen, M., Tjønneland, A., Overvad, K., Wallin, H., Raaschou-Nielsen, O., Vogel, U., 2007. *XPA* A23G, *XPC* Lys939Gln, *XPD* Lys751Gln and *XPD* Asp312Asn polymorphisms, interactions with smoking, alcohol and dietary factors, and risk of colorectal cancer. *Mutation Research* 619, 68–80.
- He, J., Shi, T.-Y., Zhu, M.-L., Wang, M.-Y., Li, Q.-X., Wei, Q.-Y., 2013. Associations of Lys939Gln and Ala499Val polymorphisms of the *XPC* gene with cancer susceptibility: a meta-analysis. *International Journal of Cancer* online, Epub ahead of print.
- Hecht, S.S., Lin, D., Castonguay, A., 1983. Effects of alpha-deuterium substitution on the mutagenicity of 4-(methyl-nitrosamino)-1-(3-pyridyl)-1-butanone (NNK). *Carcinogenesis* 4, 305–10.
- Henríquez-Hernández, L.A., Murias-Rosales, A., Hernández González, A., Cabrera De León, A., Díaz-Chico, B.N., Mori De Santiago, M., Fernández Pérez, L., 2009. Gene polymorphisms in *TYMS*, *MTHFR*, *p53* and *MDR1* as risk factors for breast cancer: A case-control study. *Oncology Reports* 22, 1425–1433.
- Herazo-Maya, J.D., Kaminski, N., 2012. Personalized medicine: applying “omics” to lung fibrosis. *Biomarkers in Medicine* 6, 529–40.
- Hill, C.E., Affatato, A.A., Wolfe, K.J., Lopez, M.S., Hallberg, C.K., Canistro, D., Abdel-Rahman, S.Z., 2005. Gender differences in genetic damage induced by the tobacco-specific nitrosamine NNK and the influence of the Thr241Met polymorphism in the *XRCC3* gene. *Environmental and Molecular Mutagenesis* 46, 22–9.
- Hirayama, T., 1981. Non-smoking wives of heavy smokers have a higher risk of lung cancer: a study from Japan. *British Medical Journal* 282, 183–5.
- Hoelzl, C., Glatt, H., Meisl, W., Sontag, G., Haidinger, G., Kundi, M., Simic, T., Chakraborty, A., Bichler, J., Ferk, F., Angelis, K., Nersesyan, A., Knasmüller, S., 2008. Consumption of Brussels sprouts protects peripheral human lymphocytes against 2-amino-1-methyl-6-phenylimidazo[4,5-b]pyridine (PhIP) and oxidative DNA-damage: results of a controlled human intervention trial. *Molecular Nutrition & Food Research* 52, 330–41.
- Hollox, E.J., 2012. The challenges of studying complex and dynamic regions of the human genome. *Methods in Molecular Biology* 838, 187–207.
- Hooker, S., Bonilla, C., Akereyeni, F., Ahaghotu, C., Kittles, R., 2008. *NAT2* and *NER* genetic variants and sporadic prostate cancer susceptibility in African Americans. *Prostate Cancer and Prostatic Disease* 11, 349–56.

- Hsia, T.-C., Liu, C.-J., Lin, C.-H., Chang, W.-S., Chu, C.-C., Hang, L.-W., Lee, H.-Z., Lo, W.-C., Bau, D.-T., 2011. Interaction of *CCND1* genotype and smoking habit in Taiwan lung cancer patients. *Anticancer Research* 31, 3601–5.
- Hsu, N.-Y., Wang, H.-C., Wang, C.-H., Chang, C.-L., Chiu, C.-F., Lee, H.-Z., Tsai, C.-W., Bau, D.-T., 2009. Lung cancer susceptibility and genetic polymorphism of DNA repair gene *XRCC4* in Taiwan. *Cancer Biomarkers* 5, 159–65.
- Hsu, T.C., Johnston, D.A., Cherry, L.M., Ramkissoon, D., Schantz, S.P., Jessup, J.M., Winn, R.J., Shirley, L., Furlong, C., 1989. Sensitivity to genotoxic effects of bleomycin in humans: possible relationship to environmental carcinogenesis. *Journal International du Cancer* 43, 403–9.
- Hsu, T.C., Spitz, M.R., Schantz, S.P., Shantz, S.P., 1991. Mutagen sensitivity: a biological marker of cancer susceptibility. *Cancer Epidemiology Biomarkers & Prevention* 1, 83–89.
- Hu, Z., Shao, M., Yuan, J., Xu, L., Wang, F., Wang, Yi, Yuan, W., Qian, J., Ma, H., Wang, Ying, Liu, H., Chen, W., Yang, L., Jin, G., Huo, X., Chen, F., Jin, L., Wei, Q., Huang, W., Lu, D., Wu, T., Shen, H., 2006. Polymorphisms in *DNA damage binding protein 2 (DDB2)* and susceptibility of primary lung cancer in the Chinese: a case-control study. *Carcinogenesis* 27, 1475–80.
- Hu, Z., Wang, H., Shao, M., Jin, G., Sun, W., Liu, H., Wang, Y., Ma, H., Qian, J., Wei, Q., Lu, D., Huang, W., Shen, H., 2007. Genetic variants in *MGMT* and risk of lung cancer in Southeastern Chinese: a haplotype based analysis. *Human Mutation* 28, 431–40.
- Huang, W.-Y., Berndt, S.I., Kang, D., Chatterjee, N., Chanock, S.J., Yeager, M., Welch, R., Bresalier, R.S., Weissfeld, J.L., Hayes, R.B., 2006. Nucleotide excision repair gene polymorphisms and risk of advanced colorectal adenoma: *XPC* polymorphisms modify smoking-related risk. *Cancer Epidemiology Biomarkers & Prevention* 15, 306–11.
- Huang, Y.-T., Chang, C.-J., Chao, K.-M., 2011. The extent of linkage disequilibrium and computational challenges of single nucleotide polymorphisms in genome-wide association studies. *Current Drug Metabolism* 12, 498–506.
- IARC, 1976. IARC monographs on the evaluation of the carcinogenic risk of chemicals to man: some naturally occurring substances. *IARC Monographs on the Evaluation of the Carcinogenic Risk of Chemicals to Man* 10, 1–342.
- Iodice, S., Gandini, S., Maisonneuve, P., Lowenfels, A.B., 2008. Tobacco and the risk of pancreatic cancer: a review and meta-analysis. *Langenbeck's Archives of Surgery* 393, 535–45.

- ISCN, 1985. An International System for Human Cytogenetic Nomenclature (1985)
ISCN 1985. Report of the Standing Committee on Human Cytogenetic
Nomenclature. Birth Defects Original Article Series 21, 1–117.
- Jackson, S.P., Bartek, J., 2009. The DNA-damage response in human biology and
disease. *Nature* 461, 1071–8.
- Jhappan, C., Noonan, F.P., Merlino, G., 2003. Ultraviolet radiation and cutaneous
malignant melanoma. *Oncogene* 22, 3099–112.
- Ji, G., Lin, Y., Cao, S., Li, L., Chen, X.-L., Sun, B.-M., Chen, C.-J., Ma, H.-X., 2012.
XPC 939A>C and 499C>T polymorphisms and skin cancer risk: a meta-analysis.
Asian Pacific Journal of Cancer Prevention 13, 4983–8.
- Johnson, J.A., Burkley, B.M., Langaee, T.Y., Clare-Salzler, M.J., Klein, T.E., Altman,
R.B., 2012. Implementing personalized medicine: development of a cost-effective
customized pharmacogenetics genotyping array. *Clinical Pharmacology and
Therapeutics* 92, 437–9.
- Kang, S., Sun, H.-Y., Zhou, R.-M., Wang, N., Hu, P., Li, Y., 2013. DNA repair gene
associated with clinical outcome of epithelial ovarian cancer treated with platinum-
based chemotherapy. *Asian Pacific Journal of Cancer Prevention* 14, 941–6.
- Kanzaki, H., Ouchida, M., Hanafusa, H., Sakai, A., Yamamoto, H., Suzuki, H., Yano, M.,
Aoe, M., Imai, K., Date, H., Nakachi, K., Shimizu, K., 2007. Single nucleotide
polymorphism in the *RAD18* gene and risk of colorectal cancer in the Japanese
population. *Oncology Reports* 18, 1171–5.
- Keller, B., Martini, S., Sedor, J., Kretzler, M., 2010. Linking variants from genome wide
association analysis to function via transcriptional network analysis. *Seminars in
Nephrology* 30, 177–84.
- Khan, S.G., Muniz-Medina, V., Shahlavi, T., Baker, C.C., Inui, H., Ueda, T., Emmert, S.,
Schneider, T.D., Kraemer, K.H., 2002. The human *XPC* DNA repair gene:
arrangement, splice site information content and influence of a single nucleotide
polymorphism in a splice acceptor site on alternative splicing and function. *Nucleic
Acids Research* 30, 3624–31.
- Khan, S.G., Oh, K.-S., Shahlavi, T., Ueda, T., Busch, D.B., Inui, H., Emmert, S., Imoto,
K., Muniz-Medina, V., Baker, C.C., DiGiovanna, J.J., Schmidt, D., Khadavi, A.,
Metin, A., Gozukara, E., Slor, H., Sarasin, A., Kraemer, K.H., 2006. Reduced *XPC*
DNA repair gene mRNA levels in clinically normal parents of xeroderma
pigmentosum patients. *Carcinogenesis* 27, 84–94.
- Kim, J.G., Sohn, S.K., Chae, Y.S., Song, H.S., Kwon, K.-Y., Do, Y.R., Kim, M.K., Lee,
K.H., Hyun, M.S., Lee, W.S., Sohn, C.-H., Jung, J.S., Kim, G.C., Chung, H.Y., Yu,

- W., 2009. *TP53* codon 72 polymorphism associated with prognosis in patients with advanced gastric cancer treated with paclitaxel and cisplatin. *Cancer Chemotherapy and Pharmacology* 64, 355–60.
- Kim, M., Kang, H.-G., Lee, S.Y., Lee, H.C., Lee, E.B., Choi, Y.Y., Lee, W.K., Cho, S., Jin, G., Jheon, H.-S., Son, J.W., Lee, M.-H., Jung, D.K., Cha, S.I., Kim, C.H., Kang, Y.M., Kam, S., Jung, T.H., Jheon, S., Park, J.Y., 2010. Comprehensive analysis of DNA repair gene polymorphisms and survival in patients with early stage non-small-cell lung cancer. *Cancer Science* 101, 2436–42.
- Kinslow, C.J., El-Zein, R.A., Hill, C.E., Wickliffe, J.K., Abdel-Rahman, S.Z., 2008. Single nucleotide polymorphisms 5' upstream the coding region of the *NEIL2* gene influence gene transcription levels and alter levels of genetic damage. *Genes, Chromosomes & Cancer* 47, 923–32.
- Kiyohara, C., Takayama, K., Nakanishi, Y., 2006. Association of genetic polymorphisms in the base excision repair pathway with lung cancer risk: a meta-analysis. *Lung Cancer* 54, 267–83.
- Klassen, C.D., 2001. Casarett & Doull's Toxicology: The Basic Science of Poisons, 6th Edition, 6th ed. McGraw-Hill Medical Publishing Division, New York, NY.
- Kohno, T., Sakiyama, T., Kunitoh, H., Goto, K., Nishiwaki, Y., Saito, D., Hirose, H., Eguchi, T., Yanagitani, N., Saito, R., Sasaki-Matsumura, R., Mimaki, S., Toyama, K., Yamamoto, S., Kuchiba, A., Sobue, T., Ohta, T., Ohki, M., Yokota, J., 2006. Association of polymorphisms in the *MTH1* gene with small cell lung carcinoma risk. *Carcinogenesis* 27, 2448–54.
- Komar, A.A., 2007. SNPs, silent but not invisible. *Science* 315, 466–7.
- Krynetskiy, E., McDonnell, P., 2007. Building individualized medicine: prevention of adverse reactions to warfarin therapy. *The Journal of Pharmacology and Experimental Therapeutics* 322, 427–34.
- Lacoste, S., Castonguay, A., Drouin, R., 2007. Repair kinetics of specific types of nitroso-induced DNA damage using the comet assay in human cells. *Mutation Research* 624, 18–30.
- Ladeiras-Lopes, R., Pereira, A.K., Nogueira, A., Pinheiro-Torres, T., Pinto, I., Santos-Pereira, R., Lunet, N., 2008. Smoking and gastric cancer: systematic review and meta-analysis of cohort studies. *Cancer Causes & Control* 19, 689–701.
- Lam, C.-W., Lau, K.-C., Tong, S.-F., 2010. Microarrays for personalized genomic medicine. *Advances in Clinical Chemistry* 52, 1–18.

- Langie, S. a S., Knaapen, A.M., Brauers, K.J.J., van Berlo, D., van Schooten, F.-J., Godschalk, R.W.L., 2006. Development and validation of a modified comet assay to phenotypically assess nucleotide excision repair. *Mutagenesis* 21, 153–8.
- Langie, S.A.S., Wilms, L.C., Hämäläinen, S., Kleinjans, J.C.S., Godschalk, R.W.L., van Schooten, F.J., 2010. Modulation of nucleotide excision repair in human lymphocytes by genetic and dietary factors. *The British Journal of Nutrition* 103, 490–501.
- Law, A.J., Kleinman, J.E., Weinberger, D.R., Weickert, C.S., 2007. Disease-associated intronic variants in the *ErbB4* gene are related to altered *ErbB4* splice-variant expression in the brain in schizophrenia. *Human Molecular Genetics* 16, 129–41.
- Lee, G.Y., Jang, J.-S., Lee, S.Y., Jeon, H.-S., Kim, K.M., Choi, J.E., Park, J.M., Chae, M.H., Lee, W.K., Kam, S., Kim, I.-S., Lee, J.-T., Jung, T.H., Park, J.Y., 2005. *XPC* polymorphisms and lung cancer risk. *Journal International du Cancer* 115, 807–13.
- Leng, S., Bernauer, A., Stidley, C.A., Picchi, M.A., Sheng, X., Frasco, M.A., Van Den Berg, D., Gilliland, F.D., Crowell, R.E., Belinsky, S.A., 2008. Association between common genetic variation in *Cockayne syndrome A* and *B* genes and nucleotide excision repair capacity among smokers. *Cancer Epidemiology, Biomarkers & Prevention* 17, 2062–9.
- Leng, S., Stidley, C.A., Liu, Y., Edlund, C.K., Willink, R.P., Han, Y., Landi, M.T., Thun, M., Picchi, M.A., Bruse, S.E., Crowell, R.E., Van Den Berg, D., Caporaso, N.E., Amos, C.I., Siegfried, J.M., Tesfayzi, Y., Gilliland, F.D., Belinsky, S.A., 2012. Genetic determinants for promoter hypermethylation in the lungs of smokers: a candidate gene-based study. *Cancer Research* 72, 707–15.
- Li, C., Hu, Z., Lu, J., Liu, Z., Wang, L.-E., El-Naggar, A.K., Sturgis, E.M., Spitz, M.R., Wei, Q., 2007. Genetic polymorphisms in DNA base-excision repair genes ADPRT, XRCC1, and APE1 and the risk of squamous cell carcinoma of the head and neck. *Cancer* 110, 867–75.
- Liang, C., Marsit, C.J., Houseman, E.A., Butler, R., Nelson, H.H., Mcclean, M.D., Kelsey, K.T., 2012. Gene–environment interactions of novel variants associated with head and neck cancer. *Head & Neck* 34, 1111–8.
- Lin, S.-L., Miller, J.D., Ying, S.-Y., 2006. Intronic microRNA (miRNA). *Journal of Biomedicine & Biotechnology* 2006, 26818 (1–13).
- Liu, C., Yin, Q., Hu, J., Li, L., Zhang, Y., Wang, Y., 2013. A meta-analysis of evidences on *XPC* polymorphisms and lung cancer susceptibility. *Tumor Biology* 34, 1205–13.
- Lo, Y.-L., Hsiao, C.-F., Jou, Y.-S., Chang, G.-C., Tsai, Y.-H., Su, W.-C., Chen, Y.-M., Huang, M.-S., Chen, H.-L., Yang, P.-C., Chen, C.-J., Hsiung, C.A., 2010. ATM

- polymorphisms and risk of lung cancer among never smokers. *Lung Cancer* 69, 148–54.
- Lodovici, M., Bigagli, E., 2009. Biomarkers of induced active and passive smoking damage. *International Journal of Environmental Research and Public Health* 6, 874–88.
- Long, X.-D., Ma, Y., Huang, Y.-Z., Yi, Y., Liang, Q.-X., Ma, A.-M., Zeng, L.-P., Fu, G.-H., 2010a. Genetic Polymorphisms in DNA Repair Genes *XPC*, *XPB*, and *XRCC4*, and Susceptibility to Helicobacter pylori Infection-Related Gastric Antrum Adenocarcinoma in Guangxi Population, China. *Molecular Carcinogenesis* 49, 611–8.
- Long, X.-D., Ma, Y., Zhou, Y.-F., Ma, A.-M., Fu, G.-H., 2010b. Polymorphism in *xeroderma pigmentosum complementation group C* codon 939 and aflatoxin B1-related hepatocellular carcinoma in the Guangxi population. *Hepatology* 52, 1301–9.
- Lui, G., Zhou, W., Christiani, D.C., Liu, G., 2005. Molecular epidemiology of non-small cell lung cancer. *Seminars in Respiratory and Critical Care Medicine* 26, 265–72.
- Luijsterburg, M.S., Lindh, M., Acs, K., Vrouwe, M.G., Pines, A., van Attikum, H., Mullenders, L.H., Dantuma, N.P., 2012. DDB2 promotes chromatin decondensation at UV-induced DNA damage. *The Journal of Cell Biology* 197, 267–81.
- Ma, H., Hu, Z., Wang, H., Jin, G., Wang, Y., Sun, W., Chen, D., Tian, T., Jin, L., Wei, Q., Lu, D., Huang, W., Shen, H., 2009. *ERCC6/CSB* gene polymorphisms and lung cancer risk. *Cancer Letters* 273, 172–6.
- Ma, H., Yu, H., Liu, Z., Wang, L.-E., Sturgis, E.M., Wei, Q., 2012. Polymorphisms of *XPG/ERCC5* and risk of squamous cell carcinoma of the head and neck. *Pharmacogenetics and Genomics* 22, 50–7.
- Ma, J., Zhang, J., Ning, T., Chen, Z., Xu, C., 2012. Association of genetic polymorphisms in *MDM2*, *PTEN* and *P53* with risk of esophageal squamous cell carcinoma. *Journal of Human Genetics* 57, 261–4.
- Maekawa, K., Fukushima-Uesaka, H., Tohkin, M., Hasegawa, R., Kajio, H., Kuzuya, N., Yasuda, K., Kawamoto, M., Kamatani, N., Suzuki, K., Yanagawa, T., Saito, Y., Sawada, J., 2006. Four novel defective alleles and comprehensive haplotype analysis of *CYP2C9* in Japanese. *Pharmacogenetics and Genomics* 16, 497–514.
- Mahmoudi, T., Karimi, K., Mohebbi, S.R., Fatemi, S.R., Zali, M.R., 2011. Start codon *FokI* and intron 8 *BsmI* variants in the vitamin D receptor gene and susceptibility to colorectal cancer. *Molecular Biology Reports* 38, 4765–70.

- Maillard, O., Solyom, S., Naegeli, H., 2007. An aromatic sensor with aversion to damaged strands confers versatility to DNA repair. *PLoS Biology* 5, 717–28.
- Manuguerra, M., Saletta, F., Karagas, M.R., Berwick, M., Veglia, F., Vineis, P., Matullo, G., 2006. *XRCC3* and *XPD/ERCC2* single nucleotide polymorphisms and the risk of cancer: a HuGE review. *American Journal of Epidemiology* 164, 297–302.
- Martin, J.S., Halvorsen, M., Davis-Neulander, L., Ritz, J., Gopinath, C., Beauregard, A., Laederach, A., 2012. Structural effects of linkage disequilibrium on the transcriptome. *RNA* 18, 77–87.
- McWilliams, R.R., Bamlet, W.R., Cunningham, J.M., Goode, E.L., de Andrade, M., Boardman, L.A., Petersen, G.M., 2008. Polymorphisms in DNA repair genes, smoking, and pancreatic adenocarcinoma risk. *Cancer Research* 68, 4928–35.
- McWilliams, R.R., Bamlet, W.R., de Andrade, M., Rider, D.N., Cunningham, J.M., Petersen, G.M., 2009. Nucleotide excision repair pathway polymorphisms and pancreatic cancer risk: evidence for role of *MMS19L*. *Cancer Epidemiology, Biomarkers & Prevention* 18, 1295–302.
- Melis, J.P.M., Luijten, M., Mullenders, L.H.F., van Steeg, H., 2011. The role of XPC: implications in cancer and oxidative DNA damage. *Mutation Research* 728, 107–17.
- Melis, J.P.M., van Steeg, H., Luijten, M., 2013. Oxidative DNA damage and nucleotide excision repair. *Antioxidants & Redox Signaling* 18, 2409–19.
- Millikan, R.C., Hummer, A., Begg, C., Player, J., de Cotret, A.R., Winkel, S., Mohrenweiser, H., Thomas, N., Armstrong, B., Krickler, A., Marrett, L.D., Gruber, S.B., Culver, H.A., Zanetti, R., Gallagher, R.P., Dwyer, T., Rebbeck, T.R., Busam, K., From, L., Mujumdar, U., Berwick, M., 2006. Polymorphisms in nucleotide excision repair genes and risk of multiple primary melanoma: the Genes Environment and Melanoma Study. *Carcinogenesis* 27, 610–18.
- Ming, M., Feng, L., Shea, C.R., Soltani, K., Zhao, B., Han, W., Smart, R.C., Trempus, C.S., He, Y., 2012. PTEN positively regulates UVB-induced DNA damage repair. *Cancer Research* 71, 5287–95.
- Mir, K.U., 2009. Sequencing genomes: from individuals to populations. *Briefings in Functional Genomics & Proteomics* 8, 367–78.
- Mittal, R.D., Gangwar, R., Mandal, R.K., Srivastava, P., Ahirwar, D.K., 2012. Gene variants of *XRCC4* and *XRCC3* and their association with risk for urothelial bladder cancer. *Molecular Biology Reports* 39, 1667–75.
- Mohelnikova-duchonova, B., Marsakova, P.L., Vrana, D., Holcatova, I., Ryska, M., Smerhovsky, Z., Slamova, A., Schejbalova, M., Soucek, P., 2011. Superoxide

- dismutase and nicotinamide adenine dinucleotide phosphate quinone oxidoreductase polymorphisms and pancreatic cancer risk. *Pancreas* 40, 72–8.
- Morita, M., Kumashiro, R., Kubo, N., Nakashima, Y., Yoshida, R., Yoshinaga, K., Saeki, H., Emi, Y., Kakeji, Y., Sakaguchi, Y., Toh, Y., Maehara, Y., 2010. Alcohol drinking, cigarette smoking, and the development of squamous cell carcinoma of the esophagus: epidemiology, clinical findings, and prevention. *International Journal of Clinical Oncology* 15, 126–34.
- Mucha, L., Stephenson, J., Morandi, N., Dirani, R., 2006. Meta-analysis of disease risk associated with smoking, by gender and intensity of smoking. *Gender Medicine* 3, 279–91.
- Musunuru, K., Strong, A., Frank-Kamenetsky, M., Lee, N.E., Ahfeldt, T., Sachs, K. V., Li, X., Li, H., Kuperwasser, N., Ruda, V.M., Pirruccello, J.P., Muchmore, B., Prokunina-Olsson, L., Hall, J.L., Schadt, E.E., Morales, C.R., Lund-Katz, S., Phillips, M.C., Wong, J., Cantley, W., Racie, T., Ejebe, K.G., Orho-Melander, M., Melander, O., Koteliansky, V., Fitzgerald, K., Krauss, R.M., Cowan, C.A., Kathiresan, S., Rader, D.J., 2010. From noncoding variant to phenotype via SORT1 at the 1p13 cholesterol locus. *Nature* 466, 714–9.
- Nagira, T., Narisawa, J., Teruya, K., Katakura, Y., Shim, S.-Y., Kusumoto, K.-I., Tokumaru, S., Tokumaru, K., Barnes, D.W., Shirahata, S., 2002. Suppression of UVC-induced cell damage and enhancement of DNA repair by the fermented milk, Kefir. *Cytotechnology* 40, 125–37.
- Nakagawa, A., Kobayashi, N., Muramatsu, T., Yamashina, Y., Shirai, T., Hashimoto, M.W., Ikenaga, M., Mori, T., 1998. Three-dimensional visualization of ultraviolet-induced DNA damage and its repair in human cell nuclei. *Journal of Investigative Dermatology* 110, 143–8.
- Nakai, K.K., Tsuboi, J., Okabayashi, H., Fukuhiro, Y., Oka, T., Habano, W., Fukushima, N., Obara, W., Fujioka, T., Suwabe, A., Gurwitz, D., Fukushima, N., 2007. Ethnic differences in the *VKORC1* gene polymorphism and an association with warfarin dosage requirements in cardiovascular surgery patients. *Pharmacogenomics* 8, 713–9.
- Nakken, S., Alseth, I., Rognes, T., 2007. Computational prediction of the effects of non-synonymous single nucleotide polymorphisms in human DNA repair genes. *Neuroscience* 145, 1273–9.
- Nam, M.-H., Won, H.-H., Lee, K.-A., Kim, J.-W., 2007. Effectiveness of in silico tagSNP selection methods: virtual analysis of the genotypes of pharmacogenetic genes. *Pharmacogenomics* 8, 1347–57.

- Nelson, D., Cox, M., 2005. *Lehninger's Principles of Biochemistry*, 4th Edition., 4th ed. W. H. Freeman and Company, New York, NY.
- Ng, P.C., Kirkness, E.F., 2010. Whole genome sequencing. *Methods in Molecular Biology* 628, 215–26.
- Nishi, R., Okuda, Y., Watanabe, E., Mori, T., Iwai, S., Masutani, C., Sugasawa, K., Hanaoka, F., 2005. Centrin 2 stimulates nucleotide excision repair by interacting with Xeroderma pigmentosum group C protein. *Molecular and Cellular Biology* 25, 5664–74.
- Nohl, H., Kozlov, A. V, Gille, L., Staniek, K., 2003. Cell respiration and formation of reactive oxygen species: facts and artefacts. *Biochemical Society Transactions* 31, 1308–11.
- Nowsheen, S., Yang, E.S., 2012. The intersection between DNA damage response and cell death pathways. *Experimental Oncology* 34, 243–54.
- Ona, K.R., Courcelle, C.T., Courcelle, J., 2009. Nucleotide excision repair is a predominant mechanism for processing nitrofurazone-induced DNA damage in *Escherichia coli*. *Journal of Bacteriology* 191, 4959–65.
- Pabalan, N., Francisco-Pabalan, O., Jarjanazi, H., Li, H., Sung, L., Ozcelik, H., 2012. Racial and tissue-specific cancer risk associated with *PARP1* (*ADPRT*) Val762Ala polymorphism: a meta-analysis. *Molecular Biology Reports* 39, 11061–72.
- Panagiotou, O.A., Evangelou, E., Ioannidis, J.P.A., 2010. Genome-wide significant associations for variants with minor allele frequency of 5% or less--an overview: A HuGE review. *American Journal of Epidemiology* 172, 869–89.
- Park, C.-J., Choi, B.-S., 2006. The protein shuffle. Sequential interactions among components of the human nucleotide excision repair pathway. *FEBS Journal* 273, 1600–8.
- Park, S.L., Bastani, D., Goldstein, B.Y., Chang, S.-C., Cozen, W., Cai, L., Cordon-Cardo, C., Ding, B., Greenland, S., He, N., Hussain, S.K., Jiang, Q., Lee, Y.-C.A., Liu, S., Lu, M.-L., Mack, T.M., Mao, J.T., Morgenstern, H., Mu, L.-N., Oh, S.S., Pantuck, A., Papp, J.C., Rao, J., Reuter, V.E., Tashkin, D.P., Wang, H., You, N.-C.Y., Yu, S.-Z., Zhao, J.-K., Zhang, Z.-F., 2010. Associations between NBS1 polymorphisms, haplotypes and smoking-related cancers. *Carcinogenesis* 31, 1264–71.
- Parliament, M.B., Murray, D., 2010. Single nucleotide polymorphisms of DNA repair genes as predictors of radioresponse. *Seminars in Radiation Oncology* 20, 232–40.
- Paszowska-Szczur, K., Scott, R., Serrano-Fernandez, P., Mirecka, A., Gapska, P., Górski, B., Cybulski, C., Maleszka, R., Sulikowski, M., Nagay, L., Lubinski, J.,

- Dębniak, T., 2013. Xeroderma pigmentosum genes and melanoma risk. *Journal international du cancer online*, Epub ahead of print.
- Peng, T., Shen, H.-M., Liu, Z.-M., Yan, L.-N., Peng, M.-H., Li, L.-Q., Liang, R.-X., Wei, Z.-L., Halliwell, B., Ong, C.N., 2003. Oxidative DNA damage in peripheral leukocytes and its association with expression and polymorphisms of *hOGG1*: A study of adolescents in a high risk region for hepatocellular carcinoma in China. *World Journal of Gastroenterology* 9, 2186–93.
- Pereira, C., Pimentel-Nunes, P., Brandão, C., Moreira-Dias, L., Medeiros, R., Dinis-Ribeiro, M., 2010. COX-2 polymorphisms and colorectal cancer risk: a strategy for chemoprevention. *European Journal of Gastroenterology & Hepatology* 22, 607–13.
- Peterson, L.A., 2010. Formation, repair, and genotoxic properties of bulky DNA adducts formed from tobacco-specific nitrosamines. *Journal of Nucleic Acids* 2010, 284935 (1–11).
- Pfeifer, G.P., 1997. Formation and processing of UV photoproducts: effects of DNA sequence and chromatin environment. *Photochemistry and Photobiology* 65, 270–83.
- Popescu, A., Miron, S., Blouquit, Y., Duchambon, P., Christova, P., Craescu, C.T., 2003. Xeroderma pigmentosum group C protein possesses a high affinity binding site to human centrin 2 and calmodulin. *Journal of Biological Chemistry* 278, 40252–61.
- Pryor, W.A., 1997. Cigarette smoke radicals and the role of free radicals in chemical carcinogenicity. *Environmental Health Perspectives* 105 Suppl, 875–82.
- Qiao, B., Ansari, A.-H., Scott, G.B., Sak, S.C., Chambers, P.A., Elliott, F., Teo, M.T.W., Bentley, J., Churchman, M., Hall, J., Taylor, C.F., Bishop, T.D., Knowles, M.A., Kiltie, A.E., 2011a. In vitro functional effects of *XPC* gene rare variants from bladder cancer patients. *Carcinogenesis* 32, 516–21.
- Qiao, B., Scott, G.B., Elliott, F., Vaslin, L., Bentley, J., Hall, J., Bishop, D.T., Knowles, M.A., Kiltie, A.E., 2011b. Functional assays to determine the significance of two common *XPC* 3'UTR variants found in bladder cancer patients. *BMC Medical Genetics* 12, 84 (1–7).
- Qiao, Y., Spitz, M.R., Guo, Z., Hadeyati, M., Grossman, L., Kraemer, K.H., Wei, Q., 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. *Mutation Research* 509, 165–74.
- Qiu, Y., Wang, W., Wang, T., Sun, P., Wu, F., Zhu, S., Qian, J., Jin, L., Au, W., Xia, Z., 2011. DNA repair gene polymorphisms and micronucleus frequencies in Chinese

- workers exposed to vinyl chloride monomer. *International Journal of Hygiene and Environmental Health* 214, 225–30.
- Rastogi, R.P., Richa, Kumar, A., Tyagi, M.B., Sinha, R.P., 2010. Molecular mechanisms of ultraviolet radiation-induced DNA damage and repair. *Journal of Nucleic Acids* 2010, 592980 (1–32).
- Ray, A., Milum, K., Battu, A., Wani, G., Wani, A.A., 2013. NER initiation factors, DDB2 and XPC, regulate UV radiation response by recruiting ATR and ATM kinases to DNA damage sites. *DNA Repair* 12, 273–83.
- Renaud, E., Miccoli, L., Zagal, N., Biard, D.S., Craescu, C.T., Rainbow, A.J., Angulo, J.F., 2011. Differential contribution of XPC, RAD23A, RAD23B and CENTRIN 2 to the UV-response in human cells. *DNA Repair* 10, 835–47.
- Rieder, M.J., Reiner, A.P., Gage, B.F., Nickerson, D.A., Eby, C.S., McLeod, H.L., Blough, D.L., Thummel, K.E., Veenstra, D.L., Rettie, A.E., 2005. Effect of *VKORC1* haplotypes on transcriptional regulation and warfarin dose. *New England Journal of Medicine* 352, 2285–93.
- Roberts, M.R., Shields, P.G., Ambrosone, C.B., Nie, J., Marian, C., Krishnan, S.S., Goerlitz, D.S., Modali, R., Seddon, M., Lehman, T., Amend, K.L., Trevisan, M., Edge, S.B., Freudenheim, J.L., 2011. Single-nucleotide polymorphisms in DNA repair genes and association with breast cancer risk in the web study. *Carcinogenesis* 32, 1223–30.
- Rothman, N., Poirier, M.C., Baser, M.E., Hansen, J.A., Gentile, C., Bowman, E.D., Strickland, P.T., Hanson, J., 1990. Formation of polycyclic aromatic hydrocarbon-DNA adducts in peripheral white blood cells during consumption of charcoal-broiled beef. *Carcinogenesis* 11, 1241–3.
- Rouzaud, F., Oulmouden, A., Kos, L., 2010. The untranslated side of hair and skin mammalian pigmentation: Beyond coding sequences. *IUBMB Life* 62, 340–6.
- Rukov, J.L., Shomron, N., 2011. MicroRNA pharmacogenomics: post-transcriptional regulation of drug response. *Trends in Molecular Medicine* 17, 412–23.
- Russo, P., Nastrucci, C., Alzetta, G., Szalai, C., 2011. Tobacco habit: historical, cultural, neurobiological, and genetic features of people's relationship with an addictive drug. *Perspectives in Biology and Medicine* 54, 557–77.
- Rzhetsky, A., Nei, M., 1992. A simple method for estimating and testing minimum-evolution trees. *Molecular Biology Evolution* 9, 945–67.

- Sak, S.C., Barrett, J.H., Paul, A.B., Bishop, D.T., Kiltie, A.E., 2006. Comprehensive analysis of 22 *XPC* polymorphisms and bladder cancer risk. *Cancer Epidemiology, Biomarkers & Prevention* 15, 2537–41.
- Sakoda, L.C., Loomis, M.M., Doherty, J. a, Julianto, L., Barnett, M.J., Neuhouser, M.L., Thornquist, M.D., Weiss, N.S., Goodman, G.E., Chen, C., 2012. Germ line variation in nucleotide excision repair genes and lung cancer risk in smokers. *International Journal of Molecular Epidemiology and Genetics* 3, 1–17.
- Sasco, A.J., Secretan, M.B., Straif, K., 2004. Tobacco smoking and cancer: a brief review of recent epidemiological evidence. *Lung Cancer* 45 Suppl 2, S3–9.
- Schuch, A.P., Menck, C.F.M., 2010. The genotoxic effects of DNA lesions induced by artificial UV-radiation and sunlight. *Journal of Photochemistry & Photobiology. B, Biology* 99, 111–6.
- Schwartz, A.G., Cote, M.L., Wenzlaff, A.S., Land, S., Amos, C.I., 2009. Racial Differences in the Association Between SNPs on 15q25.1, Smoking Behavior, and Risk of Non-small Cell Lung Cancer. *Journal of Thoracic Oncology* 4, 1195–201.
- Seldin, M.F., Pasaniuc, B., Price, A.L., 2011. New approaches to disease mapping in admixed populations. *Nature Reviews, Genetics* 12, 523–8.
- Sengupta, S., Harris, C.C., 2005. p53: traffic cop at the crossroads of DNA repair and recombination. *Nature Reviews, Molecular Cell Biology* 6, 44–55.
- Sharma, R., Ahuja, M., Panda, N.K., Khullar, M., 2011. Interactions among genetic variants in tobacco metabolizing genes and smoking are associated with head and neck cancer susceptibility in North Indians. *DNA and Cell Biology* 30, 611–6.
- Shen, H., Spitz, M.R., Qiao, Y., Guo, Z., Wang, L.-E., Bosken, C.H., Amos, C.I., Wei, Q., 2003. Smoking, DNA repair capacity and risk of nonsmall cell lung cancer. *Journal International du Cancer* 107, 84–8.
- Shen, J., Terry, M.B., Gammon, M.D., Gaudet, M.M., Teitelbaum, S.L., Eng, S.M., Sagiv, S.K., Neugut, A.I., Santella, R.M., 2005. MGMT genotype modulates the associations between cigarette smoking, dietary antioxidants and breast cancer risk. *Carcinogenesis* 26, 2131–7.
- Shen, M., Berndt, S.I., Rothman, N., Demarini, D.M., Mumford, J.L., He, X., Bonner, M.R., Tian, L., Yeager, M., Welch, R., Chanock, S., Zheng, T., Caporaso, N., Lan, Q., 2005. Polymorphisms in the DNA nucleotide excision repair genes and lung cancer risk in Xuan Wei, China. *Journal International du Cancer* 116, 768–73.
- Shriner, D., Adeyemo, A., Ramos, E., Chen, G., Rotimi, C.N., 2011. Mapping of disease-associated variants in admixed populations. *Genome Biology* 12, 223 (1–8).

- Simonelli, V., Mazzei, F., D'Errico, M., Dogliotti, E., 2012. Gene susceptibility to oxidative damage: from single nucleotide polymorphisms to function. *Mutation Research* 731, 1–13.
- Sinha, R.P., Häder, D.-P., 2002. UV-induced DNA damage and repair: a review. *Photochemical & Photobiological Sciences* 1, 225–36.
- Sliwinski, T., Przybylowska, K., Markiewicz, L., Rusin, P., Pietruszewska, W., Zelinska-Blizniewska, H., Olszewski, J., Morawiec-Sztandera, A., Mlynarski, W., Majsterek, I., 2011. *MUTYH* Tyr165Cys, *OGGI* Ser326Cys and *XPB* Lys751Gln polymorphisms and head neck cancer susceptibility: a case control study. *Molecular Biology Reports* 38, 1251–61.
- Slyskova, J., Naccarati, A., Pardini, B., Polakova, V., Vodickova, L., Smerhovsky, Z., Levy, M., Lipska, L., Liska, V., Vodicka, P., 2012. Differences in nucleotide excision repair capacity between newly diagnosed colorectal cancer patients and healthy controls. *Mutagenesis* 27, 519–22.
- Slyskova, J., Naccarati, A., Polakova, V., Pardini, B., Vodickova, L., Stetina, R., Schmuczerova, J., Smerhovsky, Z., Lipska, L., Vodicka, P., 2011. DNA damage and nucleotide excision repair capacity in healthy individuals. *Environmental & Molecular Mutagenesis* 52, 511–7.
- Smedby, K.E., Lindgren, C.M., Hjalgrim, H., Humphreys, K., Schöllkopf, C., Chang, E.T., Roos, G., Ryder, L.P., Falk, K.I., Palmgren, J., Kere, J., Melbye, M., Glimelius, B., Adami, H.-O., 2006. Variation in DNA repair genes ERCC2, XRCC1, and XRCC3 and risk of follicular lymphoma. *Cancer Epidemiology, Biomarkers & Prevention* 15, 258–65.
- Smith, G.B.J., Castonguay, A., Donnelly, P.J., Reid, K.R., Petsikas, D., Massey, T.E., 1999. Biotransformation of the tobacco-specific carcinogen 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone (NNK) in freshly isolated human lung cells. *Carcinogenesis* 20, 1809–18.
- Song, F.-J., Chen, K.-X., 2011. Single-nucleotide polymorphisms among microRNA: big effects on cancer. *Chinese Journal of Cancer* 30, 381–91.
- Song, X., Sturgis, E.M., Jin, L., Wang, Z., Wei, Q., Li, G., 2013. Variants in nucleotide excision repair core genes and susceptibility to recurrence of squamous cell carcinoma of the oropharynx. *Journal International du Cancer* 133, 695–704.
- Spitz, M.R., Bondy, M.L., 1993. Genetic susceptibility to cancer. *Cancer* 72, 991–5.
- Spitz, M.R., Fueger, J.J., Beddingfield, N.A., Annegers, J.F., Hsu, T.C., Newell, G.R., Schantz, S.P., 1989. Chromosome sensitivity to bleomycin-induced mutagenesis, an

- independent risk factor for upper aerodigestive tract cancers. *Cancer Research* 49, 4626–8.
- Spitz, M.R., Hsu, T., Wu, X., Fueger, J.J., Amos, C.I., Roth, J.A., 1995. Mutagen sensitivity as a biological marker of lung cancer risk in African Americans. *Cancer Epidemiology, Biomarkers & Prevention* 4, 99–103.
- Srkar-Roy, N., Mondal, D., Bhattacharya, P., Majumder, P., 2011. A novel statistical algorithm for enhancing the utility of HapMap data to design genomic association studies in non-HapMap populations. *International Journal of Data Mining and Bioinformatics* 5, 706–16.
- Stephens, M., Donnelly, P., 2003. A comparison of bayesian methods for haplotype reconstruction from population genotype data. *American Journal of Human Genetics* 73, 1162–9.
- Stern, M.C., Lin, J., Figueroa, J.D., Kelsey, K.T., Kiltie, A.E., Yuan, J.-M., Matullo, G., Fletcher, T., Benhamou, S., Taylor, J.A., Placidi, D., Zhang, Z.-F., Steineck, G., Rothman, N., Kogevinas, M., Silverman, D., Malats, N., Chanock, S., Wu, X., Karagas, M.R., Andrew, A.S., Nelson, H.H., Bishop, D.T., Sak, S.C., Choudhury, A., Barrett, J.H., Elliot, F., Corral, R., Joshi, A.D., Gago-Dominguez, M., Cortessis, V.K., Xiang, Y.-B., Gao, Y.-T., Vineis, P., Sacerdote, C., Guarrera, S., Polidoro, S., Allione, A., Gurzau, E., Koppova, K., Kumar, R., Rudnai, P., Porru, S., Carta, A., Campagna, M., Arici, C., Park, S.S.L., Garcia-Closas, M., 2009. Polymorphisms in DNA repair genes, smoking, and bladder cancer risk: findings from the international consortium of bladder cancer. *Cancer Research* 69, 6857–64.
- Stern, M.C., Siegmund, K.D., Conti, D. V, Corral, R., Haile, R.W., 2006. *XRCC1*, *XRCC3*, and *XPB* polymorphisms as modifiers of the effect of smoking and alcohol on colorectal adenoma risk. *Cancer Epidemiology, Biomarkers & Prevention* 15, 2384–90.
- Strom, S.S., Estey, E.H., Otschoorn, U.M., Guillermo, G.-M., 2010. AML outcomes: role of nucleotide excision repair polymorphisms in intermediate risk patients. *Leukemia & Lymphoma* 51, 598–605.
- Sugasawa, K., 2011. Multiple DNA damage recognition factors involved in mammalian nucleotide excision repair. *Biochemistry* 76, 16–23.
- Sugasawa, K., Ng, J.M., Masutani, C., Iwai, S., van der Spek, P.J., Eker, A.P.M., Hanaoka, F., Bootsma, D., Hoeijmakers, J.H.J., 1998. Xeroderma pigmentosum group C protein complex is the initiator of global genome nucleotide excision repair. *Molecular Cell* 2, 223–32.

- Sugasawa, K., Okuda, Y., Saijo, M., Nishi, R., Matsuda, N., Chu, G., Mori, T., Iwai, S., Tanaka, Keiji, Tanaka, Kiyoji, Hanaoka, F., 2005. UV-induced ubiquitylation of XPC protein mediated by UV-DDB-ubiquitin ligase complex. *Cell* 121, 387–400.
- Sun, X., Li, F., Sun, N., Shukui, Q., Baoan, C., Jifeng, F., Lu, C., Zuhong, L., Hongyan, C., YuanDong, C., Jiazhong, J., Yingfeng, Z., 2009. Polymorphisms in *XRCC1* and *XPG* and response to platinum-based chemotherapy in advanced non-small cell lung cancer patients. *Lung Cancer* 65, 230–6.
- Takeuchi, F., McGinnis, R., Bourgeois, S., Barnes, C., Eriksson, N., Soranzo, N., Whittaker, P., Ranganath, V., Kumanduri, V., McLaren, W., Holm, L., Lindh, J., Rane, A., Wadelius, M., Deloukas, P., 2009. A genome-wide association study confirms *VKORC1*, *CYP2C9*, and *CYP4F2* as principal genetic determinants of warfarin dose. *PLoS Genetics* 5, e1000433 (1–9).
- Tamura, K., Dudley, J., Nei, M., Kumar, S., 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Molecular Biology & Evolution* 24, 1596–99.
- Tchounwou, P.B., Yedjou, C.G., Patlolla, A.K., Sutton, D.J., 2012. Heavy metal toxicity and the environment. *EXS* 101, 133–64.
- Ter-Minassian, M., Asomaning, K., Zhao, Y., Chen, F., Su, L., Carmella, S.G., Lin, X., Hecht, S.S., Christiani, D.C., 2012. Genetic variability in the metabolism of the tobacco-specific nitrosamine 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone (NNK) to 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanol (NNAL). *Journal International du Cancer* 130, 1338–46.
- Thomas, F.J., McLeod, H.L., Watters, J.W., 2004. Pharmacogenomics: the influence of genomic variation on drug response. *Current Topics in Medicinal Chemistry* 4, 1399–409.
- Thorgeirsson, T.E., Stefansson, K., 2010. Commentary: gene-environment interactions and smoking-related cancers. *International Journal of Epidemiology* 39, 577–9.
- Tian, C., Kosoy, R., Nassir, R., Lee, A., Villoslada, P., Klareskog, L., Hammarström, L., Garchon, H., Pulver, A.E., Ransom, M., Gregersen, P.K., Seldin, M.F., 2009. European population genetic substructure: further definition of ancestry informative markers for distinguishing among diverse European ethnic groups. *Molecular Medicine* 15, 371–83.
- Trego, K.S., Turchi, J.J., 2006. Pre-steady-state binding of damaged DNA by XPC-hHR23B reveals a kinetic mechanism for damage discrimination. *Biochemistry* 45, 1961–9.

- Uchida, A., Sugasawa, K., Masutani, C., Dohmae, N., Araki, M., Yokoi, M., Ohkuma, Y., Hanaoka, F., 2002. The carboxy-terminal domain of the XPC protein plays a crucial role in nucleotide excision repair through interactions with transcription factor IIIH. *DNA Repair* 1, 449–461.
- Veenstra, D.L., You, J.H.S., Rieder, M.J., Farin, F.M., Wilkerson, H.-W., Blough, D.K., Cheng, G., Rettie, A.E., 2005. Association of *Vitamin K epoxide reductase complex I (VKORC1)* variants with warfarin dose in a Hong Kong Chinese patient population. *Pharmacogenetics & Genomics* 15, 687–91.
- Veglia, F., Matullo, G., Vineis, P., 2003. Bulky DNA adducts and risk of cancer: a meta-analysis. *Cancer Epidemiology, Biomarkers & Prevention* 12, 157–60.
- Vodicka, P., Kumar, R., Stetina, R., Sanyal, S., Soucek, P., Haufroid, V., Dusinska, M., Kuricova, M., Zamecnikova, M., Musak, L., Buchancova, J., Norppa, H., Hirvonen, A., Vodickova, L., Naccarati, A., Matousu, Z., Hemminki, K., 2004. Genetic polymorphisms in DNA repair genes and possible links with DNA repair rates, chromosomal aberrations and single-strand breaks in DNA. *Carcinogenesis* 25, 757–63.
- Vogel, U., Overvad, K., Wallin, H., Tjønneland, A., Nexø, B. a, Raaschou-Nielsen, O., 2005. Combinations of polymorphisms in *XPD*, *XPC* and *XPA* in relation to risk of lung cancer. *Cancer Letters* 222, 67–74.
- Wadelius, M., Chen, L.Y., Lindh, J.D., Eriksson, N., Ghori, M.J.R., Bumpstead, S., Holm, L., McGinnis, R., Rane, A., Deloukas, P., 2009. The largest prospective warfarin-treated cohort supports genetic forecasting. *Blood* 113, 784–92.
- Wang, F., He, Y., Guo, H., Li, J., Yang, Y., Wu, Z., Zheng, H., Wu, T., 2010. Genetic variants of nucleotide excision repair genes are associated with DNA damage in coke oven workers. *Cancer Epidemiology, Biomarkers & Prevention* 19, 211–8.
- Wang, H.-T., Hu, Y., Tong, D., Huang, J., Gu, L., Wu, X.-R., Chung, F.-L., Li, G.-M., Tang, M., 2012. Effect of carcinogenic acrolein on DNA repair and mutagenic susceptibility. *Journal of Biological Chemistry* 287, 12379–86.
- Wang, L.-E., Gorlova, O.Y., Ying, J., Qiao, Y., Weng, S.-F., Lee, A.T., Gregersen, P.K., Spitz, M.R., Amos, C.I., Wei, Q., 2013. Genome-wide association study reveals novel genetic determinants of DNA repair capacity in lung cancer. *Cancer Research* 73, 256–64.
- Wang, N., Dong, X.-J., Zhou, R.-M., Guo, W., Zhang, X.-J., Li, Y., 2009. An investigation on the polymorphisms of two DNA repair genes and susceptibility to ESCC and GCA of high-incidence region in northern China. *Molecular Biology Reports* 36, 357–64.

- Wang, Y., 2008. Bulky DNA lesions induced by reactive oxygen species. *Chemical Research in Toxicology* 21, 276–81.
- Wang, Y., Spitz, M.R., Lee, J.J., Huang, M., Lippman, S.M., Wu, X., 2007. Nucleotide excision repair pathway genes and oral premalignant lesions. *Clinical Cancer Research* 13, 3753–8.
- Wang, Z., Yao, H., Lin, S., Zhu, X., Shen, Z., Lu, G., Poon, W.S., Xie, D., Lin, M.C., Kung, H., 2013. Transcriptional and epigenetic regulation of human microRNAs. *Cancer Letters* 331, 1–10.
- Wei, Q., Cheng, L., Hong, W.K., Spitz, M.R., 1996. Reduced DNA repair capacity in lung cancer patients. *Cancer Research* 56, 4103–7.
- Weiss, J.M., Weiss, N.S., Ulrich, C.M., Doherty, J.A., Chen, C., 2006. Nucleotide excision repair genotype and the incidence of endometrial cancer: effect of other risk factors on the association. *Gynecologic Oncology* 103, 891–6.
- Weiss, J.N., Karma, A., MacLellan, W.R., Deng, M., Rau, C.D., Rees, C.M., Wang, J., Wisniewski, N., Eskin, E., Horvath, S., Qu, Z., Wang, Y., Lusic, A.J., 2012. “Good enough solutions” and the genetics of complex diseases. *Circulation Research* 111, 493–504.
- Wickliffe, J.K., Galbert, L.A., Ammenheuser, M.M., Herring, S.M., Xie, J., Masters, O.E., Friedberg, E.C., Lloyd, R.S., Ward, J.B., 2006. 3,4-Epoxy-1-butene, a reactive metabolite of 1,3-butadiene, induces somatic mutations in *Xpc*-null mice. *Environmental & Molecular Mutagenesis* 47, 67–70.
- Woo, S.W., Kang, T.S., Park, H.J., Lee, J.-E., Roh, J., 2009. Comparison of linkage disequilibrium patterns and haplotype structure of eight single nucleotide polymorphisms across the *CYP1A2* gene between the Korean, and other populations registered in the International HapMap database. *Journal of Clinical Pharmacy and Therapeutics* 34, 429–36.
- Wu, H.-C., Chang, C.-H., Tsai, R.-Y., Lin, C.-H., Wang, R.-F., Tsai, C.-W., Chen, K.-B., Yao, C.-H., Chiu, C.-F., Bau, D.-T., Lin, C.-C., 2010. Significant association of methylenetetrahydrofolate reductase single nucleotide polymorphisms with prostate cancer susceptibility in taiwan. *Anticancer Research* 30, 3573–7.
- Wu, X., Delclos, G.L., Annegers, J.F., Bondy, M.L., Honn, S.E., Henry, B., Hsu, T., Spitz, M.R., 1995. A case-control study of wood dust exposure, mutagen sensitivity, and lung cancer risk. *Cancer Epidemiology, Biomarkers & Prevention* 4, 583–8.
- Wu, X., Gu, J., Grossman, H.B., Amos, C.I., Etzel, C., Huang, M., Zhang, Q., Millikan, R.E., Lerner, S., Dinney, C.P., Spitz, M.R., 2006. Bladder cancer predisposition: a

- multigenic approach to DNA-repair and cell-cycle-control genes. *American Journal of Human Genetics* 78, 464–79.
- Wu, X., Ye, Y., Rosell, R., Amos, C.I., Stewart, D.J., Hildebrandt, M.A.T., Roth, J.A., Minna, J.D., Gu, J., Lin, J., Buch, S.C., Nukui, T., Ramirez Serrano, J.L., Taron, M., Cassidy, A., Lu, C., Chang, J.Y., Lippman, S.M., Hong, W.K., Spitz, M.R., Romkes, M., Yang, P., 2011. Genome-wide association study of survival in non-small cell lung cancer patients receiving platinum-based chemotherapy. *Journal of the National Cancer Institute* 103, 817–25.
- Yan, L., Yanan, D., Donglan, S., Na, W., Rongmiao, Z., Zhifeng, C., 2009. Polymorphisms of XRCC1 gene and risk of gastric cardiac adenocarcinoma. *Diseases of the Esophagus* 22, 396–401.
- Yang, Z., 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. *Computer Applications in the Biosciences* 13, 555–6.
- Yen, C.-Y., Liu, S.-Y., Chen, C.-H., Tseng, H.-F., Chuang, L.-Y., Yang, C.-H., Lin, Y.-C., Wen, C.-H., Chiang, W.-F., Ho, C.-H., Chen, H.-C., Wang, S.-T., Lin, C.-W., Chang, H.-W., 2008. Combinational polymorphisms of four DNA repair genes *XRCC1*, *XRCC2*, *XRCC3*, and *XRCC4* and their association with oral cancer in Taiwan. *Journal of Oral Pathology & Medicine* 37, 271–7.
- You, J.-S., Wang, M., Lee, S.-H., 2003. Biochemical analysis of the damage recognition process in nucleotide excision repair. *Journal of Biological Chemistry* 278, 7476–7485.
- Zabaleta, J., Schneider, B.G., Ryckman, K., Hooper, P.F., Camargo, M.C., Piazuolo, M.B., Sierra, R.A., Fonham, E.T.H., Correa, P., Williams, S.M., Ochoa, A.C., 2008. Ethnic differences in cytokine gene polymorphisms: potential implications for cancer development. *Cancer Immunology, Immunotherapy* 57, 107–14.
- Zeegers, M.P., Tan, F.E.S., Dorant, E., van Den Brandt, P.A., 2000. The impact of characteristics of cigarette smoking on urinary tract cancer risk: a meta-analysis of epidemiologic studies. *Cancer* 89, 630–9.
- Zelle, B., Lohman, P., 1979. Repair of UV-endonuclease-susceptible sites in the 7 complementation groups of xeroderma pigmentosum A through G. *Mutation Research* 62, 363–8.
- Zhang, L., Bacares, R., Boyar, S., Hudis, C., Nafa, K., Offit, K., 2009. cDNA analysis demonstrates that the *BRC42* intronic variant IVS4-12del5 is a deleterious mutation. *Mutation Research* 663, 84–9.
- Zhang, L., Gao, G., Li, X., Ren, S., Li, A., Xu, J., Zhang, J., Zhou, C., 2012. Association between single nucleotide polymorphisms (SNPs) and toxicity of advanced non-

- small-cell lung cancer patients treated with chemotherapy. *PloS One* 7, e48350 (1–6).
- Zhang, X., Miao, X., Liang, G., Hao, B., Wang, Y., Tan, W., Li, Y., Guo, Y., He, F., Wei, Q., Lin, D., 2005. Polymorphisms in DNA base excision repair genes *ADPRT* and *XRCC1* and risk of lung cancer. *Cancer Research* 65, 722–6.
- Zhou, B., Wu, H., Li, W., Liu, W., Luo, D., 2011. Microarray analysis of microRNA expression in skin of *Xpc*^{+/-} mice and wild-type mice. *Irish Journal of Medical Science* 180, 721–6.
- Zhu, X., Sun, X., Chen, B., Sun, N., Cheng, H.-Y., Li, F., Zhang, H., Feng, J., Qin, S., Cheng, L., Lu, Z., Chen, H., 2010. *XPC* Lys939Gln polymorphism is associated with the decreased response to platinum based chemotherapy in advanced non-small-cell lung cancer. *Chinese Medical Journal* 123, 3427–32.
- Zhu, Y., Lai, M., Yang, H., Lin, J., Huang, M., Grossman, H.B., Dinney, C.P., Wu, X., 2007. Genotypes, haplotypes and diplotypes of *XPC* and risk of bladder cancer. *Carcinogenesis* 28, 698–703.
- Zhu, Y., Yang, H., Chen, Q., Lin, J., Grossman, H.B., Dinney, C.P., Wu, X., Gu, J., 2008. Modulation of DNA damage/DNA repair capacity by *XPC* polymorphisms. *DNA Repair* 7, 141–8.
- Zuker, M., 2003. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Research* 31, 3406–15.

Curriculum Vita

NAME: Catherine Michelle Rondelli

DATE: 7/15/2013

PRESENT POSITION AND ADDRESS:

Doctoral Candidate
Neuroscience and Cell Biology Department NIH Ruth L Kirschstein NSRA F31
Predoctoral Fellow for NIEHS
301 University Blvd.
11.108 Medical Research Building, Rout 1062
Galveston, Texas 77555
cmrondel@utmb.edu

BIOGRAPHICAL:

Born: 09/15/1977 St. Joseph, Michigan
US Citizen
6605 GolfCrest Dr.
Galveston, Texas 77551

EDUCATION:

College:

2006-pres. University of Texas Medical Branch (Galveston, TX)
BBSC/GSBS Graduate Student
PhD candidate
Expected Graduation August 2013
Cell Biology Graduate Program

1995-1999 Michigan State University (East Lansing, MI)
Bachelor of Science
May 1999
Lyman Briggs School
Biochemistry, specialization Biotechnology, honors coursework

Additional Coursework/Continuing Education:

2003-2005 Michigan State University (East Lansing, MI)
Lifelong Continuing Education Coursework

1999-2000 Michigan State University (East Lansing, MI)
Graduate Level Coursework

1994, 1996 University of Notre Dame (South Bend, IN)
General Undergraduate Level Coursework

PROFESSIONAL EXPERIENCE:

2006-present Graduate Student
Environmental Toxicology Fellow
Department of Neuroscience and Cell Biology
University of Texas Medical Branch, Galveston, TX

2006-2012 Student Lecturer
Student Course Administration
Special Topics in Toxicology for Pathology
Texas A&M University at Galveston, Pelican Island, TX

2011 Student Lecturer
Center for Career Development Fellowship Writing Workshop Series
University of Texas Medical Branch, Galveston, TX

2007-2009 GSBS tutor
Basic Biological Sciences Curriculum
University of Texas Medical Branch, Galveston, TX

2004- 2006 Laboratory Technician
with some Administration Duties for Program Project Grant
Department of Pharmacology and Toxicology
Michigan State University, East Lansing, MI

2003 Laboratory Technician
Department of Microbiology and Molecular Biology
Michigan State University, East Lansing, MI

2000-2003 Laboratory Technician
with some Administration Duties
Department of Pharmacology and Toxicology
Michigan State University, East Lansing, MI

2000 Laboratory Technician
Plant Research Laboratory
Michigan State University, East Lansing, MI

1999-2000 Graduate Rotations/Post- Baccalaureate Studies
Department of Biochemistry
Michigan State University, East Lansing, MI

1999 Intern
Research and Engineering
Whirlpool Corporation, Benton Harbor, MI

1995-1999 Undergraduate Research Student
Michigan Space Grant Consortium Undergraduate Fellow

Department of Biochemistry
NIH Mass Spectrometry Facility
Michigan State University, East Lansing, MI

1996 Part Time Intern
Corporate Information Technology
Whirlpool Corporation, Benton Harbor, MI

RESEARCH ACTIVITIES:

Current: 2007-present

Description:

Dr. Sherif Abdel-Rahman's laboratory focuses on understanding how inherited genetic characteristics influence the susceptibility of individuals to environmental agents. My work focuses on the underlying genetic variation combinations of the DNA repair gene XPC, which detects damaged DNA and initiates NER. My work encompasses specific advanced bioinformatics techniques for data extraction and modeling, molecular biology analysis, basic biochemical analysis, and statistical analysis.

Grant Support:

1. **Principle Investigator:** NIEHS NIH Ruth L Kirschstein NSRA F31 ES019053-01 "XPC Haplotypes Alter DNA Repair Capacity and Levels of Genetic Damage" \$83,140; 4/15/10-4/15/13
2. **Awarded:** NIEHS T32 predoctoral fellowship; "XPC Haplotypes Alter DNA Repair Capacity and Levels of Genetic Damage" tuition and stipend; 8/2007-8/2009

Previous:

Descriptions:

2003-2006: Focused on understanding the underlying differences between arteries and veins using normal and experimentally hypertensive rats. Identified differences in biochemical makeup and physiological responses of isolated arteries and veins. My work encompassed using contractility as physiological model with ex vivo activation and inhibition, biochemical analysis of tissues, microscopy analysis, basic molecular biology analysis, and statistical analysis.

2003: Focused on using the sequences found in the *Gallus Gallus* genome to model vertebrate genome evolution. My work for the functional genomics studies included molecular biology analysis and data analysis for contig assembly.

2000-2003: Focused on understanding the mechanisms underlying hepatotoxic responses to inflammation and environmental toxicants. My work encompassed using both primary cells and rat models for hepatotoxicity for both in vitro and in vivo activation and inhibition, biochemical analysis of pathway endpoints and expected components, and statistical analysis.

2000: Focused on the biochemical assembly of plant cell walls. My work looked at the plant fucosyltransferases involved in cell wall biogenesis and Golgi trafficking using cloning techniques, molecular biology analysis and biochemical analysis.

1997-1999: Focused on the analysis of mechanisms required for plant nutrient transport using the drought resistant plant *Wollastonia bioflora*. My work encompassed protein chemistry analysis and biochemical analysis with focus on a novel cofactors for the methyltransferase involved in S-methylmethionine generation.

Broad Areas of Interest:

1. Mechanisms of action of environmental contaminants.
2. Analysis of sub-populations responses to environmental contaminants.
3. Mechanistic analysis of non-sensitive sub-populations to environmental contaminants.
4. Possible bioremediation strategies for contaminated sites.
5. The mechanisms of action for failures of bioremediation for environmental contaminants and the toxicity of byproducts of the process(es).

COMMITTEE RESPONSIBILITIES:

External:

2013-present Graduate Student Representative to the Hispanic Organization of Toxicologists for the national Society of Toxicology

UTMB:

2011 Student Lecturer for Center for Career Development Fellowship Writing Workshop series

2010-present Student Representative to the Toxicology Training Program Curriculum Committee

2010-2011 Secretary of UTMB's Society of Cell Biology Graduate Student Organization

2009-present Sealy Center for Cancer Biology Journal Club member

2009-2010 President of UTMB's Society of Cell Biology Graduate Student Organization

2008-2009 Vice President of UTMB's Society of Cell Biology Graduate Student Organization

2007-present Member of UTMB's Society of Cell Biology Graduate Student Organization

2007-present Student Guide and Aid for Prospective Graduate Students

2006-2012 Student Member of Team Teaching Toxicology for Special Topics in Toxicology at Texas A&M University at Galveston – various positions of direct course administration

- lecture schedule reminders
- grade book keeping
- term report committee
- project team
- exam leader

2006-present Toxicology Journal Club member
2006-present GSO Member and Participant

MSU:

1996-2001 MSU Nanotechnology Discussion Group Member
1999-2000 Staff Representative for Weekly Open Extemporaneous Acting
1999 Human Genome Project Community Forum Member
1998-1999 MSU Residence Housing Association Representative
1996 Member of Welcome Weekend Holmes Hall
1995-1999 MSU Honors College Member

TEACHING RESPONSIBILITIES AT UTMB:

2011 Student Lecturer for Center for Career Development Fellowship Writing
Workshop series
2008-present Rotation Student Laboratory Mentor
2007-2009 BBSC tutor
2006-2012 Teaching Toxicology for Special Topics in Toxicology at Texas A&M
University at Galveston
Respiratory Toxicology
Cardiovascular Toxicology
Cigarette Smoke as a Toxic Model
Smoking Hazards and Carcinogenicity
Mechanisms of Toxicology
Mechanisms of Tissue Injury

MEMBERSHIP IN SCIENTIFIC SOCIETIES:

2008-present Society of Toxicology
Hispanic Organization of Toxicologists
Women In Toxicology
Mechanisms Specialty Section
In vitro and Alternative Methods Specialty Section
Molecular Biology Specialty Section
2008-present Genetic Toxicology Association
1997-present American Chemical Society
1997-present American Association for the Advancement of Science

HONORS:

2013 Society of Toxicology Graduate Student Travel Award
2012 GSBS Associates Christina Fleischmann Travel Awards Award
2011 2nd Place Society of Toxicology Regional Chapter Poster Presentation
Award
2011 Mary Kanz Travel Award for Environmental Toxicology
2011 GSBS Associates Scholarship Award
2010-present NIEHS/NIH Ruth L Kirschstein NSRA F31 Predoctoral Scholarship
2010 Betty Williams PhD Tuition Scholarship Award
2010 James A. Hokanson, Ph.D. Endowed Scholarship Award

2009	Arthur V. Simmang Academic Scholarship Award
2008	Ann & John Hamilton Endowed Scholarship Award
2008	Genetic Toxicology Travel Award
2007-2009	NIEHS Environmental Toxicology UTMB Pre-doctoral Toxicology Training Grant
2006-2007	UTMB Toxicology Scholar Award winner
1999	Dean's List for Michigan State University
1998-1999	Michigan Space Grant Consortium Grant Fellowship
1998	National Science Foundation Summer Fellowship Program
1997-present	Golden Key National Honor Society member
1995	MSU Spartan Scholarship winner
1995	MSU Distinguished Freshman Minority Scholarship winner
1995	Air Force ROTC Scholarship winner

PUBLISHED:

ARTICLES IN PEER REVIEWED JOURNALS:

1. Wickliffe JK, Abdel-Rahman SZ, Lee C, Kormos-Hallberg C, Sood G, **Rondelli CM**, Grady JJ, Desnick RJ, Anderson KE. CYP1A2*1F and GSTM1 alleles are associated with susceptibility to porphyria cutanea tarda. *Mol Med.* 17:241-7; 2011.
2. **Rondelli CM**, El-Zein RA, Wickliffe JK, Etzel CJ, Abdel-Rahman SZ. A comprehensive haplotype analysis of the XPC genomic sequence reveals a cluster of variants associated with sensitivity to tobacco-smoke mutagens. *Tox. Sci.* 115:41-50; 2010.
3. Kinslow CJ, El-Zein RA, **Rondelli CM**, Hill CE, Wickliffe JK, Abdel-Rahman SZ. Regulatory regions responsive to oxidative stress in the promotor of the human DNA glycosylase gene NEIL2. *Mutagenesis* 25:171-7; 2009.
4. Watts SW, Thakali K, Smark C, **Rondelli C**, Fink GD. Big ET-1 processing into vasoactive peptides in arteries and veins. *Vascul. Pharmacol.* 47:302-12, 2007.
5. **Rondelli CM**, Szasz IT, Kayal A, Thakali K, Watson RE, Rovner AS, Eddinger TJ, Fink GD, Watts SW. Preferential Myosin Heavy Chain Isoform B Expression May Contribute to the Faster Velocity of Contraction in Veins versus Arteries. *J. Vasc. Res.* 44:264-272; 2007.
6. Watts SW, **Rondelli C**, Thakali K, Li X, Uhal B, Pervaiz MH, Watson RE, Fink GD. Morphological and biochemical characterization of remodeling in aorta and vena cava of DOCA-salt hypertensive rats. *Am. J. Physiol. Heart Circ. Physiol.* 292:H2438-48; 2007.
7. Hillier LW, Miller W, Birney E, Warren W, Hardison RC, Ponting CP, Bork P, Burt DW, Groenen MAM, Delany ME, Dodgson JB, Chinwalla AT, Cliften PF, Clifton SW, Delehaunty KD, Fronick C, Fulton RS, Graves TA, Kremitzki C, Layman D, Magrini V, McPherson JD, Miner TL, Minx P, Nash WE, Nhan MN, Nelson JO, Oddy LG, Pohl CS, Randall-Maher J, Smith SM, Wallis JW, Yang SP, Romanov MN, **Rondelli CM**, Paton B, Smith J, Morrice D, Daniels L, Tempest HG, Robertson L, Masabanda JS, Griffin DK, Vignal A, Fillon V, Jacobsson L, Kerje S, Andersson L, Crooijmans RPM, Aerts J, van der Poel JJ, Ellegren H, Caldwell RB, Hubbard SJ, Grafham DV, Kierzek AM, McLaren SR, Overton IM, Arakawa H, Beattie KJ, Bezzubov Y, Boardman PE, Bonfield JK, Croning MDR, Davies RM, Francis MD,

- Humphray SJ, Scott CE, Taylor RG, Tickle C, Brown WRA, Rogers J, Buerstedde JM, Wilson SA, Stubbs L, Ovcharenko I, Gordon L, Lucas S, Miller MM, Inoko H, Shiina T, Kaufman J, Salomonsen J, Skjoedt K, Wong GKS, Wang J, Liu B, Wang J, Yu J, Yang H, Nefedov M, Koriabine M, Dejong PJ, Goodstadt L, Webber C, Dickens NJ, Letunic I, Suyama M, Torrents D, von Mering C, Zdobnov EM, Makova K, Nekrutenko A, Elnitski L, Eswara P, King DC, Yang S, Tyekucheva S, Radakrishnan A, Harris RS, Chiaromonte F, Taylor J, He J, Rijnkels M, Griffiths-Jones S, Ureta-Vidal A, Hoffman MM, Severin J, Searle SMJ, Law AS, Speed D, Waddington D, Cheng Z, Tuzun E, Eichler E, Bao Z, Flicek P, Shteynberg DD, Brent MR, Bye JM, Huckle EJ, Chatterji S, Dewey C, Pachter L, Kouranov A, Mourelatos Z, Hatzigeorgiou AG, Paterson AH, Ivarie R, Brandstrom M, Axelsson E, Backstrom N, Berlin S, Webster MT, Pourquie O, Reymond A, Ucla C, Antonarakis SE, Long M, Emerson JJ, Betrán E, Dupanloup I, Kaessmann H, Hinrichs AS, Bejerano G, Furey TS, Harte RA, Raney B, Siepel A, Kent WJ, Haussler D, Eyraas E, Castelo R, Abril JF, Castellano S, Camara F, Parra G, Guigo R, Bourque G, Tesler G, Pevzner PA, Smit A, Fulton LA, Mardis ER, Wilson RK. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature* 432:695-716; 2004.
8. Copple BL, **Rondelli CM**, Maddox JF, Hoglen NC, Ganey PE, Roth RA. Modes of cell death in rat liver after monocrotaline exposure. *Toxicol Sci.* 77:172-82; 2004.

ABSTRACTS

1. **Rondelli CM**, Etzel CJ, Cross C, Xu M, Abdel-Rahman SZ. XPC haplotypes alter nucleotide excision repair and levels of UV-induced genetic damage. Abstract ID 1691. *The Toxicologist: An official journal of the Society of Toxicology*, vol 132, #1, March 2013.
2. Cross CE, **Rondelli CM**, Xu M, Abdel-Rahman SZ. Effect of increased reactive oxygen species on microRNA expression in first trimester placental trophoblasts. Abstract ID 548. *The Toxicologist: An official journal of the Society of Toxicology*, vol 132, #1, March 2013.
3. Xu M, Nekhayeva I, Cross CE, **Rondelli C**, and Abdel-Rahman SZ. MGMT haplotypes alter MGMT expression and affect response to alkylating agents. Abstract ID 195. *The Toxicologist: An official journal of the Society of Toxicology*, vol 132, #1, March 2013.
4. **Rondelli CM**, Etzel C, Abdel-Rahman SZ. DNA repair capacity is modulated by XPC haplotypes. Abstract ID 637. *The Toxicologist: An official journal of the Society of Toxicology*, vol 126, #1, March 2012.
March 2012 51st Annual Society of Toxicology Meeting
5. **Rondelli CM**, Wickliffe JK, Abdel-Rahman SZ. Modulation of DNA repair capacity by XPC haplotypes. Abstract ID 1360. *The Toxicologist: An official journal of the Society of Toxicology*, vol 120, suppl. 2, March 2011.
March 2011 50th Annual Society of Toxicology Meeting
October 2011 GC-SOT Joint Regional Annual Meeting
6. **Rondelli CM**, Wickliffe JK, El-Zein RA, Etzel C, Abdel-Rahman SZ. Modulation of genetic damage and DNA repair capacity by genetic variations in the nucleotide

- excision repair gene XPC. Abstract ID 361. *The Toxicologist: An official journal of the Society of Toxicology*, vol 108, #1, March 2010.
- . March 2010 49th Annual Society of Toxicology Meeting
7. **Rondelli CM**, Wickliffe JK, El-Zein RA, Etzel C, Abdel-Rahman SZ. Haplotype analysis of the full XPC genomic sequence reveals a cluster of variants associated with sensitivity to the genotoxic effects of tobacco smoke. Abstract ID 1392. *The Toxicologist: An official journal of the Society of Toxicology*, vol 102, vol s-1, March 2009.

June 2009	5th Annual Summer NCB Research Retreat
April 2009	50th Annual NSRF at UTMB
March 2009	48th Annual Society of Toxicology Meeting
	September 2008 Genetic Toxicology Association Annual Approaches and Genotoxic Impurities
 8. Thakali K, **Rondelli C**, Fink GD, Watts SW. Chymase-Dependent Processing of Big ET-1 in Arteries but not Vein. *Hypertension* 48 (4), E44, 2006.
 9. Kayal A, Thakali K, **Rondelli CM**, Watson RE, Rovner AS, Fink GD, Watts SW. P-18: Myosin heavy chain B isoform is more predominant in rat veins than arteries. *Am J Hypertens* 18, 15A–15A, 2005.
 10. Szasz IT, **Rondelli CM**, Fink GD, Watts SW. Reactive oxygen species metabolism in veins and arteries from rat: why is it different? *FASEB J*, 20:A725 2006.
 11. **Rondelli C**, Shen TL, Gage D. Influence Of Plants On The Global Atmospheric Chemistry: Biosynthesis Of The Precursor Of Plant-Derived Dimethyl Sulfide. Michigan Space Grant Consortium 4th Annual Conference, Ann Arbor, Michigan, October, 1999.

UNPUBLISHED:

ARTICLES SUBMITTED TO PEER REVIEWED JOURNALS:

1. **Rondelli CM**, Etzel CJ, Cross CE, Abdel-Rahman SZ. XPC haplotypes alter DNA nucleotide excision repair and influence levels of genetic damage induced by exposure to low levels of UV radiation. *SUBMITTED*.
2. Xu M, Nekhayeva I, Cross CE, **Rondelli CM**, Wickliffe JK, Abdel-Rahman SZ. Influence of promoter/enhancer region haplotypes on *MGMT* transcriptional regulation: A potential biomarker for human sensitivity to alkylating agents. *SUBMITTED*.

Other Required Information

Permanent address: 6605 GolfCrest Dr., Galveston, TX, 77550

This dissertation was typed by Catherine Michelle Rondelli.