GENOTYPE-BY-SMOKING INTERACTION AND THE RISK OF ATHEROSCLEROSIS AND ITS CLINICAL SEQUELAE

Christy L. Avery

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the School of Public Health (Epidemiology).

Chapel Hill
2007

Approved by
Advisor: Kari North
Reader: Gerardo Heiss
Reader: Andrew Olshan
Reader: Charles Poole
Reader: Ethan Lange

ABSTRACT

Christy L. Avery: Genotype-by-smoking interaction and the risk of atherosclerosis and its clinical sequelae
(Under the direction of Dr. Kari North)

Although the association between cigarette smoking and atherosclerosis is well established, the mechanisms by which smoking initiates vascular disease remain poorly understood. As heritable differences in DNA repair ability can influence the effect of environmental exposures such as cigarette smoke, we evaluated how 36 DNA repair variants from five genes (*XRCC1*, *APEX1, hOgg1, XPD,* and *XRCC3*) modified the association between ever-smoking and two atherosclerosis outcomes in Atherosclerosis Risk in Communities (ARIC) Study participants: intimal-medial thickness (IMT) and incident coronary heart disease (CHD).

The incident CHD analysis was conducted using all cases 1987-1998 (N=1,086) and a random sample (N=1,065) selected from the entire ARIC cohort at baseline (cohort random sample, CRS). Incidence rate ratios were estimated by piecewise constant models and departures from additivity were measured with interaction contrast ratios. When priors for genetic and environmental effects were added to the first-stage model, tagSNPs rs3213282 (*XRCC1*), rs50871 (*XPD*), and rs3212024 (*XRCC3*) were associated with an increase in the estimated effect of ever-smoking on incident CHD while tagSNPs rs1799782 (*XRCC1*) and rs861531 (*XRCC3*) were associated with a decrease.

We also evaluated the association between DNA repair variants, cigarette smoking, and baseline mean IMT using linear regression models in ARIC participants selected into the CRS. When priors for genetic and environmental effects were added to the first-stage linear regression model, tagSNPs rs3213282 (*XRCC1*), rs3213245 (*XRCC1*), rs3212024 (*XRCC3*), and rs3136814 (*APEX1*) were associated with increases in the estimated effect of ever-smoking on baseline mean IMT while tagSNPs rs3136817 (*APEX1*) and rs1799794 (*XRCC3*) were associated with decreases.

Few population-based studies examining the relationship between DNA repair variants, cigarette smoking and atherosclerosis have been published. Our results can stimulate inquiries into potential mechanisms linking cigarette smoke exposure and atherosclerotic diseases and help bridge the gap between observed trends and CHD biology. Future studies in different populations will undoubtedly be required to validate our results and improve our understanding of the complex relationships between DNA repair variants, cigarette smoking, and atherothrombotic disease.

ACKNOWLEDGEMENTS

I gratefully acknowledge my Committee Chair, Dr. Kari North, not only for her outstanding mentorship and untiring patience, but also her friendship. This work would not have been possible without contributions from the following committee members: Drs. Gerardo Heiss, Ethan Lange, Andrew Olshan, and Charles Poole. I have genuinely enjoyed working with them, and their input has been invaluable. I also wish to recognize the constant support I have received from my parents, Catherine and Michael Carr and Robert and Connie Avery, my grandparents Burt and Mary Palmer, my sister Jennifer Avery, as well as Nicholas Fettman and Caroline Hoffman.

TABLE OF CONTENTS

LIST OF TABLES

Page

LIST OF FIGURES

Page

LIST OF ABBREVIATIONS

AFD AFR	23 samples of African American descent from the Coriell Cell Repository were selected from the human variation panel of 50 African Americans. The samples were selected by the SeattleSNPs Program for Genomic Applications
AFD EUR	24 samples from the Coriell Cell Repository are primarily of European American descent. Samples consist of 23 unrelated CEPH parents selected by the SeattleSNPs Program for Genomic Applications, plus one sample from Coriell's human variation panel of 50 Caucasians
AFR1	Human individual DNA from 24 individuals of self-described African/African American heritage
AGI ASP	Samples from Coriell Cell Repositories Apparently Normal Collection of Caucasian and African-American females
APEX1	Apurinic/apyrimidinic endonuclease
AP	Apurinic/apyrimidinic
ARIC	Atherosclerosis Risk in Communities Study
BER	Base excision repair
BMI	Body mass index
BPDE	Benzo(a)pyrene diol epoxide
cDNA	Complementary DNA
CEPH	Centre d'Etude du Polymorphisms Human
CEU	30 mother-father-child trios from the CEPH collection (Utah residents with ancestry from northern and western Europe), representing one of the populations studied in the International HapMap project
CHD	Coronary heart disease
CI	Confidence interval
CK	Creatine kinase
CK-MB	Creatine kinase, mb fraction
CLR	Confidence limit ratio
CRS	Cohort representative sample
CVD	Cardiovascular disease
DAG	Directed acyclic graph
DNA	Deoxyribonucleic acid
DSB	Double strand break / recombination repair
EB	Empirical-Bayes
ECG	Echocardiogram (also abbreviated as EKG)
ETS	Environmental tobacco smoke
FHS	Framingham Heart Study
GGR	Global genomic repair
GscTr12003	British Phenotype: 96 BRCA1 and BRCA2 negative breast cancer index cases.
HCB	45 unrelated Han Chinese in Beijing, China, representing one of the populations studied in the International HapMap project
HDL	High-density lipoprotein
hOgg1	8 – hydroxy-2' – deoxyguanosine-glycosylase/apurinic lyase
HR	homologous recombination
HWE	Hardy-Weinberg equilibrium
ICR	Interaction contrast ratio
IHD	Ischemic heart disease
IMT	Intimal medial thickness

LDL	Low-density lipoprotein
LD	Linkage disequilibrium
LOH	Loss of heterozygosity
MAF	Minor allele frequency
MALDI-TOF	Matrix-assisted laser desorption/ ionization time-of-flight mass spectrometry
MI	Myocardial infarction
MMS	Methyl methanesulfonate
MONICA	Multinational **MONI**toring of trends and determinants in **CA**rdiovascular disease
MRFIT	Multiple Risk Factor Intervention Trial
NADPH	Nicotinamide adenine dinucleotide phosphase
NER	Nucleotide excision repair
NHEJ	Non-homologous end joining
NIHPDR	NIH Polymorphism Discovery Resource
NLS	Nuclear localization signal
NNK	Nitrosamine 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone
NO	Nitric oxide
PAD	Peripheral arterial disease
PAH	Polycyclic aromatic hydrocarbons
PDR90	The NIH Polymorphism Discovery Resource (NIHPDR) 90 individual screening subset
PVD	Peripheral vascular disease
polβ	DNA polymerase β
PARP	Poly ADP-ribose polymerase
ROS	Reactive oxygen species
SB	Semi-Bayes
SSBR	Single-strand break repair
SCD	Sudden cardiac death
SCE	Sister chromatid exchange
SMC	Smooth muscle cells
SNP	Single nucleotide polymorphism
SOD2	Superoxide dismutase
SSB	Single strand DNA break
TCR	Transcription-coupled repair
TFIIH	Transcription factor IIH
TIA	Transient ischemic attacks
WHO	World Health Organization
XP	Xeroderma pigmentosum
XRCC1	X-ray repair cross complementing, group 1
XRCC3	X-ray repair complementing defective repair in Chinese hamster cells 3
XPD	xeroderma pigmentosum D; excision repair cross-complementing rodent repair deficiency, complementation group 2 *ERCC2*
YRI	30 Yoruba mother-father-child trios in Ibadan, Nigeria, representing one of the populations studied in the International HapMap project
3-meA	3-methyladenine
8-oxo-G	7,8-dihydro-8-oxoguanine

CHAPTER I

INTRODUCTION

Although evidence linking cigarette smoking with atherosclerosis and its clinical sequelae is well established and consistent across age, sex, racial, and geographic strata[1-8], the mechanisms by which smoking initiates vascular disease are poorly understood. Cigarette smoke contains approximately 4,800 chemicals[9], of which more than 100 of the compounds have been identified as carcinogenic and/or mutagenic. Studies investigating the role of DNA damage in atherogenesis found higher levels of aromatic DNA adducts, stable complexes formed between reactive chemical species and sites within the DNA molecule, in vascular tissues (e.g. abdominal aorta and cardiac) of smokers[10, 11]. Experimental animal research also demonstrated that the tobacco smoke mutagens benzo(a)pyrene and 1,3-butadiene can induce and stimulate a proliferative vascular smooth muscle cell (SMC) phenotype[12, 13]. As differences in human toxicological responses to mutagen exposure have been attributed in part to heritable variation in DNA repair capacity[14], the identification of susceptibility genes that modify the relationship between cigarette smoking and atherosclerosis could provide new insight into the etiology of this major disease.

The present study, conducted under approval from the University of North Carolina at Chapel Hill Institutional Review Board (see Appendix A), addresses the dearth of population-based studies examining the relationship between DNA repair pathway variants, cigarette smoke exposure, and atherosclerosis and its clinical sequelae. Identifying genes that modify the relationship between cigarette smoke exposure and atherosclerosis and

associated clinical endpoints provides new opportunities to evaluate mechanistic laboratory models of CHD and further our understanding of the link between observed epidemiologic trends and CHD biology. No previous study has yet to perform a comprehensive analysis of the role of DNA repair genes with regards to CHD or subclinical atherosclerosis or considered their role as biologically plausible mediators of the effect of cigarette smoke. Here, we assess the relationship between cigarette smoking, DNA repair pathway variation, and two atherosclerosis endpoints (incident CHD and baseline IMT) using data from the ARIC Study, a community-based prospective investigation of 15,792 males and females. The two manuscripts prepared for fulfillment of the Epidemiology doctoral program requirements are as follows:

Manuscript 1: We conducted a series of case-cohort analyses to examine how variation in five DNA repair genes (*hOgg1, APEX1, XRCC1*, *XPD* and *XRCC3*) modified the association between ever-smoking and incident CHD in the ARIC cohort. All incident CHD cases 1987-98 (n=1,086) and a random sample (n=1,065) were selected from the entire cohort (n=15,792). Analyses were race-stratified and adjusted for age, sex, study center, alcohol intake, and physical activity. Incidence rate ratios (IRR) were estimated by piecewise constant exponential models and departures from additivity were measured with interaction contrast ratios (ICR). Hierarchical modeling was used to improve estimation by incorporating priors into models including all tagSNPs and models extended to examine gene-by-smoking interaction. This study addresses Aims 1, 2, and 3 of the dissertation (see Section II).

Manuscript 2: We examined how variation in five DNA repair genes (*hOgg1, APEX1, XRCC1*, *XPD* and *XRCC3*) modified the association between ever-smoking and baseline mean IMT in the ARIC cohort. A stratified random sample of 698 Caucasians and 367

African Americans was selected from all eligible participants (n=14,255). Analyses were race-stratified and adjusted for age, sex, study center, alcohol intake, and physical activity. Differences in baseline mean IMT were estimated using hierarchical linear models. This study addresses dissertation Aims 4, 5, and 6 (see Section II).

CHAPTER II
SPECIFIC AIMS

Our goal was to measure the extent to which common polymorphisms of five DNA repair genes (*XRCC1, XRCC3, XPD, hOGG1, APEX1*), in combination with cigarette smoke exposure, influence two CVD outcomes: incident CHD and subclinical atherosclerosis, as measured by IMT. The study is ancillary to the ARIC study, an ongoing, bi-racial population-based longitudinal investigation of cardiovascular and pulmonary disease in 15,792 males and females selected from four U.S. communities. Manuscript 1 addresses Aims 1, 2 and 3, and Manuscript 2 addresses Aims 4, 5, and 6.

The specific aims are:

1) To estimate the association between polymorphisms of the DNA repair genes *XRCC1, XRCC3, APEX1, hOgg1,* and *XPD* and incident CHD.
    a. *Single*-SNP analyses: Piecewise constant exponential regression modeling in which the association between each SNP and the rate of incident CHD is estimated.
2) To estimate the extent to which polymorphisms of the DNA repair genes *XRCC1, XRCC3, APEX1, hOgg1,* and *XPD* modify the association between cigarette smoking and incident CHD.

a. *Single*-SNP analyses: Piecewise constant exponential regression modeling in which DNA repair SNPs are evaluated as modifiers of the cigarette smoking – incident CHD association.

3) To incorporate information from multiple genes and cigarette smoke exposure as higher level priors into analyses investigating the relationship between DNA repair variants, cigarette smoking, and incident CHD.

   a. *Hierarchical modeling*: An extension of Aim 2 in which models are extended to incorporate higher-level priors.

4) To estimate the association between polymorphisms of the DNA repair genes *XRCC1, XRCC3, APEX1, hOgg1,* and *XPD* and baseline mean IMT.

   a. *Single*-SNP analyses: Linear regression modeling in which the association between each SNP and baseline mean IMT is estimated.

5) To estimate the extent to which polymorphisms of the DNA repair genes *XRCC1, XRCC3, APEX1, hOgg1,* and *XPD* modify the association between cigarette smoke exposure and baseline mean IMT.

   a. *Single*-SNP analyses: Linear regression modeling in which DNA repair SNPs are evaluated as modifiers of the tobacco exposure –baseline mean IMT association.

6) To incorporate information from multiple genes and cigarette smoke exposure as higher level priors into analyses investigating the relationship between DNA repair variants, cigarette smoking, and baseline mean IMT.

   a. *Hierarchical modeling*: An extension of Aim 5 in which models are extended to incorporate higher-level priors

CHAPTER III
BACKGROUND AND SIGNIFICANCE

A. Atherosclerosis

Atherosclerosis is a disease of the vasculature characterized by plaques in the innermost layer of the artery (atheromas) and is the main pathophysiological process responsible for cardiovascular diseases (CVD) such as myocardial and cerebral infarctions and peripheral vascular disease[15]. Atherosclerosis is an ancient disease, being detected in the arteries of Egyptian mummies[16] and recognized as a pathologic condition for >150 years[17]. Atherosclerosis is now virtually ubiquitous among Western populations[18].

The initiation of atherosclerosis begins early in life upon the inception of a diet rich in cholesterol and saturated fat. Fatty streaks, aggregations of lipid-rich leukocytes within the innermost layer of the artery (the intima), are precursors of atherosclerotic lesions and are typically present in children 10-14 years of age[19, 20]. While any artery may become affected, the aorta, coronary, carotid, iliac, and femoral arteries are the vessels most likely to develop atherosclerosis.

Pathophysiology of atherosclerosis

An intricate sequence of cellular events occurs during the initiation and evolution of an atherosclerotic plaque (Figure 1), which include 1) extracellular lipid accumulation, 2) leukocyte recruitment, 3) intracellular lipid accumulation and foam cell formation, 4) smooth

muscle cell migration, proliferation, and development of the arterial extracellular matrix, and 5) plaque angiogenesis.

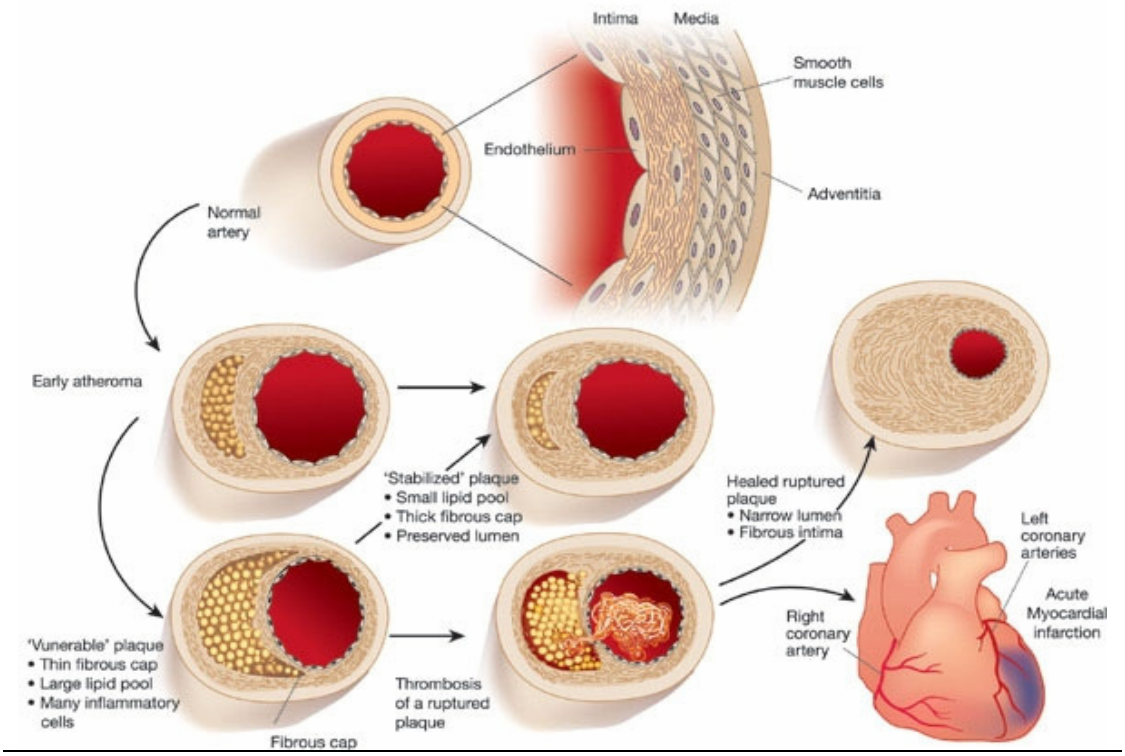
Briefly, the first stage of atherogenesis, while somewhat conjectural[21], is believed to involve the accumulation and retention of low-density lipoprotein (LDL) molecules in the intima, usually after initiation of a diet high in fat and cholesterol[22]. Once bound and thus retained by proteoglycans in the intima, the lipoproteins exhibit increased susceptibility to chemical modification and oxidation. The modified lipids are able to induce the expression of adhesion molecules, chemokines, and proinflammatory cytokines in macrophages and vascular wall cells, thus favoring the recruitment and retention of leukocytes. Oxidized lipids also have cytotoxic effects on endothelial and smooth muscle cells[23], although the molecular mechanisms underlying LDL-mediated cell apoptosis are not fully understood[24].

The next morphologically identifiable event in atherogenesis is leukocyte recruitment and accumulation within the intima. Circulating monocytes adhere to the endothelium in clusters and diapedese (pass through intact capillary walls into surrounding body tissue) between endothelial cells and enter the intima. Once in the intima, monocytes are converted to macrophages that accumulate lipids and transform into foam cells. Foam cells, or lipid-laden macrophages[26], are apparent macroscopically as fatty streaks.

The evolution of the atheroma into a more complex plaque involves leukocyte production of chemoattractants that recruit SMCs from the medial layer of the artery (tunica media) to the intima. SMCs produce the main matrix macromolecules that accumulate in atheromas, including collagens and proteoglycans that facilitate plaque stability[27] and transform the fatty streak into a mature fibrofatty atheroma. Plaque microcirculation is

formed by endothelial cells that promote plaque growth by circumventing diffusion limitation of oxygen and nutrients[28].

Figure 1. Schematic of the natural history of an atheroma.



Adapted from Libby (2002)[25]

The development of an atherosclerotic plaque occurs over many years, during which the affected individual is typically devoid of symptoms. Arterial vessels can enlarge or constrict in size, referred to as geometric remodeling, in order to compensate for the expanding atheroma[17]. However, once the plaque burden surpasses the ability of the artery to remodel outward, encroachment on the arterial lumen and consequent narrowing ensues.

Generally stenoses exceeding 60% of the arterial lumen can cause blood flow limitations and ischemia during periods of increased cardiac demand[21]. While the clinical manifestations for lower-extremity peripheral arterial disease (PAD) include claudication and limb ischemia[29], transient ischemic attacks (TIA) and stroke are the principal clinical

presentations of carotid artery ischemia associated with obstructive atherosclerosis[30]. Obstructive lesions in the coronary arteries may produce angina, chest pain caused by transient myocardial ischemia or myocardial infarctions.

Atherosclerosis risk factors

Given that atherosclerosis is a generalized macrovascular disease, lesions in one vascular locale predict lesions in other areas. Likewise, risk factor profiles are similar among populations with coronary, peripheral, and carotid atherosclerosis[31]. Although upwards of 200 risk factors for atherosclerosis are recognized[32], numerous long-term population-based prospective studies have identified the following major risk factors discussed below: hyperlipidemia, elevated blood pressure, age, cigarette smoking, type 2 diabetes, obesity, male sex, physical activity, alcohol intake and family history.

Hyperlipidemia

Data from animal, epidemiologic, experimental pathology, and family-based studies, as well as randomized clinical trials of statins (or HMG-CoA reductase inhibitors, a class of hypolipidemic agents used to lower cholesterol levels), all support an association between increased serum cholesterol and atherosclerosis[33]. Approximately 7% of the body's cholesterol circulates in the plasma, predominantly in the form of LDL. LDL is considered the major atherogenic component of total cholesterol and chronic hyperlipidemia results in the accumulation of LDL in the intima. In contrast, an inverse relationship between high-density lipoprotein (HDL) and atherosclerosis has been established, as HDL mobilizes cholesterol from the atheroma and transports it to the liver for excretion[18].

Elevated blood pressure

Elevated blood pressure is believed to influence the development of atherosclerotic lesions in numerous ways. Animal studies suggest that elevated blood pressure induces the infiltration of monocytes and macrophages into the vessel walls of the kidney and heart[34-37]. Arterial remodeling, structural changes in the vessel wall, may also be associated with the hypertensive process. Arterial remodeling results in an increased lumen size and reduced arterial wall shear stress[38]. As vessels most vulnerable to the development of atherosclerosis are those experiencing conditions promoting a weak net hemodynamic shear stress[39], an increased lumen size may accelerate the atherosclerotic process.

Age

While atherosclerosis is typically not clinically apparent until middle age when lesions produce organ injury, the disease is manifested in childhood and progresses slowly over the following decades[40]. Death rates from CHD and ischemic stroke, two diseases greatly influenced by atherosclerosis, rise with each decade into advanced age. For example, the annual rates per 1,000 person-years for incident MIs in non-African American males are 19.2, 28.3, and, 50.6 for ages 65-74, 75-84, and 85 and older, respectively[41].

Cigarette smoking

Cigarette smoking causes numerous hemodynamic changes prompting the development of atherosclerosis, including coronary vasoconstriction and acute increases in blood pressure and heart rate. Nicotine in the blood also promotes arterial endothelial injury and prothrombotic changes, such as increased platelet adhesiveness and aggregation[42, 43], elevated fibrinogen[44, 45], and decreased fibrinolysis[46-48]. Chronic exposure to cigarette smoke has been shown to promote SMC proliferation. The association between cigarette smoking and atherosclerosis is further reviewed in chapter III.

Type 2 diabetes

Diabetes, reflecting a state of chronic hyperglycemia and resistance to the effects of insulin, is closely related to hypertension, obesity, and insulin resistance. Decreased insulin sensitivity results in substantial protein modification including the glycation of amino acid residues and glycated amino acids can stimulate proliferation of the fibromuscular components of atherosclerotic plaques[38]. Hyperglycemia also inhibits nitric oxide (NO) production by blocking *eNOS* synthase activation[49]. NO has many anti-atherogenic properties including inhibiting platelet activation, reducing leukocyte adhesion and migration, and diminishing vascular smooth muscle cell proliferation and migration[50-52], thus a reduction in NO promotes atherogenesis.

Obesity

Obesity and overweight influence the development of atherosclerosis through their effects on the systemic vasculature, endothelial function, and the vasomotor function of insulin. The elevated levels of free fatty acids observed in obesity blunt insulin-mediated glucose uptake and NO-dependent blood flow[53], in addition to inducing oxidative stress and proinflammatory signaling[54]. The increased expression of nicotinamide adenine dinucleotide phosphase (NADPH) oxidase associated with obesity also causes deregulated production of adipokines, fat cells that produce and secrete numerous hormones including adiponectin, PAI-1, and monocyte chemoattractant protein. NADPH oxidase inhibition has been shown to reduce ROS production and improve glucose metabolism[55].

Male sex

Males are much more prone to atherosclerosis and its ramifications than females, and at any given age the prevalence of CHD in males is higher than that in females[56]. For

example, the lifetime risk for developing CHD after age 40 is 49% for males and 32% for females[41]. MIs and other complications of atherosclerosis are uncommon in premenopausal females unless a predisposition to diabetes, hyperlipidemia, or severe hypertension exists[18].

Physical activity

Exercise influences metabolic and atheromatous processes, fibrinolytic activity, blood lipid patterns, oxygen uptake, BMI, myocardial function, pulse rate, and blood pressure. A reduced risk of CHD relates both to occupational and leisure-time physical activity, the relationship is dose-dependent, and the findings are consistent regardless of age, sex, and population studied[57]. Experiments in primates (*Macaca fascicularis*) also support the concept that long-term moderate exercise delays the development of atherosclerosis, despite the administration of an atherogenic diet[58].

Alcohol intake

Multiple prospective studies have reported an inverse association between low-to-moderate alcohol consumption and CHD[59-62]. Although the biologic mechanisms underlying the relationship between alcohol intake and the atherosclerosis endpoint CHD are not completely understood, experimental studies suggest that ethanol in any form increases HDL levels. Increases in HDL are thought to result from the effect of alcohol on hepatic production and secretion of apolipoproteins and the lipolysis of triglyceride-rich particles, which increase the transfer of cholesterol from very low density lipoproteins, a precursor of LDL, to HDL. Moderate alcohol consumption is also associated with decreased clotting proteins and platelet aggregability and increased fibrinolytic activity[63].

Family history

A familial contribution to atherosclerosis has been acknowledged for over 100 years when Sir William Osler recognized that angina could recur in families[64]. Now, hundreds of articles highlight the importance of family history in the prediction of CHD. Family history is considered an independent risk factor for CHD[65-69] and a surrogate for coronary risk factors[70, 71] .

The well-established familial predisposition to atherosclerosis is related both to the clustering of risk factors within families, such as hyperlipidemia and hypertension, and inherited genetic variation. Examples of monogenic causes of atherosclerosis include lipoprotein metabolism disorders such as familial hypercholesterolemia, which results in excessively high blood lipid levels in carriers of the variant alleles. However, atherosclerosis is most likely an oligogenic (in which a small number of loci with major effects determine disease susceptibility) or polygenic disease, wherein numerous genes, each with a small-to-moderate effect that may be modified by environmental exposures, influence the development and progression of the disease[72].

Coronary atherosclerosis

Coronary atherosclerosis refers to atherosclerosis affecting the sub-epicardial vasculature supplying blood to the myocardium. The result of advanced coronary atherosclerosis is myocardial ischemia, an imbalance reflecting an insufficiency of oxygen and a reduced availability of nutrients. Syndromes caused by reduced cardiac muscle blood supply, generically designated CHD or ischemic heart disease (IHD), include myocardial infarction (MI), sudden cardiac death, and angina pectoris.

Angina pectoris is an indicator of complex atherosclerosis typified by paroxysmal and at times recurrent attacks of visceral chest pain. Angina is caused by transitory (15 seconds to 15 minutes) myocardial ischemia that fails to induce the cellular necrosis that defines MI and is aggravated by increased myocardial demand and decreased myocardial perfusion[18].

Sudden cardiac death (SCD) is most commonly defined as death from cardiac causes within one hour after or without the onset of symptoms. SCD often is the first clinical manifestation of advanced coronary athersclerosis[73], with ventricular tachyarrhythmias accounting for the majority of these cases[74]. Approximately 90% of SCD victims have at least two coronary arteries with ≥ 90% occlusion[75].

MI, also known as heart attack, is the major underlying manifestation of CHD[18]. Although a MI may initially present as angina pectoris, the chest discomfort is usually more severe and prolonged. MIs are manifestations of atherosclerotic coronary disease complicated by plaque rupture and coronary thrombosis. If a thrombus limits or occludes blood flow in the vessel, ischemia develops. Failure to restore blood flow leads to myocardial necrosis, with acute damage followed by scarring and permanent injury to the heart muscle.

MIs diagnoses are typically informed by cardiac signs and symptoms, serum biomarkers and ECGs (i.e. echocardiograms, EKG). Chest, epigastric, neck, jaw, or arm pain are typical symptoms of acute MIs while cardiac signs include acute congestive heart failure or cardiogenic shock. ECGs are a graphic procedure that may indicate acute or previous myocardial damage or ischemia[76]. Although the first ECG is uninformative in approximately

50% of patients with an acute MI, the diagnostic yield increases substantially with each serial ECG[56].

Serum biomarkers of myocyte necrosis are also an important diagnostic tool for MI. Troponin, for example, enters the bloodstream four to eight hours after MI onset, and has nearly absolute myocardial tissue specificity. Increased troponin is conventionally defined as that which exceeds the 99th percentile of troponin values measured in a reference control group[76]. Additional biomarkers of myocardial necrosis include myoglobin and creatine kinase (CK), measured as either total CK or the MB fraction of CK (CK-MB). While biomarkers indicate myocardial damage, myocardial necrosis and MI are not necessarily synonymous, as elevated biomarkers in the absence of clinical evidence of MI may reflect another cause of myocardial injury, such as myocarditis, cardiac trauma, congestive heart failure, or renal failure[77].

CHD case definitions

Table 1. Classification of MI based on biomarker findings, ECG findings, and cardiac symptoms or signs.

	Biomarker Findings*							
	Cardiac Symptoms or Signs Present				Cardiac Symptoms or Signs Absent			
ECG Findings	Diagnostic	Equivocal	Missing	Normal	Diagnostic	Equivocal	Missing	Normal
Evolving diagnostic	Definite	Definite	Definite	Definite	Definite	Definite	Definite	Definite
Positive	Definite	Probable	Probable	No	Definite	Probable	Possible	No
Nonspecific	Definite	Possible	No	No	Definite	Possible	No	No
Normal or other ECG findings	Definite	Possible	No	No	Definite	No	No	No

Definite indicates definite MI; Probable, probable MI; Possible, possible MI; and No, no MI. Classification of case is at highest level allowed by combinations of three characteristics. *Adapted from Luepker et al., 2003[22]

CHD case definitions for epidemiology studies and clinical trials were first based upon World Health Organization (WHO) (1959) and American Heart Association (1964)

reports, as well as the WHO European Acute Myocardial Infarction Registry criteria[78, 79]. While the protocols established by these reports have been widely used, variable interpretations have resulted in a lack of comparability between and within studies[22]. More recent criteria based on WHO MONICA (Multinational **MONI**toring of trends and determinants in **CA**rdiovascular disease) analyses[80] and surveillance studies such as the Lipid Research Clinics[81] further specified the original WHO CHD definition.

Table 2. Case classifications for CHD.

Event*	Characterization	Indication
1.Non-fatal events	A. Definite MI	1. Evolving diagnostic ECG
		2. Diagnostic biomarkers
	B. Probable MI	1. Positive ECG findings plus cardiac symptoms or signs plus missing biomarkers
		2. Positive ECG findings plus equivocal biomarkers
	C. Possible MI	1. Equivocal biomarkers plus nonspecific ECG findings
		2. Equivocal biomarkers plus cardiac symptoms or signs
		3. Missing biomarkers plus positive ECG
	D. Unrecognized MI	1. Appearance, in a non-acute setting, of a new diagnostic Q wave with or without ST-T-wave depression or ST elevation
	E. Medical procedure-related event	1. Cardiac events after (up to 28 days) a medical procedure (e.g., general surgery) with criteria for definite, probable, and possible MI identical to those described above
		2. May be reported separately as procedure-related cardiac events or combined with overall event rates.
		3. If the medical procedure was performed for the treatment of acute ischemia (e.g., angioplasty, coronary bypass surgery), an event should be classified as described above
	F. Unstable angina pectoris	1. New cardiac symptoms and positive ECG findings with normal biomarkers 2. Changing symptoms pattern and positive ECG findings with normal biomarkers
	G. Stable angina pectoris	1. Cardiac symptoms in a pattern that remains constant in presentation, frequency, character, and duration over time
2. Fatal events (hospitalized patients)	A. Definite fatal MI	1. Death within 28 days of hospital admission in 1A.
		2. Postmortem findings consistent with MI within 28 days
	B. Probable fatal MI	1. Death within 28 days of hospital admission in cases defined in I.B 2. Death within 6 hours of hospital admission with cardiac symptoms and/or signs. Other confirmatory data (biomarkers, ECG) are absent or not diagnostic.
	C. Possible fatal coronary event	1. Death within 28 days of hospital admission in cases defined in I.C, I.F, and I.G 2. Postmortem findings show old infarct and/or ≥ 50% atherosclerotic narrowing of coronary arteries.

*Adapted from Luepker et al., 2003[22]

The availability of new diagnostic tests, namely improved biomarkers of myocardial injury, shifting disease presentation, and an increasing number of survivors prompted further improvements in CHD case definitions. Thus, the current CHD consensus criteria was developed in 2003 by scientists representing the American Heart Association, the World Heart Federation Council on Epidemiology and Prevention, the European Society of Cardiology, the Centers for Disease Control and Prevention, the WHO, and the National Heart, Lung and Blood Institute. As shown in Table 1 and Table 2, the current definition of CHD is based on symptoms and signs, biomarkers, ECG, and/or autopsy findings. The extent and quality of the available data is then used to define definite, probable, and possible cases of fatal and nonfatal MI, procedure-related events, and angina pectoris[22]. The specific criteria used by ARIC investigators for the classification of incident CHD is discussed in chapter IV.

Epidemiology of CHD

CHD poses a substantial public health burden, as it is the main cause of death in Western societies and has been predicted to remain so for future decades[82]. Framingham Heart Study (FHS) investigators estimated a lifetime risk of developing CHD after age 40 of 49% for males and 32% for females[83]. Current estimates indicate that 565,000 Americans will experience their first MI in 2005, whereas 300,000 will experience a recurrent MI[41].

Age-adjusted CHD incidence rate estimates (95% confidence interval (CI)) per 1,000 person-years for Caucasian male, African-American male, Caucasian female, and African American female ARIC participants were 12.5 (11.5 – 13.7), 10.6 (8.9 - 12.7), 4.0 (3.5 - 4.6), and 5.1 (4.2 – 6.2), respectively[84]. Among American Indian male and female Strong Heart Study participants, incidence rates for CHD mortality per 1,000 person-years were 8.0 (6.1,

10.0) and 3.3 (2.3, 4.3) respectively, almost two-times as high as those estimated in the ARIC Study[85]. Annual CHD incidence rates (per 1,000 person-years) over ten years of follow-up among males of Japanese descent residing on Oahu in 1965 (Honolulu Heart Program) were 4.6, 6.0, 7.2, 8.8, and 10.5 for ages 45-49, 50-54, 55-59, 60-64, and 65-68, respectively[86].

As indicated by the above estimates, CHD imparts a substantial burden on the United States' health care system. Coronary atherosclerosis and acute MI were the two most expensive diagnoses treated in US hospitals in 2002, costing $38.4 and $27.8 billion dollars in healthcare costs respectively. Overall, CHD accounts for $142.1 billion dollars annually in direct and indirect costs (http://www.hcup.ahrq.gov/) and this estimate will only increase as the U.S. population ages.

Atherosclerosis paradigms

While the proliferation of SMCs in the intima is a fundamental mechanism in the pathophysiology of atherosclerosis[87], there is disagreement on the exact role SMC proliferation plays in the development of atherosclerosis. One paradigm, the "response to injury" or "inflammation" hypothesis, posits that the joint action of growth factors, proteolytic agents, and extracellular matrix molecules, produced by a dysfunctional endothelium and inflammatory cells, induces SMC migration from the media and their consequent proliferation in the intima[88]. Thus, initiating factors are cytokines and growth factors and SMC proliferation is only a reactive process[88, 89].

Another theory, the "monoclonal" hypothesis, contends that media SMCs can experience phenotypic modulation and that a predisposed SMC population, or even a single cell, is responsible for consequent SMC proliferation and intimal thickening[88]. Introduced in

1974 by Benditt and Benditt[90] upon the discovery that atherosclerotic plaques have features of a monoclonal lesion, a finding that has been verified by several laboratories[91-93], the monoclonal hypothesis suggest that the initiation of atherosclerotic plaques requires a mutation or viral infection that transforms a single SMC into the progenitor of a proliferative clone, analogous to the evolution of neoplastic cells[90]. Likewise, an increased mutation rate and extensive microsatellite instability has been reported in human atherosclerotic plaques[94, 95]. DNA extracted from atherosclerotic plaques also had a transforming ability when transfected into NIH3T3 cells[96] and SMCs cultured from plaques retain transforming potential throughout many cell generations[97].

Rather than alternatives, the response to injury and monoclonal hypotheses of atherogenesis may be complementary. Initial events leading to plaque formation may reflect the "response to injury" hypothesis, whereas clone formation and expansion, transforming an inflammatory process into a neoplastic process, requires a longer time. As a corollary the first stage of atheroma formation may be more readily reversible than the following phase involving clone formation and expansion[98].

B. Cigarette smoke exposure and atherosclerosis

One factor that ties the response to injury and monoclonal hypotheses of atherosclerosis together is exposure to mitotic / proliferative agents, for example compounds found in cigarette smoke. Cigarette smoke contains approximately 4,800 chemicals[9], of which more than 100 have been identified as carcinogenic and/or mutagenic[14], and has been associated with numerous clinical atherosclerotic symptoms including stable angina, acute coronary syndromes, sudden death, stroke, and aortic and peripheral atherosclerosis[99]. While evidence of increased risk for CHD associated with cigarette smoking is well

established and consistent across age, sex, racial, and geographic strata[1-8], the mechanisms by which smoking initiates vascular disease are poorly understood.

Cigarette smoke can be divided into two phases: a tar or particulate phase and a gas phase. The tar phase is arbitrarily defined as all material that is retained by a Cambridge glass-fiber filter which traps 99.9% of particles larger than 0.1 micron. All other material is considered the gas phase of cigarette smoke[100]. The tar phase of cigarette smoke contains $>10^{17}$ free radicals/gram, whereas the gas phase contains $>10^{15}$ free radicals per puff. Free radicals are species containing one or more unpaired electron. Examples of free radicals include the hydrogen atom (H·), nitric oxide (NO·), and nitrogen dioxide (NO_2·)[101]. Although the generation of free radicals occurs at a continuous low-level in the human body, an imbalance can cause DNA, protein, lipid, and carbohydrate damage[102, 103].

Free radicals contained in the tar phase are relatively stable, lasting hours to months, while radicals associated with the gas phase have a lifespan of seconds[104]. However, a paradox exists. Although individual free radicals in the gas phase of cigarette smoke have short lifetimes, free radical concentrations overall are sustained at high levels for > 10 minutes, actually increasing in concentration as smoke is aged[105, 106]. A possible explanation is that radicals in the gas phase of cigarette smoke exist in a steady state, being continuously formed and destroyed[105-108].

The pulmonary circulation is the first site exposed to the gas phase of cigarette smoke[109]. In addition to the high concentrations of free radicals present at proximal exposure sites, the gas phase of cigarette smoke contains many stable compounds capable of inducing the production of free radicals in vascular fields away from the primary exposure site[110]. For

example, α,β-unsaturated aldehydes, α,β –unsaturated ketones, and certain saturated aldehydes have been shown to react with thiol groups regulating two potentially inducible intracellular reserves of free radicals;[110] NADPH oxidase, an enzyme present in the vasculature, and xanthine oxidase, a form of the ubiquitous enzyme xanthine dehydrogenase[111, 112].

Each cigarette smoked also deposits upwards of 20 mg of tar in the lungs of smokers, or as much as one gram/day[104]. Tar contains > 5000 organic compounds, which are in contact with pulmonary fluids that extract the water-soluble components. Catechol and hydroquinone, two major components of cigarette tar, are water soluble and can produce free radicals[104]. Indeed, incubation of bacteriophage DNA with aqueous tar extracts produced single-strand DNA breaks[113, 114].

Cigarette smoke affects the initiation and progression of atherosclerosis through its effects on vasomotor dysfunction, inflammation, and lipid modification[115], factors that proceed any apparent structural and clinicopathologic disease manifestations[88, 116]. Nicotine, possibly the most-studies component of tobacco smoke, also has deleterious effects on the vasculature. Nicotine in the blood promotes arterial endothelial injury and prothrombotic changes, such as increased platelet adhesiveness and aggregation[42, 43], elevated fibrinogen[44, 45], and decreased fibrinolysis[46-48].

Studies in both human and animals also demonstrate that active and passive cigarette smoke exposure is associated with a diminished vasodilatory function[117-124]. Cigarette smoking is related to decreased NO availability[122, 123, 125], a free radical responsible for vasodilatory endothelial functioning that also helps regulate inflammation, leukocyte

adhesion, platelet activation, and thrombosis[126, 127]. Cigarette smoke is also associated with an increased level of inflammatory markers[128-131] and smokers have higher serum cholesterol, triglyceride, and LDL levels and lower HDL levels when compared to non-smokers[132].

Chronic exposure to cigarette smoke has also been shown to promote SMC proliferation. Experimental animal research showed that the tobacco smoke mutagens benzo(a)pyrene and 1,3-butadiene can induce and stimulate a proliferative vascular SMC phenotype[12, 13] and studies investigating the role of DNA damage in atherogenesis found higher levels of aromatic DNA adducts, stable complexes formed between reactive chemical species and sites within the DNA molecule, in vascular tissues (e.g. abdominal aorta and cardiac) of smokers[10, 11]. Plasma cotinine levels (as a measure of smoking behavior) were also predictive of bulky DNA adduct levels in humans[133]. Thus, the pattern of somatic DNA damage in atherosclerotic lesions may reflect the mutagenicity of tobacco smoke.

Mutagens found in cigarette smoke or the activated metabolites of cigarette smoke constituents may cause genetic damage by binding to or interacting with DNA. Interactions between mutagens and DNA can cause lesions or a disruption of the genetic structure resulting in gross chromosomal alterations such as aneuploidy, breaks, translocations, amplifications, or deletions. However, the successful binding of a mutagen to DNA does not always result in chromosomal damage, as pathways such as DNA repair must fail before a mutagenic event occurs[14]. The repair of DNA damage is further reviewed in section IIIC.

Cohort studies of cigarette smoking

A review of 11 prospective studies that examined the association between cigarette smoking and atherosclerotic endpoints is presented in Table 3. Seven studies investigated

incident CHD, of which five used the original or a supplemented version of the WHO MONICA classification of MI[6, 134-137]. Briefly, non-fatal CHD events classified by the WHO MONICA definition were those that included any report of MI accompanied by at least two of the following criteria: 1) history of severe prolonged chest pain; 2) ECG evidence of MI; and 3) cardiac enzyme changes associated with MI. However, Price *et al.,* and Goldberg *et al.,* also classified participants as having a MI if diagnostic ECG codes were present in the absence of elevated enzyme levels or pain in an attempt to capture silent MIs. Kawatchi *et al* also included events if medical records were not available, but hospitalization was required and confirmatory information was obtained by interview or letter, which may reflect the study population (nurses). With regard to fatal events, Wannamethee and colleagues depended on ICD-9 coding alone, whereas Price *et al*, Baba *et al*, Goldberg *et al*, and Kawachi *et al* utilized ICD codes supplemented by post-mortem findings and medical records.

Of the two studies of incident MI not employing the WHO criteria, ICD-9 codes (410-414, Table 19) alone were used in the Singapore Cardiovascular Cohort Study and Wilson and colleagues omitted information pertinent to outcome classification. Of the three studies that examined CHD deaths, all relied solely on ICD codes obtained from death certificates.

The measurement of cigarette exposure in the studies reviewed in Table 3 was variable and the motivations behind exposure classifications were often not described in detail. The two exceptions were Neaton and colleagues, who dichotomized pack-years based on a plot of age-adjusted CHD death rates by the number of cigarettes smoked/day and Price et al, who divided pack-years of smoking into two approximately equal groups, labeling the

higher group as "heavy smokers" and the lower group as "moderate smokers". Adjustment strategies were also inconsistent, although age and sex were typically included in the multivariable models.

Cigarette composition and construction also differs between countries, including tobacco type, filter and paper type, and additives, and may influence exposure yield per cigarette. For example, approximately 97% of cigarettes sold in the U.S. and other developed countries contain a filter, with the exception of France, where acceptance of the filter cigarette has been delayed[138]. The type of filter also varies by country. Although smoke from cigarettes with charcoal filters was less ciliatoxic *in vitro* when compared to other filter types[139, 140], only 1-2% of cigarettes on the U.S. market contain charcoal filters, compared to Japan, Venezuela, South Korea, and Hungary where over 90% of cigarettes have charcoal filters[141, 142]. Instead, cellulose acetate filters, which retain upwards of 80% of semivolatile phenols, are marketed in the remainder of developing countries[138].

Estimates of the association between cigarette smoking and incident CHD and related traits in diverse populations over three to 40 years follow-up with never smokers as referent ranged from 2.3 – 4.2 for current smokers, 1.1 – 2.8 for former smokers, and 0.7 – 3.9 for the cigarette/day or packs-year dichotomizations. However, the estimate of 3.9 reported by Price *et al* for >25 pack-years contrast was obtained using a logistic model and thus is an over-estimate of the incidence rate ratios reported by other investigators. The estimated IRR for the relationship between ≥ 20 PY among current vs. never smokers and incident CHD in the Singapore Cardiovascular Cohort Study was imprecise (confidence limit ratio (CLR) = 3.3) compared to other estimates, as was the estimate of 2.8 for the former vs. never contrast

estimated in the Japan Public Health Centre study (CLR = 10.1), perhaps explaining the high estimates obtained by these studies.

Other considerations are the inconsistent adjustment strategies, especially the fact that numerous studies adjusted for factors affected by the exposure, such as body mass index (BMI), blood pressure, and lipoproteins, complicating interpretation. Overall, the literature suggests a moderate relationship between cigarette smoking and CHD.

Table 3. Review of 11 prospective studies examining the relationship between cigarette smoking and CHD and related traits.

Author (year)	Study	Study population	Length of follow-up	Outcome	Number of events by smoking status	Measure estimated	Estimate (95% CI)	Covariate adjustments
Current vs. Former								
Howard (1998)[143]	ARIC	US males and females aged 45-64 years	3 years	Carotid IMT progression	2956/3193*	Mean difference (SE)	4.9 (2.5)	Age, race, sex, baseline IMT
Current vs. never								
Baba (2006)[137]	Japan Public Health Centre (JPHC)	Japanese males and females aged 40-59 years	12 years	Incident CHD	13022/227558†	IRR	3.1 (1.5, 6.4) females	Age, alcohol, fruit, vegetable, and fish intake, hypertension, diabetes, treated hyperlipidemia, education, and public health center.
Baba (2006)[137]	Japan Public Health Centre (JPHC)	Japanese males and females aged 40-59 years	12 years	Incident CHD	114,527/53420†	IRR	2.8 (2.0, 4.1) males	Age, alcohol, fruit, vegetable, and fish intake, hypertension, diabetes, treated hyperlipidemia, education, and public health center.
Kawachi (1997)[136]	Nurses' Health Study	US female nurses 30 to 55 age range	12 years	Incident CHD	377,171/593,02†	IRR	4.2 (3.6, 5.0)	Age, follow-up period, BMI, hypertension, cholesterol, diabetes, history of MI before age 60, HRT use, menopausal status, past use of OC, age at smoking initiation
Wannamethee(1999)[134]	British Regional Heart Study (BRHS)	UK males aged 40-59 years	15 years	Incident CHD	983/6752‡	IRR	2.3 (1.9, 2.9)	Age, BMI, SBP, cholesterol, physical activity, alcohol intake, diabetes, and family history of heart disease.
Former vs. Never								
Baba (2006)[137]	Japan Public Health Centre (JPHC)	Japanese males and females aged 40-59 years	12 years	Incident CHD	49204/53420†	IRR	1.1 (0.7, 1.7) males	Age, alcohol, fruit, vegetable, and fish intake, hypertension, diabetes, treated hyperlipidemia, education, and public health center.
Baba (2006)[137]	Japan Public Health Centre (JPHC)	Japanese males and females aged 40-59 years	12 years	Incident CHD	4030/227558†	IRR	2.8 (0.9, 9.1) females	Age, alcohol, fruit, vegetable, and fish intake, hypertension, diabetes, treated hyperlipidemia, education, and public health center.
Howard (1998)[143]	ARIC	US males and females aged 45-64 years	3 years	Carotid IMT progression	3193/4765*	Mean difference (SE)	5.9 (2.3)	Age, race, sex, baseline IMT
Hrubec(1997)[144]	.	US males veterans enlisted between 1917 and 1940	26 years	CHD death	10369/13257	IRR	1.2 (1.2, 1.2)	Age and calendar time

Kawachi (1997)[136]	Nurses' Health Study	US female nurses 30 to 55 age range	12 years	Incident CHD	404359/59302†	IRR	1.5(1.2, 1.8)	Age, follow-up period, BMI, hypertension, cholesterol, diabetes, history of MI before age 60, HRT use, menopausal status, past use of OC, age at smoking initiation
Lee (2001)[145]	Singapore Cardiovascular Cohort Study	Singapore males and females	8.9 years	Incident CHD	2649.4/12972.0†	IRR	1.3 (0.8, 2.2)	Age and ethnic group
Wannamethee(1999)[134]	British Regional Heart Study (BRHS)	UK males aged 40-59 years	15 years	Incident CHD	983/6752‡	IRR	1.5 (1.2, 2.8)	Age, BMI, SBP, cholesterol, physical activity, alcohol intake, diabetes, and family history of heart disease.
Cigarettes/day scaled at 10								
Menotti (2004)	US Railroad study	Males aged 40-59, US railroad company employees	40 years	CHD death	627/2376	IRR	1.2 (1.1, 1.3)	Age, SBP, cholesterol
Wilson (1991)[146]	Framingham offspring study	US males and females	12 years	Incident CHD	55/1659‡	IRR females	1.2 (1.0, 1.4)	HDL, LDL, VLDL, hypertension, BMI, glucose, and age.
Wilson (1991)[146]	Framingham offspring study	US males and females	12 years	Incident CHD	156/1507‡	IRR males	1.2 (1.1, 1.3)	HDL, LDL, VLDL, hypertension, BMI, glucose, and age.
PY > 25 vs. never smokers								
Price (1999)[6]	Edinburgh Artery Study	Scottish males and females 55-74 age range	5 years	Incident PAD	46/782	IOR	3.9 (2.0, 7.6)	Age and sex
Price (1999)[6]	Edinburgh Artery Study	Scottish males and females 55-74 age range	5 years	Incident CHD	82/746	IOR	1.7 (1.1, 2.6)	Age and sex
PY ≥ 20 among current smokers vs. never smokers								
Lee (2001)[145]	Singapore Cardiovascular Cohort Study	Singapore males and females	8.9 years	Incident CHD	3016.5/12972.0†	IRR	0.7 (0.4, 1.3)	Age and ethnic group
PY ≤ 25 vs. never								
Price (1999)[6]	Edinburgh Artery Study	Scottish males and females 55-74 age range	5 years	Incident CHD	88/802	IOR	1.6 (1.0, 2.4)	Age and sex
Price (1999)[6]	Edinburgh Artery Study	Scottish males and females 55-74 age range	5 years	Incident PAD	31/859	IOR	1.9 (0.1, 3.8)	Age and sex

PY < 20 among current smokers vs. never smokers								
Lee (2001)[145]	Singapore Cardiovascular Cohort Study	Singapore males and females	8.9 years	Incident CHD	6347.9/12972.0[†]	IRR	1.3 (0.8, 1.3)	Age and ethnic group
≥26 cigarettes/day vs. nonsmokers								
Neaton (1992)[147]	MRFIT	US males 35-57 years old	12 years	CHD death	1932/2874	IRR	2.9[¥]	Age, cholesterol, SBP, and DBP
<26 cigarettes/day vs. nonsmokers								
Neaton (1992)[147]	MRFIT	US males 35-57 years old	12 years	CHD death	1521/2874	IRR	2.1[¥]	Age, cholesterol, SBP, and DBP
>20 cigarettes/day vs. nonsmoker								
Goldberg (1995)[135]	Honolulu Heart Program	Males of Japanese descent residing on Oahu in 1965, 45-65 years of age	20 years	Incident definite CHD	352/2108[‡]	IRR	2.3 (1.6, 3.3)	Ventricular rate, BMI, SBP, cholesterol, triglycerides, glucose, uric acid, hematocrit, FEV , physical activity, and alcohol
Goldberg (1995)[135]	Honolulu Heart Program	Males of Japanese descent residing on Oahu in 1965, 45-65 years of age	20 years	Total atherosclerotic events[§]	602/2108[‡]	IRR	2.2 (1.6, 2.8)	Ventricular rate, BMI, SBP, cholesterol, triglycerides, glucose, uric acid, hematocrit, FEV , physical activity, and alcohol
20 cigarettes/day vs. nonsmoker								
Goldberg (1995)[135]	Honolulu Heart Program	Males of Japanese descent residing on Oahu in 1965, 45-65 years of age	20 years	Incident definite CHD	352/2108[‡]	IRR	2.1 (1.6, 3.0)	Ventricular rate, BMI, SBP, cholesterol, triglycerides, glucose, uric acid, hematocrit, FEV , physical activity, and alcohol
Goldberg (1995)[135]	Honolulu Heart Program	Males of Japanese descent residing on Oahu in 1965, 45-65 years of age	20 years	Total atherosclerotic events[§]	602/2108[‡]	IRR	2.1 (1.7, 2.7)	Ventricular rate, BMI, SBP, cholesterol, triglycerides, glucose, uric acid, hematocrit, FEV , physical activity, and alcohol
<20 cigarettes/day vs. nonsmoker								
Goldberg (1995)[135]	Honolulu Heart Program	Males of Japanese descent residing on Oahu in 1965, 45-65 years of age	20 years	Incident definite CHD	352/2108[‡]	IRR	1.7 (1.1, 2.6)	Ventricular rate, BMI, SBP, cholesterol, triglycerides, glucose, uric acid, hematocrit, FEV , physical activity, and alcohol
Goldberg (1995)[135]	Honolulu Heart Program	Males of Japanese descent residing on Oahu in 1965, 45-65 years of age	20 years	Total atherosclerotic events[§]	602/2108[‡]	IRR	1.7 (1.2, 2.3)	Ventricular rate, BMI, SBP, cholesterol, triglycerides, glucose, uric acid, hematocrit, FEV , physical activity, and alcohol

Former smokers (who reported 1-9 cigarettes/day as highest amount smoked) vs. never								
Hrubec(1997)[144]	.	US males veterans enlisted between 1917 and 1940	26 years	CHD death	1966/13257	IRR	1.1 (1.0, 1.1)	Age and calendar time
Former smokers (who reported 10-20 cigarettes/day as highest amount smoked) vs. never								
Hrubec(1997)[144]	.	US males veterans enlisted between 1917 and 1940	26 years	CHD death	4685/13257	IRR	1.2 (1.1, 1.2)	Age and calendar time
Former smokers (who reported 21-39 cigarettes/day as highest amount smoked) vs. never								
Hrubec(1997)[144]	.	US males veterans enlisted between 1917 and 1940	26 years	CHD death	2723/13257	IRR	1.3 (1.3, 1.4)	Age and calendar time
Former smokers (who reported >40 cigarettes/day as highest amount smoked) vs. never								
Hrubec(1997)[144]	.	US males veterans enlisted between 1917 and 1940	26 years	CHD death	995/13257	IRR	1.4 (1.3, 1.5)	Age and calendar time

*Presented as no. exposed/no. unexposed; †Presented as total person-years at risk by smoking status; ‡Number of events and non-events overall; § Incident CHD, angina, aortic aneurysms, and thromboembolic stroke; ¥ CI not reported; BMI, body mass index; CHD, coronary heart disease; CI, confidence interval; DBP, diastolic blood pressure; FEV, forced expiratory volume; HDL, high-density lipoprotein; HRT, hormone replacement therapy; IRR, incidence rate ratio; IMT, intimal-medial thickness; LDL, low-density lipoprotein; MI, myocardial infarction; OC, oral contraceptive; PAD, peripheral arterial disease; PY, pack-years; SBP, systolic blood pressure; SE, standard error; VLDL, very low density lipoprotein

Classification of cigarette smoke exposure

While the relationship between cigarette smoking and CHD is well-established (Table 3), the magnitude and shape of the dose-response relationship remains unresolved[115]. Early attempts to characterize the dose-relationship utilized published data on Caucasian males aged 45 to 54 years from four studies to estimate dose-response curves: The Veterans Administration Study, The American Cancer Society Study, the Study of British Physicians, and The Pooling Project, which combined five U.S. epidemiologic studies. Briefly, a dose response model was fit to the observed data and goodness-of-fit measures and smoothed death rates were estimated. The smoothed death rates were then applied to the population distribution by amount smoked in 1965 and 1976 in an attempt to explain the effect of decreased smoking among males on the declining CHD mortality rate, yielding two "expected" CHD death rates.

Results indicated that a wide variety of models (linear, logarithmic, and exponential) were consistent with the decline in CHD mortality[148]. However, the differences may have reflected the imprecision rather than true variation[149], as well as the assumption of a univariate dose-response relationship, the low statistical power, and the reliance upon grouped data, not individual data. The use of midpoints may also have underestimated the effects of lighter smoking and inflate the effects of heavy smoking[150].

Framingham investigators initially utilized numerous classifications for cigarette smoking including the seven-level categorization of never-smoked, <10, 10, 11-19, 20, 21-39, and ≥ 40 cigarettes/day. However, variation in history formats resulted in a condensed classification of none, <20, 20, and >20 cigarettes/day, which appears in most Framingham publications[151].

As demonstrated in Table 3, researchers used numerous classifications to measure cigarette smoking exposure, including current, former, or never smoking, pack-years, and cigarettes/day. While pack-years of smoking is a composite variable intended to capture the duration and magnitude of cigarette smoke exposure, there is little consistency in the choice of cutpoints between studies. The lack of comparability may reflect the fact that some researchers chose cutpoints based on means or medians of the pack-years distribution, not biology. The relationship between the duration and magnitude cigarette smoking and the rate of CHD may not be linear, thus questioning the suitability of pack-years of smoking or cigarettes/day for exposure measurement.

Pack-years and other measures of cumulative smoking exposure also do not take account of inter-individual variation in mode of smoking, as the type of cigarette smoked, the frequency and depth of inhalation, and amount of stub remaining all contribute to variation in exposure to cigarette smoke[152]. Additional sources of variability include faulty recall of cigarette smoking history or terminal digit preference corresponding to packs of cigarettes (e.g. 20 or 30)[153]. The feasibility of obtaining an optimized measure is also questionable considering social pressures that may cause an underreporting of true exposure.

Previous research suggests that former smokers were exposed to fewer cigarettes/day and initiated smoking at an older age when compared to current smokers. Thus, at any age of cessation, former smokers are believed to have less cumulative exposure to cigarettes than current smokers[136]. However, former smokers may have quit smoking due to smoking-related respiratory and CVD symptoms and may actually have a higher burden of disease than their currently smoking counterparts[143]. In whole, the optimal classification for cigarette smoke exposure in CVD research remains unclear.

Short term and long term effects of smoking

While numerous studies have demonstrated short-term benefits of smoking cessation, disagreement exists as to the long-term effects of smoking on CHD risk. Studies of British physicians[154] and American nurses[136] have demonstrated that the mortality rates for vascular diseases in former smokers remain intermediate to vascular disease mortality rates among never and current smokers and the British Regional Heart Study found little risk reduction for ex-smokers after seven years follow-up[155].

Research in the ARIC cohort also suggests that the effect of smoking on atherosclerosis progression may be cumulative and long-lasting. Howard et al., (1998) examined the influence of cigarette smoking and environmental tobacco smoke (ETS) exposure on the progression of atherosclerosis. They demonstrated a relationship between increased cigarette smoke exposure and carotid atherosclerosis progression (Table 3), after adjustment for demographic and cardiovascular risk factors. Notably, atherosclerosis progression was higher among former smokers than never smokers, but no differences were identified between past and current smokers after accounting for pack-years of exposure[143].

Inconsistencies between studies examining the risk reduction associated with smoking cessation may reflect other factors such as age of smoking cessation and/or initiation, CHD risk factors, and disease severity[156], or even our incomplete understanding of vascular biology. Considering that DNA damage induced by tobacco smoke can result in the formation of DNA adducts that, if not repaired, may cause uncontrolled cellular proliferation and that cigarette smoke is capable of promoting SMC proliferation, plausible mechanisms linking cigarette smoke, DNA repair capacity, and atherosclerotic disease exist. While we cannot directly evaluate somatic alterations in SMCs, measures of smoking exposure and

inherited DNA repair polymorphisms are available, which could inform future investigators of the influence of genetic factors in the relationship between cigarette smoking and atherosclerosis-related diseases.

C. Repair of DNA damage

The human genome with its approximately three billion base pairs is vulnerable to an array of DNA-damaging agents of both endogenous and exogenous origin. The integrity of DNA molecules is maintained primarily by DNA repair mechanisms, which continually excise and replace nucleotides thus reducing the burden of potentially mutagenic and cytotoxic products. Reduced DNA repair capacity has been associated with cancers[157-159], neurodegeneration[160], and premature aging[161].

DNA damage occurs via a variety of mechanisms. Oxidative damage, for example, occurs when a cell is exposed to an increased amount of reactive oxygen species (ROS)[162, 163], compounds containing partially reduced oxygen that possess high reactivity with biomolecules (e.g. DNA, proteins, and lipids)[163-166]. Ubiquitous in aerobic life, ROS include both oxygen radicals and derivatives of O_2 that do not contain unpaired electrons[101] and are derived from exogenous sources and normal by-products of metabolic processes. Examples of ROS include hydroxyl radicals (·OH), singlet oxygen ($^1O_2$), superoxide ($O_2^{\cdot-}$), and hydrogen peroxide (H_2O_2)[167]. The effects of ROS are typically counteracted by plasma antioxidants that scavenge the ROS[168-170] and the enzyme superoxide dismutase (*SOD2*), which reduces ROS to excretable endproducts[170, 171]. ROS that escape the effects of plasma antioxidants and *SOD2* are capable of damaging DNA.

Numerous oxidative base lesions are mutagenic, regardless of whether they are formed in situ or arise by misincorporation from the deoxynucleotide pool[172]. Many compounds can generate ROS capable of damaging DNA including benzo(a)pyrene, benzene, low wavelength ionizing radiation, and UV light. The effects of ROS are not limited to mutation however. Lesions can block replication or cause DNA deletions and/or microsatellite instability / loss of heterozygosity (LOH). Epigenetic effects have also been associated with oxidant exposure[172].

Other causes of DNA damage include methylation and hydrolysis. DNA methylation results in the addition of a methyl group to DNA and is an example of epigenetics, reversible heritable alterations in genes that occur without modifying the DNA sequence. While some lesions are harmless in that they do not result in miscodings or have cytotoxic properties, 3-methyladenine (3-meA), promotes spontaneous mutagenesis if not repaired[173]. Urinary excretion of 3-meA was increased after controlled exposure to cigarette smoke in smokers compared to non-smokers, although the DNA-reactive agents responsible for the increase remain unknown[174].

Base loss, or hydrolysis, is the most frequent type of damage in human cells. Approximately 2,000 – 10,000 abastic sites (apurinic or apyrimidinic sites resulting from the loss of a purine or pyrimidine residue) are generated daily in a mammalian cell under physiologic conditions[175]. Brain tissue is the most affected, followed by the colon and myocardium[176]. Apurinic/apyrimidinic (AP) sites are formed by the spontaneous hydrolysis of the *N*-glycosylic bond or during the DNA repair process (further discussed in section IIIC). Nitrosamine 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone (NNK), a potent

cigarette smoke carcinogen formed by the nitrosation of nicotine[177, 178], was shown to increase base loss in human non-small cell lung cancer cells using the comet assay[179].

DNA damage is often manifest as "non - bulky" DNA adducts[180], single-base modifications, such as oxidized, fragmented, or reduced bases, or lesions produced by methylating agents. 8-oxo-7, 8 – dihydroguanine (8-oxo-G), is the most common and mutagenic ROS-induced non-bulky DNA lesion[181-183] and is a marker of cellular oxidative stress[184]. The failure to repair this DNA adduct could result in a mutation during cell division (G – T transversion), transmitting the damage to successive cell generations[185]. Cigarette smoking has been consistently associated with a 30-50% increase in urinary 8-oxo-G expression[186-188].

DNA damage caused by "bulky" adducts result in distortions of the helix and is caused by pyrimidine dimers, photo-products, or cross-links[189]. BPDE, the activated form of the cigarette smoke carcinogen benzo(a)pyrene, is one of the many compounds capable of inducing bulky DNA adduct formation without further activation[190] and blocks the transcription of essential genes if unrepaired[191]. Indeed, lymphocytes from breast cancer patients were more sensitive to *in vitro* BPDE exposure than controls, suggesting suboptimal DNA repair capacity[192].

Oxidative stress, as well as ionizing radiation and the overlap of excision repair tracts[193], can also produce double strand DNA breaks, potentially the most dangerous type of DNA damage. Repair of a double strand DNA break is considerably more difficult than the repair of a DNA adduct, reflecting the lack of an undamaged DNA template[194]. Research has demonstrated that concentrations of cigarette smoke condensate far below those contained in

a single cigarette can induce double strand DNA damage in cultured cells and purified DNA[195].

Approximately 130 human genes are involved in repairing the types of DNA damage reviewed above[196]. While loss-of-repair variants are infrequent, studies using lymphocyte-based assays suggest that 10-20% of the human population have a heritable reduced capacity to repair DNA damage induced by exposures including gamma radiation, bleomycin, and BPDE[197-199]. Such heritable differences in DNA repair genes are most often captured by the SNP, the most common type of genetic variation in the human genome occurring at an estimated density of one in 1,000 base pairs[200-203]. SNPs are typically bi-allelic, are variably distributed throughout the human genome, and have frequencies that differ between race/ethnic groups. While a fraction of SNPs encode amino acid changes (non-synonymous SNPs) that potentially change protein structure and function, the majority of SNPs are intronic.

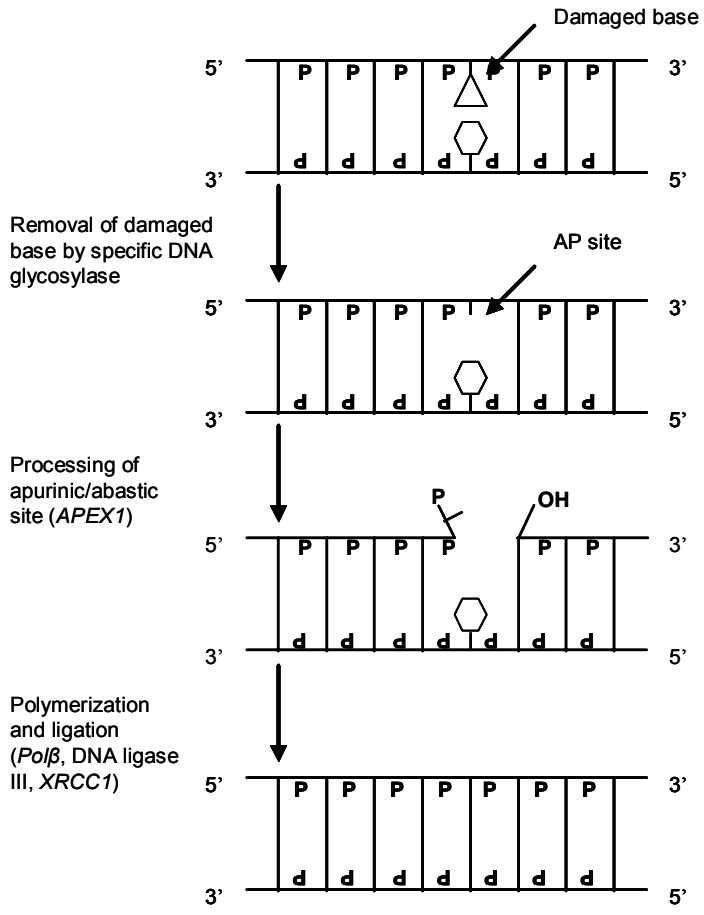
The numerous compounds contained in cigarette smoke, each with a unique chemical structure, demonstrably cause a variety of DNA damage and necessitate distinct DNA repair pathways and genes[189]. As inherited genetic variants at one or more loci can affect disease susceptibility and/or influence the effect of environmental exposures such as cigarette smoke, we examined three genes in the base excision repair (BER) pathway (*hOgg1, APEX1,* and *XRCC1*), one gene in the nucleotide excision repair (NER) pathway (*XPD*), and one gene in the double-strand break (DSB) repair pathway (*XRCC3*), the DNA repair pathways of most importance for the repair of tobacco-related DNA damage[204]. The process of SNP selection used for this dissertation was informed by functional data, SNP type with preference for non-

synonymous SNPs, published literature, and patterns of pair-wise linkage disequilibrium (LD).

1. Base excision repair (BER)

The BER pathway operates on small lesions (Figure 2) such as oxidized or reduced bases, fragmented or non-bulky adducts produced by methylating agents, and (AP) sites, all of which may arise during inflammatory responses, spontaneously within the cell, or from exogenous agents. While this type of damage causes minor changes in the helical DNA structure, BER is one of the most highly conserved DNA repair mechanisms[205], emphasizing its importance in maintaining genomic integrity.

Figure 2. Diagram of the short-patch BER pathway.



Adapted from Goode et al., 2002[184]

Briefly, BER first involves cleavage of the damaged nucleotide by DNA glycosylases, generating an abastic site. *APEX1* then nicks the damaged DNA strand

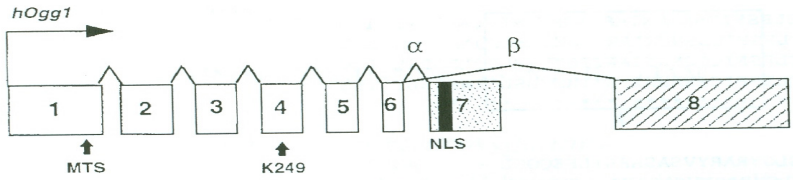
upstream of the AP site, creating a 3'-OH terminus adjacent to the abastic sites. BER is completed by the replacement of one (short-patch pathway) or multiple (long-patch pathway) nucleotides at the 3'-OH terminus by DNA polymerases and sealing of the incision by DNA ligase[180, 206-208], restoring DNA to its unmodified state.

As discussed above, tobacco smoke contains numerous carcinogens capable of causing DNA damage that is repaired by the BER pathway and an excess of ROS has been associated with cellular damage and atherogenesis in numerous studies[209, 210]. For example, strong 8-oxo-G immunoreactivity was demonstrated in all atherosclerosis plaque cell types obtained from 13 human carotid endarterectomy specimens of patients with a carotid stenosis exceeding 70%, but not in the underlying media or non-atherosclerotic mammary arteries[211]. As BER is major repair mechanism for the type of oxidative damage caused by tobacco exposure and sequence variants in DNA repair genes are believed to modulate DNA repair capacity, a mechanistic basis exists for evaluating the role of BER variants in the relationship between cigarette smoking and CHD.

hOgg1

hOgg1 is located on 3p25 and is one of the five DNA glycosylases that participate in the first steps of BER. *hOgg1* is expressed as 12 alternatively-spliced isoforms, although the 1α-form is the only isoform with a nuclear localization signal[213, 214] (Figure 3).

Figure 3. Schematic of the *hOgg1* protein.



Adapted from Boiteus and Radicella, (2000)[212]. MTS, mitochondrial targeting signal; NLS, nuclear localization sequence.

The *hOgg1* gene catalyzes the removal of 8-oxo-G from DNA[207, 215, 216], an oxidized derivative of guanine that is one of the most mutagenic DNA lesions, as it mispairs with adenine during DNA replication, resulting in G to T transversions (G:C to T:A) *in vivo* and *in vitro*[217]. *hOgg1* initiates the repair of 8-oxo-G lesions by cleaving the *N*-glycosyl oxidized guanine-deoxyribose backbone bond, releasing the modified base and producing an apurinic / apyrimidinic site (Figure 2). *hOgg1* does not require any additional cofactors to recognize 8-oxo-G DNA lesions or to initiate enzymatic activity[180]. ROS-mediated DNA damage is hypothesized to cause mutations associated with the initiation or progression of human cancers, as such mutations may activate oncogenes or inactivate tumor suppressor genes[218]. Indeed, *hOgg1* frequently shows LOH, the loss of a single parent's contribution to part of the cell's genome, in several human cancers[215, 219].

Numerous functional studies have investigated the role of *hOgg1* in DNA repair. Inactivation of the *E .coli hOgg1* homologue *fpg* has been shown to lead to a spontaneous mutator phenotype typified by an increase in G:C to T:A transversions[217, 220]. Deletion of the *hOgg1* gene in yeast also creates a mutator phenotype specific for G:C to T:A transversions[221].

Nishimura (2002) demonstrated that mice with a targeted disruption of the *hOgg1* gene had elevated rates of spontaneous mutagenesis and high levels of 8-oxo-G[216] and DNA 8-oxo-G content was shown to be higher in lung cancer patients than in controls[222]. *hOgg1* expression was also investigated in 13 human carotid endarterectomy specimens obtained from patients with a carotid stenosis exceeding 70%. However, western blots suggested similar *hOgg1* protein levels when carotid endarterectomy and non-atherosclerotic vessel specimens were compared[211].

In addition, the effect of peroxynitrite on *hOgg1* activity was examined in a human cholangiocarcinoma cell line. Peroxynitrite is a potent and versatile oxidant that can attack a wide range of biological targets and is formed by the reaction of the ROS superoxide with NO. Jaiswal et al. demonstrated that cellular *hOgg1*-mediated BER activity was inhibited during peroxynitrite exposure, suggesting that peroxynitrite not only caused oxidative DNA damage, but also prevented DNA repair[223]. Although cigarette smoke contains large quantities of peroxynitrites and other ROS[104, 105], these substances are short lived and rapidly degrade in aqueous solution[104, 224]. However, chemically stable compounds present in the gas phase of cigarette smoke, such as α,β-unsaturated aldehyde acrolein, were shown to increase the production of the peroxynitrite precursor superoxide in cultured bovine pulmonary artery endothelial cells in a dose-dependent manner[109].

hOgg1 genetic variants

Previous studies have identified seven nonsynonymous *hOgg1* variants (Table 4), although minor allele frequencies (MAF) preclude evaluating most polymorphisms in population-based studies. The C/G polymorphism at position 1235 (Ser326Cys) in the 1α-specific exon seven[219] is the most studied *hOgg1* variant and Cys/Cys carriers are postulated to have a decreased capability in repairing oxidative DNA damage compared to Ser/Ser or Ser/Cys carriers. One functional study examining the Ser326Cys variant demonstrated that the 326Ser- containing *hOgg1* had a seven-fold higher activity for repairing 8-oxo-G than 326Cys-containing *hOgg1* using a complementation assay of an *E. coli* mutant defective in the repair 8-oxo-G[219]. Cys326-initiated BER was also shown to be transiently impaired, compared to Ser326, following pro-oxidant treatment in transgenic mouse embryonic fibroblast cell lines[167]. However, no mean differences in DNA repair activity by Ser326Cys

polymorphism was detected in human cryoconserved lymphocytes[225], although the study was limited to samples from 34 healthy donors. The function of one other *hOgg1* variants was also evaluated as Chevillard *et al.*,(1998) demonstrated that mutation of Arg229 to Gln in cDNA abolished the ability of the *hOgg1* protein to repair 8-oxo-G[226].

Table 4. Characterization of seven known *hOgg1* nonsynonymous SNPs.

SNP*	Protein residue	MAF	Functional data?	Studied in human populations?
rs11548133	Thr27Pro	.	.	.
rs17050550	Ser85Ala	0.04 (CEPH)	.	.
rs1805373	Gln229Arg	0.00 (CEU) 0.10 (YRI)	Sunaga et al., (2002)[227]	.
rs3219012	Val288Ala	0.01 (CEU)	.	.
rs1801128	Thr320Ser	.	.	.
rs3219014	Asn322Asp	0.006 (PDR90)	.	.
rs1052133	Cys326Ser	0.30 (NIHPDR) 0.10 (AFR1)	Sunaga et al., (2002)[227] Smart et al., (2006)[167] Janssen et al., (2001)[225]	Table 5 and Table 6

*Information obtained from dbSNP (http://www.ncbi.nlm.nih.gov/projects/SNP/); **AFR1**, Human individual DNA from 24 individuals of self-described African/African American heritage; **CEPH**, Genomic DNA samples obtained for a panel of 92 unrelated individuals chosen from Centre d'Etude du Polymorphisme Human (CEPH) pedigrees. The genomic DNA comprised of UTAH (93%), French (4%), and Venezuelan (3%) samples were purchased from Coriell Cell Repository; **CEU**, 30 mother-father-child trios from the CEPH collection (Utah residents with ancestry from northern and western Europe), representing one of the populations studied in the International HapMap project; **NIHPDR**, The NIH Polymorphism Discovery Resource (NIHPDR) contains cell lines and DNA from 450 anonymous, unrelated individuals with equal numbers of females and males. The sample has sampled non- European regions at frequencies higher than the general U.S. Population to enrich the genetic variability of the resource; **PDR90**, The NIH Polymorphism Discovery Resource (NIHPDR) 90 individual screening subset; **YRI**, 30 Yoruba mother-father-child trios in Ibadan, Nigeria, representing one of the populations studied in the International HapMap project.

Although *hOgg1* studies in human populations are often focused on cancers, Wang et al (2006) investigated the relationship between Ser326Cys and insulin sensitivity, as oxidative stress may impair insulin action[228-231]. Briefly, 297 Taiwanese males and females (mean age 45.3 years) with fasting plasma glucose < 100 mg/dl and no reported history of cancer were genotyped for the Ser326 variant. Mean levels of fasting insulin (μIU/ml, SE) for the Ser/Ser, Ser/Cys, and Cys/Cys participants were 4.9 (2.5), 5.5(2.8), and 6.9(4.0), respectively[232], suggesting that insulin sensitivity is decreased in participants carrying the Cys/Cys genotype.

The observational studies reviewed in Table 5 and have not provided consistent evidence for an association between *hOgg1* variants and cancers and related traits. For example, estimates of the Ser326Cys variant and disease ranged from 0.8 – 1.3 for breast cancer, 0.6 – 2.1 for lung cancer, and 0.6 – 1.1 for stomach cancer. As expected, the most extreme estimate, 7.6 (1.8, 31) for the Cys/Cys vs. Ser/Ser contrast and prevalent prostate cancer[233], was also the most imprecise (confidence limit ratio (CLR) = 17). When stratified by smoking status the marked imprecision of the estimates precluded evaluation. Furthermore, studies of the same outcome were few, largely focused on the Ser326Cys *hOgg1* variant, typically underpowered to detect modest effects, especially modification by cigarette smoking, and often examined prevalent disease. No studies evaluated additive interaction.

Table 5. Review of 19 case control studies examining the relationship between *hOGG1* polymorphisms and cancers and related traits stratified by cancer and polymorphism.

Variant	Outcome	Author (year)	Study population	No. cases/controls	Genotype contrast	OR (95% CI)	Covariate adjustments
Ser326Cys							
	Breast cancer	Cai (2006)[234]	Female Chinese aged 25-64 years*	568/630	Cys/Cys vs. Ser/Ser	1.1 (0.8, 1.4)	Age, education level, menopausal status, and age at first birth.
				720/751	Cys/Ser vs. Ser/Ser	1.2 (0.9, 1.5)	Age, education level, menopausal status, and age at first birth.
		Choi (2003)[235]	Korean and Japanese women*	224/239	Cys/Cys vs. Ser/Ser	1.3 (0.9, 1.9)	Age, BMI, family history of breast cancer, and parity
				355/332	Cys/Ser vs. Ser/Ser	1.0 (0.7, 1.5)	Age, BMI, family history of breast cancer, and parity
		Vogel (2003)[236]	Denmark women aged 50-64 years*	42/501	Cys/Cys vs. Ser/Ser	1.0 (0.5, 1.9)	Unadjusted
				316/501	Cys/Cys vs. Ser/Ser	0.8 (0.6, 1.1)	Unadjusted
	Colon cancer	Kim (2003)[237]	Korean males and females†	116/59	Cys/Cys vs. Ser/Ser	1.2 (0.6, 2.4)	Unadjusted
				183/90	Cys/Ser vs. Ser/Ser	1.1 (0.6, 2.0)	Unadjusted
	Lung cancer	Le Marchand (2002)[238]	Oahu, Hawaii residents aged 18-79*	233/352	Cys/Ser vs. Ser/Ser	0.7 (0.5, 1.1)	Age, sex, race, smoking, smoking years, smoking years^2, cigarettes/day, saturated fat and vegetable intake.
				188/230	Cys/Cys vs. Ser/Ser	2.1 (1.2, 3.7)	Age, sex, race, smoking, smoking years, smoking years^2, cigarettes/day, saturated fat and vegetable intake.
		Liang (2005)[239]	Chinese males and females, 30 – 86 years of age†	227/227	Cys/Cys, Cys/Ser vs. Ser/Ser	0.9 (0.5, 1.6)	Sex, age, and smoking
		Sugimura (1999)[240]	Male Japanese*	126/90	Cys/Cys vs. Ser/Ser	1.3 (0.6, 2.6)	Age and smooking habit
				200/170	Cys/Ser vs. Ser/Ser	0.6 (0.4, 1.1)	Age and smooking habit
		Sunaga (2002)[227]	Japanese†	92/86	Cys/Cys vs. Ser/Ser	0.9 (0.5, 1.7)	Sex, age, smoking
				160/116	Cys/Ser vs. Ser/Ser	1.3 (0.8, 2.2)	Sex, age, smoking
	Stomach cancer	Hanaoka (2001)[241]	Japanese and non-Japanese Brazilians, aged 37-89 years*	49/100	Cys/Ser vs. Ser/Ser among Japanese	1.1 (0.6, 2.3)	Unadjusted
				29/71	Cys/Cys vs. Ser/Ser among Japanese	0.7 (0.3, 1.8)	Unadjusted
				200/197	Cys/Ser vs. Ser/Ser among non-Japanese	0.8 (0.6, 1.3)	Unadjusted
				141/131	Cys/Cys vs. Ser/Ser among non-Japanese	0.9 (0.3, 2.5)	Unadjusted

		Takezaki (2002)[242]	Chinese males and females*	40/78	Cys/Cys vs. Ser/Ser	0.6 (0.3, 1.4)	Unadjusted
				81/150	Cys/Ser vs. Ser/Ser	0.8 (0.4, 1.5)	Unadjusted
	Nasopharyngeal carcinoma	Cho (2003)[243]	Taiwanese males and females†	158/154	Cys/Cys vs. Ser/Ser	1.4 (0.9, 2.4)	Age, sex, and ethnicity
				211/175	Cys/Ser vs. Ser/Ser	1.8 (1.1, 2.9)	Age, sex, and ethnicity
	Orolaryngeal cancer	Elahi (2002)[244]	U.S. males and females aged 25-87 years†	113/255	Cys/Cys vs. Ser/Ser	3.6 (1.2, 10)	Unadjusted
				158/325	Cys/Ser vs. Ser/Ser	1.7 (1.1, 2.6)	Unadjusted
		Xing (2001)[245]	Chinese males and females*	196/201	Cys/Cys vs. Cys/Ser, Ser/Ser	1.9 (1.3, 2.6)	Sex, age, cigarette smoking
		Zhang(2004)[246]	University of TX patients and residents of greater Houston TX*	486/808	Cys/Cys vs. Ser/Ser	1.0 (0.6, 1.5)	Age, sex, smoking, and alcohol
				667/1127	Cys/Ser vs. Ser/Ser	0.9 (0.8, 1.1)	Age, sex, smoking, and alcohol
	Prostate cancer	Chen (2003)[233]	U.S. males aged 42-82†	55/188	Cys/Cys vs. Ser/Ser	7.6 (1.8, 31)	Unadjusted
				78/248	Cys/Ser vs. Ser/Ser	1.7 (1.0, 3.0)	Unadjusted
		Nam (2005)[247]	Canadian males†	646/706	Cys/Cys vs. Ser/Ser	0.6 (0.5, 0.8)	Unadjusted
				943/1003	Cys/Cys vs. Ser/Ser	0.9 (0.8, 1.1)	Unadjusted
		Xu (2002)[248]	U.S. males†	111/128	Cys/Cys vs. Ser/Ser	3.2 (1.2, 8.5)	Unadjusted
				159/193	Cys/Ser vs. Ser/Ser	1.1 (0.7, 1.7)	Unadjusted
	Pterygium	Kau (2004)[249]	Taiwanese males and females†	70/86	Cys/Cys vs. Cys/Ser, Ser/Ser	2.2 (1.1, 4.5)	Age, sex, and occupation type
	Oral cleft	Olshan (2005)[250]	California infants born 1983-1986†	60/184	Cys/Cys vs. Ser/Ser	0.2 (0.1, 0.8)	Race/ethnicity
				113/280	Cys/Ser vs. Ser/Ser	1.4 (0.8, 2.1)	Race/ethnicity
	Spina bifida	Olshan (2005)[250]	California infants born 1983-1986†	77/184	Cys/Cys vs. Ser/Ser	1.5 (0.7, 3.1)	Race/ethnicity
				110/280	Cys/Ser vs. Ser/Ser	0.8 (0.5, 1.3)	Race/ethnicity
11657‡							
	Prostate cancer	Nam (2005)[247]	Canadian males†	730/653	GG vs. AA	0.9 (0.6, 1.3)	Unadjusted
				960/1056	GG vs. AA	1.1 (0.9, 1.3)	Unadjusted
		Xu (2002)[248]	U.S. males†	140/169	CC/AA	0.1 (0, 0.8)	Unadjusted
				286/214	AG/AA	3 (2, 4.4)	Unadjusted
3402‡							

	Prostate cancer	Xu (2002)[248]	U.S. males†	102/102	GG/AA	1.4 (0.7, 2.6)	Unadjusted
				154/162	AG/AA	1.1 (0.7, 1.6)	Unadjusted
3574‡							
	Prostate cancer	Xu (2002)[248]	U.S. males†	115/138	GG/AA	1.4 (0.6, 3.3)	Unadjusted
				162/182	AG/AA	1.3 (0.8, 2.1)	Unadjusted
6170‡							
	Prostate cancer	Xu (2002)[248]	U.S. males†	114/138	GG/CC	2.1 (0.8, 5.2)	Unadjusted
				161/199	CG/CC	1.1 (0.7, 1.7)	Unadjusted
7143‡							
	Prostate cancer	Xu (2002)[248]	U.S. males†	132/165	GG/AA	0.2 (0, 0.9)	Unadjusted
				182/212	AG/AA	1 (0.7, 1.6)	Unadjusted
9110‡							
	Prostate cancer	Xu (2002)[248]	U.S. males†	122/144	AA/GG	2.5 (0.9, 6.9)	Unadjusted
				172/204	GA/GG	1.2 (0.8, 1.8)	Unadjusted
10629‡							
	Prostate cancer	Xu (2002)[248]	U.S. males†	106/105	GG/CC	1.1 (0.6, 1.8)	Unadjusted
				126/138	CG/CC	0.9 (0.5, 1.4)	Unadjusted
10660‡							
	Prostate cancer	Xu (2002)[248]	U.S. males†	123/148	AA/TT	1.9 (0.7, 4.8)	Unadjusted
				170/209	TA/TT	1.1 (0.7, 1.7)	Unadjusted
11826‡							
	Prostate cancer	Xu (2002)[248]	U.S. males†	122/145	TT/AA	2.2 (0.8, 5.6)	Unadjusted
				170/202	AT/AA	1.2 (0.8, 1.8)	Unadjusted

*Study of incident disease; †Study of prevalent disease; ‡Celera Genomics – Celera Human Reference SNP Database notation: http://www.celera.com/; Results were extracted with preference for unadjusted estimates

Table 6. Review of 10 case control studies examining the relationship between the Ser326Cys *hOGG1* polymorphisms and cancers and related traits, stratified by smoking status.

Variant	Outcome	Author (year)	Study population	No. cases/ controls	Genotype contrast	OR (95% CI)	Covariate adjustments
Ser326Cys							
	Colon cancer	Kim (2003)[237]	Korean males and females †	81/141	Cys/Cys vs. Ser/Ser and Ser/Cys among smokers	1.5 (0.8, 2.8)	Unadjusted
				44/106	Cys/Ser vs. Ser/Ser and Ser/Cys among non-smokers	0.6 (0.3, 1.4)	Unadjusted
	Orolaryngeal cancer	Xing (2001)[245]	Chinese*	73/118	Cys/Cys smokers vs. Cys/Ser and Ser/Ser non-smokers	4.8 (2.0, 11)	Sex and age
		Elahi (2002)[244]	U.S. males and females aged 25-87 years†	21/83	Cys/Cys vs. Ser/Ser among never smokers	2.3 (0.2, 28)	Age, sex, and alcohol
				30/105	Cys/Ser vs. Ser/Ser among never smokers	1.6 (0.6, 4.1)	Age, sex, and alcohol
				92/171	Cys/Cys vs. Ser/Ser among ever smokers	4.8 (1.3, 18)	Age, sex, PY and alcohol
				128/219	Cys/Ser vs. Ser/Ser among ever smokers	1.6 (1.0, 2.8)	Age, sex, PY and alcohol
	Gastric cancer	Hanaoka (2001)[241]	Japanese and non-Japanese Brazilians 37-89 years*	88/63	Ser/Cys, Cys/Cys vs. Ser/Ser among non-Japanese ever-smokers	0.7 (0.3, 1.3)	Unadjusted
				120/140	Ser/Cys, Cys/Cys vs. Ser/Ser among non-Japanese never-smokers	1.0 (0.6, 1.7)	Unadjusted
		Takezaki (2002)[242]	Chinese males and females*	41/143	Cys/Cys vs. Ser/Ser among ever-smokers	0.7 (0.4, 1.3)	Unadjusted
				26/84	Cys/Cys vs. Ser/Ser among never-smokers	0.9 (0.4, 2.1)	Unadjusted
	Lung cancer	Sugimura (1999)[240]	Male Japanese*	44/58	Cys/Cys vs. Ser/Ser, <800 cigarette-years	1.1 (0.5, 2.5)	Age
				62/100	Cys/Cys vs. Cys/Ser, <800 cigarette-years	0.7 (0.4, 1.4)	Age
				82/19	Cys/Cys vs. Ser/Ser, ≥ 800 cigarette-years	1.7 (0.5, 5.8)	Age
				138/49	Cys/Cys vs. Cys/Ser, ≥ 800 cigarette-years	0.6 (0.3, 1.3)	Age
		Liang (2005)[239]	Chinese males and females, 30 – 86 years of age†	102/135	Cys/Cys and Cys/Ser vs. Ser/Ser among nonsmokers	0.8 (0.4, 1.6)	Sex, age, and smoking
				132/100	Cys/Cys and Cys/Ser smokers vs. Ser/Ser among nonsmokers	0.9 (0.5, 1.6)	Sex, age, and smoking
		Wikman (2000)[251]	German males and females†	105/105	Cys/Cys and Cys/Ser vs. Ser/Ser among smokers	0.7 (0.4, 1.3)	Unadjusted
		Chen (2003)[233]	U.S. males aged 42-82†	43/178	Cys/Cys and Cys Ser vs. Ser/Ser among smokers	2.7 (1.3, 5.3)	Unadjusted
				35/73	Cys/Cys and Cys Ser vs. Ser/Ser among never smokers	1.7 (0.7, 4.0)	Unadjusted
		Sunaga (2002)[227]	Japanese males and females†	94/76	Cys/Cys vs. Ser/Ser among smokers	1.1 (0.6, 2.1)	Sex and age
				62/52	Cys/Cys vs. Cys/Ser among smokers	1.2 (0.5, 2.5)	Sex and age
				28/30	Cys/Cys vs. Ser/Ser among nonsmokers	0.5 (0.1, 1.8)	Sex and age
				65/35	Cys/Cys vs. Cys/Ser among nonsmokers	2.0 (0.7, 5.2)	Sex and age

*Study of incident disease; †Study of prevalent disease; PY, pack-years of smoking; Results were extracted with preference for unadjusted estimates

APEX1

APEX1 resides on 14q11.2 – 14q12 and processes the AP sites or single strand breaks (SSB) remaining after the damaged base has been excised by DNA glycosylases. Considered the rate-limiting step in BER[252, 253], *APEX1* hydrolyzes the DNA backbone 5' of the abastic site, producing a 3' hydroxyl group and a 5'-deoxyribose phosphate group (Figure 2)[254-256]. As AP sites are cytotoxic and mutagenic, a decreased AP site repair capacity could compromise genomic integrity. *APEX1* also removes the 3'-blocking groups remaining after SSBs have been processed by DNA glycosylases[257]. Other functions of *APEX1,* which are not further discussed, are unrelated to BER and include transcription factor stimulation by a redox-dependent mechanism[258, 259].

APEX1 has been characterized in numerous functional studies. The *E. coli APEX1* homologue *xthA* demonstrated hypersensitivity to hydrogen peroxide and UV light when AP activity was eliminated[260, 261] and yeast defective in AP site repair exhibited an elevated spontaneous mutator phenotype[262]. Depletion of *APEX1* by the overexpression of antisense mRNA resulted in hypersensitivity to the DNA damaging agents hydrogen peroxide and methyl methane-sulfonate (MMS) in a human cell line[263] and oxidative stress induced *APEX1* overexpression in Chinese hamster cells[264]. In addition, *APEX* activity is required for cultured human cells to remain viable[265, 266] and Ramana et al., (1998) demonstrated that human cells exposed to sublethal doses of oxidizing agents showed an increase in both the amount of *APEX1* as well as *APEX1* activity[267]. Mice engineered to lack *APEX1* do not survive embryogenesis, although specific dietary manipulation of pregnant females with antioxidants rescued a fraction of the litter[268-270].

APEX1 has also been shown to interact with *XRCC1*[271] and *APEX1* overexpression can compensate for *XRCC1*-deficient cells in the repair of single-strand DNA breaks (SSB) induced by oxidative DNA damage, both *in vivo* and in whole-cell extracts[272]. *APEX1* is also upregulated in animal models of atherosclerosis[273] and hypertension[274].

APEX1 genetic variants

Previous studies have identified five nonsynonymous *APEX1* variants (Table 7), although most are too rare for population-based studies. Also, the carboxy terminus of *APEX1* contains the endonuclease activity required for DNA repair and spans residues 61-318[256], whereas residues 1-127 comprise the redox domain[256, 275] (Figure 4). Thus, studies examining the role of *APEX1* in BER-mediated cancers have largely focused on SNPs in the carboxy terminus.

Table 7. Characterization of five known *APEX1* nonsynonymous SNPs.

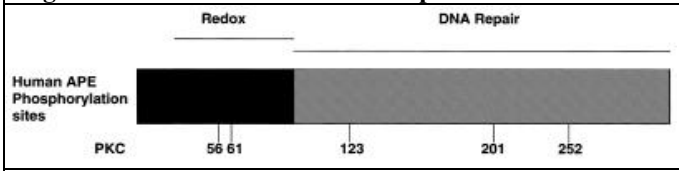
SNP*	Protein residue	MAF (Population)	Functional data?	Studied in human populations?
rs1048945	His51Gln	0.03 (CEU) 0.00 (YRI)	.	Table 8
rs2307486	Val64Ile	0.00 (CEU) 0.00 (YRI) 0.06 (HCB)	.	Table 8
rs1130409	Glu148Asp	0.49 (CEU) 0.27 (YRI)	Nishimura et al., (2002)[216] Hu et al., (2001)[276]	Table 8, Table 9
rs1803120	Ser311Pro	.	.	.
rs1803118	Val317Ala	.	.	.

*Information obtained from dbSNP (http://www.ncbi.nlm.nih.gov/projects/SNP/); **CEU**, 30 mother-father-child trios from the CEPH collection (Utah residents with ancestry from northern and western Europe), representing one of the populations studied in the International HapMap project; **HCB**, 45 unrelated Han Chinese in Beijing, China, representing one of the populations studied in the International HapMap project; **YRI**, 30 Yoruba mother-father-child trios in Ibadan, Nigeria, representing one of the populations studied in the International HapMap project;

The Asp148Glu polymorphism has been associated with hypersensitivity to ionizing radiation in genomic DNA isolated from peripheral lymphocytes[216, 276]. However, molecular modeling and amino acid conservation analyses among the ExoIII family (which consists of *E. coli* exonuclease III, *Drosophila melanogaster* Rrp1, *Arabidopsis thaliana* Arp,

mouse *APEX1*, and human *APEX1*) suggested that the Glu148Asp is unlikely to impact protein structure or function, given that it was not conserved among the ExoIII family and its position between helices. Likewise, no direct impact on endonuclease or DNA binding activities was observed for Glu148Asp, although the authors postulated that the variant may be associated with a reduced ability to communicate with other BER proteins[277] as even a slight change in DNA repair capacity could be detrimental and many functional studies are underpowered to detect subtle changes.

Figure 4. Schematic of the *APEX1* protein.



Adapted from Hsieh et al., (2001)[278]

The limited epidemiologic data examining the relationship between *APEX1* polymorphisms and cancers and related traits, has suggested a weak to null effect[279-281] (Table 8, Table 9), although multiple studies of the same outcome are few. Many investigators focused on the Asp148Glu variant, as it resides in the carboxy terminus. However, variants outside the *APEX1* DNA repair domain, such as residues 61-318 or promoter regions, could be markers for the disease causing SNP and may be informative. Also, credible modifiers were often analyzed as confounders, ignoring biologically plausible mechanisms of disease. However, most studies were underpowered to detect main effects, let alone joint effects.

Table 8. Review of seven case-control studies examining the relationship between *APEX1* polymorphisms and cancers and related traits stratified by cancer and polymorphism.

Variant	Outcome	Author (year)	Study population	No. cases/ controls	Genotype contrast	OR (95% CI)	Covariate adjustments
Gln51His							
	Lung cancer	Zienolddiny (2006)[282]	Norwegian males and females*	287/310	His/His vs. Gln/Gln	0.9 (0.3, 2.2)	Age, sex, and PY
				304/324	Gln/His vs. Gln/Gln	1.1 (0.6, 2.0)	Age, sex, and PY
Ile64Val							
	Esophageal squamous cell carcinoma	Hao (2004)[283]	Chinese males and females†	414/479	Val/Val, Val/Ile vs. Ile/Ile	1.1 (0.7, 1.7)	Age, sex, and smoking
	Lung cancer	Zienolddiny (2006)[282]	Norwegian males and females*	340/410	Val/Val, Val/Ile vs. Ile/Ile	0.6 (0.4, 0.8)	Age, sex, and PY
Asp148Glu							
	Breast cancer	Zhang (2006)[284]	U.S. females aged 20-74 years*	777/617	Glu/Asp vs. Asp/Asp	1.0 (0.8, 1.3)	Age and study site
				1156/917	Glu/Glu vs. Asp/Asp	1.0 (0.9, 1.2)	Age and study site
	Esophageal squamous cell carcinoma	Hao (2004)[283]	Chinese males and females†	198/244	Glu/Asp vs. Asp/Asp	1.2 (0.8, 1.7)	Age, sex, and smoking
				337/383	Glu/Glu vs. Asp/Asp	1.2 (0.9, 1.8)	Age, sex, and smoking
	Lung cancer	Ito (2004)[280]	Japanese males and females*	94/223	Glu/Glu vs. Asp/Asp	1.3 (0.8, 2.2)	Unadjusted
				146/385	Glu/Asp vs. Asp/Asp	1.0 (0.7, 1.4)	Unadjusted
		Popanda (2004)[281]	German males and females aged 28-84 years*	98/224	Glu/Glu vs. Asp/Asp	0.8 (0.5, 1.1)	Unadjusted
				165/351	Glu/Asp vs. Asp/Asp	0.9 (0.6, 1.2)	Unadjusted
		Zienolddiny (2006)[282]	Norwegian males and females*	197/160	Glu/Glu vs. Asp/Asp	0.8 (0.5, 1.1)	Age, sex, and PY
				184/198	Glu/Asp vs. Asp/Asp	1.2 (0.8, 1.9)	Age, sex, and PY
	Pancreatic cancer	Jiao (2006)[285]	U.S. males and females†	187/156	Glu/Glu vs. Asp/Asp	0.9 (0.6, 1.3)	Age, sex, race, and PY
				288/259	Glu/Asp vs. Asp/Asp	0.8 (0.6, 1.2)	Age, sex, race, and PY
	Oral Cleft	Olshan (2005)[250]	California infants born 1983-1986†	58/165	Glu/Glu vs. Asp/Asp	1.1 (0.6, 2.1)	Race/ethnicity
				91/254	Glu/Asp vs. Asp/Asp	1.1 (0.7, 1.9)	Race/ethnicity
	Spina bifida	Olshan (2005)[250]	California infants born 1983-1986†	72/165	Glu/Glu vs. Asp/Asp	0.5 (0.3, 1.0)	Race/ethnicity
				105/254	Glu/Asp vs. Asp/Asp	0.6 (0.4, 1.0)	Race/ethnicity

*Study of incident disease; †Study of prevalent disease; PY, pack-years of smoking; Results were extracted with preference for unadjusted estimates

Table 9. Review of four case control studies examining the relationship between *APEX1* variants and cancers and related traits, stratified by smoking status.

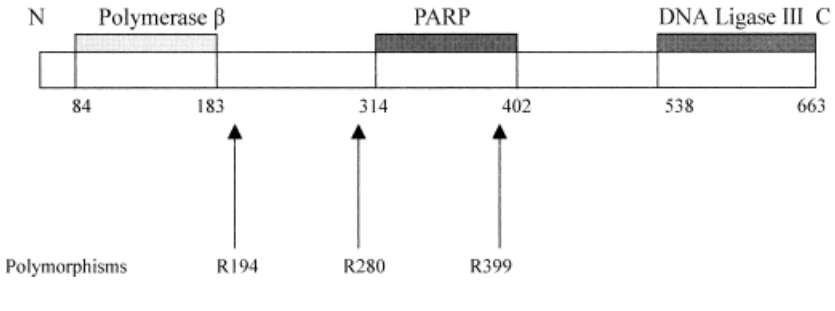
Variant	Outcome	Author (year)	Study population	No. cases/controls	Genotype contrast	OR (95% CI)	Covariate adjustments
Asp148Glu							
	Bladder cancer	Matullo (2006)[286]	European males and females*	55/568	Glu/Glu vs. Asp/Asp among nonsmokers	1.0 (0.5, 1.9)	Unadjusted
				100/835	Glu/Asp vs. Asp/Asp among nonsmokers	1.4 (0.8, 2.4)	Unadjusted
		Terry (2006)[287]	U.S. males and females†	81/62	Glu/Glu vs. Asp/Asp among ever smokers	0.9 (0.4, 1.7)	Age, sex, and race
				157/105	Glu/Asp vs. Asp/Asp among ever smokers	1.2 (0.7, 2.0)	Age, sex, and race
	Lung cancer	Ito (2004)[280]	Japanese males and females*	70/132	Glu/Glu vs. Asp/Asp among ever smokers	1.7 (0.9, 3.2)	Unadjusted
				24/91	Glu/Glu vs. Asp/Asp among never smokers	0.6 (0.2, 1.4)	Unadjusted
				98/232	Glu/Asp vs. Asp/Asp among ever smokers	0.9 (0.6, 1.4)	Unadjusted
				48/153	Glu/Asp vs. Asp/Asp among never smokers	1.0 (0.5, 2.0)	Unadjusted
		Matullo (2006)[286]	European males and females*	60/568	Glu/Glu vs. Asp/Asp among nonsmokers	0.8 (0.4, 1.5)	Unadjusted
				99/835	Glu/Asp vs. Asp/Asp among nonsmokers	0.6 (0.4, 1.1)	Unadjusted
		Misra (2003)[279]	Male Finns 50-69 years of age*	143/142	Glu/Glu vs. Asp/Asp among ever smokers	0.9 (0.6, 1.5)	Smoking years and cigarettes/day
				231/225	Glu/Asp vs. Asp/Asp among ever smokers	1.1 (0.7, 1.7)	Smoking years and cigarettes/day

*Study of incident disease; †Study of prevalent disease; Results were extracted with preference for unadjusted estimates

XRCC1

XRCC1 is a SSB binding protein that maps to 19q13.2[288] and was the first mammalian gene implicated in cellular sensitivity to ionizing radiation[289]. While *XRCC1* has no known catalytic activity, it recognizes and binds single-strand DNA breaks[290] and is thought to complex with other BER components during short-patch DNA repair via its role as a chaperone or central scaffolding protein for DNA ligase III (responsible for sealing the nick)[291, 292], DNA polymerase β (*pol* β; polymerase that fills in nucleotide sequence gaps)[293], and *PARP* (poly ADP-ribose polymerase)[291, 293] (Figure 5). Research also supports a role for *XRCC1* in the single-strand break repair (SSBR) pathway[294, 295] and the maintenance of genetic stability in noncycling and postmitotic cell cycle stages[296].

Figure 5. Schematic of the *XRCC1* protein.



Adapted from Lee et al., (2001)[297]

Elevated frequencies (10-fold or higher) of spontaneous sister chromatid exchange (SCE, the exchange of genetic material between two identical sister chromatids) were observed in EM9 hamster cells that lacked two-thirds of the normal hamster *XRCC1* sequence compared to wild type cells, which was thought to reflect a deficiency in rejoining SSBs. Similar SCE phenotypes were observed when EM9 cells were exposed to compounds capable of inducing SSB, such as alkylating agents and ionizing radiation[298]. EM9 cells were also unable to grow under conditions in which 20% of thymine bases are replaced with chlorouracil, a well-known mutagen, whereas wild-type cells remained viable[299]. *XRCC1* mRNA and protein levels

were also elevated in malignant prostate cells when compared to normal epithelial cells. Despite the increased *XRCC1* expression, the malignant cells also exhibited a defective oxidative base and SSB repair phenotype, suggesting that prostate tumorgenesis may reflect aberrant DNA repair capacity[300].

XRCC1 upregulation was also associated with induced atherosclerotic plaques in male New Zealand White rabbits[273] and Rossi et al., (2004) demonstrated increased *XRCC1* expression in tissue from stable angina plaques, compared to acute coronary syndrome atherectomies[301]. *XRCC1* transcription levels were also elevated in diabetic patients when compared to non-diabetic patients, suggesting that the ROS metabolites produced under the hyperglycemic state are mediated by *XRCC1* expression[302].

XRCC1 genetic variants

While multiple *XRCC1* polymorphisms have been identified (Table 10), the functional consequences are not well understood and population-based and laboratory research has largely focused on the Arg194, Arg280 and Arg399 variants. The Arg194 and Arg280 variants reside between the binding domain of *pol* β and *PARP* while codon 399 is positioned near the *PARP* binding domain[303] (Figure 5). Savas and colleagues (2004) performed protein conservation analysis on *XRCC1* nonsynonymous polymorphisms in an attempt to predict whether an amino acid substitution may impact protein function and predicted that the Leu7Arg, Ala72Val, Leu161Pro, Arg280His, Met381Val, and Arg399Gln variants "possibly damaged" protein function[304].

Table 10. Characterization of 21 known *XRCC1* nonsynonymous SNPs.

SNP*	Protein residue	MAF (Population)	Functional data?	Studied in human populations?
rs2307177	Thr576Asn	0.02 (NIHPDR)	.	.
rs2682557	Tyr576Asn	.	.	.
rs2307166	Trp560Arg	0.001 (NIHPDR)	.	.
rs2307167	Gln559Arg	0.001 (NIHPDR)	.	.
rs25474	Leu514Pro	0.00 (CEU) 0.00 (YRI)	.	.
rs2307184	Tyr485Ser	0.00 (CEU) 0.00 (YRI)	.	.
rs25487	Arg399Gln	0.42 (CEPH) 0.10 (YRI)	Qu et al., (2005)[303] Takanami et al., (2005)[305] Abdel-Rahman et al., (2000)[306] Cornetta et al., (2006)[307] Wang et al., (2003)[308] Savas et al., (2004)[304] Pachkowski et al., (2006)[309]	Table 11, Table 12
rs2271980	Met381Val	.	Savas et al., (2004)[304]	.
rs25491	Ser309Pro	0.00 (CEU) 0.00 (YRI)	.	.
rs25490	Ala304Thr	0.00 (CEPH) 0.04 (AGI ASP)	.	.
rs2307188	Asn298Lys	0.004 (NIHPDR)	.	.
rs25489	His280Arg	0.00 (CEU) 0.025 (YRI)	Qu et al., (2005)[303] Takanami et al., (2005)[305] Savas et al., (2004)[304] Pachkowski et al., (2006)[309]	Table 11, Table 12
rs1799782	Trp194Arg	0.05 (CEPH) 0.039 (AGI ASP)	Qu et al., (2005)[303] Takanami et al., (2005)[305] Wang et al., (2003)[308]	Table 11, Table 12
rs2307191	Leu161Pro	0.011 (HCB) 0.00 (YRI)	Savas et al., (2004)[304]	.
rs2307180	Lys157Glu	0.001 (NIHPDR)	.	.
rs2228487	His107Arg	0.021 (AFD EUR) 0.00 (AFD AFR)	.	.
rs25496	Ala72Val	0.02 (NIHPDR) 0.063 (AFR1)	Savas et al., (2004)[304]	.
rs25495	X51Lys	0.00 (CEPH) 0.00 (AFR1)	.	.
rs2307171	Met10Val	0.001 (NIHPDR)	.	.
rs2307186	Leu7Arg	0.003 (NIHPDR)	Savas et al., (2004)[304]	.
rs11553659	His5Arg	.	.	.

*Information obtained from dbSNP (http://www.ncbi.nlm.nih.gov/projects/SNP/); **AFD AFR**, 23 samples of African American descent from the Coriell Cell Repository selected from the human variation panel of 50 African Americans by the SeattleSNPs Program for Genomic Applications; **AFD EUR**, 24 samples from the Coriell Cell Repository are primarily of European American descent and consist of 23 unrelated CEPH parents selected by the SeattleSNPs Program for Genomic Applications, plus one sample from Coriell's human variation panel of 50 Caucasians; **AFR1**, Human individual DNA from 24 individuals of self-described African/African-American heritage; **AGI ASP,** Samples from Coriell Cell Repositories Apparently Normal Collection of Caucasian and African-American females; **CEPH**, Genomic DNA samples obtained for a panel of 92 unrelated individuals chosen from Centre d'Etude du Polymorphisme Human (CEPH) pedigrees comprised of UTAH (93%), French (4%), and Venezuelan (3%) samples purchased from Coriell Cell Repository; **CEU**, 30 mother-father-child trios from the CEPH collection, one of the populations studied in the HapMap project; **HCB**, 45 unrelated Han Chinese in Beijing, China, one of the populations studied in the HapMap project; **NIHPDR**, The NIH Polymorphism Discovery Resource (NIHPDR) contains cell lines and DNA from 450 anonymous, unrelated individuals with equal numbers of females and males. Non- Europeans were sampled at frequencies higher than the general U.S. population to enrich the genetic variability; **YRI**, 30 Yoruba mother-father-child trios in Ibadan, Nigeria, one of the populations studied in the HapMap project.

As *XRCC1* acts as a scaffold for numerous proteins, changes in the amino acid structure could enhance or reduce protein binding. For example, the Arg194Trp variant is located in an area rich in proline, serine, arginine, and lysine residues. Thus, a mutation from arginine to tryptophan would exchange a positively charged arginine for a hydrophobic tryptophan, possibly affecting protein binding and DNA repair efficacy[310]. Knock-in mouse models of the Arg194 and Arg280 variants are viable, but have not yet been characterized while the Arg399Gln mouse model is still under construction (Ladiges *et al.,* unpublished data).

One study of Arg194Trp, Arg280His, and Arg399Gln polymorphisms in normal human *XRCC1* cDNA and EM9 hamster cells, which lack the full DNA sequence necessary for *XRCC1* function, suggested that cDNA containing the Arg194Trp and Arg280His variants fully restored the phenotype, while the *XRCC1* cDNA containing the Arg399Gln variant did not[303]. However, studies of *XRCC1* variants in EM9 cells are inconsistent, as one analysis of the 280His and 399Gln variant proteins demonstrated that only the 280His variant accumulated SSB after exposure to hydrogen peroxide, methanesulfonate, or camptothecin, chemicals chosen to mimic the genotoxicity of cigarette smoke exposure[309] and another examining the same variants and the alkylating agent MMS found that Arg280His only partially restored the MMS sensitivity[305].

In studies of human lymphocytes, cells with the 399Gln polymorphism were slightly more sensitive to SCE induction by the tobacco-specific NNK[306] and had a higher percentage of damaged DNA following X-ray irradiation[307]. Mutation assays using bleomycin, an agent that mimics the effects of radiation by generating ROS, and benzo(a)pyrene-diol-epoxide (BPDE), a highly toxic intermediate of benzo(a)pyrene inactivation, suggested that cells

homozygous for the 194Arg or 399Gln variant had higher numbers of chromosomal breaks per cell when either agent was applied[308]. Duell et al., (2000) also demonstrated that the 399Gln variant was associated with a modestly elevated frequency of SCE among healthy smokers[311] and carriers of the 399Gln variant had an elevated SCE frequency in a study of male resin synthesis employees[312].

The relationship between *XRCC1* variants and cancers and related traits has been contradictory (Table 11). A meta-analysis of 16 published studies examining the Arg194Trp polymorphism in tobacco-related cancers (lung, upper aerodigestive tract, and bladder) estimated a summary OR (95% CI) of 0.86 (0.77, 0.95) for the 194Trp contrast and a case-only interaction odds ratio for tobacco smoking, the 194Trp contrast, and tobacco-related cancers of 0.80 (0.56, 1.16) using five studies. Studies examining the Arg280His or Arg399Gln variants across numerous cancers including lung, upper aerodigestive tract, bladder, breast, and skin neoplasms appeared too heterogeneous to warrant a summary measure[313].

One notable observation is the marked variation in *XRCC1* MAFs observed between populations. For example, while Shen et al., (2005)[314] and Zhang et al., (2006)[284] had sample sizes exceeding 2,000 U.S. females for the investigation of the Arg194Trp variant and breast cancer, only 0.6% of their samples consisted of Trp/Trp homozygotes. However, the frequency of Trp/Trp homozygotes in studies of Korean[315] and Chinese[283] populations exceeded 10%. As differences in allele frequencies are not uncommon between populations, it is entirely possible that susceptibility variants and/or disease markers differ between populations, which may help explain the inconsistent results.

While 11 studies examined modification by cigarette smoking (Table 12), power was generally limited and adjustment strategies and exposure classifications varied. For example, Hung et al., (2005)[316], who had one of the largest samples for the investigation of the Arg399 variant and lung cancer, analyzed modification by smoking by dividing pack-years of exposure into three equal groups. The results, although imprecise, appeared relatively homogeneous across strata. As with studies of other BER variants, ICR estimates were not reported.

Table 11. Review of 45 case control studies examining the relationship between the *XRCC1* polymorphisms and cancers and related traits stratified by cancer and polymorphism.

Variant	Outcome	Author (year)	Study population	No. cases/ controls	Genotype contrast	OR (95% CI)	Covariate adjustments
Arg194Trp							
	Bladder cancer	Matullo (2005)[317]	Italian males aged 34-76 years*	315/313	Trp/Trp, Trp/Arg vs. Arg/Arg	0.7 (0.5, 1.1)	Age and smoking
		Stern (2001)[318]	African American and Caucasian U.S. males and females†	222/210	Trp/Trp, Trp/Arg vs. Arg/Arg	0.6 (0.3, 1.0)	Age, sex, and race
	Breast cancer	Chacko (2005)[319]	Indian females†	88/100	Trp/Trp vs. Arg/Arg	2.7 (0.8, 9.2)	Unadjusted
				118/119	Trp/Arg vs. Arg/Arg	1.8 (1.0, 3.4)	Unadjusted
		Kim (2002)[315]	Korean females†	111/119	Trp/Trp vs. Arg/Arg	0.9 (0.5, 1.7)	Unadjusted
				182/178	Trp/Arg vs. Arg/Arg	1.1 (0.8, 1.7)	Unadjusted
		Patel (2005)[320]	U.S. females 50-74 years at enrollment*	485/485	Trp/Trp, Trp/Arg vs. Arg/Arg	0.7 (0.5. 1.0)	Unadjusted
		Shen (2005)[321]	U.S. females*	1066/1108	Trp/Trp, Trp/Arg vs. Arg/Arg	0.9 (0.7, 1.2)	Age
		Smith (2003)[322]	U.S. female§*	246/266	Trp/Trp, Trp/Arg vs. Arg/Arg	1.6 (0.9, 2.9)	Age, family history, age at first live birth, and BMI
		Thyagarajan (2006)[323]	U.S. females*	432/322	Trp/Trp ,Trp/Arg vs. Arg/Arg	1.2 (0.8, 1.8)	Unadjusted
		Zhang (2006)[284]	U.S. females aged 20-74 years*	1391/1097	Trp/Trp vs. Arg/Arg	0.5 (0.2, 1.4)	Age and study site
				1573/1086	Trp/Arg vs. Arg/Arg	0.9 (0.8, 1.2)	Age and study site
	Colon adenoma	Skjelbred (2006)[324]	Norwegian males and females*	983/399	Trp/Trp, Trp/Arg vs. Arg/Arg	1.1 (0.7, 1.6)	Age
		Stern (2005)[325]	U.S. males and females aged 50-74 years*	598/649	Trp/Trp vs. Arg/Arg	0.7 (0.2, 1.9)	Age, sex, race, clinic, and exam date
				732/778	Trp/Arg vs. Arg/Arg	1.1 (0.8, 1.5)	Age, sex, race, clinic, and exam date
	Colon carcinoma	Skjelbred (2006)[324]	Norwegian males and females*	156/399	Trp/Trp, Trp/Arg vs. Arg/Arg	1.0 (0.4, 2.3)	Age
	Esophageal squamous cell carcinoma	Hao (2004)[283]	Chinese males and females†	249/269	Trp/Trp vs. Arg/Arg	1.2 (0.7, 2.0)	Age, sex, and smoking
				369/440	Trp/Arg vs. Arg/Arg	0.8 (0.6, 1.1)	Age, sex, and smoking
		Lee (2001)[297]	Taiwanese males and females*	58/144	Trp/Trp vs. Arg/Arg	0.5 (0.2, 1.5)	Unadjusted
				101/245	Trp/Arg vs. Arg/Arg	0.9 (0.6, 1.4)	Unadjusted
		Ratnasinghe (2004)[326]	Chinese males and females*	131/454	Trp/Trp and Trp/Arg vs. Arg/Arg	0.9 (0.6, 1.4)	Sex, age, smoking, drinking, and center
		Xing (2002)[327]	Chinese males and females†	269/296	Trp/Trp vs. Arg/Arg	1.9 (1.2, 3.0)	Unadjusted

			375/487	Trp/Arg vs. Arg/Arg	0.9 (0.7, 1.2)	Unadjusted
	Yu (2004)[328]	Chinese males and females*	86/92	Trp/Trp vs. Arg/Arg	1.1 (0.4, 3.3)	Unadjusted
			131/148	Trp/Arg vs. Arg/Arg	0.9 (0.5, 1.4)	Unadjusted
Gastric cancer	Lee (2002)[329]	South Korean males and females*	115/86	Trp/Trp vs. Arg/Arg	0.8 (0.4, 1.8)	Age and sex
			174/158	Trp/Arg vs. Arg/Arg	0.6 (0.4, 1.0)	Age and sex
	Shen (2000)[330]	Chinese males and females*	111/89	Trp/Trp vs. Arg/Arg	0.6 (0.3, 1.2)	Unadjusted
			173/147	Trp/Arg vs. Arg/Arg	0.7 (0.5, 1.1)	Unadjusted
Lung cancer	Chen (2002)[331]	Chinese males and females*	59/62	Trp/Trp vs. Arg/Arg	1.6 (0.5, 5.6)	PYs and *GSTM1* genotype
			92/97	Trp/Arg vs. Arg/Arg	1.6 (0.8, 3.0)	PYs and *GSTM1* genotype
	Hung (2005)[316]	Eastern European males and females*	1888/1840	Trp/Trp vs. Arg/Arg	1.2 (0.5, 2.9)	Country, age at diagnosis, sex, and PY
			2137/2120	Trp/Arg vs. Arg/Arg	0.9 (0.7, 1.1)	Country, age at diagnosis, sex, and PY
	Ratnasinghe (2001)[332]	Chinese male and female tin miners*	61/106	Trp/Trp vs. Arg/Arg	0.7 (0.3, 1.6)	Unadjusted
			99/189	Trp/Arg vs. Arg/Arg	0.7 (0.4, 1.2)	Unadjusted
	Zienolddiny (2006)[282]	Norwegian males and females*	316/405	Trp/Trp, Trp/Arg vs. Arg/Arg	0.9 (0.5, 1.5)	Age, sex, and PY
Melanoma	Han (2004)[333]	U.S. females aged 30-55 years*	215/863	Trp/Trp, Trp/Arg vs. Arg/Arg	1.2 (0.8, 1.9)	Age and race
Nasopharyngeal carcinoma	Cao (2006)[334]	Chinese males and females†	251/278	Trp/Trp vs. Arg/Arg	0.4 (0.3, 0.8)	Unadjusted
			398/452	Trp/Arg vs. Arg/Arg	0.8 (0.6, 1.0)	Unadjusted
Pancreatic cancer	Jiao (2006)[285]	U.S. males and females†	133/304	Trp/Trp vs. Arg/Arg	0.9 (0.2, 4.6)	Age, sex, race, and PY
			179/335	Trp/Arg vs. Arg/Arg	1.4 (0.9, 2.3)	Age, sex, race, and PY
Renal cell carcinoma	Hirata (2006)[335]	Japanese males and females aged 29-84*	69/121	Trp/Trp vs. Arg/Arg	1.1 (0.5, 2.4)	Unadjusted
			98/158	Trp/Arg vs. Arg/Arg	1.3 (0.8, 2.2)	Unadjusted
SCCHN	Demokan (2005)[336]	Turkish males and females*	81/90	Trp/Trp vs. Arg/Arg	1.7 (0.3, 10)	Unadjusted
			92/96	Trp/Arg vs. Arg/Arg	2.0 (0.8, 5)	Unadjusted
	Olshan (2002)[337]	U.S. male and female Caucasians	98/161	Trp/Trp and Trp/Arg vs. Arg/Arg	1.3 (0.6, 2.9)	Age and sex
	Sturgis (1999)[338]	U.S. male and female African Americans, Caucasians, and Latinos	203/424	Trp/Trp and Trp/Arg vs. Arg/Arg	0.8 (0.4, 1.3)	Age, sex, race, and alcohol
Pro206Pro						

	Bladder cancer	Matullo (2005)[317]	Italian males aged 34-76 years*	147/165	AA vs. GG	1.6 (1.0, 2.7)	Age and smoking
				226/241	AG vs. GG	1.8 (1.2, 2.7)	Age and smoking
Arg280His							
	Bladder cancer	Stern (2001)[318]	African American and Caucasian U.S. males and females†	233/208	His/His and His/Arg vs. Arg/Arg	1.2 (0.6, 2.6)	Age, sex, and race
	Breast cancer	Chacko (2005)[319]	Indian females†	103/91	His/His vs. Arg/Arg	1.8 (0.3, 10)	Unadjusted
				119/121	His/Arg vs. Arg/Arg	0.6 (0.3, 1.1)	Unadjusted
		Metsola (2005)[339]	Finnish females*	480/479	His/His and His/Arg vs. Arg/Arg	1.2 (0.8, 1.7)	Age, age at menarche, age at first birth, number of pregnancies, history of breast disease, WHR, family history of breast cancer, smoking and alcohol
		Zhang (2006)[284]	U.S. females aged 20-74 years*	1407/1125	His/His vs. Arg/Arg	1.0 (0.2, 4.3)	Age and study site
				1560/1235	His/Arg vs. Arg/Arg	1.1 (0.8, 1.1)	Age and study site
	Colon adenoma	Skjelbred (2006)[324]	Norwegian males and females*	983/399	His/Arg vs. Arg/Arg	1.7 (1.0, 2.9)	Age
	Colon carcinoma	Skjelbred (2006)[324]	Norwegian males and females*	157/399	His/Arg vs. Arg/Arg	1.7 (0.6, 5)	Age
	Esophageal squamous cell carcinoma	Hao (2004)[283]	Chinese males and females†	348/384	His/His vs. Arg/Arg	1.5 (0.2, 9.3)	Age, sex, and smoking
				412/478	His/Arg vs. Arg/Arg	0.8 (0.6, 1.1)	Age, sex, and smoking
		Lee (2001)[297]	Taiwanese males and females*	105/264	His/His, His/Arg vs. Arg/Arg	1.4 (0.8, 2.4)	Unadjusted
	Gastric cancer	Lee (2002)[329]	South Korean males and females*	190/172	His/His, His/Arg vs. Arg/Arg	1.5 (0.9, 2.4)	Age and sex
	Lung cancer	Hung (2005)[316]	Eastern European males and females*	1907/1902	His/His vs. Arg/Arg	1.3 (0.4, 4.1)	Country, age at diagnosis, sex, and PY
				2082/2086	His/Arg vs. Arg/Arg	0.9 (0.7, 1.2)	Country, age at diagnosis, sex, and PY
		Ratnasinghe (2001)[332]	Chinese male and female tin miners*	106/209	His/His, His/Arg vs. Arg/Arg	1.6 (0.9, 2.9)	Unadjusted
		Vogel (2004)[340]	Danish males and females 50-67 years at enrollment*	256/289	His/His, His/Arg vs. Arg/Arg	1.1 (0.6, 1.9)	Unadjusted
		Zienolddiny (2006)[282]	Norwegian males and females*	329/377	His/His, His/Arg vs. Arg/Arg	1.5 (0.9, 2.7)	Age, sex, and PY
	Nasopharyngeal carcinoma	Cho (2003)[243]	Taiwanese males and females†	332/283	His/His, His/Arg vs. Arg/Arg	0.6 (0.4, 1.0)	Age, sex, and ethnicity
Arg399Gln							
	Bladder cancer	Kelsey (2004)[341]	U.S. males and females*	168/314	Gln/Gln vs. Arg/Arg	0.8 (0.5, 1.2)	Age, sex, and PY

			219/458	Gln/Arg vs. Arg/Arg	1.4 (1.0, 1.9)	Age, sex, and PY
	Matullo (2005)[317]	Italian males aged 34-76 years*	176/167	Gln/Gln vs. Arg/Arg	0.7 (0.3, 1.8)	Age and smoking
			171/165	Gln/Arg vs. Arg/Arg	0.8 (0.5, 1.2)	Age and smoking
	Sanyal (2004)[342]	Swedish males and females ages 33-96 years†	156/136	Gln/Gln vs. Arg/Arg	1.3 (0.7, 2.4)	Unadjusted
			279/223	Gln/Arg vs. Arg/Arg	1.3 (0.9, 1.9)	Unadjusted
	Shen(2003)[343]	Italian males aged 20-80 years*	114/116	Gln/Gln vs. Arg/Arg	0.9 (0.5, 1.7)	Age
			180/190	Gln/Arg vs. Arg/Arg	0.9 (0.6, 1.3)	Age
	Stern (2001)[318]	African American and Caucasian U.S. males and females†	117/114	Gln/Gln vs. Arg/Arg	0.7 (0.4, 1.4)	Age, sex, and race
			212/184	Gln/Arg vs. Arg/Arg	1.1 (0.7, 1.6)	Age, sex, and race
Breast cancer	Chacko (2005)[319]	Indian females†	73/88	Gln/Gln vs. Arg/Arg	2.7 (1.1, 6)	Unadjusted
			106/114	Gln/Arg vs. Arg/Arg	2.0 (1.2, 3.5)	Unadjusted
	Figueiredo (2004)[344]	White Canadian females 25-54 years*	223/217	Gln/Gln vs. Arg/Arg	0.9 (0.6, 1.4)	Unadjusted
			347/345	Gln/Arg vs. Arg/Arg	0.9 (0.7, 1.2)	Unadjusted
	Kim (2002)[315]	Korean females†	72/66	Gln/Gln vs. Arg/Arg	3.8 (1.4, 10)	Unadjusted
			104/109	Gln/Arg vs. Arg/Arg	1.2 (0.7, 2.1)	Unadjusted
	Metsola (2005)[339]	Finnish females*	283/293	Gln/Gln vs. Arg/Arg	1.4 (0.8, 2.3)	Age, age at menarche, age at first birth, number of pregnancies, history of breast disease, WHR, family history of breast cancer, smoking and alcohol
			433/441	Gln/Arg vs. Arg/Arg	1.2 (0.9, 1.7)	Age, age at menarche, age at first birth, number of pregnancies, history of breast disease, WHR, family history of breast cancer, smoking and alcohol
	Patel (2005)[320]	U.S. females 50-74 years at enrollment*	257/250	Gln/Gln vs. Arg/Arg	1.1 (0.7, 1.6)	Unadjusted
			389/396	Gln/Arg vs. Arg/Arg	1.0 (0.7, 1.3)	Unadjusted
	Shen (2005)[321]	U.S. females*	528/574	Gln/Gln vs. Arg/Arg	1.0 (0.7, 1.3)	Age
			951/980	Gln/Arg vs. Arg/Arg	1.1 (0.9, 1.3)	Age
	Shu (2003)[345]	Chinese females aged 25-64 years*	646/684	Gln/Gln vs. Arg/Arg	1.2 (0.9, 1.7)	Age
			1003/1108	Gln/Arg vs. Arg/Arg	0.9 (0.8, 1.1)	Age
	Smith (2003)[322]	U.S. females*	129/144	Gln/Gln vs. Arg/Arg	1.1 (0.6, 2.0)	Age, family history, age at first live birth, and BMI

			221/238	Gln/Arg vs. Arg/Arg	1.0 (0.7, 1.5)	Age, family history, age at first live birth, and BMI
	Thyagarajan (2006)[323]	U.S. females*	117/182	Gln/Gln vs. Arg/Arg	0.9 (0.5, 1.7)	Unadjusted
			133/175	Gln/Arg vs. Arg/Arg	1.3 (0.9, 2.0)	Unadjusted
	Zhang (2006)[284]	U.S. females aged 20-74 years*	1606/1414	Gln/Gln vs. Arg/Arg	0.9 (0.8, 1.1)	Age and study site
			2647/2227	Gln/Arg vs. Arg/Arg	1.1 (0.9, 1.2)	Age and study site
Colon adenoma	Skjelbred (2006)[324]	Norwegian males and females*	540/212	Gln/Gln vs. Arg/Arg	0.9 (0.6, 1.3)	Age
			834/335	Gln/Arg vs. Arg/Arg	0.8 (0.6, 1.1)	Age
	Stern (2005)[325]	U.S. males and females aged 50-74 years*	402/459	Gln/Gln vs. Arg/Arg	0.7 (0.5, 1.0)	Age, sex, race, clinic, and exam date
			676/688	Gln/Arg vs. Arg/Arg	1.1 (0.9, 1.3)	Age, sex, race, clinic, and exam date
Colon carcinoma	Skjelbred (2006)[324]	Norwegian males and females*	87/212	Gln/Gln vs. Arg/Arg	0.9 (0.4, 2.0)	Age
			133/335	Gln/Arg vs. Arg/Arg	0.7 (0.4, 1.3)	Age
Esophageal squamous cell carcinoma	Cai (2006)[346]	Chinese males and females*	75/174	Gln/Gln vs. Arg/Arg	1.7 (0.9, 3.0)	Age, sex, education, BMI, smoking, and alcohol
			179/350	Gln/Arg vs. Arg/Arg	1.6 (1.1, 2.4)	Age, sex, education, BMI, smoking, and alcohol
	Hao (2004)[283]	Chinese males and females†	257/282	Gln/Gln vs. Arg/Arg	1.2 (0.7, 2.1)	Age, sex, and smoking
			377/446	Gln/Arg vs. Arg/Arg	0.8 (0.6, 1.0)	Age, sex, and smoking
	Lee (2001)[297]	Taiwanese males and females*	72/156	Gln/Gln vs. Arg/Arg	0.7 (0.3, 1.6)	Unadjusted
			97/240	Gln/Arg vs. Arg/Arg	0.6 (0.4, 1.0)	Unadjusted
	Ratnasinghe (2004)[326]	Chinese males and females*	131/454	Gln/Gln and Gln/Arg vs. Arg/Arg	0.7 (0.5, 1.1)	Sex, age, smoking, drinking, and center
	Xing (2002)[327]	Chinese males and females†	286/328	Gln/Gln vs. Arg/Arg	0.8 (0.5, 1.3)	Unadjusted
			398/475	Gln/Arg vs. Arg/Arg	0.8 (0.6, 1.0)	Unadjusted
	Yu (2004)[328]	Chinese males and females*	84/93	Gln/Gln vs. Arg/Arg	5.2 (2.4, 11)	Unadjusted
			106/147	Gln/Arg vs. Arg/Arg	1.2 (0.7, 1.9)	Unadjusted
Gastric cancer	Huang (2005)[347]	Polish males and females aged 21-79 years*	160/211	Gln/Gln vs. Arg/Arg	1.0 (0.6, 1.7)	Age, sex, and smoking
			245/345	Gln/Arg vs. Arg/Arg	1.1 (0.7, 1.7)	Age, sex, and smoking
	Lee (2002)[329]	South Korean males and females*	119/103	Gln/Gln vs. Arg/Arg	0.9 (0.3, 2.3)	Age and sex

			181/163	Gln/Arg vs. Arg/Arg	0.9 (0.6, 1.4)	Age and sex
	Shen (2000)[330]	Chinese males and females*	105/107	Gln/Gln vs. Arg/Arg	1.0 (0.4, 2.3)	Unadjusted
			178/153	Gln/Arg vs. Arg/Arg	1.5 (1.0, 2.3)	Unadjusted
Lung cancer	Chen (2002)[331]	Chinese males and females*	60/59	Gln/Gln vs. Arg/Arg	1.0 (0.5, 1.8)	PYs and *GSTM1* genotype
			98/92	Gln/Arg vs. Arg/Arg	0.3 (0.03, 3.2)	PYs and *GSTM1* genotype
	Hung (2005)[316]	Eastern European males and females*	1098/1134	Gln/Gln vs. Arg/Arg	0.9 (0.8, 1.2)	Country, age at diagnosis, sex, and PY
			1795/1755	Gln/Arg vs. Arg/Arg	1.1 (0.9, 1.2)	Country, age at diagnosis, sex, and PY
	Ito (2004)[280]	Japanese males and females*	112/279	Gln/Gln vs. Arg/Arg	1.4 (0.7, 2.8)	Crude
			164/422	Gln/Arg vs. Arg/Arg	1.0 (0.7, 1.4)	Unadjusted
	Park (2002)[348]	South Korean males*	117/87	Gln/Gln vs. Arg/Arg	2.3 (0.9, 6)	Unadjusted
			175/1292	Gln/Arg vs. Arg/Arg	1.3 (0.8, 2.0)	Unadjusted
	Popanda (2004)[281]	German males and females aged 28-84 years*	111/238	Gln/Gln vs. Arg/Arg	0.9 (0.5, 1.5)	Unadjusted
			175/393	Gln/Arg vs. Arg/Arg	0.9 (0.6, 1.3)	Unadjusted
	Ratnasinghe (2001)[332]	Chinese male and female tin miners*	67/128	Gln/Gln vs. Arg/Arg	1.4 (0.5, 3.7)	Unadjusted
			99/197	Gln/Arg vs. Arg/Arg	1.0 (0.6, 1.6)	Unadjusted
	Vogel (2004)[340]	Danish males and females 50-67 years at enrollment*	152/148	Gln/Gln vs. Arg/Arg	0.9 (0.5, 1.6)	Unadjusted
			221/229	Gln/Arg vs. Arg/Arg	0.9 (0.6, 1.3)	Unadjusted
	Zhang (2005)[349]	Chinese males and females*	637/620	Gln/Gln vs. Arg/Arg	1.2 (0.8, 1.6)	Age, sex, smoking and PY
			898/911	Gln/Arg vs. Arg/Arg	0.9 (0.8, 1.1)	Age, sex, smoking and PY
	Zienolddiny (2006)[282]	Norwegian males and females*	160/205	Gln/Gln vs. Arg/Arg	0.7 (0.4, 1.1)	Unadjusted
			300/337	Gln/Arg vs. Arg/Arg	1.1 (0.8, 1.5)	Unadjusted
Melanoma	Han (2004)[333]	U.S. females aged 30-55 years*	105/464	Gln/Gln vs. Arg/Arg	1.1 (0.7, 1.8)	Age and race
			175/696	Gln/Arg vs. Arg/Arg	1.3 (0.9, 1.8)	Age and race
Pancreatic cancer	Jiao (2006)[285]	U.S. males and females†	183/182	Gln/Gln vs. Arg/Arg	1.2 (0.7, 1.9)	Age, sex, race, and PY
			327/307	Gln/Arg vs. Arg/Arg	1.2 (0.9, 1.7)	Age, sex, race, and PY

	Prostate cancer	Ritchey (2005)[350]	Chinese males >18 years of age*	102/144	Gln/Gln vs. Arg/Arg	2.2 (1.0, 4.8)	Age
				138/231	Gln/Arg vs. Arg/Arg	0.8 (0.5, 1.3)	Age
	Renal cell carcinoma	Hirata (2006)[335]	Japanese males and females aged 29-84*	80/112	Gln/Gln vs. Arg/Arg	2.5 (1.1, 6.0)	Unadjusted
				96/170	Gln/Arg vs. Arg/Arg	0.8 (0.4, 1.3)	Unadjusted
	SCCHN	Demokan (2005)[336]	Turkish males and females*	54/52	Gln/Gln vs. Arg/Arg	0.9 (0.4, 2.0)	Unadjusted
				83/85	Gln/Arg vs. Arg/Arg	0.8 (0.5, 1.5)	Unadjusted
		Olshan (2002)[337]	U.S. males and female Caucasians	48/79	Gln/Gln vs. Arg/Arg	0.1 (0.04, 0.6)	Age and sex
				95/144	Gln/Arg vs. Arg/Arg	0.8 (0.4, 1.1)	Age and sex
		Sturgis (1999)[338]	U.S. male and female African Americans, Caucasians, and Latinos		Gln/Gln vs. Arg/Arg, Gln/Arg	1.6 (1.0, 2.6)	Age, sex, race, and alcohol
	Nasopharyngeal carcinoma	Cao (2006)[334]	Chinese males and females†	273/300	Gln/Gln vs. Arg/Arg	1.2 (0.7, 2.0)	Unadjusted
				393/471	Gln/Arg vs. Arg/Arg	0.8 (0.6, 1.1)	Unadjusted
		Cho (2003)[243]	Taiwanese males and females†	332/283	Gln/Gln vs. Arg/Arg	1.3 (0.7, 2.4)	Age, sex, and ethnicity
					Gln/Arg vs. Arg/Arg	1.0 (0.7, 1.5)	Age, sex, and ethnicity
	Oral cleft	Olshan (2005)[250]	California infants born 1983-1986†	64/170	Gln/Gln vs. Arg/Arg	0.8 (0.4, 1.8)	Race/ethnicity
				107/290	Gln/Arg vs. Arg/Arg	0.9 (0.6, 1.4)	Race/ethnicity
	Spina bifida	Olshan (2005)[250]	California infants born 1983-1986†	73/170	Gln/Gln vs. Arg/Arg	1.3 (0.6, 2.6)	Race/ethnicity
				108/290	Gln/Arg vs. Arg/Arg	0.8 (0.5, 1.3)	Race/ethnicity
Gln632Gln							
	Melanoma	Han (2004)[333]	U.S. females aged 30-55 years*	110/447	GG vs. AA	0.9 (0.6, 1.4)	Age and race
				175/683	GA vs. AA	1.0 (0.7, 1.4)	Age and race
rs3213245							
	Esophageal squamous cell carcinoma	Hao (2004)[283]	Chinese males and females†	311/389	CC vs. TT	1.5 (0.4, 5.0)	Age, sex, and smoking
				399/473	CT vs. TT	1.4 (1.0, 1.9)	Age, sex, and smoking

*Study of incident disease; †Study of prevalent disease; PY, pack-years of smoking; SCCHN, squamous cell carcinoma of the head and neck; WHR, waist-hip ratio; Results were extracted with preference for unadjusted estimates

Table 12. Review of 11 case control studies examining the relationship between the *XRCC1* polymorphisms and cancers and related traits, stratified by smoking.

Variant	Outcome	Author (year)	Study population	No. cases/ controls	Genotype contrast	OR (95% CI)	Covariate adjustments
Arg194Trp							
	Bladder cancer	Matullo (2005)[317]	Italian males aged 34-76 years*	182/110	Trp/Trp or Trp/Arg vs. Arg/Arg among current smokers	0.7 (0.4, 1.4)	Age
				103/104	Trp/Trp or Trp/Arg vs. Arg/Arg among former smokers	0.8 (0.4, 1.9)	Age
				30/99	Trp/Trp,Trp/Arg vs. Arg/Arg in never smokes	0.4 (0.1, 1.9)	Age
		Matullo (2006)[286]	European males and females*	124/1092	Arg/Trp vs. Arg/Arg among nonsmokers	1.0 (0.5, 2.0)	Unadjusted
	Breast cancer	Patel (2005)[320]	U.S. females 50-74 years at enrollment*	243/273	Arg/Trp vs. Arg/Arg among never smokers	0.8 (0.5, 1.3)	Unadjusted
				237/207	Arg/Trp vs. Arg/Arg among ever smokers	0.5 (0.3, 1.0)	Unadjusted
	Lung cancer	Hung (2005)[316]	Eastern European males and females*	161/718	Trp/Trp or Trp/Arg vs. Arg/Arg among never smokers	1.5 (0.9, 2.4)	Country, age and sex
				145/357	Trp/Trp or Trp/Arg vs. Arg/Arg among 0<PY<15	0.9 (0.5, 1.6)	Country, age and sex
				956/694	Trp/Trp or Trp/Arg vs. Arg/Arg among 14<PY<38	1.0 (8, 1.4)	Country, age and sex
				878/355	Trp/Trp or Trp/Arg vs. Arg/Arg among PY>38	0.7 (0.5, 0.9)	Country, age and sex
		Matullo (2006)[286]	European males and females*	114/1092	Arg/Trp vs. Arg/Arg among nonsmokers	0.9 (0.5, 1.9)	Unadjusted
	SCCHN	Olshan (2002)[337]	U.S. males and female Caucasians*	32/155	Arg/Trp vs. Arg/Arg among ever-smokers	1.1 (0.5, 2.3)	Unadjusted
Pro206Pro							
	Bladder cancer	Matullo (2006)[286]	European males and females*	68/585	G/G vs. A/A among nonsmokers	0.8 (0.5, 1.5)	Unadjusted
				97/850	A/G vs. A/A among nonsmokers	0.9 (0.5, 1.5)	Unadjusted
		Matullo (2005)[317]	Italian males aged 34-76 years*	87/67	GG vs. AA among current smokers	1.6 (0.8, 3.4)	Age
				134/88	GA vs. AA among current smokers	2.0 (1.1, 3.6)	Age
				49/55	GG vs. AA among former smokers	2.0 (0.8, 4.7)	Age
				75/89	GA vs. AA among former smokers	1.8 (0.9, 3.6)	Age
				11/43	GG vs. AA among never smokers	1.0 (0.4, 4.6)	Age

				14/64	GA vs. AA among never smokers	1.0 (0.3, 3.2)	Age
	Lung cancer	Matullo (2006)[286]	European males and females*	58/585	G/G vs. A/A among nonsmokers	0.8 (0.4, 1.6)	Unadjusted
				94/850	A/G vs. A/A among nonsmokers	1.5 (0.9, 2.6)	Unadjusted
Arg280His							
	Lung cancer	Hung (2005)[316]	Eastern European males and females*	158/699	His/His or His/Arg vs. Arg/Arg among never smokers	1.2 (0.6, 2.2)	Country, age and sex
				141/359	His/His or His/Arg vs. Arg/Arg among 0<PY<15	0.9 (0.4, 1.8)	Country, age and sex
				941/683	His/His or His/Arg vs. Arg/Arg among 14<PY<38	1.2 (0.8, 1.7)	Country, age and sex
				842/344	His/His or His/Arg vs. Arg/Arg among PY>38	0.6 (0.4, 0.9)	Country, age and sex
		Misra (2003)[279]	Male Finns 50-69 years of age*	309/302	His/His and Arg/His vs. Arg/Arg in ever smokers	1.0 (0.7, 1.6)	Country, age and sex
Arg399Gln							
	Bladder cancer	Kelsey (2004)[341]	U.S. males and females*	30/89	Gln/Gln vs. Arg/Arg among never smokers	0.9 (0.4, 2.4)	Age and sex
				58/151	Gln/Arg vs. Arg/Arg among never smokers	1.8 (0.9, 3.3)	Age and sex
				138/225	Gln/Gln vs. Arg/Arg among ever smokers	0.7 (0.4, 1.2)	Age, sex, and PY
				257/320	Gln/Arg vs. Arg/Arg among ever smokers	1.3 (0.9, 1.9)	Age, sex, and PY
		Matullo (2005)[317]	Italian males aged 34-76 years*	100/62	AA vs. GG among current smokers	0.5 (0.2, 1.2)	Age
				160/94	AG vs. GG among current smokers	0.8 (0.5, 1.4)	Age
				58/50	AA vs. GG among former smokers	0.8 (0.3, 1.9)	Age
				86/89	AG vs. GG among former smokers	0.7 (0.4, 1.4)	Age
				18/58	AA vs. GG among never smokers	1.3 (0.4, 4.6)	Age
				25/82	AG vs. GG among never smokers	0.8 (0.3, 2.1)	Age
		Matullo (2006)[286]	European males and females*	71/612	Gln/Gln vs. Arg/Arg among nonsmokers	1.2 (0.6, 2.4)	Unadjusted
				107/966	Gln/Arg vs. Arg/Arg among nonsmokers	1.0 (0.6, 1.6)	Unadjusted
		Shen(2003)[343]	Italian males aged 20-80 years*	71/43	Gln/Gln vs. Arg/Arg in ≥ 26 PY	0.4 (0.1, 1.0)	Age
				32/41	Gln/Gln vs. Arg/Arg 26	1.8 (0.6, 5.6)	Age

				>PY >0		
			11/32	Gln/Gln vs. Arg/Arg nonsmokers	3.1 (0.7, 14.8)	Age
			120/70	Gln/Arg vs. Arg/Arg ≥ 26 PY	0.7 (0.4, 1.4)	Age
			47/72	Gln/Arg vs. Arg/Arg among 26 >PY >0	0.9 (0.4, 2.0)	Age
			13/48	Gln/Arg vs. Arg/Arg nonsmokers	1.1 (0.3, 3.8)	Age
Breast cancer	Patel (2005)[320]	U.S. females 50-74 years at enrollment*	130/139	Gln/Gln vs. Arg/Arg among never smokers	0.6 (0.4, 1.2)	Unadjusted
			207/216	Gln/Arg vs. Arg/Arg among never smokers	0.9 (0.6, 1.3)	Unadjusted
			126/108	Gln/Gln vs. Arg/Arg among ever smokers	1.9 (1.0, 3.5)	Unadjusted
			180/177	Gln/Arg vs. Arg/Arg among ever smokers	1.0 (0.7, 1.6)	Unadjusted
Esophageal squamous cell carcinoma	Yu (2004)[328]	Chinese males and females*	38/53	Gln/Gln vs. Arg/Arg among never smokers	2.5 (1.1, 5.7)	Unadjusted
			56/82	Gln/Arg vs. Arg/Arg among never smokers	1.1 (0.7, 1.8)	Unadjusted
			46/40	Gln/Gln vs. Arg/Arg among ever smokers	8.3 (4, 18)	Unadjusted
			60/65	Gln/Arg vs. Arg/Arg among never smokers	1.2 (0.8, 2.0)	Unadjusted
Lung cancer	Hung (2005)[316]	Eastern European males and females*	91/388	Gln/Gln vs. Arg/Arg among ever smokers	0.8 (0.5, 1.5)	Country, age and sex
			137/582	Gln/Arg vs. Arg/Arg among never smokers	1.0 (0.7, 1.5)	Country, age and sex
			77/189	Gln/Gln vs. Arg/Arg among 0<PY<15	0.9 (0.4, 1.8)	Country, age and sex
			129/197	Gln/Arg vs. Arg/Arg among 0<PY<15	1.2 (0.8, 1.8)	Country, age and sex
			508/374	Gln/Gln vs. Arg/Arg among 14<PY<38	1.0 (0.7, 1.4)	Country, age and sex
			807/577	Gln/Arg vs. Arg/Arg among 14<PY<38	1.1 (0.9, 1.4)	Country, age and sex
			420/179	Gln/Gln vs. Arg/Arg among PY>38	0.9 (0.6, 1.4)	Country, age and sex
			717/292	Gln/Arg vs. Arg/Arg among PY>38	1.1 (0.8, 1.4)	Country, age and sex
	Ito (2004)[280]	Japanese males and females*	78/80	Gln/Gln vs. Arg/Arg among ever smokers	0.4 (0.2, 1.0)	Unadjusted
			31/105	Gln/Gln vs. Arg/Arg among never smokers	1.3 (0.3, 5.2)	Unadjusted
			117/248	Gln/Arg vs. Arg/Arg among ever smokers	1.1 (0.7, 1.8)	Unadjusted

			47/174	Gln/Arg vs. Arg/Arg among never smokers	0.8 (0.4, 1.6)	Unadjusted
	Matullo (2006)[286]	European males and females*	58/612	Gln/Gln vs. Arg/Arg among nonsmokers	0.5 (0.2, 1.2)	Unadjusted
			109/966	Gln/Arg vs. Arg/Arg among nonsmokers	1.4 (0.8, 2.2)	Unadjusted
	Misra (2003)[279]	Male Finns 50-69 years of age*	175/182	Gln/Gln vs. Arg/Arg among ever smokers	0.8 (0.4, 1.4)	Years of smoking a cigarettes/day
			291/184	Gln/Arg vs. Arg/Arg among ever smokers	1.1 (0.8, 135)	Years of smoking a cigarettes/day
	Park (2002)[348]	South Korean males*	43/70	Gln/Gln vs. Arg/Arg among <41 PY	5.9 (1.5, 23)	Unadjusted
			63/105	Gln/Arg vs. Arg/Arg among <41 PY	1.5 (0.8, 2.8)	Unadjusted
			22/17	Gln/Gln vs. Arg/Arg among >40 PY	1.4 (0.3, 6.7)	Unadjusted
			22/17	Gln/Arg vs. Arg/Arg among >40 PY	1.4 (0.5, 4.0)	Unadjusted
SCCHN	Olshan (2002)[337]	U.S. males and female Caucasians*	90/97	Gln/Gln, Gln/Arg vs. Arg/Arg in ever -smokers	1.5 (0.8, 2.7)	Unadjusted
			7/64	Gln/Gln, Gln/Arg vs. Arg/Arg in never -smokers	0.9 (0.2, 4.4)	Unadjusted

*Study of incident disease; †Study of prevalent disease; PY, pack-years of smoking; SCCHN, squamous cell carcinoma of the head and neck; Results were extracted with preference for unadjusted estimates

2. Nucleotide excision repair (NER)

NER operates primarily on bulky helix-distorting DNA lesions such as pyrimidine dimers, photo-products, larger chemical adducts, and cross-links[189]. The NER pathway was first identified in individuals with xeroderma pigmentosum (XP), an inherited autosomal recessive NER defect in which patients exhibit extreme sensitivity to UV radiation and a substantially (1000-fold) increased risk of skin cancers[351] and (>10-fold) internal tumors[352]. Other NER disorders include Cockayne's syndrome and trichothiodystrophy, diseases that are not characterized by the sun sensitivity that marks XP and instead are distinguished by postnatal developmental delay, microcephaly, skeletal abnormalities, progressive mental degeneration, ataxia, and hypogonadism, as well as features suggestive of premature aging (e.g. brittle hair and nails and scaling skin)[353-356].

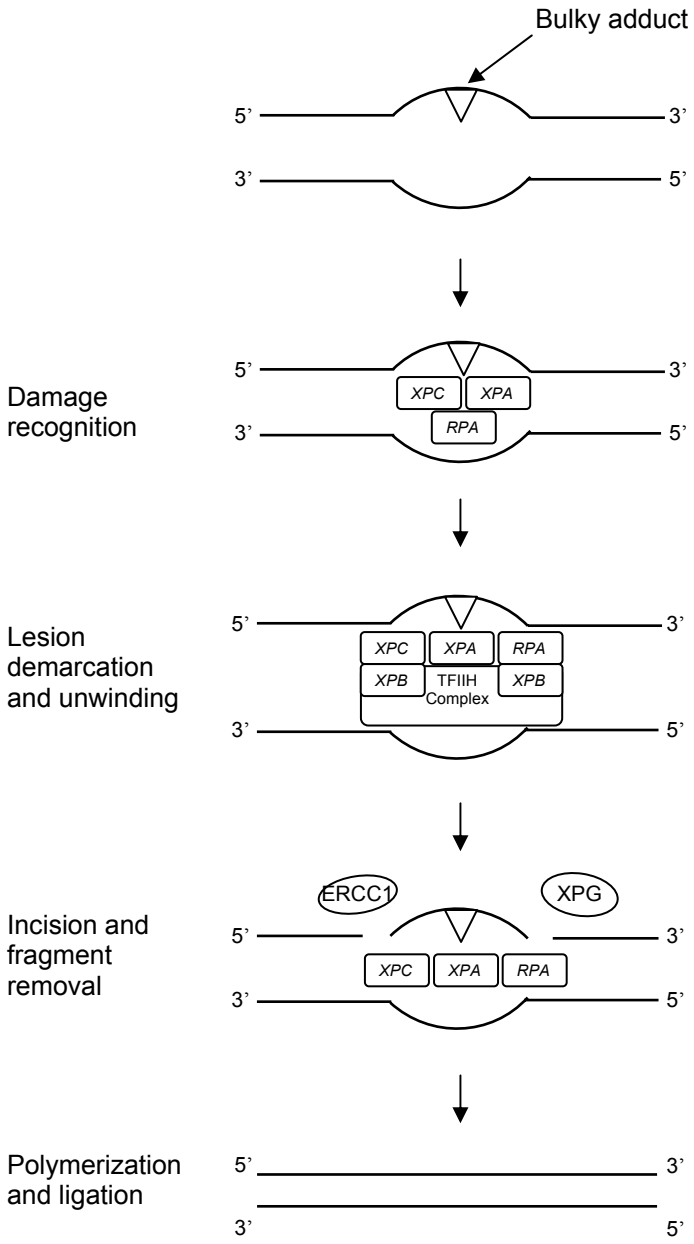
The NER pathway repairs DNA strand damage in a 'cut-and-paste' manner involving five stages (Figure 6): 1) recognition of DNA damage via a damage recognition factor, 2) unwinding of the DNA duplex around the damaged site by the transcription factor IIH (TFIIH) complex, 3) dual incisions 3' and 5' to the damaged site by the 16-unit excinuclease complex and release of the damaged strand, 4) gap repair, and 5) ligation[357].

NER is typically divided into two processes, global genomic repair (GGR) and transcription-coupled repair (TCR)[358] that differ mainly in the means by which DNA damage is recognized. In GGR, the *XPC* protein detects helix deformations due to bulky adducts and initiates repair, whereas in TCR lesions on the transcribed strand block RNA polymerase II and thus signal the repair process[359]. Separation of the double helix is a major step in both processes, necessitating the presence of the transcription factor TFIIH[360]. Mutations in genes

encoding three of the ten TFIIH components – *XPB, XPD,* and *p8/TTD-A* – are related to a broad spectrum of clinical manifestations[361, 362].

As discussed above, cigarette smoke contains numerous carcinogens, including polycyclic aromatic hydrocarbons, aromatic amines, and N-nitroso compounds that can form bulky DNA adducts repaired by the NER pathway. We examined the *XPD* NER enzyme, as *XPD* is absolutely necessary for efficient NER[363] and functional studies have suggested that *XPD* variants can influence DNA repair capacity.

Figure 6. Schematic of Nucleotide – Excision Repair.



Adapted from Goode et al., 2002[184] and Hung et al., 2005[313]

XPD

XPD (also known as *ERCC2*) resides on 19q13.2 and is an ATP-dependent 5'-3' helicase, one of 10 subunits of the TFIIH complex. As discussed above, TFIIH locally unwinds the DNA helix, allowing the NER machinery access to the lesion, and is essential

for both RNA polymerase II transcription initiation and NER[364]. The inherited point mutations in *XPD* that cause XP, trichothiodystrophy, and Cockayne's syndrome are located in the C-terminus (Figure 7) where *XPD* interacts with the TFIIH complex[365].

Mice that lack *XPD* do not survive post-implantation[366], reflecting the essential transcription initiation function of the TFIIH complex. However, a viable mouse mutant was created by mimicking a point mutation identified in a trichothiodystrophy patient displayed many of the characteristics of the disease, including premature aging and a reduced life span[367, 368]. Functionally, the mutation caused a partial defect in both GGR and TCR NER pathways, although an elevated DNA mutation frequency was not observed[369]. Research also suggests that *XPD* mRNA levels are regulated by the expression of the insulin receptor, suggesting that prolonged exposure to elevated glucose levels reduces insulin-dependent regulation of DNA repair[370].

XPD genetic variants

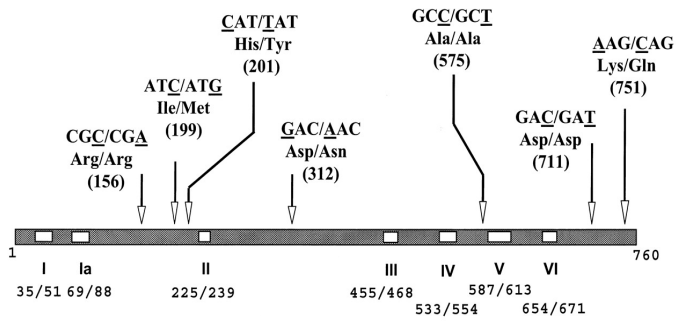
Table 13. Characterization of four known *XPD* nonsynonymous SNPs.

SNP*	Protein residue	MAF (Population)	Functional data?	Studied in human populations?
rs13181	Gln751Lys	0.26 (CEPH) 0.17 (AFR1)	Lunn et al., (2000)[371] Vodicka et al, (2004)[372]	Table 14, Table 15
rs1799793	Asn312Asp	0.31 (CEU) 0.065 (AFR1)	Seker et al., (2001)[373]	Table 14, Table 15
rs1799792	Tyr201His	.		Table 14
rs1799791	Met199Ile	.	.	.

*Information obtained from dbSNP (http://www.ncbi.nlm.nih.gov/projects/SNP/); **AFR1**, Human individual DNA from 24 individuals of self-described AFRICAN/AFRICAN AMERICAN heritage; **CEPH**, Genomic DNA samples obtained for a panel of 92 unrelated individuals chosen from Centre d'Etude du Polymorphisme Human (CEPH) pedigrees. The genomic DNA comprised of UTAH (93%), French (4%), and Venezuelan (3%) samples were purchased from Coriell Cell Repository; **CEU**, 30 mother-father-child trios from the CEPH collection (Utah residents with ancestry from northern and western Europe), representing one of the populations studied in the International HapMap project

Several *XPD* polymorphisms have been reported, but few with a MAF > 0.01[374] (Table 13). Two variants of interest are Asp312Asn (30%MAF) and Lys751Gln (32% MAF), which have MAFs high enough to facilitate clinical and epidemiologic investigations.

Figure 7. Schematic of the *XPD* protein.



Adapted from Benhamou and Sarasin (2005)[363]. Roman numerals denote the seven DNA helicase domains.

The 751Gln variant produces a conformational change at the domain where *XPD* interacts with its helicase activator p44[375] and has been associated with a reduced DNA repair proficiency, as measured by a cytogenetic assay that detects X-ray induced chromatid aberrations[371] and an elevated frequency of chromosomal aberrations in peripheral lymphocytes of a central European population[372]. Codon 312 has been conserved in vertebrates, suggesting the mutation may be functional[376]. BPDE DNA adduct levels also appeared to be elevated in 312Asn homozygotes in a study of 67 Polish coke oven workers[377] and a study of apoptic rates in lymphoblastoid cell lines demonstrated that cells lines homozygous for the Asn had an elevated increase in apoptosis following UV exposure, relative to cells carrying Asp312[373].

The relationship between *XPD* variants and cancers and related traits suggests a subtle to null effect (Table 14). For example, while the three estimates for the Asp312Asn allele and bladder cancer association are relatively precise and suggest a small increase in the odds of cancer for carriers of the Asn allele, the four breast cancer studies generally suggest an inverse association. However, the bladder cancer studies were conducted in European

populations or those of European descent, whereas the breast cancer studies were carried out in European, U.S. and Asian populations, perhaps complicating interpretation. Of the ten studies that evaluated the association between lung cancer and the Lys751Gln allele, all but Shen *et al.,* (2005)[314] reported an increased odds for carriers of the Gln allele, notwithstanding ICR estimates that ranged from 1.7 to 6. It is also difficult to asses whether publication bias may have influenced the reporting of these results, as researchers may be more inclined to publish biologically plausible associations.

As seen in *XRCC1,* there also were marked differences in MAFs between populations, possibly suggesting that susceptibility variants and/or disease markers differ between populations, which may help explain the inconsistent results. For example, in studies of lung cancer and the Asp312Asn variant, 12% of study participants were Asn/Asn homozygotes when U.S. Caucasians were examined[378], whereas only 0.6% of Chinese males and females were identified as such[379]. The marked imprecision of the smoking-stratified estimates (Table 15) complicated interpretation, although results presented by Schabath et al, (2005) suggested a increasing odds of bladder cancer with increasing pack-years of smoking for the Asp312Asn variant[380].

Table 14. Review of 25 case control studies examining the relationship between the *XPD* polymorphisms and cancers and related traits stratified by cancer and polymorphism.

Variant	Outcome	Author (year)	Study population	No. cases/ controls	Genotype contrast	OR (95% CI)	Covariate adjustments
Arg156Arg							
	Bladder cancer	Garcia-Closas (2006)[381]	Spanish males and females aged 21 to 80 years*	562/602	CC vs. AA	1.1 (0.8, 1.4)	Sex, age, region and smoking
				949/947	AC vs. AA	1.2 (1.0, 1.4)	Sex, age, region and smoking
	Ovarian cancer	Costa (2006)[382]	Portuguese females†	118/187	CC, AC vs. AA	1.3 (0.7, 2.3)	Unadjusted
	Lung cancer	Shen (2005)[314]	Chinese males and females*	53/61	CC vs. AA	1.0 (0.5, 2.0)	Age, sex, and current fuel type
				94/74	AC vs. AA	0.5 (0.2, 1.1)	Age, sex, and current fuel type
		Yin (2005)[383]	Chinese males and females*	72/72	CC vs. AA	0.9 (0.5, 1.8)	Unadjusted
				111/98	AC vs. AA	1.1 (0.6, 2.1)	Unadjusted
His201Tyr							
	Lung cancer	Zienolddiny (2006)[282]	Norwegian males and females*	339/405	Tyr/Tyr, Tyr/His vs. His/His	1.1 (0.8, 1.6)	Age, sex, and PY
Asp312Asn							
	Bladder cancer	Garcia-Closas (2006)[381]	Spanish males and females aged 21 to 80 years*	655/655	Asn/Asn vs. Asp/Asp	1.2 (0.9, 1.6)	Sex, age, region and smoking
				990/1028	Asn/Asp vs. Asp/Asp	1.1 (0.9, 1.3)	Sex, age, region and smoking
		Matullo (2005)[317]	Italian males aged 34-76 years*	139/150	Asn/Asn vs. Asp/Asp	1.1 (0.6, 2.0)	Age and smoking
				245/158	Asn/Asp vs. Asp/Asp	1.1 (0.7, 1.6)	Age and smoking
		Schabath (2005)[380]	U.S. males and females*	282/298	Asn/Asn vs. Asp/Asp	1.1 (0.9, 1.4)	Age, sex, race, and smoking
				440/427	Asn/Asp vs. Asp/Asp	1.1 (0.8, 1.4)	Age, sex, race, and smoking
	Breast cancer	Jorgensen (2006)[384]	U.S. males and females*	132/131	Asn/Asn vs. Asp/Asp	0.7 (0.4, 1.3)	Unadjusted
				238/244	Asn/Asp vs. Asp/Asp	0.8 (0.6, 1.2)	Unadjusted
		Justenhoven (2004)[385]	German females*	394/355	Asn/Asn vs. Asp/Asp	0.5 (0.3, 0.7)	Age, smoking, family history of breast cancer, HRT, and parity
				220/334	Asn/Asp vs. Asp/Asp	0.8 (0.6, 1.3)	Age, smoking, family history of breast cancer, HRT, and parity
		Lee (2005)[386]	Korean females*	478/104	Asn/Asn vs. Asp/Asp	0.8 (0.1, 3.2)	BMI, estrogen exposure, education, and family history
				525/442	Asn/Asp vs. Asp/Asp	1.2 (0.7, 1.8)	BMI, estrogen exposure, education, and family history
		Zhang (2005)[387]	Chinese females aged 26-60 years*	109/170	Asn/Asn vs. Asp/Asp	0.5 (0.3, 0.9)	Unadjusted

			200/259	Asn/Asp vs. Asp/Asp	1.0 (0.7, 1.5)	Unadjusted
Esophageal squamous cell carcinoma	Xing (2002)[327]	Chinese males and females[†]	433/524	Asn/Asn, Asn/Asp vs. Asp/Asp	1.0 (0.7, 1.5)	Unadjusted
	Ye (2006)[388]	Swedish males and females[*]	40/233	Asn/Asn vs. Asp/Asp	0.8 (0.3, 2.0)	Age, sex, SES, BMI, smoking, symptomatic gastroesophageal reflux, alcohol, and fruit and vegetable intake
			71/413	Asn/Asp vs. Asp/Asp	1.2 (0.7, 2.3)	Age, sex, SES, BMI, smoking, symptomatic gastroesophageal reflux, alcohol, and fruit and vegetable intake
Lung cancer	Butkiewicz (2001)[376]	Polish males[†]	61/46	Asn/Asn vs. Asp/Asp	1.4 (0.6, 3.2)	Age and PY
			53/65	Asn/Asp vs. Asp/Asp	0.7 (0.3, 1.5)	Age and PY
	Hu (2006)[389]	Chinese males and females[*]	970/986	Asn/Asn, Asn/Asp vs. Asp/Asp	1.1 (0.9, 1.5)	Age, sex, PY, and family history of cancer
	Liang (2003)[379]	Chinese males and females[*]	881/890	Asn/Asn vs. Asp/Asp	11 (1.5, 87)	Unadjusted
			881/1019	Asn/Asp vs. Asp/Asp	1.0 (0.8, 1.3)	Unadjusted
	Popanda (2004)[281]	German males and females aged 28-84 years[*]	113/264	Asn/Asn vs. Asp/Asp	1.2 (0.7, 1.9)	Unadjusted
			169/388	Asn/Asp vs. Asp/Asp	1.1 (0.8, 1.6)	Unadjusted
	Shen (2005)[314]	Chinese males and females[*]	118/113	Asn/Asn, Asn/Asp vs. Asp/Asp	0.6 (0.2, 1.4)	Age, sex, and current fuel type
	Spitz (2001)[390]	U.S. males and females[*]	123/153	Asn/Asn vs. Asp/Asp	1.5 (0.8, 3.1)	Unadjusted
			174/239	Asn/Asp vs. Asp/Asp	0.9 (0.6, 1.4)	Unadjusted
	Zhou (2002)[378]	U.S. male and female Caucasians[*]	613/668	Asn/Asn vs. Asp/Asp	1.4 (1.1, 1.8)	Unadjusted
			942/1115	Asn/Asp vs. Asp/Asp	1.0 (0.8, 1.2)	Unadjusted
	Zienolddiny (2006)[282]	Norwegian males and females[*]	173/169	Asn/Asn vs. Asp/Asp	1.1 (0.7, 1.7)	Age, sex, and PY
			221/241	Asn/Asp vs. Asp/Asp	0.8 (0.6, 1.2)	Age, sex, and PY
Melanoma	Millikan (2006)[391]	Australian, Canadian, Italian, and U.S. males and females[*]	644/1299	Asn/Asn vs. Asp/Asp	1.3 (1.1, 1.7)	Unadjusted
			1014/2137	Asn/Asp vs. Asp/Asp	1.0 (0.9, 1.2)	Unadjusted
Ovarian cancer	Costa (2006)[382]	Portuguese females[†]	114/199	Asn/Asn, Asn/Asp vs. Asp/Asp	2.5 (1.2, 5.0)	Unadjusted
Pancreatic cancer	Jiao(2006)[392]	U.S. males and females[*]	176/195	Asn/Asn vs. Asp/Asp	0.8 (0.5, 1.3)	Age and sex
			303/316	Asn/Asp vs. Asp/Asp	1.0 (0.7, 1.4)	Age and sex
SCCHN	Buch (2005)[393]	U.S. males and females[*]	127/111	Asn/Asn vs. Asp/Asp	1.7 (0.9, 3.4)	Unadjusted
			187/181	Asn/Asp vs. Asp/Asp	1.5 (0.9, 2.2)	Unadjusted
	Yu (2004)[394]	Chinese males and females[*]	135/152	Asn/Asn, Asn/Asp vs. Asp/Asp	1.0 (0.5, 2.1)	Unadjusted

	Oral Cleft	Olshan (2005)[250]	California infants born 1983-1986†		Asn/Asn vs. Asp/Asp	0.7 (0.3, 1.7)	Race/ethnicity
					Asn/Asp vs. Asp/Asp	1.5 (0.9, 2.3)	Race/ethnicity
	Spina bifida	Olshan (2005)[250]	California infants born 1983-1986†		Asn/Asn vs. Asp/Asp	0.8 (0.3, 2.2)	Race/ethnicity
					Asn/Asp vs. Asp/Asp	1.5 (0.9, 2.3)	Race/ethnicity
Lys751Gln							
	Bladder cancer	Garcia-Closas (2006)[381]	Spanish males and females aged 21 to 80 years*	636/638	Gln/Gln vs. Lys/Lys	1.2 (0.9, 1.6)	Sex, age, region and smoking
				990/999	Gln/Lys vs. Lys/Lys	1.1 (0.9, 1.3)	Sex, age, region and smoking
		Matullo (2005)[317]	Italian males aged 34-76 years*	149/148	Gln/Gln vs. Lys/Lys	1.3 (0.7, 2.2)	Age and smoking
				264/266	Gln/Lys vs. Lys/Lys	1.0 (0.7, 1.6)	Age and smoking
		Sanyal (2004)[342]	Swedish males and females ages 33-96 years†	166/132	Gln/Gln vs. Lys/Lys	1.3 (0.8, 2.2)	Unadjusted
				251211	Gln/Lys vs. Lys/Lys	1.1 (0.7, 1.6)	Unadjusted
		Schabath (2005)[380]	U.S. males and females*	264/253	Gln/Gln vs. Lys/Lys	1.2 (0.9, 1.4)	Age, sex, race, and smoking
				280/410	Gln/Lys vs. Lys/Lys	1.0 (0.8, 1.4)	Age, sex, race, and smoking
		Shen (2003)[343]	Italian males aged 20-80 years*	114/116	Gln/Gln vs. Lys/Lys	1.0 (0.6, 1.8)	Age
				166/178	Gln/Lys vs. Lys/Lys	0.9 (0.6, 1.4)	Age
		Stern (2002)[395]	U.S. males and females*	109/111	Gln/Gln vs. Lys/Lys	0.8 (0.4, 1.4)	Age, sex, and ethnicity
				167/183	Gln/Lys vs. Lys/Lys	1.0 (0.7, 1.5)	Age, sex, and ethnicity
	Breast cancer	Brewster (2006)[396]	U.S. females*	.	Gln/Gln vs. Lys/Lys	1.1 (0.6, 1.9)	Unadjusted
				.	Gln/Lys vs. Lys/Lys	1.4 (1.0, 2.0)	Unadjusted
		Justenhoven (2004)[385]	German females*	321/351	Gln/Gln vs. Lys/Lys	1.3 (0.9, 1.9)	Age, smoking, family history of breast cancer, HRT, and parity
				489/556	Gln/Lys vs. Lys/Lys	1.1 (0.9, 1.4)	Age, smoking, family history of breast cancer, HRT, and parity
		Metsola (2005)[339]	Finnish females*	243/243	Gln/Gln vs. Lys/Lys	1.1 (0.7, 1.6)	Age, age at menarche, age at first birth, number of pregnancies, history of breast disease, WHR, family history of breast cancer, smoking and alcohol
				385/392	Gln/Lys vs. Lys/Lys	1.0 (0.8, 1.4)	Age, age at menarche, age at first birth, number of pregnancies, history of breast disease, WHR, family history of breast cancer, smoking and alcohol
		Zhang (2005)[387]	Chinese females aged 26-60 years*	108/145	Gln/Gln vs. Lys/Lys	1.0 (0.6, 1.8)	Unadjusted
				186/264	Gln/Lys vs. Lys/Lys	0.9 (0.6, 1.4)	Unadjusted
	Colon adenoma	Huang (2006)[397]	U.S. males and females aged 55-74 years at enrollment*	395/427	Gln/Gln vs. Lys/Lys	0.9 (0.7, 1.2)	Sex, race, and age
				648/647	Gln/Lys vs. Lys/Lys	1.1 (0.9, 1.4)	Sex, race, and age
		Skjelbred (2006)[324]	Norwegian males and females*	519/225	Gln/Gln vs. Lys/Lys	1.6 (1.1, 2.3)	Age
				827/348	Gln/Lys vs. Lys/Lys	1.4 (1.0, 1.8)	Age
	Colon carcinoma	Skjelbred (2006)[324]	Norwegian males and females*	80/225	Gln/Gln vs. Lys/Lys	0.8 (0.4, 1.9)	Age
				134/348	Gln/Lys vs. Lys/Lys	0.9 (0.5, 1.6)	Age
	Esophageal squamous cell	Xing (2002)[327]	Chinese males and females†	433/524	Gln/Gln, Gln/Lys vs. Lys/Lys	1.1 (0.8, 1.6)	Unadjusted

carcinoma						
	Ye (2006)[388]	Swedish males and females*	37/269	Gln/Gln vs. Lys/Lys	1.8 (07, 4.4)	Age, sex, SES, BMI, smoking, symptomatic gastroesophageal reflux, alcohol, and fruit and vegetable intake
			67/401	Gln/Lys vs. Lys/Lys	2.0 (1.1, 3.9)	Age, sex, SES, BMI, smoking, symptomatic gastroesophageal reflux, alcohol, and fruit and vegetable intake
Gastric cancer	Huang (2005)[347]	Polish males and females aged 21-79 years*	153/218	Gln/Gln vs. Lys/Lys	0.8 (0.5, 1.2)	Age, sex, and smoking
			233/308	Gln/Lys vs. Lys/Lys	1.0 (0.7, 1.5)	Age, sex, and smoking
Lung cancer	Chen (2002)[331]	Chinese males and females*	62/61	Gln/Gln vs. Lys/Lys	2.2 (1.0, 5.3)	Unadjusted
			58/68	Gln/Lys vs. Lys/Lys	1.8 (0.8, 4.1)	Unadjusted
	David-Beabes (2001)[398]	U.S. Caucasian and African-American males and females	191/398	Gln/Gln vs. Lys/Lys	1.3 (0.8, 2.2)	Age, sex, smoking, and race
			296/616	Gln/Lys vs. Lys/Lys	1.0 (0.7, 1.4)	Age, sex, smoking, and race
	Hu (2006)[389]	Chinese males and females*	975/997	Gln/Gln, Gln/Lys vs. Lys/Lys	1.2 (0.9, 1.5)	Age, sex, PY, and family history of cancer
	Liang (2003)[379]	Chinese males and females*	853/854	Gln/Gln vs. Lys/Lys	2.4 (0.9, 6)	Unadjusted
			992/1014	Gln/Lys vs. Lys/Lys	0.9 (0.7, 1.2)	Unadjusted
	Popanda (2004)[281]	German males and females aged 28-84 years*	110/252	Gln/Gln vs. Lys/Lys	1.6 (1.0, 2.6)	Unadjusted
			165/395	Gln/Lys vs. Lys/Lys	1.2 (0.8, 1.7)	Unadjusted
	Shen (2005)[314]	Chinese males and females*	118/108	Gln/Gln, Gln/Lys vs. Lys/Lys	0.4 (0.2, 0.9)	Age, sex, and current fuel type
	Spitz (2001)[390]	U.S. males and females*	188/198	Gln/Gln vs. Lys/Lys	1.4 (0.8, 2.2)	Unadjusted
			194/321	Gln/Lys vs. Lys/Lys	1.1 (0.8, 1.5)	Unadjusted
	Yin (2006)[399]	Chinese males and females*	147/145	Gln/Gln, Gln/Lys vs. Lys/Lys	2.8 (1.1, 6.8)	Unadjusted
	Zhou (2002)[378]	U.S. male and female Caucasians*	594/665	Gln/Gln vs. Lys/Lys	1.2 (0.9, 1.5)	Unadjusted
			926/1074	Gln/Lys vs. Lys/Lys	1.0 (0.9, 1.2)	Unadjusted
	Zienolddiny (2006)[282]	Norwegian males and females*	216/265	Gln/Gln vs. Lys/Lys	1.6 (1.1, 2.3)	Age, sex, and PY
			228/304	Gln/Lys vs. Lys/Lys	1.2 (0.8, 1.7)	Age, sex, and PY
Melanoma	Millikan (2006)[391]	Australian, Canadian, Italian, and U.S. males and females*	636/1308	Gln/Gln vs. Lys/Lys	1.3 (1.1, 1.6)	Unadjusted
			1017/2109	Gln/Lys vs. Lys/Lys	1.1 (1.0, 1.3)	Unadjusted
Ovarian cancer	Costa (2006)[382]	Portuguese females†	126/202	Gln/Gln, Gln/Lys vs. Lys/Lys	3.4 (1.6, 7)	Unadjusted
Prostate cancer	Jiao(2006)[392]	U.S. males and females*	154/179	Gln/Gln vs. Lys/Lys	1.1 (0.6, 1.9)	Age and sex
			308/350	Gln/Lys vs. Lys/Lys	1.1 (0.8, 1.5)	Age and sex
	Ritchey (2005)[350]	Chinese males >18 years of age*	160/247	Gln/Gln, Gln/Lys vs. Lys/Lys	0.8 (0.5, 1.5)	Age

	SCCHN	Buch (2005)[393]	U.S. males and females*	185/148	Gln/Gln vs. Lys/Lys	2.0 (1.2, 3.4)	Unadjusted
				237/229	Gln/Lys vs. Lys/Lys	2.2 (1.5, 3.1)	Unadjusted
		Yu (2004)[394]	Chinese males and females*	119/135	Gln/Gln vs. Lys/Lys	6.7 (1.8, 26)	Unadjusted
				124/150	Gln/Lys vs. Lys/Lys	1.2 (0.5, 2.4)	Unadjusted
	Oral Cleft	Olshan (2005)[250]	California infants born 1983-1986†	71/126	Gln/Gln vs. Lys/Lys	1.3 (0.6, 2.6)	Race/ethnicity
				109/302	Gln/Lys vs. Lys/Lys	1.5 (1.0, 2.4)	Race/ethnicity
	Spina bifida	Olshan (2005)[250]	California infants born 1983-1986†	63/126	Gln/Gln vs. Lys/Lys	0.3 (0.1, 1.2)	Race/ethnicity
				120/302	Gln/Lys vs. Lys/Lys	1.9 (1.2, 2.9)	Race/ethnicity
rs1618536							
	Lung cancer	Hu (2006)[389]	Chinese males and females*	483/497	GG vs. AA	1.2 (0.9, 1.5)	Age, sex, PY, and family history of cancer
				669/687	GA vs. AA	1.1 (0.9, 1.4)	Age, sex, PY, and family history of cancer
rs1799786							
	Lung cancer	Hu (2006)[389]	Chinese males and females*	965/986	CT/TT vs. CC	1.2 (0.9, 1.5)	Age, sex, PY, and family history of cancer
rs1799787							
	Lung cancer	Shen (2005)[314]	Chinese males and females*	117/111	TT, CT vs. CC	0.4 (0.2, 0.9)	Age, sex, and current fuel type

*Study of incident disease; †Study of prevalent disease; Results were extracted with preference for unadjusted estimates

Table 15. Review of seven case control studies examining the relationship between the *XPD* polymorphisms and cancers and related traits, stratified by smoking.

Variant	Outcome	Author (year)	Study population	No. cases/ controls	Genotype contrast	OR (95% CI)	Covariate adjustments
Asp312Asn							
	Bladder cancer	Matullo (2006)[286]	European males and females*	64/588	Asn/Asn vs. Asp/Asp among nonsmokers	0.9 (0.4, 1.8)	Unadjusted
				108/924	Asn/Asp vs. Asp/Asp among nonsmokers	1.0 (0.6, 1.6)	Unadjusted
		Schabath (2005)[380]	U.S. males and females*	128/230	Asn/Asn, Asn/Asp vs. Asp/Asp among never smokers	1.1 (0.7, 1.8)	Unadjusted
				95/124	Asn/Asn, Asn/Asp vs. Asp/Asp among 1-19 PY	1.2 (0.7, 2.1)	Unadjusted
				271/123	Asn/Asn, Asn/Asp vs. Asp/Asp among ≥ 20 PY	1.5 (1.0, 2.3)	Unadjusted
	Lung cancer	Misra (2003)[279]	Male Finns 50-69 years of age*	186/165	Asn/Asn vs. Asp/Asp among ever smokers	0.9 (0.6, 1.6)	Years of smoking a cigarettes/day
				270/272	Asn/Asp vs. Asp/Asp among ever smokers	0.7 (0.5, 1.0)	Years of smoking a cigarettes/day
		Matullo (2006)[286]	European males and females*	64/588	Asn/Asn vs. Asp/Asp among nonsmokers	0.9 (0.4, 1.8)	Unadjusted
				108/924	Asn/Asp vs. Asp/Asp among nonsmokers	1.0 (0.6, 1.6)	Unadjusted
		Zhou (2002)[378]	U.S. male and female Caucasians*	41/227	Asn/Asn vs. Asp/Asp among nonsmokers	4.7 (2.3, 9.6)	Unadjusted
				88/216	Asn/Asn vs. Asp/Asp among 0<PY<26	1.6 (0.9, 2.8)	Unadjusted
				199/161	Asn/Asn vs. Asp/Asp among 25<PY<55	1.3 (0.8, 2.4)	Unadjusted
				285/64	Asn/Asn vs. Asp/Asp among PY>55	0.7 (0.4, 1.2)	Unadjusted
				53/396	Asn/Asp vs. Asp/Asp among nonsmokers	1.4 (0.3, 0.8)	Unadjusted
				124/346	Asn/Asp vs. Asp/Asp among 0<PY<26	1 (0.2, 0.6)	Unadjusted
				340/262	Asn/Asp vs. Asp/Asp among 25<PY<55	1.2 (0.2, 0.9)	Unadjusted
				425/111	Asn/Asp vs. Asp/Asp among PY>55	0.6 (0.2, 0.4)	Unadjusted
Lys751Gln							
	Bladder cancer	Matullo (2005)[317]	Italian males aged 34-76 years*	90/54	CC vs. AA among current smokers	2.2 (1.0, 5.2)	Age
				149/97	AC vs. AA among current smokers	1.1 (0.6, 2.0)	Age
				43/50	CC vs. AA among former smokers	0.9 (0.4, 2.2)	Age
				89/86	AC vs. AA among former	1.3 (0.6, 2.5)	Age

				smokers		
			16/44	CC vs. AA among never smokers	0.8 (0.2, 3.1)	Age
			26/83	AC vs. AA among never smokers	0.6 (0.2, 1.6)	Age
	Matullo (2006)[286]	European males and females*	66/590	Gln/Gln vs. Lys/Lys among nonsmokers	0.6 (0.3, 1.1)	Unadjusted
			108/901	Gln/Lys vs. Lys/Lys among nonsmokers	0.8 (0.5, 1.3)	Unadjusted
	Schabath (2005)[380]	U.S. males and females*	124/222	Gln/Gln, Gln/Lys vs. Lys/Lys in never smokers	1.1 (0.7, 1.6)	Unadjusted
			90/120	Gln/Gln, Gln/Lys vs. Lys/Lys among 1-19 PY	1.3 (0.7, 2.3)	Unadjusted
			263/119	Gln/Gln, Gln/Lys vs. Lys/Lys among ≥ 20 PY	1.2 (0.8, 1.9)	Unadjusted
	Shen(2003)[343]	Italian males aged 20-80 years*	80/42	Gln/Gln vs. Lys/Lys among ≥ 26 PY	1.0 (0.4, 2.3)	Age
			29/50	Gln/Gln vs. Lys/Lys among < 26 PY	1.8 (0.7, 4.7)	Age
			5/24	Gln/Gln vs. Lys/Lys nonsmokers	0.3 (0.1, 3.5)	Age
			108/71	Gln/Lys vs. Lys/Lys ≥ 26 PY	0.6 (0.3, 1.1)	Age
			42/64	Gln/Lys vs. Lys/Lys <26 PY	2.1 (0.9, 4.7)	Age
			16/43	Gln/Lys vs. Lys/Lys nonsmokers	1.4 (0.4, 5.2)	Age
	Stern (2002)[395]	U.S. males and females*	36/72	Gln/Gln, Gln/Lys vs. Lys/Lys in never smokers	0.4 (0.1, 1.1)	Unadjusted
			174/125	Gln/Gln, Gln/Lys vs. Lys/Lys in ever smokers	1.9 (0.9, 3.6)	Unadjusted
Lung cancer	Misra (2003)[279]	Male Finns 50-69 years of age*	165/149	Gln/Gln vs. Lys/Lys among ever smokers	1.0 (0.6, 1.7)	Years of smoking a cigarettes/day
			257/156	Gln/Lys vs. Lys/Lys among ever smokers	0.8 (0.6, 1.2)	Years of smoking a cigarettes/day
	Matullo (2006)[286]	European males and females*	58/590	Gln/Gln vs. Lys/Lys among nonsmokers	1.2 (0.6, 2.4)	Unadjusted
			95/901	Gln/Lys vs. Lys/Lys among nonsmokers	1.5 (0.9, 2.6)	Unadjusted
	Zhou (2002)[378]	U.S. male and female Caucasians*	39/228	Gln/Gln vs. Lys/Lys among nonsmokers	2 (1, 4.1)	Unadjusted
			90/216	Gln/Gln vs. Lys/Lys among 0<PY<26	1.2 (0.7, 2.1)	Unadjusted
			197/161	Gln/Gln vs. Lys/Lys among 25<PY<55	1.4 (0.9, 2.3)	Unadjusted
			268/60	Gln/Gln vs. Lys/Lys among PY>55	0.5 (0.3, 0.9)	Unadjusted

57/376	Gln/Lys vs. Lys/Lys among nonsmokers	1.2 (0.3, 0.7)	Unadjusted
126/340	Gln/Lys vs. Lys/Lys among 0<PY<26	0.9 (0.2, 0.6)	Unadjusted
325/253	Gln/Lys vs. Lys/Lys among 25<PY<55	1.3 (0.2, 0.9)	Unadjusted
418/105	Gln/Lys vs. Lys/Lys among PY>55	0.6 (0.2, 0.4)	Unadjusted

*Study of incident disease; †Study of prevalent disease; PY, pack-years of smoking; Results were extracted with preference for unadjusted estimates

3. Double strand break / recombination repair

Unlike BER and NER, which excise and repair DNA damage using the intact complementary strand for a template, DSBs affect both DNA strands so no intact template is available to direct repair. Thus, DSB lesions are particularly genotoxic as they effectively induce chromosomal aberrations (CA) such as deletions and inversions within a chromosome or rearrangements between multiple chromosomes[400]. CA may also alter gene expression profiles[401]. Causes of DSBs include ionizing ration and oxidative insults, as well as somatic recombination or the overlap of excision repair tracts[193]. If unrepaired, DSBs can block replication and transcription and exposed chromosomal fragments are prone to nuclease attack and subsequent destruction (Figure 8). Even one DSB can kill a cell if it inactivates a crucial gene[402].

Two DSB repair mechanisms exist to eliminate chromosome aberrations before cell division occurs; homologous recombination (non error-prone, HR) and non-homologous end joining (error prone, NHEJ)[403-406]. Deficiencies in either pathway cause a chromosomal instability phenotype characterized by increased CA, serious physiological defects, cancer susceptibility, and premature aging[407-409]. DSB-deficient diseases include the diseases Nijmegen breakage syndrome, Fanconi anemia, Bloom syndrome, Werner syndrome, and ataxia telangiectasia[410]. In addition, the well known *BRCA1* and *BRCA2* breast cancer genes also reflect compromised DSB repair[411], as homozygous loss of either causes a marked increased chromosomal instability.

HR is considered 'non error-prone' because the undamaged homologous sequence of the sister chromatid is used to repair the damaged duplex by gene conversion, with or without an associated crossover[412]. While crossovers can produce deletions, inversions (chromosome

rearrangement where a chromosomal segment is reversed end to end), translocations (abnormal chromosomal rearrangements), and LOH[413], HR is usually a high-fidelity repair process since crossovers are usually suppressed in mitotic cells[414, 415]. Cells with HR protein defects (e.g. *RAD51, XRCC2, XRCC3, RAD51B-RAD51D, BRCA1*, and *BRCA2*) exhibit chromosomal instability, producing aneuploidy (a reduction in the number of chromosomes due to extra or missing chromosomes), chromosome breaks, translocations, and fusions, possibly reflecting a shift in DSB repair from HR to NEHJ[412].

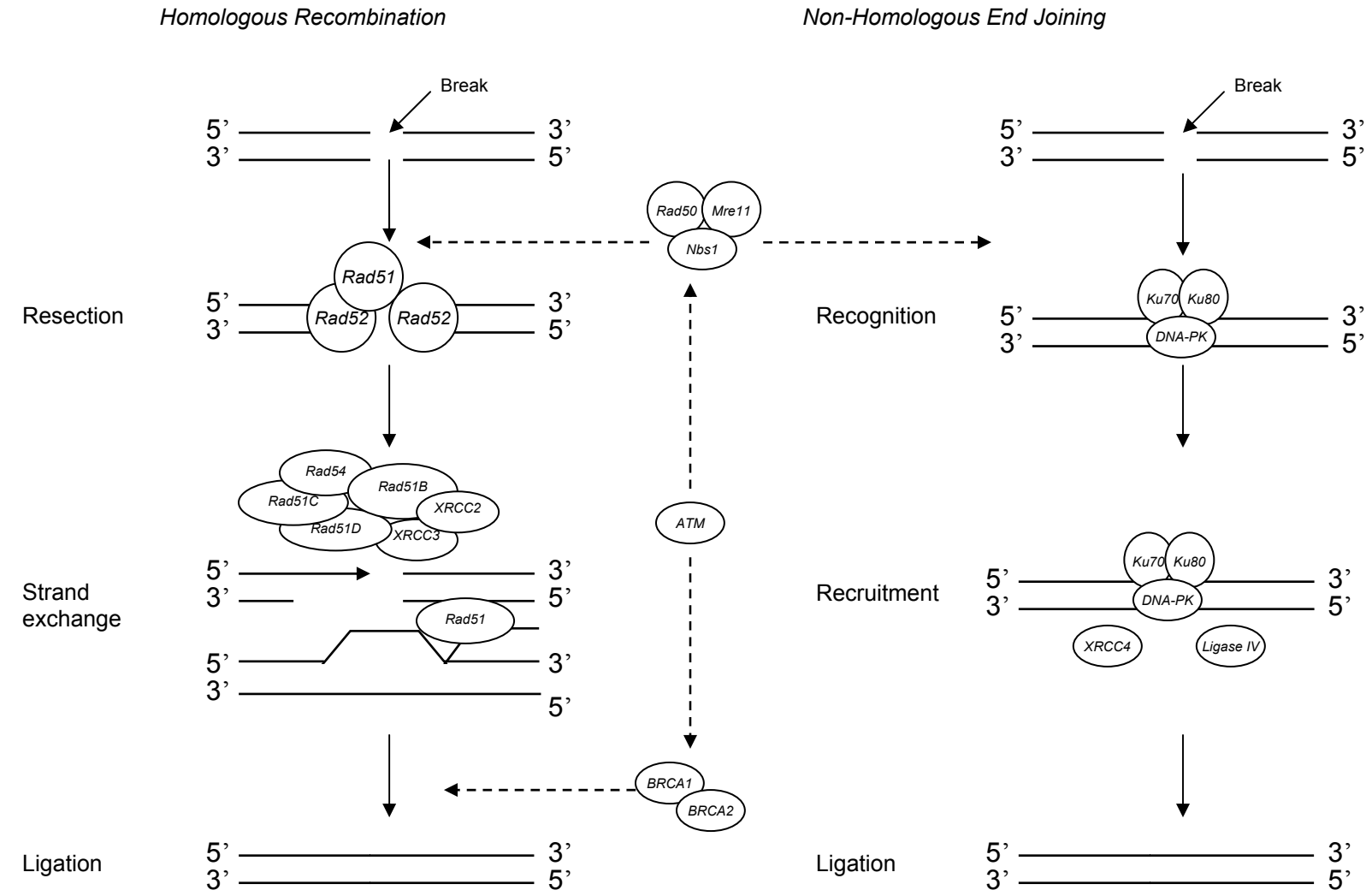
NHEJ is a homology-independent process and simply rejoins any two DSBs end-to-end. Thus, NHEJ often generates small sequence modifications at break sites unless the original sequence is precisely re-ligated. However, consequences of NHEJ repair may be tolerated if the number of DSBs is small so that the originally connected DNA strands are rejoined with high preference[400]. The low proportion of coding sequences in the mammalian genome (1%) results in a low probability that a DSB would occur within a functional region. Furthermore, mammalian cells are diploid, thus the intact allele may be able to compensate for LOH.

The initial step for DSB repair is damage recognition and signaling by the *NBS1, MRE11,* and *RAD50* protein complex[416]. Briefly, HR proceeds by: 1) 5' resection by exonuclease to expose the 3' ends, 2) strand invasion in which the RecA-like proteins facilitate the identification of complementary genomic regions in sister chromosomes, 3) new DNA synthesis using the 3' ends as primers and the sister chromatid as the template, and 4) unwinding from the template, annealing, and ligation, in essence transferring sequence information from the intact donor to the defective recipient, yielding two intact DNA copies[400]. The basic NHEJ steps following recognition of the DSB are: 1) binding of the

DSB by *KU70* and *KU80* to the two DNA ends where it serves as an end-bridging and alignment factor, 2) recruitment and activation of additional processing enzymes by DNA-protein kinase (*DNA-PK)* as well as DNA end-processing and gap-filling when DNA ends are non-complementary or damaged, and 3) direct joining of the two ends by DNA ligase[417, 418].

Research has shown that levels of cigarette smoke condensate far below those contained in a single cigarette can induce DSB in cultured cells and purified DNA[195]. As deficiencies in the DSB repair system may be critical in the generation and persistence of CA, and increased CA may reflect a shift in DSB repair from HR to NEHJ[412], this dissertation examined the *XRCC3* enzyme, which plays a central role in HR of DSB.

Figure 8. Schematic of Double – Strand – Break Repair.



*Adapted from Goode et al., 2002[184] and Kuschel et al., (2002)[193]

XRCC3

XRCC3 is located on 14q32.3 and is involved in the repair of DSB by HR, as it induces replication fork slowing and facilitates recruitment of *RAD51* to DSB sites[419, 420]. In addition to playing a central role in HR, *XRCC3* is also important for maintaining the correct centrosome number in mammalian cells[421, 422].

Fluorescence-based assays in Chinese hamster ovary cells demonstrated that *XRCC3* promoted the repair of DSBs by HR[422] and hamster and human cell lines containing with mutations of *XRCC3* showed 25-fold decrease in HR[422], while constitutive *XRCC3* expression conferred resistance to DNA-damaging agents. *XRCC3* mRNA and protein levels were elevated in malignant prostate cells when compared to normal epithelial cells. Despite the increased *XRCC3* expression, the malignant cells exhibited a defective DNA break repair phenotype, suggesting that prostate tumorgenesis may reflect aberrant DNA repair capacity[300].

XRCC3 genetic variants

While four nonsynonymous *XRCC3* SNPs have been identified in the dbSNP database (Table 16), only one has been examined in human populations or functional studies (Met241Thr). Yoshihara *et al.,* (2004) investigated the role of Met241Thr in human *XRCC3*-/- cell lines that exhibited a phenotype characterized by a two-fold increase in sensitivity to DNA cross-linking agents, a reduction in sister chromatid exchange, increased CA, and impaired *RAD51* function. While expression of the wild-type *XRCC3* cDNA rescued the phenotype, expression of the Met241Thr variant was unable to restore the increased endoreduplication (duplication of the genome without mitosis)[423]. Savas and colleagues (2004) also performed protein conservation analysis on *XRCC3* nonsynonymous

polymorphisms in an attempt to predict functional consequences of amino acid substitutions and predicted that the Met241Thr variant "possibly damaged" protein function[304].

Table 16. Characterization of four known *XRCC3* nonsynonymous SNPs.

SNP*	Protein residue	MAF	Functional data?	Studied in human populations?
rs28903081	His302Arg	0.005 (GscTr12003)	.	.
rs28903080	Arg271Gly	0.005 (GscTr12003)	.	.
rs861539	Met241Thr	0.44 (CEPH) 0.146 (AFR1)	Savas et al (2004)[304] Yoshihara et al., (2004)[423]	Table 17, Table 16
rs3212057	His94Arg	0.032 (PDR90)	.	.

*Information obtained from dbSNP (http://www.ncbi.nlm.nih.gov/projects/SNP/); **AFR1**, Human individual DNA from 24 individuals of self-described AFRICAN/AFRICAN AMERICAN heritage; **CEPH**, Genomic DNA samples obtained for a panel of 92 unrelated individuals chosen from Centre d'Etude du Polymorphisme Human (CEPH) pedigrees. The genomic DNA comprised of UTAH (93%), French (4%), and Venezuelan (3%) samples were purchased from Coriell Cell Repository; **GscTr12003**, British Phenotype: 96 BRCA1 and BRCA2 negative breast cancer index cases; **PDR90**, The NIH Polymorphism Discovery Resource (NIHPDR) 90 individual screening subset

The limited epidemiologic data examining the relationship between *XRCC3* polymorphisms and cancers and related traits, has suggested a weak to null effect (Table 17), although multiple studies of the same outcome are few. For example, in the five studies of breast cancer, the most extreme estimates of 0.6[387] and 1.4[424] are also those with the largest CLRs. The four studies of the Met241Thr variant and lung cancer produced estimates ranging from 0.6 to 1.5[425], all with generally comparable precision. The smoking stratified estimates were too imprecise to warrant interpretation (Table 18).

Table 17. Review of 20 case control studies examining the relationship between the *XRCC3* polymorphisms and cancers and related traits stratified by cancer and polymorphism.

Variant	Outcome	Author (year)	Study population	No. cases/ controls	Genotype contrast	OR (95% CI)	Covariate adjustments
5'UTR-4541							
	Bladder cancer	Matullo (2005)[317]	Italian males aged 34-76 years*	218/213	CC vs. TT	0.7 (0.3, 1.3)	Age and smoking
				305/303	CT vs. TT	0.8 (0.5, 1.2)	Age and smoking
	Breast cancer	Han (2004)[426]	U.S. females 30-55 years an enrollment*	669/919	CC vs. TT	1.0 (0.6, 1.5)	Age, menopausal status, HRT, data and time of blood draw, and fasting status
				952/1237	CT vs. TT	1.2 (1.0, 1.4)	Age, menopausal status, HRT, data and time of blood draw, and fasting status
	Lung cancer	Jacobsen (2004)[425]	Danish males and females aged 50-65 at enrollment*	185/190	CC vs. TT	0.6 (0.2, 1.6)	Age and smoking
				280/257	CT vs. TT	0.8 (0.5, 1.2)	Age and smoking
IVS6-14 17893							
	Bladder cancer	Matullo (2005)[317]	Italian males aged 34-76 years*	192/185	GG vs. AA	0.9 (0.4, 4.8)	Age and smoking
				288/292	AG vs. AA	1.0 (0.7, 1.4)	Age and smoking
		Han (2004)[426]	U.S. females 30-55 years an enrollment*	534/721	GG vs. AA	1.2 (0.9, 1.6)	Age, menopausal status, HRT, data and time of blood draw, and fasting status
				869/1147	AG vs. AA	1.1 (0.9, 1.3)	Age, menopausal status, HRT, data and time of blood draw, and fasting status
	Lung cancer	Jacobsen (2004)[425]	Danish males and females aged 50-65 at enrollment*	140/142	GG vs. AA	0.7 (0.4, 1.2)	Age and smoking
				227/235	AG vs. AA	0.8 (0.6, 1.2)	Age and smoking
IVS6 C1571T							

	Lung cancer	Jacobsen (2004)[425]	Danish males and females aged 50-65 at enrollment*	216/233	CC vs. TT	4.5 (1.3, 15)	Age and smoking
				248/267	CT vs. TT	1.2 (0.7, 2.0)	Age and smoking
C18067T							
	SCCHN	Shen (2002)[427]	U.S. males and females*	208/184	TT vs. CC	1.3 (0.8, 2.0)	Unadjusted
				309/311	CT vs. CC	0.9 (0.6, 1.2)	Unadjusted
Thr241Met							
	Bladder cancer	Matullo (2005)[317]	Italian males aged 34-76 years*	162/169	Met/Met vs. Thr/Thr	1.5 (0.9, 2.4)	Age and smoking
				254/268	Met/Thr vs. Thr/Thr	1.4 (0.9, 2.1)	Age and smoking
		Sanyal (2004)[342]	Swedish males and females ages 33-96 years†	182/137	Met/Met vs. Thr/Thr	1.3 (0.8, 2.4)	Unadjusted
				260/216	Met/Thr vs. Thr/Thr	1.0 (0.7, 1.4)	Unadjusted
		Shen(2003)[343]	Italian males aged 20-80 years*	114/98	Met/Met vs. Thr/Thr	0.7 (0.4, 1.4)	Age
				176/187	Met/Thr vs. Thr/Thr	0.6 (0.4, 0.9)	Age
	Breast cancer	Figueiredo (2004)[344]	Caucasian Canadian females 25-54 years*	216/202	Met/Met vs. Thr/Thr	1.4 (0.9, 2.2)	Unadjusted
				325/346	Met/Thr vs. Thr/Thr	1.0 (0.7, 1.3)	Unadjusted
		Han (2004)[426]	U.S. females 30-55 years an enrollment*	523/638	Met/Met vs. Thr/Thr	1.0 (0.7, 1.3)	Age, menopausal status, HRT, data and time of blood draw, and fasting status
				817/1075	Met/Thr vs. Thr/Thr	0.9 (0.7, 1.1)	Age, menopausal status, HRT, data and time of blood draw, and fasting status
		Jacobsen (2003)[428]	Danish males and females aged 50-65 at enrollment*	222/225	Met/Met vs. Thr/Thr	0.9 (0.6, 1.4)	Age and smoking

			366/358	Met/Thr vs. Thr/Thr	1.0 (0.7, 1.4)	Age and smoking
	Smith (2003)[322]	U.S. females*	147/139	Met/Met vs. Thr/Thr	1.4 (0.8, 2.4)	Age, family history, age at first live birth, and BMI
			201/231	Met/Thr vs. Thr/Thr	0.9 (0.6, 1.3)	Age, family history, age at first live birth, and BMI
	Thyagarajan(2006)[323]	U.S. females*	227/166	Met/Met vs. Thr/Thr	1.3 (0.8, 2.1)	Unadjusted
			352/283	Met/Thr vs. Thr/Thr	1.0 (0.7, 1.3)	Unadjusted
	Zhang (2005)[387]	Chinese females aged 26-60 years*	140/195	Met/Met vs. Thr/Thr	0.6 (0.3, 1.0)	Unadjusted
			187/281	Met/Thr vs. Thr/Thr	0.9 (0.6, 1.4)	Unadjusted
Colon adenoma	Skjelbred (2006)[324]	Norwegian males and females*	527/198	Met/Met vs. Thr/Thr	0.9 (0.6, 1.4)	Age
			836/339	Met/Thr vs. Thr/Thr	0.8 (0.6, 1.1)	Age
	Stern (2005)[325]	U.S. males and females aged 50-74 years*	452/472	Met/Met vs. Thr/Thr	0.8 (0.6, 1.1)	Age, sex, race, clinic, and exam date
			660/695	Met/Thr vs. Thr/Thr	0.9 (0.7, 1.1)	Age, sex, race, clinic, and exam date
Colon carcinoma	Skjelbred (2006)[324]	Norwegian males and females*	84/198	Met/Met vs. Thr/Thr	1.1 (0.5, 2.6)	Age
			137/339	Met/Thr vs. Thr/Thr	1.2 (0.6, 2.1)	Age
Esophageal squamous cell carcinoma	Ye (2006)[388]	Swedish males and females*	42/254	Met/Met vs. Thr/Thr	1.2 (0.5, 3.2)	Age, sex, SES, BMI, smoking, symptomatic gastroesophageal reflux, alcohol, and fruit and vegetable intake
			70/421	Met/Thr vs. Thr/Thr	1.1 (0.6, 2.0)	Age, sex, SES, BMI, smoking, symptomatic gastroesophageal reflux, alcohol, and fruit and vegetable intake
Gastric cancer	Huang (2005)[347]	Polish males and females aged 21-79 years*	153/227	Met/Met vs. Thr/Thr	0.6 (0.4, 1.1)	Age, sex, and smoking

			256/337	Met/Thr vs. Thr/Thr	1.0 (0.7, 1.4)	Age, sex, and smoking
	Shen (2004)[429]	Chinese males and females*	188/166	Met/Met, Met/Thr vs. Thr/Thr	1.1 (0.5, 2.1)	Unadjusted
Lung cancer	David-Beabes (2001)[398]	U.S. Caucasian and African-American males and females	199/389	Met/Met vs. Thr/Thr	1.1 (0.6, 1.8)	Age, sex, smoking, and race
			298/609	Met/Thr vs. Thr/Thr	0.9 (0.7, 1.3)	Age, sex, smoking, and race
	Jacobsen (2004)[425]	Danish males and females aged 50-65 at enrollment*	132/156	Met/Met vs. Thr/Thr	1.5 (0.9, 2.5)	Age and smoking
			218/226	Met/Thr vs. Thr/Thr	1.5 (1.1, 2.3)	Age and smoking
	Popanda (2004)[281]	German males and females aged 28-84 years*	115/237	Met/Met vs. Thr/Thr	1.5 (0.9, 2.4)	Unadjusted
			160/390	Met/Thr vs. Thr/Thr	1.0 (0.7, 1.4)	Unadjusted
	Zienolddiny (2006)[282]	Norwegian males and females*	130/139	Met/Met vs. Thr/Thr	0.6 (0.3, 1.2)	Age, sex, and PY
			204/226	Met/Thr vs. Thr/Thr	0.8 (0.6, 1.2)	Age, sex, and PY
Prostate cancer	Ritchey (2005)[350]	Chinese males >18 years of age*	142/216	Met/Met vs. Thr/Thr	2.2 (0.4, 13)	Age
			156/245	Met/Thr vs. Thr/Thr	0.8 (0.5, 1.6)	Age
Spina bifida	Olshan (2005)[250]	California infants born 1983-1986†	73/208	Met/Met vs. Thr/Thr	1.1 (0.5, 2.2)	Race/ethnicity
			109/296	Met/Thr vs. Thr/Thr	1.1 (0.7, 1.8)	Race/ethnicity
Oral cleft	Olshan (2005)[250]	California infants born 1983-1986†	76/108	Met/Met vs. Thr/Thr	0.5 (0.2, 1.3)	Race/ethnicity
			109/296	Met/Thr vs. Thr/Thr	0.8 (0.5, 1.3)	Race/ethnicity

*Study of incident disease; †Study of prevalent disease; BMI, body mass index; PY, pack-years of smoking; Results were extracted with preference for unadjusted estimates

Table 18. Review of four case control studies examining the relationship between the *XRCC3* polymorphisms and cancers and related traits, stratified by smoking status.

Variant	Outcome	Author (year)	Study population	No. cases/ controls	Genotype contrast	OR (95% CI)	Covariate adjustments
5'UTR-4541							
	Bladder cancer	Matullo (2005)[317]	Italian males aged 34-76 years*	127/68	CC vs. TT among current smokers	0.9 (0.2, 3.6)	Age
				175/107	CT vs. TT among current smokers	1.0 (0.6, 1.7)	Age
				66/76	CC vs. TT among former smokers	0.5 (0.1, 3.3)	Age
				101/100	CT vs. TT among former smokers	1.5 (0.8, 2.8)	Age
				25/63	CC vs. TT among never smokers	0.5 (0.04, 5)	Age
				29/96	CT vs. TT among never smokers	0.3 (0.1, 0.8)	Age
IVS6-14 17893							
	Bladder cancer	Matullo (2005)[317]	Italian males aged 34-76 years*	115/68	GG vs. AA among current smokers	0.3 (0.1, 0.8)	Age
				169/100	GA vs. AA among current smokers	0.7 (0.4, 1.3)	Age
				60/62	GG vs. AA among former smokers	3.0 (0.8, 11)	Age
				92/101	GA vs. AA among former smokers	1.2 (0.7, 2.2)	Age
				17/55	GG vs. AA among never smokers	2.1 (0.4, 11)	Age
				27/91	GA vs. AA among never smokers	1.2 (0.5, 3.1)	Age
		Matullo (2006)[286]	European males and females*	77/645	GG vs. AA among nonsmokers	1.3 (0.6, 2.3)	Unadjusted
				107/1001	AG vs. AA among nonsmokers	0.9 (0.5, 1.4)	Unadjusted
	Lung cancer	Matullo (2006)[286]	European males and females*	62/645	GG vs. AA among nonsmokers	0.8 (0.3, 1.9)	Unadjusted
				107/1001	AG vs. AA among nonsmokers	1.0 (0.6, 1.7)	Unadjusted
Thr241Met							
	Bladder cancer	Matullo (2005)[317]	Italian males aged 34-76 years*	89/57	Met/Met vs. Thr/Thr among current smokers	2.7 (1.2, 5.8)	Age
				142/96	Met/Thr vs. Thr/Thr among current smokers	2.0 (1.1, 3.5)	Age
				59/61	Met/Met vs. Thr/Thr among former smokers	0.5 (0.2, 1.1)	Age
				92/101	Met/Thr vs. Thr/Thr among former smokers	0.8 (0.4, 1.6)	Age
				14/51	Met/Met vs. Thr/Thr	4.3 (1.1, 16)	Age

				among never smokers		
			23/89	Met/Thr vs. Thr/Thr among never smokers	2.0 (0.7, 5.7)	Age
	Matullo (2006)[286]	European males and females*	63/550	Met/Met vs. Thr/Thr among nonsmokers	1.0 (0.5, 2.0)	Unadjusted
			107/927	Met/Thr vs. Thr/Thr among nonsmokers	1.0 (0.6, 1.6)	Unadjusted
	Shen(2003)[343]	Italian males aged 20-80 years*	71/41	Met/Met vs. Thr/Thr ≥ 26 PY	0.5 (0.2, 1.1)	Age
			23/35	Met/Met vs. Thr/Thr 0<PY< 26	1.5 (0.5, 5.0)	Age
			10/22	Met/Met vs. Thr/Thr nonsmokers	1.1 (0.2, 5.9)	Age
			114/69	Met/Thr vs. Thr/Thr, ≥ 26 pack-yrs	0.5 (0.3, 0.9)	Age
			48/71	Met/Thr vs. Thr/Thr 0<PY< 26	1.2 (0.6, 2.6)	Age
			14/47	Met/Thr vs. Thr/Thr nonsmokers	0.5 (0.2, 1.8)	Age
Lung cancer	Misra (2003)[279]	Male Finns 50-69 years of age*	189/172	Met/Met vs. Thr/Thr among ever smokers	1.1 (0.6, 2.1)	Smoking years and cigarettes/day
			284/183	Met/Thr vs. Thr/Thr, among ever smokers	1.0 (0.7, 1.3)	Smoking years and cigarettes/day
	Matullo (2006)[286]	European males and females*	60/550	Met/Met vs. Thr/Thr among nonsmokers	1.4 (0.6, 3.0)	Unadjusted
			100/927	Met/Thr vs. Thr/Thr among nonsmokers	0.9 (0.5, 1.5)	Unadjusted

*Study of incident disease; †Study of prevalent disease; PY, pack-years; Results were extracted with preference for unadjusted estimates

D. Public Health Significance

Although evidence linking cigarette smoking with atherosclerosis and its clinical is well established and consistent across age, sex, racial, and geographic strata[1-8], the mechanisms by which smoking initiates vascular disease remain poorly understood. Identifying genes that in combination with cigarette smoke exposure influence the risk of atherosclerosis and atherothrombotic events could provides new opportunities to evaluate mechanistic laboratory models of CHD and further our understanding of the link between observed epidemiologic trends and CHD biology. Cigarette smoking is also a powerful model to study atherogenic mechanisms and their biology. Furthermore, considering that atherosclerosis is a generalized macrovascular disease, these results may inform research examining lesions in other vascular locales. The analysis of DNA repair variants, cigarette smoke exposure, and two atherosclerosis measures also improves upon previous studies that generally focused on a small number of polymorphisms and were too small to precisely estimate the main effects of genotype, let alone the degree to which they modified the relationship between cigarette smoke exposure and atherosclerosis/atherothrombotic events.

CHAPTER IV
RESEARCH PLAN

A. Overview

The present study utilized data collected from the ARIC Study, a community-based prospective cohort study examining cardiovascular and pulmonary disease, patterns of medical care, and disease variation over time. ARIC investigators enrolled 15,792 participants from four U.S. field centers located in NC, MN, MD, and AL. Information pertaining to cardiovascular disease risk factors, socioeconomic factors, and family medical history was obtained at the home interview whereas extensive clinical data, including serum samples for genotyping, was obtained during the clinic examination. The baseline examination was conducted between 1987 and 1989 and the three subsequent follow-up visits occurred at approximately three-year intervals through 1998.

For Manuscript 1, we conducted maximum likelihood and hierarchical analyses using a piecewise constant exponential model that assessed the degree to which DNA repair pathway variants modified the relationship between ever-smoking cigarettes and incident CHD.

For Manuscript 2, we characterized how DNA repair pathway variants modified the relationship between ever-smoking cigarettes and mean baseline IMT. Maximum likelihood models were fit using linear regression methods. Hierarchical models were fit using linear mixed effects models.

B. The ARIC Study

Study sample description

ARIC, the parent population for this study, is an ongoing population-based longitudinal investigation examining cardiovascular and pulmonary disease, patterns of medical care, and disease variation over time. The ARIC cohort was selected as a probability sample of 15,792 Caucasian and African American males and females from four United States geographic locales: Washington County, Maryland; Forsyth County, North Carolina; suburbs of Minneapolis, Minnesota; and Jackson, Mississippi. Eligible households in North Carolina were identified by area probability sampling while age-eligible participants in MS, MN, and MD were identified from driver's licenses, voter registration cards, and identification cards. The overall recruitment response proportion at cohort intake was 60%: African American males (42%), African American females (49%), Caucasian males (67%), and Caucasian females (68%).

Home interviews were administered to each potential cohort member, followed by an invitation for clinical examination. Researchers collected information pertaining to cardiovascular disease risk factors, socioeconomic factors, and family medical history at the home interview and extensive clinical data during the clinic examination. Incident medical events were identified through an annual questionnaire (telephone administered), community surveillance, and examinations at three-year intervals through 1999. ARIC investigators also conduct on-going surveillance of hospital admissions and mortality for all residents 35 to 74 years of age in the four communities from which the original cohort was recruited.

Construction of the Cohort Representative Sample

Of the 15,792 ARIC participants initially available, 14,255 participants met the following eligibility criteria: a) reported race of Caucasian and from NC, MN, or MD field centers or African American and from the NC or MS field centers (N= 103 ineligible), b) no history of CHD (N=1,102 ineligible), and c) no history of transient ischemic attack or stroke (N=332 ineligible). The first criterion maintained adequate sample sizes in each race-center category for the weighted analysis and the second and third ensured that only incident atherothrombotic events were examined. The selection of the cohort representative sample (CRS) was performed at baseline by stratifying the eligible study population into eight mutually exclusive groups based on age (≤ 55 years vs. > 55 years), sex, and race. Different sampling fractions were applied to each stratum in an attempt to oversample higher-risk participants. The CRS consisted of 1,065 participants, 85 of which were diagnosed with CHD over follow-up.

C. Outcome assessment

1. Incident CHD

ARIC investigators classified an event as incident CHD if a participant met at least one of the following four criteria: (1) definite or probable myocardial infarction distinguished by ECG, cardiac pain, and/or enzymes; (2) definite fatal CHD: combinations of (a) no known possibly lethal non-atherosclerotic or non-cardiac atherosclerotic process or event, (b) angina, (c) history of CHD (MI or chest pain), and/or (d) death certificate listing the ICD-9 codes 410-414, 427.5, 429.2, and/or 799 (Table 19); (3) silent MI detected by ECG; (4) coronary revascularization procedure (including percutaneous transluminal coronary

angioplasty and coronary artery bypass graft surgery). Through December 1998 a total of 1,086 validated incident CHD cases were identified in the ARIC cohort.

2. IMT

Baseline carotid wall thickness was measured by B-mode ultrasound using validated techniques[460], scanning protocols common to each study center, and standardized central readings[461]. The far walls of the left and right common carotid, carotid bifurcation, and the internal carotid arteries (six sites total) were measured at designated 1-cm lengths and averaged across as many 1-mm-apart intima-to-media (IMT) distances as were available.

D. Exposure assessment

ARIC has several metrics to assess cigarette smoking history, such as smoking intensity (cigarettes/day), duration, age at initiation, second hand smoke exposure (hours/week), and smoking status such as current smokers, former smokers, ever smokers (defined as > 400 cigarettes in a lifetime), and never smokers. Although validation of reported smoking status using biomarkers of tobacco exposure was not attempted in the ARIC study, each measure was ascertained at baseline and updated at cohort re-examinations via an interviewer-administered questionnaire that captures changes in exposure and inconsistencies with previously reported smoking habits. Such inconsistencies served as an exclusion criterion at the time of case and CRS selection.

Table 19. ICD-9 codes and descriptions utilized by ARIC investigation for the classification of CHD deaths.

Category	ICD-9 Code	Condition	Includes

Ischemic heart disease	410	Acute myocardial infarction	• Cardiac infarction
			• Coronary (artery):
			  o embolism
			  o occlusion
			  o rupture
			  o thrombosis
			• Infarction of heart, myocardium, or ventricle
			• Rupture of heart, myocardium, or ventricle
			• ST elevation and non-ST elevation myocardial infarction
			• Any condition classifiable to 414.1-414.9 specified as acute or with a stated duration of 8 weeks or less

	411	Other acute and subacute forms of IHD	• Postmyocardial infarction syndrome • Intermediate coronary syndrome
	412	Old myocardial infarction	• Healed myocardial infarction • Past myocardial infarction diagnosed on ECG (EKG) or other special investigation, but currently presenting no symptoms
	413	Angina pectoris	• Angina decubitus • Prinzmetal angina • Other and unspecified angina pectoris
	414	Other forms of chronic IHD	• Coronary atherosclerosis • Aneurysm and dissection of heart • Other specified forms of chronic IHD • Chromic IHD, unspecified
Other forms of heart disease	427.5	Cardiac dysrhythmias	• Cardiac arrest

	429.2	Cardiovascular disease, unspecified	• Atherosclerotic cardiovascular disease
			• Cardiovascular arteriosclerosis
			• Cardiovascular:
			  o degeneration (with mention of arteriosclerosis)
			  o disease (with mention of arteriosclerosis)
			  o sclerosis (with mention of arteriosclerosis)

Symptoms, signs, and ill-defined conditions	799	Ill-defined and unknown causes of morbidity and mortality	• Asphyxia and hypoxemia • Respiratory arrest • Cachexia • Cardiorespiratory failure

E. DNA repair variant genotyping

First stage of SNP genotyping

The first stage of SNP genotyping was performed using matrix-assisted laser desorption/ ionization time-of-flight mass spectrometry (MALDI-TOF)[430] in Dr. Molly Bray's laboratory at the University of Texas. MALDI-TOF first amplifies the region of interest using PCR techniques, followed by enzymatic digestion of unincorporated dideoxynucleotides and single-stranded primers. The PCR product is then extended by a single base pair using a mini-sequencing reaction containing oligonucleotide primers, dideoxynucleotides, and a thermostable polymerase (Thermosequenase, Amersham Pharmacia). Extension products are then purified, concentrated, and combined with a matrix chemical that absorbs the laser energy and assists in sample mass determination. Reaction byproducts are analyzed by the MGS1 software application that includes a database which identifies each sample, the mini-sequencing primers present, and the polymorphisms being genotyped. The software then determines the masses of the unextended oligos, identifies which mass signals are present in the sample spectrum and assigns the genotype. In addition, blind replicates were included to examine genotyping repeatability.

Second stage of SNP genotyping

The second genotyping stage used the BeadStation system (Illumina, Inc., San Diego, CA) and a custom oligonucleotide pool[431] and was performed in Dr. Molly Bray's laboratory. Briefly, double-stranded genomic DNA was labeled with biotin to facilitate the capture of single stranded DNA onto streptavidin-coated sepharose beads for purification of PCR template. The PCR template was created using a highly specific polymerase and ligase that extend and ligate allele-specific primers that bind to complementary sequences surrounding

the variant sites and include universal primer sequences and an "address" sequence that is ultimately hybridized to the genotyping array. The PCR template was amplified via the use of universal primers labeled with either Cy3 or Cy5 fluorescent tags, and the amplified products were hybridized to a fiber optic bundle array and imaged with the BeadArray Reader (Illumina). Allele detection and genotype calling were performed using the GenCall and GTS Reports software (Illumina).

Overall genotyping quality control

An initial quality check was performed by the Bray laboratory, in which Hardy Weinberg Equilibrium (HWE) calculations were performed for each SNP in the entire population. All SNPs demonstrated low missing frequencies and good quality scores. The quality of each DNA plate was assessed and a cluster file was developed with the seven best plates on which to "train" all other data. The identical cluster file, with minor modifications, was used for all genotype calling, ensuring no batch or plate bias in the data.

Further quality control was carried out as detailed below. Controls internal to the assay were used to assess the completion of each assay step, quality standards for allele specific extension, hybridization, and PCR uniformity were verified for each plate, and each plate contained replicate QC samples. Each plate contained numerous blank wells confirmed to have average intensity signals at or near zero. In addition, each plate was verified to have no contamination present from previous assays. Agreement across replicate samples exceeded 0.99 for every valid QC sample in the plate. These data were also compared to five randomly chosen SNPs previously genotyped in the ARIC cohort and the mismatch proportion ranged from 0.005 to 0.02. Overall 165 of the ARIC participants selected for this

study had no genotype data, of which 144 had no DNA and 21 were excluded as genotype calls were missing for most loci. Three participants did not consent to genomic studies.

F. Statistical analysis

1. Assessment of population substructure

In addition to the preliminary HWE calculations provided by the Bray lab, HWE was examined in the CRS, by race, for each DNA repair variant. For a biallelic locus in a randomly mating population, where the frequency of alleles are represented by 'p' and 'q', the distribution of genotypes in the referent population should be $p^2 + 2pq + q^2$. Deviations from HWE are assessed using a chi-square test. Significant deviations from HWE may be indicative of laboratory error[432] or a violation of the factors necessary to maintain HWE in a population, such as population admixture. While the power of HWE to detect population admixture is small, assessing HWE before analysis can generally reduce false positive findings of genes underlying complex traits[433].

2. Analysis of incident CHD

We initially employed a proportional hazards model and the pseudolikelihood, which accounted for the stratified random sampling and case-cohort design (the Barlow method). However, several covariates appeared to violate the proportional hazards assumption and/or their modest effect sizes complicated assessments of proportionality. Incidence rate ratios (IRR) were estimated with piecewise constant exponential models (piecewise by year) and empirical standard errors[434, 435].–

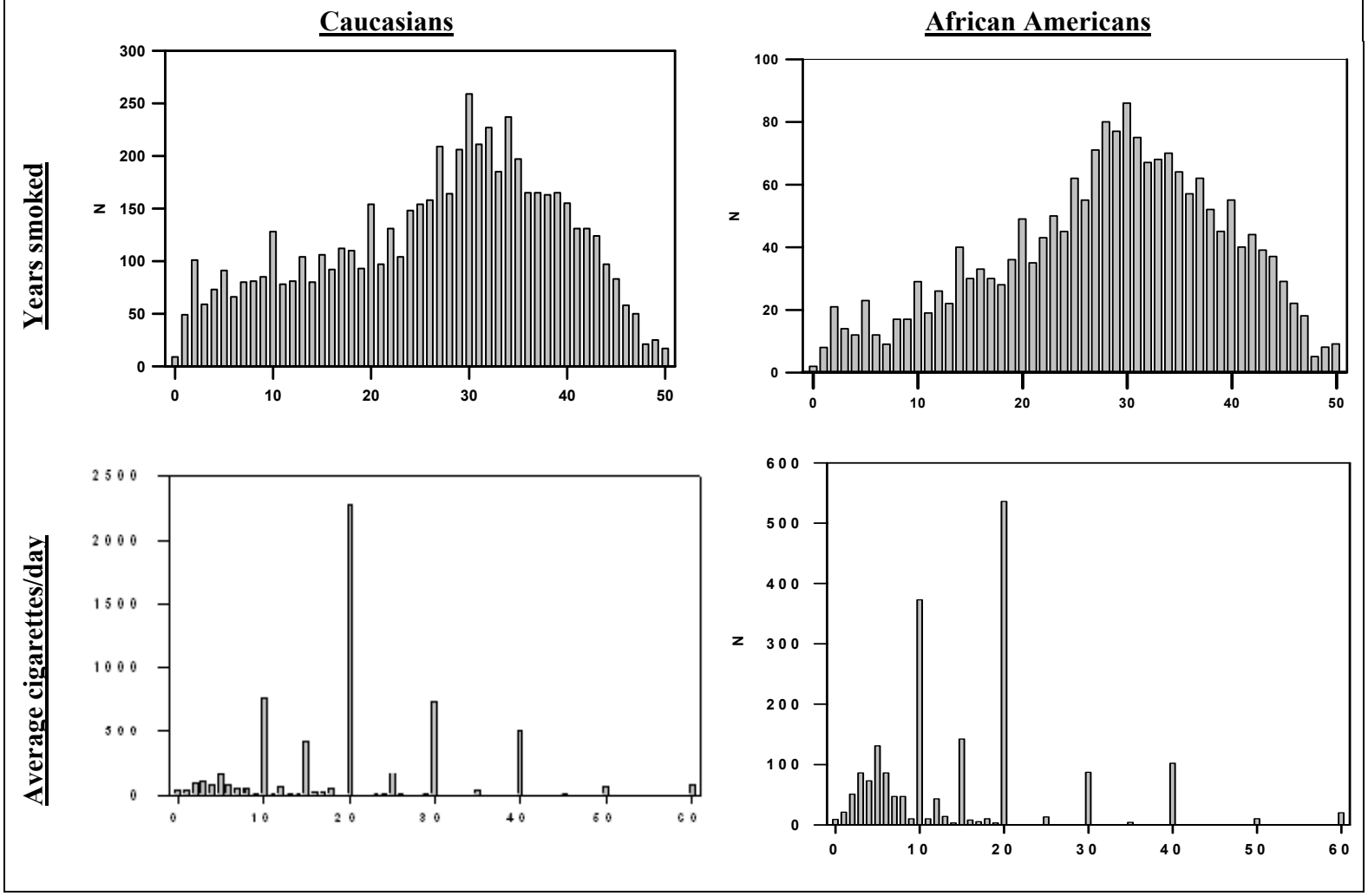
3. Analysis of mean IMT

As only 13% of study participants had full data for all six carotid artery sites, missing data were imputed for participants with at least one IMT measurement using sex- and race-specific models adjusting for age, body mass index, and arterial depth (BMDP 5V). Baseline mean IMT was then defined as the weighted IMT average at the six carotid artery sites after adjustment for measurement drift and reader differences[462]. Estimated site-specific reliability coefficients obtained from 36 ARIC participants scanned at three visits 7-14 days apart were 0.77, 0.73, and 0.70 for mean wall thickness at the carotid bifurcation, internal, and common carotid arteries, respectively[463]. A natural log transformation was applied to correct for non-normality.

4. Measurement of cigarette smoking

We measured cigarette smoke exposure using the ever-smoking metric, although other smoking measures were available including intensity, duration, age at initiation, second hand smoke exposure, and smoking status. Although ever-smoking considers all participants who reported ever smoking > 400 cigarettes as a homogeneous group, 90% of Caucasian and African American participants classified as ever-smokers reported ≥ 10 years of cigarette smoking (Figure 9). While there is sure to be some misclassification of exposure to cigarette smoke, the distribution of smoking duration and intensity indices suggest that the majority of participants reporting ever-smoking actually experienced long-term exposure. Practical constraints also limited our analytic options, as power would be reduced considerably if we considered a three-level categorization of cigarette smoke exposure and continuous parameterizations would also be infeasible given our hypothesis of modification by DNA repair variants.

Figure 9. Distribution of cigarette smoking metrics by race in 8,152 ARIC participants reporting ever-smoking.



5. Genotype analysis

All analyses were race-specific. Consistency of SNP genotypes with Hardy-Weinberg equilibrium (HWE) was evaluated among races by chi-square analysis and tagSNPs were identified using a pair-wise $r^2 \geq 0.80$. A general genetic model assuming no mode of inheritance was used when ever-smoking-tagSNP cell frequencies exceeded ten; otherwise an autosomal dominant model was assigned. Missing genotype data were imputed by race using fastPHASE[464]. Inferred genotypes were used for analyses if the posterior probability estimate exceeded 0.90.

6. Assessment of confounding

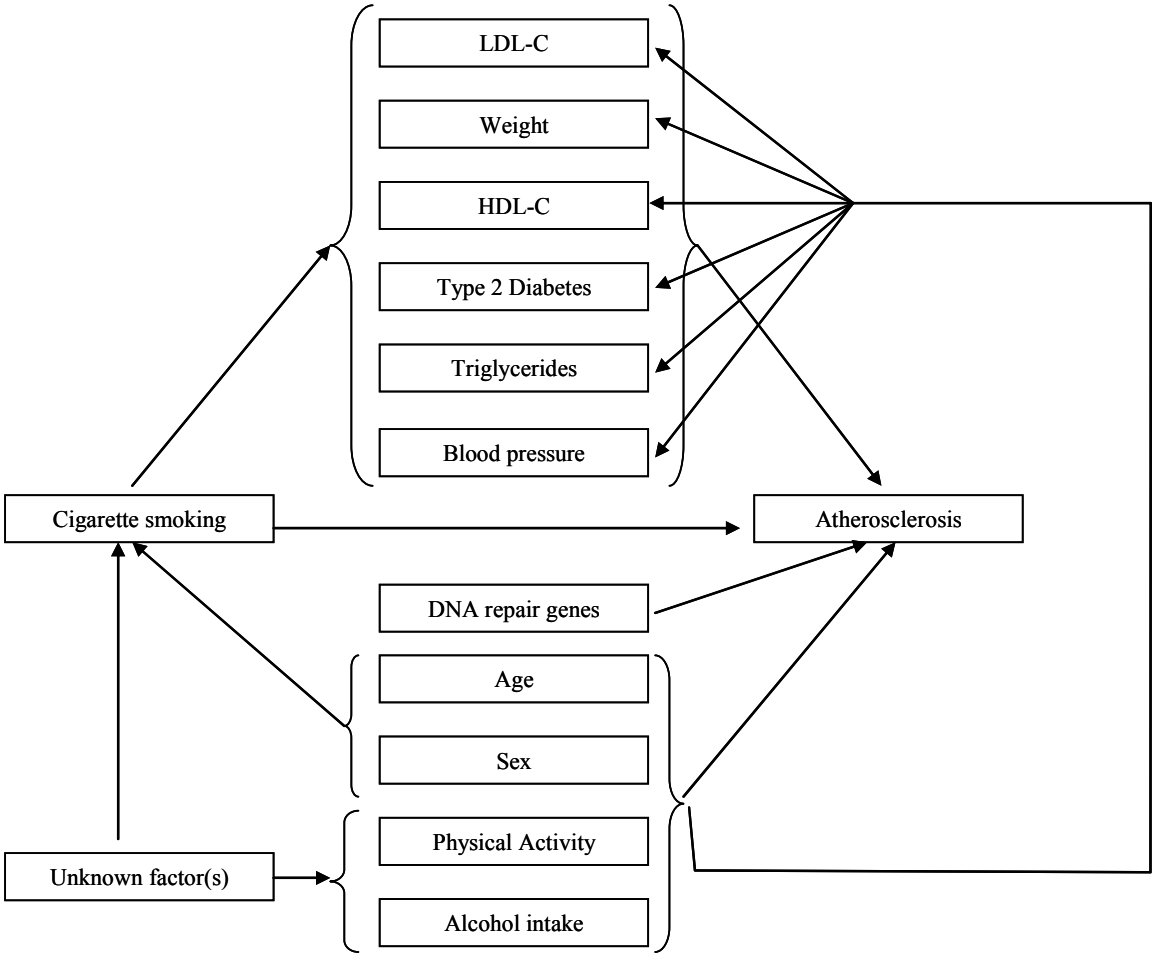
Table 20. Covariates identified as confounders by DAG analysis and existing parameterizations.

Covariate description	Coding
Age	As age (dichotomized at 55 years) was one of the stratified random sampling covariates, and was included as composite variable with sex in all analyses. Additional adjustment strategies within age strata was examined to account for residual confounding.
Sex	As sex was one of the stratified random sampling covariates, it was included as composite variable with age (dichotomized at 55 years) in all analyses.
Physical activity*	Physical activity was assessed using three ordinal variables (provided by ARIC investigators) that range from low (1) to high (5). The three physical activity indices were derived from the Baecke survey and correspond to physical activity from work, leisure, and sports activities. Possible parameterizations include categorical indicator coding as suggested by the data.
Alcohol intake*	Alcohol use is available as a continuous variable, defined as usual ethanol intake in grams/week using the original ARIC format. Non-linear alcohol use effects was examined, considering the J-shaped relationship between alcohol use and CHD. Possible parameterizations include splines or categorical indicator coding as suggested by the data.
Study center	Study center (categorical variable) was included in all analyses as the frequency of cigarette smoking and CHD may vary by study center, as might allele frequencies (for the univariable gene-CHD association analyses).

*Covariates included only in analyses examining cigarette smoking.

As the research is observational in nature, the potential for confounding exists. Confounders of the relationship between cigarette smoking and incident CHD was assessed with a directed acyclic graph (DAG)[438], informed by subject matter knowledge. The DAG representing the association between cigarette smoking and atherosclerosis is presented in Figure 10, informed from the literature. Based on Figure 10, a minimally sufficient adjustment set comprising included age, sex, physical activity, study center, and alcohol intake was identified.

Figure 10. DAG of the association between cigarette smoking and atherosclerosis.



7. Assessment of modification

A central aspect of this project is the evaluation of modification of the relationship between tobacco exposure and incident CHD by DNA repair polymorphisms. Additive interaction for generalized linear models was assessed using interaction contrast ratios (ICR)[439]. For IRR analysis and a dominant genetic model, ICR = IRR_{AB}– IRR_A – IRR_B+ 1, where IRR_{AB} represents the joint effect of cigarette smoking and the polymorphism and IRR_A and IRR_B represent the main effects of cigarette smoking and the polymorphism, respectively. Departures from zero suggest additive interaction.

8. Hierarchical regression

Genetic analyses typically involve estimating numerous SNP-disease associations using standard analytic approaches including: 1) fitting a saturated model containing all genetic variants; 2) fitting a saturated model and then reducing it using a preliminary-testing algorithm (i.e. forward, stepwise, or backward selection); or 3) fitting numerous one-variant-at-a-time models. Approach 1 is unfeasible if parameters are nonestimable and may provide biased and inefficient estimates[445]. Approach 2 excludes "nonsignificant" exposures despite biological plausibility or strength of association[446, 447] while producing biased point and variance estimates,[448, 449] and approach 3 neglects the correlation between SNPs. Furthermore, false positive associations, frequently reflecting a point estimate that is inflated and/or unstable[450] are not addressed by these methods.

Hierarchical regression methods, also known as random-coefficient or multilevel modeling, are a natural extension of the conventional analyses described in Aims 2 and 5, as we are evaluating multiple DNA repair polymorphisms, which are related spatially and possibly functionally. Hierarchical regression methods can generate estimates that are more stable than single-SNP associations, thus improving accuracy.

An improved accuracy in hierarchical regression analyses is achieved by "correcting" overestimates of the observed variance. Under independence, the variance of the distribution of estimates (V_E) is approximately equal to the sum of the variance of the distribution of true values (V_T) and random error, represented by the mean of the variances of the individual estimates (V_M)[451, 452].

$$V_E \approx V_T + V_M$$

Thus, estimates with smaller total error may be produced by "shrinking" unstable estimates towards the geometric mean of the ensemble of variants. For example, if an estimate is far from the geometric mean and has a large standard error, it may have been unduly influenced by random error and its true but unobserved effect is closer to the geometric mean than the value estimated in the first stage model[450]. Thus, we anticipate a reduction in total error by "shrinking" outlying conventional estimates toward the geometric mean if the geometric mean is reasonably close to the mean of the true values.

The degree of shrinkage is proportional to the precision of the estimate (measured in the first-stage model) and the prespecified variance of the parameter of interest (V_T), discussed below. However, the addition of a second stage will have little effect on estimates if adequate data are available[445]. The resulting group of shrunk estimates should have a distribution with variance less than the variance of the distribution of conventional estimates, and will outweigh any bias introduced by the shifts[450].

There are three assumptions inherent in hierarchical methods: 1) there is no systematic bias in the conventional estimates that compromise the validity of "shrinkage to the mean", 2) the true values and random errors are approximately normal, and 3) the true

values of the effect estimate are exchangeable[453]. Exchangeability implies an approximately equal prior for each SNP given an approximately equivalent sample size[454].

Hierarchical regression requires two models. In the first stage, incident CHD is regressed on individual variants and confounders, using the model:

$$E(\mathbf{Y}|\mathbf{X},\mathbf{W}) = \boldsymbol{\alpha} + \mathbf{X}\boldsymbol{\beta} + \mathbf{W}\boldsymbol{\gamma},$$

where **X** is a n-row matrix of DNA repair variants, **W** is a n-row matrix of confounders and ***y***, a vector of fixed coefficients, is presumed to be randomly sampled conditional on **X, W**[455].

The second stage, representing "categories of exchangeability", is added to improve the accuracy in estimating **β** and contains variables believed to determine the magnitude of, or explain some variability between, the individual target parameters[455]. For example, exchangeability could be presumed for SNPs in the same gene or genes in the same pathway if the effects are thought to arise from a common distribution with an unknown mean. Of note, exchangeability is a weaker assumption than one presuming that all effects are the same[454]. The second stage model is expressed as:

$$\boldsymbol{\beta} = \mathbf{Z}\boldsymbol{\pi} + \boldsymbol{\delta} = \boldsymbol{\mu} + \boldsymbol{\delta}$$

where **Z** is a *n*-by-*p* matrix of known prior covariates, **π** is a column vector of *p* prior coefficients corresponding to the effects of second-stage covariates on CHD, and **δ** are the random coefficient estimates, independent normal random variables with mean zero and variance τ^2 that represent effects not accounted for by the 'group' effect of the second-stage parameters. The distribution of **β** is referred to as the prior distribution, and integrates what

is known prior to observing the data. The hyperparameters $\boldsymbol{\mu}$ and $\boldsymbol{\tau^2}$ are the prior mean and prior variance (V_T) of $\boldsymbol{\beta}$[455]. The final mixed-effects generalized linear model is:

$$(E(\mathbf{Y}|\mathbf{X, Z, W})) = \alpha + \mathbf{XZ}\boldsymbol{\pi} + \mathbf{X}\boldsymbol{\delta} + \mathbf{W}\boldsymbol{\gamma},$$

V_T can be either estimated from the study data (the empirical-Bayes (EB) approach) or defined by investigators using background information (the semi-Bayes (SB) approach)[450, 456]. We used both SB and EB methods. While EB estimates all parameters from the data, this method can "overshrink" estimates, especially in the context of large datasets with numerous parameters.

Although SB can outperform EB, it requires accurate prior information regarding the parameters, which may not be available[456] . For example, in SB analyses V_T values are chosen such that at least 95% of the true associations would be captured by the interval $2(1.96)\sqrt{V_T}$. A $V_T = 0.35$ presumes, with 95% certainty, that the IRR for each SNP, after accounting for the second-level priors, would fall within a 10-fold range around its prior mean $((\ln(10))/3.92)^2 \approx 0.35)$, assuming normality. As the true value of V_T is unknown, a range of estimates are often used to determine how sensitive the results are to the choice of V_T. Large V_T values imply considerable residual effects of exposure beyond that explained by the second-stage covariates, while smaller V_T values indicate that the effects of exposure are well captured by the second-stage covariates. In addition, values of V_T can vary for different first stage exposures. While we may not have adequate information to accurately define V_T for the genetic effects, we can define V_T values to reflect plausible ranges for the effect of smoking on CHD. SAS commands PROC GLIMMIX[457] and PROC MIXED were used to fit the generalized and general linear hierarchical models.

Hierarchical analysis and interaction

Hierarchical regression methods can also be extended to incorporate effect measure modification by including priors for genetic and environmental effects. While our sample size did not permit evaluating modification of the CHD-smoking or IMT-smoking relationship for all genetic effects simultaneously (i.e. including the environmental factor, all genetic factors, and all product terms in a single model), individual and joint effects of smoking and a given variant can be assessed in models including the main effects of all other variants and confounders[458]. An example Z matrix for the interaction between cigarette smoke exposure and the *XRCC1* variant rs1475933 is presented in Table 21.

Table 21. Potential Z matrix (prior) for the interaction between cigarette smoke exposure and the *XRCC1* variant rs1475933.

Effect	Variant	Type	*XRCC1*	Cigarette smoke	Cigarette smoke and rs1475933
XRCC1	rs1475933	Intronic	**1**	0	**1**
	rs1799778	Intron	**1**	0	0
	rs1799782	Trp194Arg	**1**	0	0
	rs25486	Intron	**1**	0	0
	rs25487	Arg339Gln	**1**	0	0
	rs3213245	UTR	**1**	0	0
	rs3213282	Intron	**1**	0	0
	rs915927	Pro206Pro	**1**	0	0
	rs25489	His280Arg	**1**	0	0
	rs2228487	His107Arg	**1**	0	0
	rs2307187	UTR	**1**	0	0
	rs2307189	Thr42Thr	**1**	0	0
	rs25474	Leu514Pro	**1**	0	0
	rs25496	Ala72Val	**1**	0	0
Cigarette smoke	.	.	0	**1**	**1**

*Matrix does not account for MAF < 0.05, HWE < 0.001, or LD between SNPs within a given gene.

Construction of the hierarchical regression Z matrices

There are numerous Z matrices (e.g. priors) that may be informative when assessing the main and joint effects of DNA repair genes (Table 22). The first apparent SNP grouping

would be by gene, presented in Table 23, where the SNP effects are considered exchangeable within a given gene. SNPs could also be grouped by SNP type within a given gene, across genes (Table 24), or across DNA repair pathways. A variant of the Z matrices presented in Table 22 may also be considered by assigning a score of -1 to variants hypothesized to have opposite effects[453] (e.g. a negative coefficient), as informed by the functional literature.

While the analysis strategy presented in Table 22 is feasible for main effect estimation, the method by which the joint effects of DNA repair variants and cigarette smoking is assessed would necessitate a different model for each SNP considered in each prior matrix.

Table 22. Potential Z matrices for hierarchical models by estimation strategy.

Analysis strategy	Construction of the Z matrix (prior)
Main effect estimation	**Group SNPs by gene** • All SNPs are exchangeable within a given gene **Group SNPs by DNA repair pathway** • All SNPs are exchangeable within a given pathway **Analyze all SNPs simultaneously** • All DNA repair SNPs are considered exchangeable
Joint effect estimation	**Group SNPs by gene** • All SNPs are exchangeable within a given gene **Group SNPs by DNA repair pathway** • All SNPs are exchangeable within a given pathway **Analyze all SNPs simultaneously** • All DNA repair SNPs are considered exchangeable

Table 23. Example of Z matrix (prior) for SNPs by gene.

Gene*	Variant	Type	*XRCC1*	*XRCC3*	*XPD*	*hOGG1*	*APEX1*
XRCC1	rs1475933	Intron	**1**	0	0	0	0
	rs1799778	Intron	**1**	0	0	0	0
	rs1799782	Trp194Arg	**1**	0	0	0	0
	rs25486	Intron	**1**	0	0	0	0
	rs25487	Arg339Gln	**1**	0	0	0	0
	rs3213245	UTR	**1**	0	0	0	0
	rs3213282	Intron	**1**	0	0	0	0
	rs915927	Pro206Pro	**1**	0	0	0	0
	rs25489	His280Arg	**1**	0	0	0	0
	rs2228487	His107Arg	**1**	0	0	0	0
	rs2307187	UTR	**1**	0	0	0	0
	rs2307189	Thr42Thr	**1**	0	0	0	0
	rs25474	Leu514Pro	**1**	0	0	0	0
	rs25496	Ala72Val	**1**	0	0	0	0
XRCC3	rs1799796	Intron	0	**1**	0	0	0
	rs3212024	UTR	0	**1**	0	0	0
	rs3212057	UTR	0	**1**	0	0	0
	rs861531	Intron	0	**1**	0	0	0
	rs861539	Met241Thr	0	**1**	0	0	0
	rs1799795	Intron	0	**1**	0	0	0
	rs3212038[#]	UTR	0	**1**	0	0	0
	rs1799794	UTR	0	**1**	0	0	0
XPD	rs1052555	Asp711Asp	0	0	**1**	0	0
	rs1052559	Gln751Lys	0	0	**1**	0	0
	rs50871	Intron	0	0	**1**	0	0
	rs1799793	Asn312Asp	0	0	**1**	0	0
	rs3916874	Intron	0	0	**1**	0	0
	rs1618536	Intron	0	0	**1**	0	0
hOGG1	rs1052133	Cys326Ser	0	0	0	**1**	0
	rs1805373	Gln229Arg	0	0	0	**1**	0
	rs3219008	Intron	0	0	0	**1**	0
	rs2072668	Intron	0	0	0	**1**	0
APEX1	rs1048945	His51Gln	0	0	0	0	**1**
	rs3136817	Intron	0	0	0	0	**1**
	rs3136820	Glu148Asp	0	0	0	0	**1**
	rs3136814	UTR	0	0	0	0	**1**

*Matrix does not account for MAF < 0.05, HWE < 0.001, and LD between SNPs within a given gene.

Table 24. Examples of Z matrices (priors) for DNA repair variants by SNP type.

Gene*	Variant	Type	Nonsynonymous	Synonymous	Intronic	UTR
XRCC1	rs1475933	Intron	0	0	**1**	0
	rs1799778	Intron	0	0	**1**	0
	rs1799782	Trp194Arg	**1**	0	0	0
	rs25486	Intron	0	0	**1**	0
	rs25487	Arg339Gln	**1**	0	0	0
	rs3213245	UTR	0	0	0	**1**
	rs3213282	Intron	0	0	**1**	0
	rs915927	Pro206Pro	0	**1**	0	0
	rs25489	His280Arg	**1**	0	0	0
	rs2228487	His107Arg	**1**	0	0	0
	rs2307187	UTR	0	0	0	**1**
	rs2307189	Thr42Thr	0	**1**	0	0
	rs25474	Leu514Pro	**1**	0	0	0
	rs25496	Ala72Val	**1**	0	0	0
XRCC3	rs1799796	Intron	0	0	**1**	0
	rs3212024	UTR	0	0	0	**1**
	rs3212057	UTR	0	0	0	**1**
	rs861531	Intron	0	0	**1**	0
	rs861539	Met241Thr	**1**	0	0	0
	rs1799795	Intron	0	0	**1**	0
	rs3212038[#]	UTR	0	0	0	**1**
	rs1799794	UTR	0	0	0	**1**
XPD	rs1052555	Asp711Asp	0	**1**	0	0
	rs1052559	Gln751Lys	**1**	0	0	0
	rs50871	Intron	0	0	**1**	0
	rs1799793	Asn312Asp	**1**	0	0	0
	rs3916874	Intron	0	0	**1**	0
	rs1618536	Intron	0	0	**1**	0
hOGG1	rs1052133	Cys326Ser	**1**	0	0	0
	rs1805373	Gln229Arg	**1**	0	0	0
	rs3219008	Intron	0	0	**1**	0
	rs2072668	Intron	0	0	**1**	0
APEX1	rs1048945	His51Gln	**1**	0	0	0
	rs3136817	Intron	0	0	**1**	0
	rs3136820	Glu148Asp	**1**	0	0	0
	rs3136814	UTR	0	0	0	**1**

*Matrix does not account for MAF < 0.05, HWE < 0.001, and LD between SNPs within a given gene.

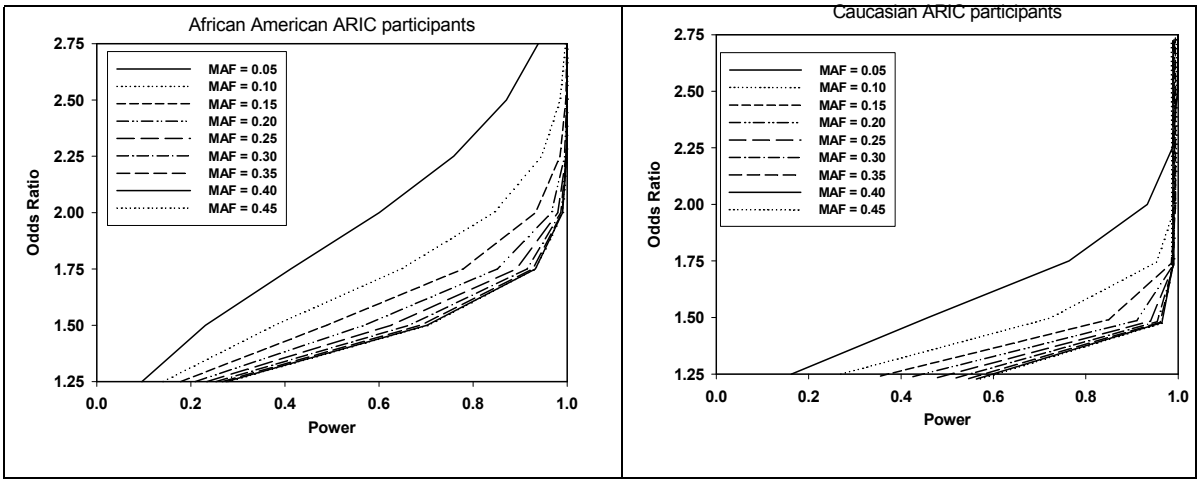
9. Multiple comparisons

One consideration is multiple comparisons. As no solid *a priori* evidence suggestive of an association between specific DNA repair variants, cigarette smoking, and incident CHD or subclinical disease exists, all DNA repair polymorphisms were examined. We did account for testing multiple hypotheses by adjusting alpha. Instead, this research focused upon estimating the main and joint effects of cigarette smoking and DNA repair polymorphisms rather than testing for statistical significance.

G. Sample size and statistical power

1. Statistical power for incident CHD

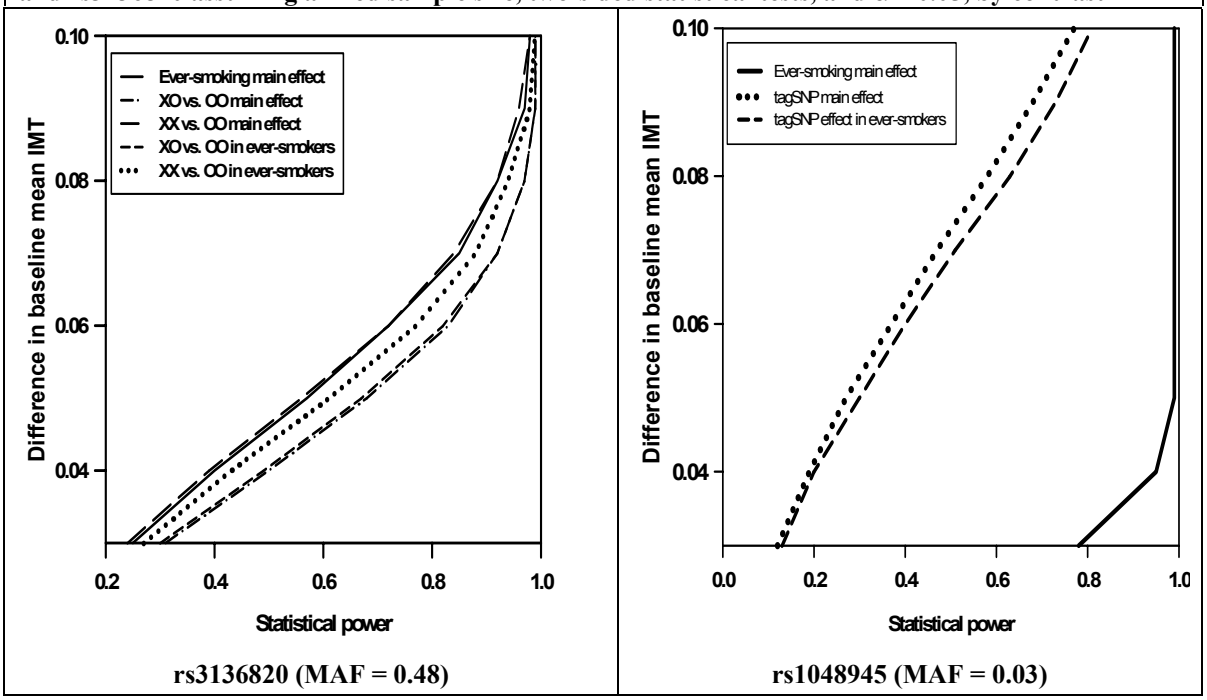
Figure 11. Statistical power for single-SNP associations in the African American and Caucasian ARIC participants for a fixed sample size assuming a two-sided statistical test and α = 0.05 by MAF.



Power curves, by race, to detect single-SNP associations for a range of MAFs are presented in Figure 11. Among Caucasians, MAFs for the DNA repair variants (Table 25) ranged between 0.001 and 0.49, with 78% having a MAF ≥ 0.20 and we have at least 80% power to detect an OR exceeding 1.50 when the MAF is ≥ 0.15. Within the African American stratum MAFs range from 0.002 to 0.43 (mean = 0.20), with 65% having a MAF ≥ 15%. Although underpowered for rare alleles or subtle effects, we have at least 80% to detect an OR exceeding 1.75 when the MAF is ≥ 15% among African Americans. However, these power curves are approximations, as they do not incorporate the sampling strategy or the weighting by person-time and presume all participants have complete data.

2. Statistical power for mean IMT

Figure 12. Statistical power for single-SNP associations in Caucasian ARIC participants for rs1048945 and rs3136820 assuming a fixed sample size, two-sided statistical tests, and α = 0.05, by contrast



Power curves for tagSNPs with the highest and lowest MAF in Caucasian CRS members for each ever-smoking – tagSNP stratum are presented in Figure 12. Standard deviation estimates for each ever-smoking stratum were estimated in ARIC Caucasian participants. The software n_Query advisor 4.0 (2000) was used to complete power calculations. With regards to the rs1048956 ever-smoking main effect, we have at least 80% power to detect a difference in baseline mean IMT exceeding 0.04 mm. As expected, the statistical power to detect differences in baseline mean IMT when considering the tagSNP main effect or modification by ever-smoking is quite dismal and does not exceed 80% even when differences in baseline mean IMT as large as 0.10 are considered. Conversely, we have at least 80% power to detect all main and joint effect estimates when a baseline mean IMT difference as low as 0.06 and tagSNP rs3136820 is considered. However, our estimated power curves are approximations, as they do not incorporate the sampling strategy, confounders, or the hierarchical regression methods

3. Statistical power for hierarchical regression

Estimating power for multilevel analyses is complicated, since a variable has effects at multiple levels. Generally the critical factor determining power for multilevel analyses is the number of higher-level units[459] (i.e., the number of participants, rather than the number of observations per participant). Since a well-established power calculation method was not available for hierarchical regression models except in overly simplified situations, we used multiple regression analysis with fixed effects instead. As our primary interest is in genotype effects, other covariates can be regarded as nuisance parameters. An index of the effect size the analyses could detect is provided by the difference in the R^2 values (the proportion of variance in the dependent variable which the model accounts for) between a full model that includes genotype and a reduced model with genotype removed.

The software n_Query advisor 4.0 (2000) was used to complete power calculations. Power was explored over a range of 0.20 – 0.30 for the multiple correlation (R^2). Based on our calculations, power approaches 0.80 with a sample as small as 623 persons when the reduced model accounts for 20% of the variance in CHD (as the case among the African American stratum), and the DNA repair variant increases the total variance by at least 1%. With a sample of 900, an absolute difference as small as 0.69% (or 0.93%) can be detected with 80% (or 90%) power. The power increases as the proportion of the total variance that is explained by a reduced model increases, assuming other conditions are held fixed. For example, if the multiple correlation coefficient of 30% (or 50%) is explained by the reduced model, a sample size as low as 544 (or 387) is sufficient to detect an 1% increase in R^2. As our total sample sizes for Caucasians and African Americans are 1,528 and 621, respectively, we expect to attain high statistical power for these analyses.

Table 25. Hardy-Weinberg Equilibrium *P* - values and minor allele frequency estimates, by race, for 36 DNA repair variants.

				Caucasians*		African Americans*	
Gene	**Cytogenic location**	**Variant**	**Type**	**HWE† *P*-value**	**MAF‡**	**HWE† *P*-value**	**MAF‡**
XRCC1	19q13.2	rs1475933	Intron	0.81	0.41	0.23	0.33
		rs1799778	Intron	0.88	0.38	0.37	0.22
		rs1799782	Trp194Arg	0.96	0.07	0.87	0.05
		rs25486	Intron	0.22	0.36	0.63	0.22
		rs25487	Arg339Gln	0.46	0.38	0.76	0.15
		rs3213245	UTR	0.51	0.40	0.70	0.41
		rs3213282	Intron	0.68	0.43	0.86	0.43
		rs915927	Pro206Pro	0.87	0.41	0.17	0.40
		rs25489	His280Arg	1.00	0.04	0.36	0.05
		rs2228487	His107Arg	.	0	1.0	0.002
		rs2307187	UTR	.	0	.	0
		rs2307189	Thr42Thr	.	0	1.0	0.002
		rs25474	Leu514Pro	1.0	0.001	1.0	0.002
		rs25496	Ala72Val	.	0	.	0
XRCC3	14q32.3	rs1799796	Intron	0.88	0.32	0.18	0.15
		rs3212024	UTR	0.69	0.31	0.18	0.15
		rs3212057	UTR	0.98	0.001	0.77	0.02
		rs861531	Intron	0.71	0.41	0.09	0.30
		rs861539	Met241Thr	0.53	0.39	0.88	0.23
		rs1799795	Intron	0.53	0.05	0.58	0.06
		rs3212038#	UTR	0.20	0.32	0.49	0.32
		rs1799794	UTR	0.44	0.20	0.68	0.23
XPD	19q13.3	rs1052555	Asp711Asp	0.20	0.35	0.75	0.11
		rs1052559	Gln751Lys	0.37	0.39	0.02	0.24
		rs50871	Intron	0.70	0.47	0.69	0.09
		rs1799793	Asn312Asp	0.21	0.25	0.74	0.35
		rs3916874	Intron	0.002	0.19	0.55	0.26
		rs1618536	Intron	0.04	0.34	1.00	0.37
hOGG1	3p26.2	rs1052133	Cys326Ser	0.24	0.24	0.47	0.17
		rs1805373	Gln229Arg	.	0	0.58	0.08
		rs3219008	Intron	0.08	0.21	0.04	0.41
		rs2072668	Intron	0.24	0.23	0.66	0.27
APEX1	14q11.2-q12	rs1048945	His51Gln	0.51	0.03	0.88	0.01
		rs3136817	Intron	0.34	0.24	0.96	0.15
		rs3136820	Glu148Asp	0.06	0.49	0.91	0.36
		rs3136814	UTR	0.94	0.06	1.00	0.04

*Metrics calculated in the cohort representative sample (CRS); †Hardy-Weinberg Equilibrium; ‡Minor allele frequency; #Poor quality score

CHAPTER V
RESULTS

A. Manuscript 1: DNA repair genes, cigarette smoking, and coronary heart disease: The Atherosclerosis Risk in Communities (ARIC) Study

ABSTRACT

Cigarette smoke contains over 50 mutagenic compounds and is associated with atherosclerotic conditions. As heritable differences in DNA repair ability can influence the effect of environmental exposures such as cigarette smoke, we conducted a series of case-cohort analyses to examine how variation in five DNA repair genes (*hOgg1, APEX1, XRCC1*, *XPD* and *XRCC3*) modified the association between ever-smoking and incident coronary heart disease (CHD) in the Atherosclerosis Risk in Communities (ARIC) cohort. All incident CHD cases identified from 1987-98 (n=1,086) and a random sample (n=1,065) were selected from eligible participants (n=14,255). Analyses were race-stratified and adjusted for sampling strategy, study center, alcohol intake and physical activity. Incidence rate ratios (IRR) were estimated by piecewise constant models. Departures from additivity were measured with interaction contrast ratios (ICR). Hierarchical modeling was used to improve estimation by incorporating priors into models examining genotype-by-smoking interaction. Although tagSNP main effects were generally null, when ever-smoking and priors for genetic (within gene) and environmental effects were added to the first-stage model, tagSNPs rs3212024 [*XRCC3,* $ICR_{XX\ vs.\ OO}$ = 1.1 (-0.7, 2.8), $ICR_{XO\ vs.\ OO}$ = 0.6 (-0.1,

1.4)] and rs50871 [*XPD,* $ICR_{XX\ vs.\ OO}$ = 1.2 (0.2, 2.2), $ICR_{XO\ vs.\ OO}$ = 0.8 (0.3, 1.4)] were associated with increases in the estimated effect of ever-smoking on incident CHD, while tagSNPs rs1799782 [*XRCC1,* $ICR_{XX,XO\ vs.\ OO}$ = -0.7 (-1.8, 0.3)] and rs861531 [*XRCC3,* $ICR_{XX\ vs.\ OO}$ = -1.1 (-2.8, 0.6), $ICR_{XO\ vs.\ OO}$ = -1.2 (-2.8, 0.3)] were associated with decreases. Though imprecise, our results suggest that DNA repair pathway variants may modify the effect of cigarette smoking on incident CHD. Further work examining these pathways is warranted.

1. Introduction

Coronary heart disease (CHD) poses a substantial public health burden, as it is the main cause of death in Western societies and has been predicted to remain so in future decades(1). Although evidence of increased risk for CHD associated with cigarette smoking is well established and consistent across age, sex, racial, and geographic strata (2-8), the mechanisms by which smoking initiates and/or precipitates vascular events remain poorly understood. Cigarette smoke contains approximately 4,800 chemicals (9), 67 of which are known to be mutagenic (10) and animal research has demonstrated that the tobacco smoke mutagens can induce and stimulate a proliferative vascular smooth muscle cell phenotype (11, 12). Elevated levels of DNA adducts have also been found in the vascular tissues of smokers (13, 14). As differences in responses to mutagen exposure have been attributed in part to heritable variation in DNA repair capacity (15), the identification of genes that influence the relationship between cigarette smoke exposure and atherosclerotic endpoints could provide new insights into the etiology of this major disease.

Although there are multiple DNA repair mechanisms, the base excision repair (BER), nucleotide excision repair (NER), and double-strand break (DSB) pathways are of the most importance for the repair of tobacco-related DNA damage(16). The BER pathway operates on small lesions that arise during inflammatory responses, spontaneously within the cell, or from exogenous agents(17), whereas NER is responsible for bulky helix-distorting DNA lesions such as pyrimidine dimers, larger chemical adducts, and cross-links(18). The DSB pathway repairs damage that affects both DNA strands (19).

While BER, NER, and DSB pathways involve over 130 genes(20), we focused on five that have been implicated in tobacco-related cancers: 8 – hydroxy-2' – deoxyguanosine-glycosylase/apurinic lyase (*hOGG1*), apurinic/apyrimidinic endonuclease (*APEX1*), X-ray repair cross complementing, group 1 (*XRCC1*), xeroderma pigmentosum D (*XPD*), and X-ray repair complementing defective repair in Chinese hamster cells 3 (*XRCC3*). *hOgg1* catalyzes the removal of 7,8-dihydro-8-oxoguanine from DNA(21-23), one of the most mutagenic DNA lesions(24). *APEX1* is considered the rate-limiting step in BER(25, 26) and processes abastic sites. While *XRCC1* has no known catalytic activity, it recognizes and binds single-strand DNA breaks(27) and complexes with other BER components(28-30). *XPD* is an ATP-dependent 5'-3' helicase (31) and *XRCC3* is involved in the homologous recombination DSB repair pathway(32).

Few population-based studies examining the relationship between DNA repair variants, cigarette smoke and athrothrombotic events have been published. Furthermore, no study has yet performed a comprehensive analysis of the role of DNA repair genes with regards to CHD or considered them as mediators of the cigarette smoke – CHD relationship. We measured the extent to which DNA repair variants modified the relationship between

cigarette smoking and incident CHD in individuals enrolled in the biracial Atherosclerosis Risk in Communities (ARIC) study.

2. Materials and methods

a. Study population and sources of data

ARIC, the parent population for this study, is a population-based longitudinal investigation of cardiovascular and pulmonary diseases in participants sampled from four locales: Washington County, Maryland (MD); Forsyth County, North Carolina (NC); Minneapolis, Minnesota (MN); and Jackson, Mississippi (MS) (33). Participants were followed via annual telephone interviews, clinic examinations approximately every three years 1987-1999, and ongoing hospital and death certificate record abstraction. The study protocol was approved by the Institutional Review Board of each center, and consent was obtained from each participant.

Of the 15,792 ARIC participants, 14,255 met the following eligibility criteria: 1) self-reported race of Caucasian from the NC, MN, or MD field centers or African American from the NC or MS field centers (N= 103 ineligible) and no history of 2) CHD (N=1,102 ineligible) or 3) transient ischemic attack or stroke (N=332 ineligible). A stratified random sample of eligible participants (cohort random sample, CRS) served as the reference group (N = 1,066, 85 of whom experienced a CHD event during follow-up). CRS selection was performed at baseline by stratifying eligible participants into eight groups based on age (≤ 55 years vs. > 55 years), sex, and race.

An event was classified as incident CHD if it met at least one of the following study criteria(34): 1) definite or probable myocardial infarction, 2) definite fatal CHD, 3) silent MI or 4) coronary revascularization procedure. Through December, 1998 a total of 1,086 incident CHD cases were identified. Cigarette smoking was ascertained at baseline using an

interviewer-administered questionnaire. We measured exposure to cigarette smoke using the ever-smoking metric defined as > 400 cigarettes in a lifetime.

We did not attempt to capture all genetic variability within the *XRCC1, APEX1, hOGG1, XPD,* and *XRCC3* genes. Instead, targeted SNP selection was informed by functional data, minor allele frequency (MAF, > 0.05), SNP type with preference for non-synonymous SNPs, association studies in the cancer literature, and patterns of pair-wise linkage disequilibrium (LD) reported in the CEPH population by the Hapmap project (http://www.hapmap.org/). We used matrix-assisted laser desorption/ ionization time-of-flight mass spectrometry(35) for the first genotyping stage and the BeadStation system (Illumina Inc., San Diego, CA) with a custom oligonucleotide pool for the second (36). Agreement across replicate samples exceeded 0.99 for every valid sample. The data were also compared to five randomly chosen SNPs previously genotyped in the ARIC cohort and the mismatch proportion ranged from 0.005 to 0.02.

b. Statistical Methods

Incidence rate ratios (IRR) were estimated with piecewise constant exponential models (37, 38). Pearson's chi-square tests were used to assess whether observed genotype distributions were consistent with expected Hardy-Weinberg equilibrium (HWE) proportions. A general genetic model assuming no mode of inheritance was used when all CHD-ever-smoking-tagSNP cell frequencies exceeded ten; otherwise an autosomal dominant model was assumed. Pair-wise LD statistics were calculated in the CRS and tagSNPs were identified using an $r^2 \geq 0.80$. Missing genotype data were imputed by CHD status and race

using fastPHASE(39). Inferred genotypes were used for analyses if the posterior probability estimate exceeded 0.90.

Confounders were identified from a directed acyclic graph(40) that considered age, sex, study center, physical activity, alcohol intake, serum lipoproteins concentrations, body composition, diabetes, and blood pressure. A minimally sufficient adjustment set comprising age, sex, physical activity, field center, and alcohol intake was identified. Physical activity was assessed using three variables that measured leisure, sport, and work-related physical activity(41). Alcohol intake was measured as usual ethanol intake (grams/week).

Additive interaction was assessed using interaction contrast ratios (ICR)(42). When considering a dominant genetic model, ICR = IRR_{AB}– IRR_A – IRR_B+ 1, where IRR_{AB} represents the effect of those exposed jointly to cigarette smoking and the polymorphism. IRR_A and IRR_B represent the effects of cigarette smoking and the polymorphism in the absence of the other, respectively. Departures from zero suggest additive interaction.

Hierarchical modeling

Genetic analyses typically evaluate numerous SNP-disease associations. Standard analytic approaches include: 1) fitting a saturated model containing all variants, 2) model building using a preliminary-testing algorithm (e.g. stepwise variable selection), or 3) fitting one-variant-at-a-time models. Approach one is unfeasible if parameters are nonestimable and may provide biased and inefficient estimates (43). Approach two excludes nonsignificant exposures despite biological plausibility or effect size (44, 45) while producing biased point and variance estimates(46), and approach three neglects the

correlation between SNPs. Furthermore, false positive associations, frequently reflecting a point estimate that is inflated and/or unstable(47) are not addressed by these strategies.

We addressed the potential for false positive associations and biased point estimates in part with hierarchical regression models, since they produce estimates with smaller total error by "shrinking" unstable estimates towards the geometric mean of the ensemble of variants. The degree of shrinkage was proportional to the precision of each estimate and a prespecified prior variance, which represented the range of effects remaining after the first- and second-level effects were estimated (48).

Hierarchical models required two stages. In the first stage, incident CHD was regressed on individual variants and covariates (48). tagSNP beta coefficients were then regressed in a second-stage linear model as a function of prior covariates (i.e. tagSNPs s are considered random observations around the second-stage prior covariates) and a pre-specified prior variance. The second stage prior covariates represented categories of exchangeability, added to improve the accuracy in parameter estimation and contained variables believed to determine the magnitude of, or explain some variability between, the individual tagSNP estimates(48). Exchangeability was presumed if a group of tagSNPs were thought to arise from a common distribution with an unknown mean and was a weaker assumption than one presuming all effects are equal(49). We evaluated three categories of exchangeability: all SNPs were exchangeable, SNPs within a given gene were exchangeable, and SNPs within a given pathway were exchangeable.

The pre-specified prior variance τ^2 was estimated both from the study data (the empirical-Bayes (EB) approach) and defined by investigators using prior information (the

semi-Bayes (SB) approach) (47, 50). τ^2 values for SB analyses were chosen such that at least 95% of the true associations were captured by the interval EXP(2(1.96)$\sqrt{\tau^2}$). $\tau^2 = 0.35$ presumed that 95% of all true IRRs lay within a 10-fold range around the prior mean. While our sample size did not permit the evaluation of modification of the CHD-smoking relationship by all tagSNPs simultaneously, interactions between cigarette smoking and each variant were assessed individually in models including the main effects of all other variants and confounders (51). Hierarchical models were fit using PROC GLIMMIX (SAS, Cary, NC)(52).

3. Results

Baseline characteristics by race and case status are presented in Table 1. Compared to eligible ARIC participants, incident CHD cases were more likely to be older, male, and ever-smokers, and reported slightly more alcohol intake and less physical activity. The CRS had a higher proportion of males and older participants than the case group, due to sample design

MAF estimates and HWE *P* – values are presented in Table 2. Genetic variation for the 36 SNPs was captured by 20 tagSNPs among Caucasians and 22 tagSNPs among African Americans. MAF estimates were generally high in Caucasians, suggesting adequate precision to measure both the main and joint effects of the tagSNPs. The smaller sample size (total N = 622, 55% with full tagSNP data) limited the power to detect effects among African Americans.

Maximum likelihood (i.e., non-hierarchical) estimates that included all tagSNPs, an indicator for ever-smoking status, and product term(s) for the tagSNP and ever-smoking

status are presented in supplemental Figures 1 (Caucasians) and 2 (African Americans). The ever-smoking estimates were relatively precise among Caucasians (95% confidence limit ratio range = 2.1 - 3.6 (CLR, defined as the upper limit of the confidence interval divided by the lower limit)). In general, elevated rates of incident CHD among ever-smokers were suggested (IRR range = 0.85 – 3.66, 95% of IRR estimates above 1.00). Numerous estimated tagSNP IRRs were considerably imprecise (e.g. the estimated joint effect of rs3213282 and ever-smoking: $IRR_{XX\ vs.\ OO} = 7.45$, 95% CLR = 55.4; the estimated main effect of rs3213245, $IRR_{XX\ vs.\ OO} = 0.22$, 95% CLR = 33.3), making these findings unpersuasive. Results among African Americans participants were even more variable and difficult to interpret.

The co-occurrence of multiple elevated effect estimates and wide variation in the estimated precision complicated interpretation of the entire panel of results presented in supplemental Figures 1 and 2. Thus, the analyses were extended by examining three categories of exchangeability (all tagSNPs are exchangeable, tagSNPs within each gene are exchangeable, and tagSNPs within each DNA repair pathway are exchangeable) and two prior variance specifications ($\tau^2 = 0.162$ and 0.35, corresponding to a 5- and 10- fold residual effect range around the prior mean, respectively). The EB method was employed, but this approach appeared to over-shrink effect estimates and often unrealistically estimated $\tau^2 = 0$ among the African American stratum (i.e. corresponding to a 95% certainty that there were no residual tagSNP effects after accounting for second-stage covariates).

Main and joint estimated IRRs specifying that tagSNPs within each gene were exchangeable and $\tau^2 = 0.35$ are presented in Figures 1 (Caucasians) and 2 (African Americans). This approach resulted in the estimation of six second-stage fixed effects for

analyses examining modification by DNA repair tagSNPs; five prior means that corresponded to each DNA repair gene (e.g. the estimated main and joint effects of rs1799782 were shrunk towards the estimated *XRCC1* fixed effect) and a sixth representing the estimated effect of ever-smoking (e.g. the main effect of ever-smoking and the joint effect of rs1799782 and ever-smoking were shrunk to the smoking fixed effect).

Incorporating a prior mean and variance improved the precision of unstable estimates (e.g. joint effect of the ever-smoking - rs3213282 XO vs. OO contrast , maximum likelihood approach: IRR = 3.99, 95% CLR = 9.1; hierarchical approach: IRR = 2.01, 95% CLR = 3.7) while producing little change in already stable estimates (e.g. estimated effect of ever-smoking for tagSNP rs3219008, maximum likelihood approach: IRR = 2.07, 95% CLR = 2.3; hierarchical approach: IRR = 1.97, 95% CLR = 2.3). This pattern that was even more apparent within the African American stratum. Although two other categories of exchangeability and $\tau^2 = 0.162$ were considered, results were generally similar although more precise when specifying $\tau^2 = 0.162$ (results not shown).

As in the maximum likelihood approach, the estimated effect of ever-smoking was associated with an increased rate of incident CHD among Caucasians, consistent with the large body of published literature on this topic. The point estimates (i.e., geometric means of the posterior probability distributions), ranged from 0.96 to 2.87 and 95% of them exceeded the null value. The estimated tagSNPs main effects were generally close to the null. Exceptions included rs861531 [$IRR_{XX \text{ vs. } OO} = 2.00$, 95% CLR = 3.6; $IRR_{XO \text{ vs. } OO} = 1.42$ 95% CLR = 6.8], rs50871 [$IRR_{XX \text{ vs. } OO} = 0.92$, 95% CLR = 3.5, $IRR_{XO \text{ vs. } OO} = 0.68$, 95% CLR = 3.0] and rs1052555 [$IRR_{XX \text{ vs. } OO} = 0.78$, 95% CLR = 4.9, $IRR_{XO \text{ vs. } OO} = 1.46$, 95% CLR = 3.0].

When evaluating additive interaction, tagSNPs rs3213282 (*XRCC1*), rs50871 (*XPD*), and rs3212024 (*XRCC3*) were associated with increases in the estimated effect of ever-smoking on incident CHD while tagSNPs rs1799782 (*XRCC1*) and rs861531 (*XRCC3*) were associated with decreases (Figure 3). Other ICR estimates were difficult to reconcile, for example the ICR estimates for rs1052133 which suggest a protective effect for heterozygotes but a causative effect for minor allele homozygotes. The marked imprecision apparent for the *hOgg1* variants may reflect the fact that there were only two *hOGG1* tagSNPs identified among Caucasians, thus limited data were available to estimate a prior mean.

Table 26. (MS1: Table 1) Selected characteristics of the 14,255 study participants, by race and case status. ARIC Study baseline examination

		Caucasians			African Americans		
		Incident CHD cases (N=831)	CRS (N=698)	All eligible ARIC Participants (N=10,428)	Incident CHD cases (N=255)	CRS (N=367)	All eligible ARIC Participants (N=3,827)
Median age at baseline (IQR)		56 (52, 61)	57 (51, 60)	54 (49, 59)	56 (50, 60)	55 (50, 59)	53 (48, 58)
Male (%)		595 (71.6)	384 (55.0)	4,741 (45.5)	136 (53.3)	194 (52.9)	1,416 (37.0)
Ever smokers (%)		600 (72.2)	423 (60.7)	6,142 (58.9)	163 (63.9)	209 (57.1)	2,010 (52.5)
Current smokers (%)		280 (33.7)	154 (22.1)	2,552 (24.5)	111 (43.9)	113 (30.9)	1,120 (29.3)
Former smokers (%)		320 (38.5)	269 (38.6)	3,590 (34.4)	50 (19.8)	96 (26.2)	887 (23.2)
Median pack-years of smoking (IQR)		21.7 (0, 41.0)	7.5 (0, 28.3)	6.0 (0, 29.0)	10.5 (0, 26.0)	3.4 (0, 20.3)	0.5 (0, 17)
Median alcohol intake, grams/week (IQR)		0 (0, 45.3)	0 (0, 60.4)	0 (0, 52.8)	0 (0, 26.4)	0 (0, 39.6)	0 (0, 13.2)
Median exercise (IQR)	**Leisure**	2.25 (2.00, 2.75)	2.50 (2.00, 2.75)	2.00 (1.75, 2.25)	2.00 (1.50, 2.50)	2.00 (1.75, 2.25)	2.00 (1.75, 2.50)
	Sport	2.50 (2.00, 3.00)	2.50 (2.00, 3.00)	2.00 (1.75, 2.50)	2.00 (1.50, 2.25)	2.00 (1.75, 2.50)	2.00 (1.75, 2.50)
	Work	2.25 (0, 3.00)	2.00 (0, 3.00)	2.50 (1.00, 3.00)	2.25 (0, 3.00)	2.50 (1.00, 3.00)	2.50 (1.00, 3.00)
Mean person days at-risk		2,132.4	3,586.3	3,578.2	2,056.7	3,391.3	3,469.7

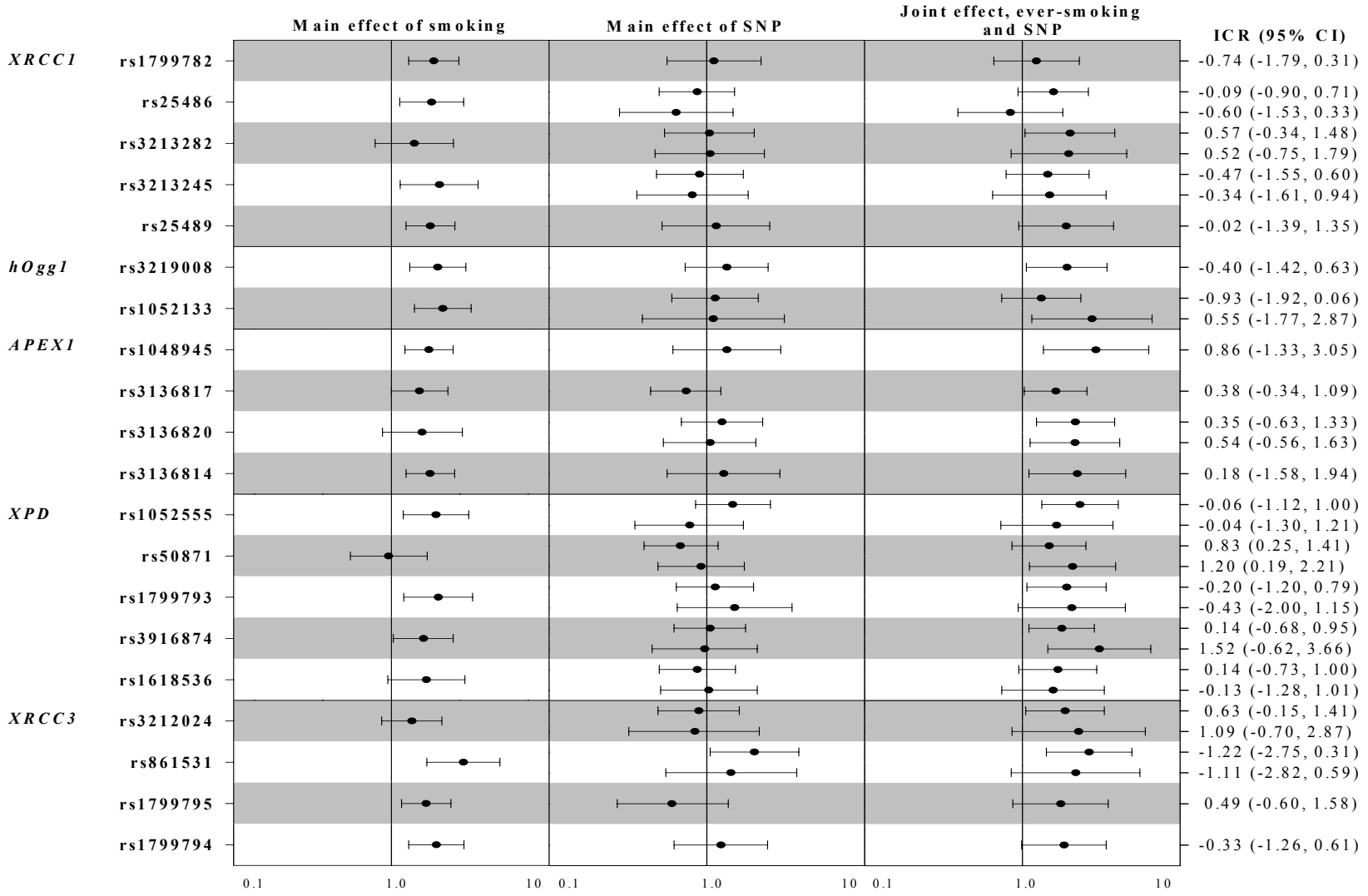
ARIC, Atherosclerosis Risk in Communities Study; CHD, coronary heart disease; CRS, cohort random sample; IQR, Interquartile range

Table 27. (MS1: Table 2) Hardy-Weinberg Equilibrium *P* - values and minor allele frequency estimates for 36 DNA repair variants in 698 Caucasian and 367 African American ARIC participants selected into the CRS.

				Caucasians		African Americans	
Gene	**Cytogenic location**	**Variant**	**SNP function**	**HWE* *P*-value**	**MAF***	**HWE* *P*-value**	**MAF***
Base Excision Repair (BER)							
XRCC1	19q13.2	rs1799782	Trp194Arg	0.08	0.07†	1.0	0.05†
		rs25489	His280Arg	0.19	0.04†	0.34	0.02¥
		rs25486	Intron	0.31	0.36†	0.54	0.23†
		rs3213282	Intron	0.68	0.44†	0.92	0.43†
		rs3213245	UTR	0.55	0.41†	0.92	0.41†
		rs1475933	Intron	0.93	0.41‡	0.17	0.34†
		rs1799778	Intron	0.98	0.37‡	0.33	0.23‡
		rs25487	Arg339Gln	0.49	0.37‡	0.79	0.15†
		rs915927	Pro206Pro	1.0	0.42‡	0.21	0.40‡
		rs2228487	His107Arg	.	0¥	.	0¥
		rs2307187	UTR	.	0§	.	0¥
		rs2307189	Thr42Thr	.	0¥	.	0¥
		rs25474	Leu514Pro	.	0¥	1.0	0.002¥
		rs25496	Ala72Val	.	0§	.	0¥
hOGG1	3p26.2	rs1052133	Cys326Ser	0.81	0.24†	0.36	0.18†
		rs3219008	Intron	0.11	0.21†	0.11	0.42†
		rs1805373	Gln229Arg	.	0¥	0.90	0.08†
		rs2072668	Intron	0.63	0.24‡	1.0	0.28†
APEX1	14q11.2-q12	rs1048945	His51Gln	1.0	0.03†	1.0	0.009¥
		rs3136820	Glu148Asp	0.12	0.48†	1.0	0.36†
		rs3136817	Intron	0.53	0.23†	0.96	0.15†
		rs3136814	UTR	1.0	0.03†	0.08	0.14†
Nucleotide Excision Repair (NER)							
XPD	19q13.3	rs1052555	Asp711Asp	0.32	0.35†	1.0	0.11‡
		rs1799793	Asn312Asp	0.87	0.36†	0.87	0.12†
		rs1618536	Intron	0.26	0.45†	1.0	0.13†
		rs3916874	Intron	0.38	0.26†	0.0002	0.07†
		rs50871	Intron	0.94	0.46†	0.87	0.09†
		rs1052559	Gln751Lys	0.46	0.40‡	0.08	0.24†
Double-Strand Break Repair (DSB)							
XRCC3	14q32.3	rs861531	Intron	0.57	0.40†	0.21	0.30†
		rs1799795	Intron	0.69	0.01†	1.0	0.02¥
		rs1799794	UTR	0.33	0.19†	0.67	0.20†
		rs3212024	UTR	0.60	0.30†	0.87	0.16†
		rs861539	Thr241Met	0.69	0.39‡	0.69	0.24†
		rs1799796	Intron	0.82	0.31‡	0.65	0.15‡
		rs3212038	UTR	<0.0001	0.38§	.	0§
		rs3212057	UTR	1.0	0.001¥	1.0	0.02¥

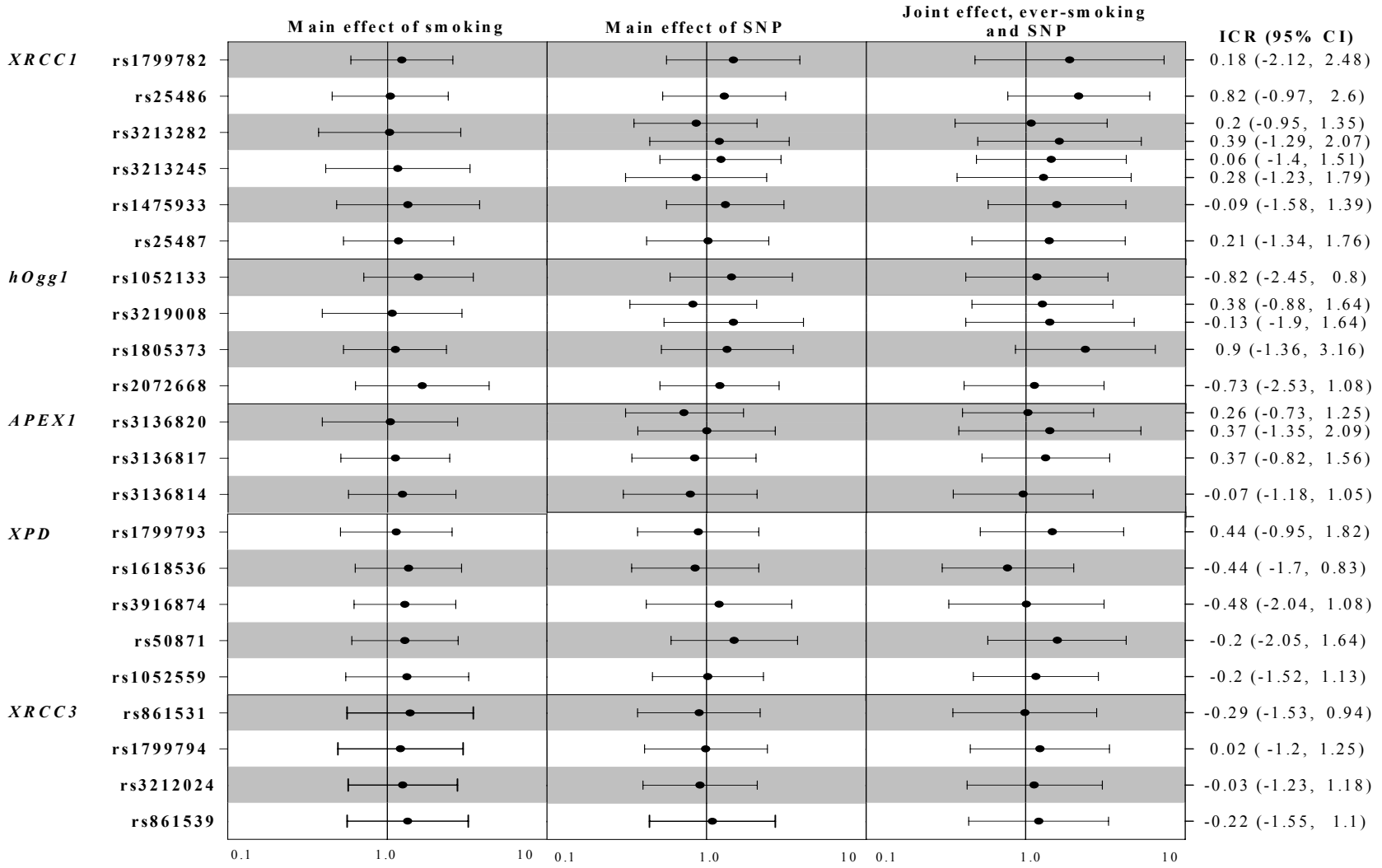
*Metrics calculated in the cohort representative sample (CRS); †tagSNP; ‡non-tag SNP, not analyzed further; §Poor quality score, not analyzed; ¥MAF too low, not analyzed; HWE, Hardy-Weinberg Equilibrium, MAF, minor allele frequency

Figure 13. (MS1: Figure 1) Main and joint estimated effects of 20 DNA repair tagSNPs and ever-smoking (IRRs and 95% confidence intervals) on the rate of incident CHD in 1,160 Caucasian ARIC participants specifying tagSNPs within each gene as exchangeable and a 10-fold 95% IRR residual effect range ($\tau^2 = 0.35$).



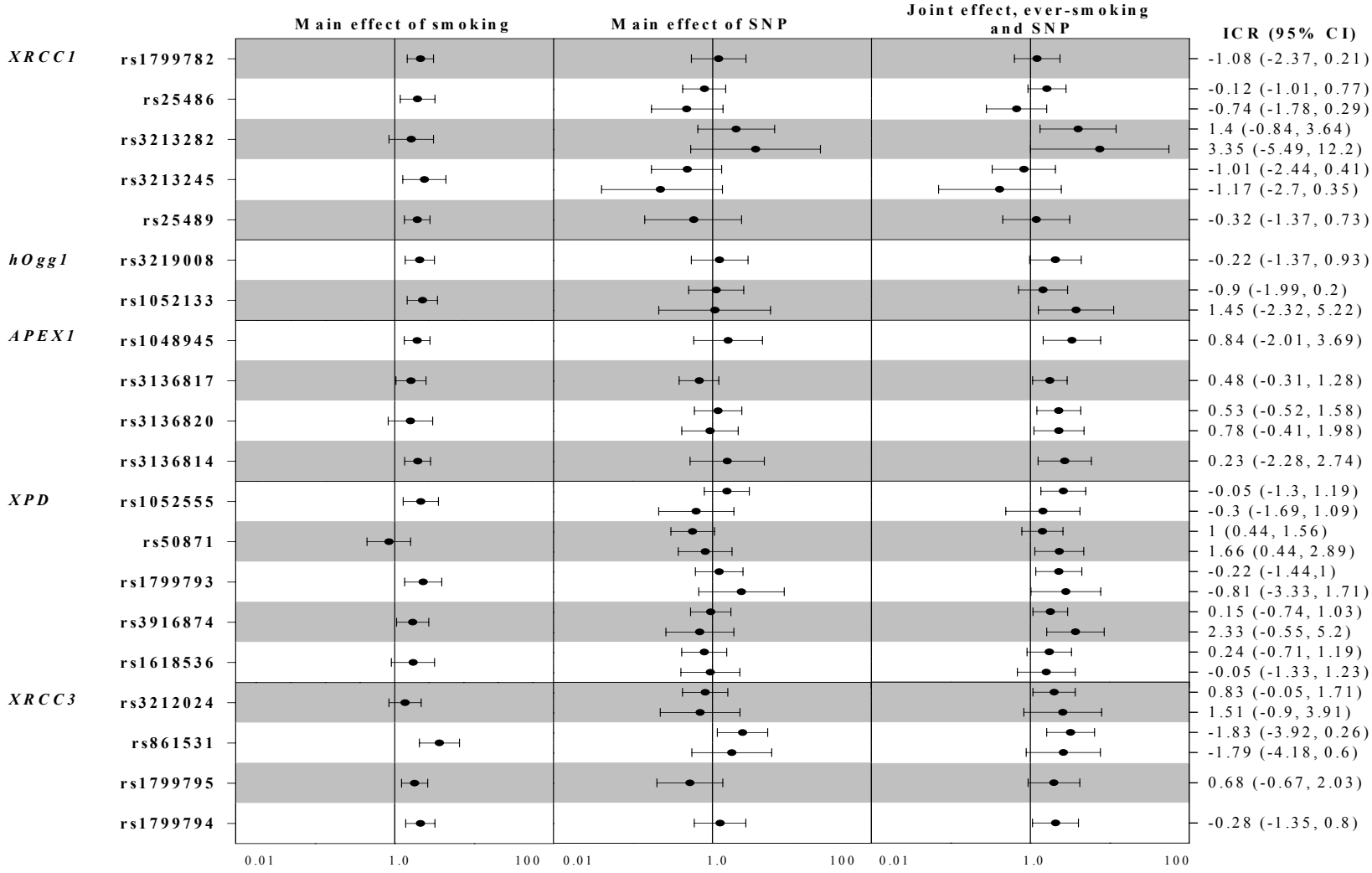
IRR, incidence rate ratio; ICR, interaction contrast ratio; CHD, coronary heart disease; the XO vs. OO contrast is presented atop of the XX vs. OO contrast when a general genetic model was used.

Figure 14. (MS1: Figure 2) Main and joint estimated effects of 22 DNA repair tagSNPs and ever-smoking (IRRs and 95% confidence intervals) on the rate of incident CHD in 345 African American ARIC participants specifying tagSNPs within each gene as exchangeable and a 10-fold 95% IRR residual effect range ($\tau^2 = 0.35$).



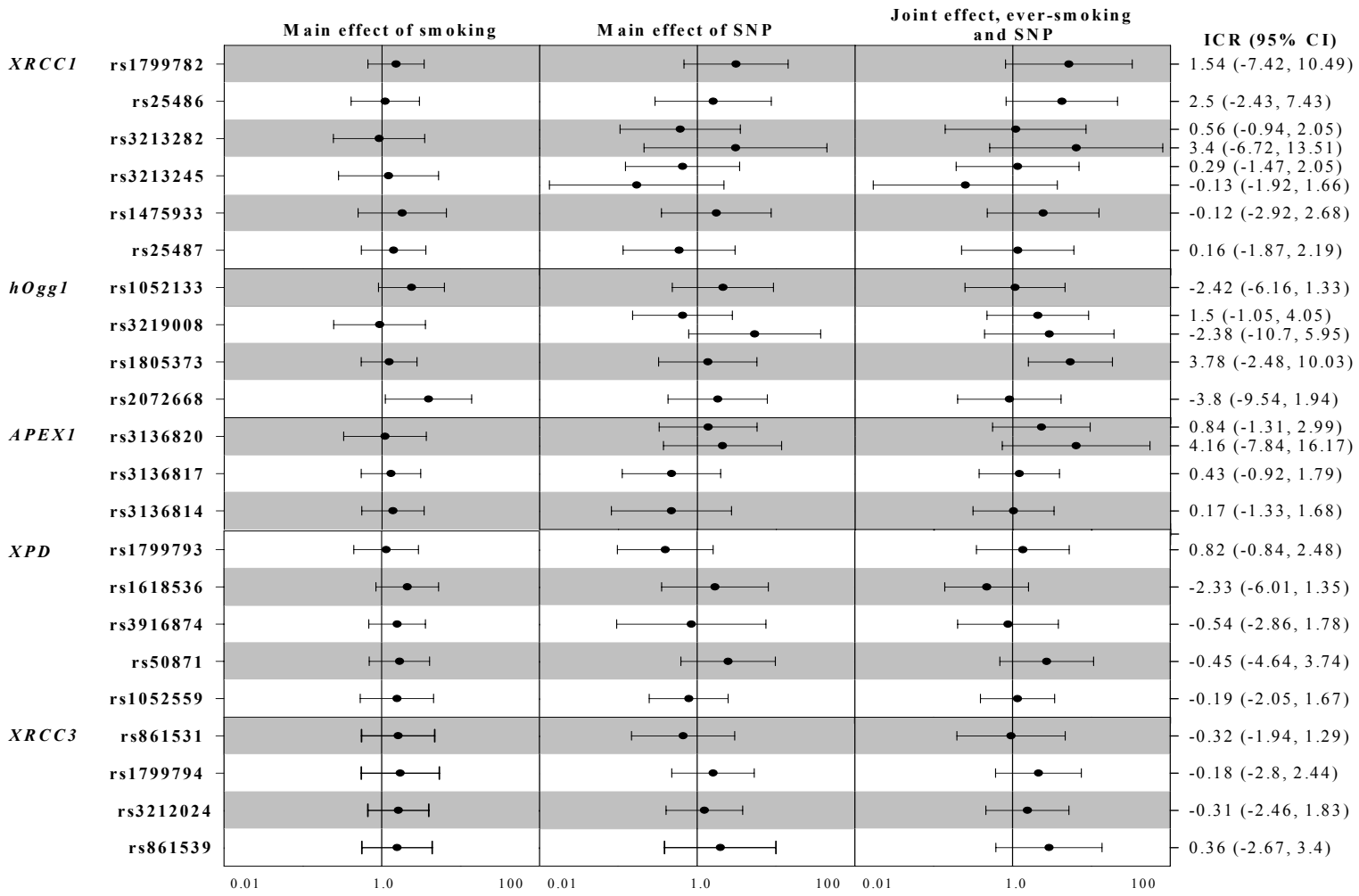
IRR, incidence rate ratio; ICR, interaction contrast ratio; CHD, coronary heart disease; the XO vs. OO contrast is presented atop of the XX vs. OO contrast when a general genetic model was used.

Figure 15. (MS1: Figure S1) Main and joint estimated effects of 20 DNA repair tagSNPs and ever-smoking (IRRs and 95% confidence intervals) on the rate of incident CHD in 1,160 Caucasian ARIC participants including a gene-smoking product term for one SNP and the main effects of all others.



IRR, incidence rate ratio; ICR, interaction contrast ratio; CHD, coronary heart disease; the XO vs. OO contrast is presented atop of the XX vs. OO contrast when a general genetic model was used.

Figure 16. (MS1: Figure S2) Main and joint estimated effects of 22 DNA repair tagSNPs and ever-smoking (IRRs and 95% confidence intervals) on the rate of incident CHD in 345 African American ARIC participants including a gene-smoking product term for one SNP and the main effects of all others.



IRR, incidence rate ratio; ICR, interaction contrast ratio; CHD, coronary heart disease; the XO vs. OO contrast is presented atop of the XX vs. OO contrast when a general genetic model was used.

4. Discussion

We examined the relationship between ever-smoking, DNA repair polymorphisms, and incident CHD. We show that modification of the association between ever-smoking and incident CHD by DNA repair variation is a potentially informative hypothesis that warrants further investigation. *A priori* implementation of prior probability distributions through hierarchical analysis adjusted implausible estimates and enhanced precision, thus facilitating the interpretation of the entire panel of results. This represents a likely improvement upon traditional analytic methods.

Mutagens in cigarette smoke may cause genetic damage by binding to, or interacting with DNA, although mutagen exposure does not invariably result in chromosomal damage, as pathways including DNA repair must fail for the latter to occur (15). If a polymorphism increases or decreases the ability of a protein to repair DNA damage, evaluating the variant in the context of cigarette smoke exposure can provide insights into the mechanisms by which cigarette smoke promotes atherothrombotic events.

While functional data are limited, some are published for certain variants we evaluated. Functional data for Arg339Gln (rs25487, tagged by rs25486 ($r^2 = 0.97$)) suggested that cells with the 399Gln polymorphism are more sensitive to chemically induced DNA damage(53, 54). In contrast, we show that the variant is associated with a decrease in the estimated effect of ever-smoking on incident CHD in a community-based study population. Several studies of smoking-related cancers also described discrepant genetic effects when smoking status was evaluated(55-57). While studies demonstrated that cells with the 280His polymorphism (rs25489) accumulated single strand DNA breaks after

exposure to genotoxic chemicals(58), our results were null, as were results for the *XPD* variants Gln751Lys (rs1052559, tagged by rs1052555, $r^2 = 0.982$) and Asn312Asp (rs1799793).

Results for the *XRCC1* Trp194Arg (rs1799782) and *XRCC3* Thr241Met (rs861539, tagged by rs861531, $r^2 = 0.982$) variants appear consistent with published literature. Mutation assays examining the Trp194Arg suggested that cells with the Trp allele had lower numbers of chromosomal breaks (54), which is consistent with our analysis showing that the Trp allele was associated with a decrease in the estimated effect of ever-smoking (ICR = -0.74 (-1.79, 0.31). When Thr241Met was examined in human *XRCC3*-/- cell lines that exhibited a phenotype of increased sensitivity to DNA cross-linking agents, expression of the minor allele did not fully restore the wild-type phenotype(59). Although our data on the main effects of Met/Thr and Met/Met indicated an increase in the rate of incident CHD (IRR = 2.00 (1.05, 3.83) and 1.42 (0.55, 3.72)) (Figure 3), ICR estimates suggest that the minor allele is associated with a decrease in the causative effect of ever-smoking. If replicated by other studies, these results can contribute to bridge the laboratory, experimental, and population-based inquiries.

Reconciling population-based findings with functional data remains a challenge, particularly since our work focused on clinically manifest, downstream outcomes of processes that putatively initiate or promote the underlying morbidity, namely the extent and severity of atherosclerosis. Although it represents a target outcome in the efforts to reduce the population burden of CVD, CHD is a complex phenotype influenced by numerous environmental and metabolic factors. Such complexities make it less likely that individual

variants have detected effects in a decades-long process and point to the need for consideration of gene-gene and gene-environmental contexts.

Although the polymorphisms we examined were carefully selected, our analysis was limited to 36 variants from five genes (six of which were monomorphic). As the pathways we examined contain over 130 genes(20), additional work to evaluate the role of DNA repair genes is needed. In addition, while the indirect candidate association approach we used is a powerful method, it assumes little allelic heterogeneity within loci and the common disease/common variant paradigm. This strategy would be unsuccessful if the genetic component of atherothrombosis involves numerous rare variants at many loci(60).

As no *a priori* evidence suggestive of an association between specific variants, cigarette smoking, and incident CHD existed, all tagSNPs were examined. We did not account for testing multiple hypotheses by adjusting the alpha level criterion, as we focused upon describing the magnitude and precision of the estimates, rather than significance testing, while exercising awareness of the potential for random error in the interpretation of results. We also did not account for uncertainty using a weighted analysis. The posterior probability estimates for the inferred genotypes above the 0.90 criterion consistently exceeded 0.99: thus, a weighted analysis would have had little effect on the results. Point estimates using a non-imputed data set or considering cases and CRS members together during imputation were comparable (results not shown).

We assessed cigarette smoke exposure using the ever-smoking metric, although this measure implies that all participants who reported ever smoking > 400 cigarettes are fairly homogeneous as a group. The ever-smoking metric is conservative as it does not account for

the magnitude or duration of cigarette smoke exposure; thus our results should be robust to other smoking measures. Indeed, results were similar when using the pack-years metric dichotomized at 20 (results not shown). Practical study size considerations also limited our ability to consider a three-level categorization of cigarette smoke exposure and a continuous parameterization of cigarette smoking exposure would be infeasible given the hypothesis of modification by DNA repair variants.

The application of hierarchical regression methods without attention to model assumptions can produce estimates that are more biased than those obtained from traditional methods(61). Attempts to improve accuracy could result in increased bias if, for example, the second-stage fixed effects are poor measures of the true means. However, the three different categories of exchangeability we considered provided consistent estimates, suggesting that the results are fairly robust. While the prior information was somewhat crude, a hierarchical model with even a simplified second stage can outperform maximum likelihood methods(62). Drawbacks include the requirement that participants have full data on all variables, which may be problematic, especially as researchers assay larger and larger regions of the genome.

Cigarette smoking is a major threat to public health and an established risk factor for CHD. Although imprecise, particularly for African Americans and variants with low MAFs, our results can stimulate inquiries into potential mechanisms linking tobacco exposure and atherothrombotic disease. Of note, no previous study has yet to perform a comprehensive analysis of the role of DNA repair genes with regards to CHD or considered their role as biologically plausible mediators of the cigarette smoke – incident CHD relationship. Future studies in different populations will be required to validate our findings and improve our

understanding of the complex relationship between DNA repair variants, cigarette smoking, and CHD.

Acknowledgements

This publication was made possible by grant HL074377 awarded to Dr. North, training grant HL007055, the American Heart Association predoctoral award 0615310U (Ms. Avery), and by a grant from the National Institute of Environmental Health Sciences (P30ES10126, Dr. Olshan). Dr. Poole received partial support from National Institute of Environmental Health Sciences grant PE0ES10126. The ARIC study is carried out as a collaborative effort supported by NHLBI contracts N01-HC-55015, N01-HC-55016, N01-HC-55018, N01-HC-55019, N01-HC-55020, N01-HC-55021, and N01-HC-55022. The authors thank the staff and participants of the ARIC study for their important contributions. These contents are solely the responsibility of the authors and do not necessarily represent the official views of the funding agencies.

5. References

1. Murray CJ, Lopez AD. Alternative projections of mortality and disability by cause 1990-2020: Global Burden of Disease Study. Lancet 1997;349:1498-504.

2. Doll R, Gray R, Hafner B, Peto R. Mortality in relation to smoking: 22 years' observations on female British doctors. Br Med J 1980;280:967-71.

3. Doll R, Hill AB. Lung cancer and other causes of death in relation to smoking; a second report on the mortality of British doctors. Br Med J 1956;12:1071-81.

4. Doll R, Peto R. Mortality in relation to smoking: 20 years' observations on male British doctors. Br Med J 1976;2:1525-36.

5. Hammond EC, Horn D. Smoking and death rates; report on forty-four monghs of follow-up of 187,783 men. II. Death rates by cause. J Am Med Assoc 1958;166:1294-308.

6. Willett WC, Green A, Stampfer MJ, et al. Relative and absolute excess risks of coronary heart disease among women who smoke cigarettes. N Engl J Med 1987;317:1303-9.

7. Price JF, Mowbray PI, Lee AJ, Rumley A, Lowe GD, Fowkes FG. Relationship between smoking and cardiovascular risk factors in the development of peripheral arterial disease and coronary artery disease: Edinburgh Artery Study. Eur Heart J 1999;20:344-53.

8. Jonas MA, Oates JA, Ockene JK, Hennekens CH. Statement on smoking and cardiovascular disease for health care professionals. American Heart Association. Circulation 1992;86:1664-9.

9. Green CR, Rodgman A. The Tobacco Chemists' Research Conference: A half-century of advances in analytical methodology of tobacco and its products. Recent Adv. Tob. Sci. 1996;22:131-304.

10. Hoffmann D, Hoffmann I, El-Bayoumy K. The less harmful cigarette: a controversial issue. a tribute to Ernst L. Wynder. Chem Res Toxicol 2001;14:767-90.

11. Zhang Y, Ramos KS. The induction of proliferative vascular smooth muscle cell phenotypes by benzo[a]pyrene does not involve mutational activation of ras genes. Mutat Res 1997;373:285-92.

12. Penn A, Snyder CA. 1,3 Butadiene, a vapor phase component of environmental tobacco smoke, accelerates arteriosclerotic plaque development. Circulation 1996;93:552-7.

13. De Flora S, Izzotti A, Walsh D, Degan P, Petrilli GL, Lewtas J. Molecular epidemiology of atherosclerosis. Faseb J 1997;11:1021-31.

14. Van Schooten FJ, Hirvonen A, Maas LM, et al. Putative susceptibility markers of coronary artery disease: association between VDR genotype, smoking, and aromatic DNA adduct levels in human right atrial tissue. Faseb J 1998;12:1409-17.

15. Anonymous. Cancer. In: Stratton K, Shetty P, Wallace R, Bondurant S, eds. Clearing the Smoke. Assessing the Science Base for Tobacco Harm Reduction. Washington, D.C.: National Academy Press, 2001:367-469.

16. Wu X, Zhao H, Suk R, Christiani DC. Genetic susceptibility to tobacco-related cancer. Oncogene 2004;23:6500-23.

17. Eisen JA, Hanawalt PC. A phylogenomic study of DNA repair genes, proteins, and processes. Mutat Res 1999;435:171-213.

18. Squire JA, Whitmore GF, Phillips RA. A Genetic Basis of Cancer. In: Tannock IF, Hill RP, eds. The Basic Science of Oncology: McGraw-Hill Companies, Inc., 1998:48-78.

19. Kuschel B, Auranen A, McBride S, et al. Variants in DNA double-strand break repair genes and breast cancer susceptibility. Hum Mol Genet 2002;11:1399-407.

20. Wood RD, Mitchell M, Sgouros J, Lindahl T. Human DNA repair genes. Science 2001;291:1284-9.

21. Lindahl T, Wood RD. Quality control by DNA repair. Science 1999;286:1897-905.

22. Nishimura S. Involvement of mammalian OGG1(MMH) in excision of the 8-hydroxyguanine residue in DNA. Free Radic Biol Med 2002;32:813-21.

23. Shinmura K, Yokota J. The OGG1 gene encodes a repair enzyme for oxidatively damaged DNA and is involved in human carcinogenesis. Antioxid Redox Signal 2001;3:597-609.

24. Grollman AP, Moriya M. Mutagenesis by 8-oxoguanine: an enemy within. Trends Genet 1993;9:246-9.

25. Chaudhry MA, Dedon PC, Wilson DM, 3rd, Demple B, Weinfeld M. Removal by human apurinic/apyrimidinic endonuclease 1 (Ape 1) and Escherichia coli exonuclease III of 3'-phosphoglycolates from DNA treated with neocarzinostatin, calicheamicin, and gamma-radiation. Biochem Pharmacol 1999;57:531-8.

26. Parsons JL, Dianova, II, Dianov GL. APE1 is the major 3'-phosphoglycolate activity in human cell extracts. Nucleic Acids Res 2004;32:3531-6.

27. Marintchev A, Mullen MA, Maciejewski MW, Pan B, Gryk MR, Mullen GP. Solution structure of the single-strand break repair protein XRCC1 N-terminal domain. Nat Struct Biol 1999;6:884-93.

28. Masson M, Niedergang C, Schreiber V, Muller S, Menissier-de Murcia J, de Murcia G. XRCC1 is specifically associated with poly(ADP-ribose) polymerase and negatively regulates its activity following DNA damage. Mol Cell Biol 1998;18:3563-71.

29. Nash RA, Caldecott KW, Barnes DE, Lindahl T. XRCC1 protein interacts with one of two distinct forms of DNA ligase III. Biochemistry 1997;36:5207-11.

30. Caldecott KW, Aoufouchi S, Johnson P, Shall S. XRCC1 polypeptide interacts with DNA polymerase beta and possibly poly (ADP-ribose) polymerase, and DNA ligase III is a novel molecular 'nick-sensor' in vitro. Nucleic Acids Res 1996;24:4387-94.

31. Drapkin R, Reardon JT, Ansari A, et al. Dual role of TFIIH in DNA excision repair and in transcription by RNA polymerase II. Nature 1994;368:769-72.

32. Brenneman MA, Wagener BM, Miller CA, Allen C, Nickoloff JA. XRCC3 controls the fidelity of homologous recombination: roles for XRCC3 in late stages of recombination. Mol Cell 2002;10:387-95.

33. The Atherosclerosis Risk in Communities (ARIC) Study: design and objectives. The ARIC investigators. Am J Epidemiol 1989;129:687-702.

34. Lee CR, North KE, Bray MS, et al. NOS3 polymorphisms, cigarette smoking, and cardiovascular disease risk: the Atherosclerosis Risk in Communities study. Pharmacogenet Genomics 2006;16:891-9.

35. Bray MS, Boerwinkle E, Doris PA. High-throughput multiplex SNP genotyping with MALDI-TOF mass spectrometry: practice, problems and promise. Hum Mutat 2001;17:296-304.

36. Shen R, Fan JB, Campbell D, et al. High-throughput SNP genotyping on universal bead arrays. Mutat Res 2005;573:70-82.

37. Volovics A, Van Den Brandt PA. Methods for the analyses of case-cohort studies. Biometrical journal 1997;39:195-214.

38. Schouten EG, Dekker JM, Kok FJ, et al. Risk ratio and rate ratio estimation in case-cohort designs: hypertension and cardiovascular mortality. Stat Med 1993;12:1733-45.

39. Scheet P, Stephens M. A fast and flexible statistical model for large-scale population genotype data: applications to inferring missing genotypes and haplotypic phase. Am J Hum Genet 2006;78:629-44.

40. Greenland S, Brumback B. An overview of relations among causal modelling methods. Int J Epidemiol 2002;31:1030-7.

41. Baecke JA, Burema J, Frijters JE. A short questionnaire for the measurement of habitual physical activity in epidemiological studies. Am J Clin Nutr 1982;36:936-42.

42. Rothman KJ. Epidemiology: An Introduction. New York: Oxford University Press, 2002.

43. Witte JS. Genetic analysis with hierarchical models. Genet Epidemiol 1997;14:1137-42.

44. Greenland S. Comment: cautions in the use of preliminary test estimators. Stat Med 1989;8:669-673.

45. Greenland S. Modeling and variable selection in epidemiologic analysis. Am J Public Health 1989;79:340-9.

46. Sclove SL, Morris CN, Radhakrishna R. Non-optimality of preliminary-test estimators for the mean of a multivariate normal distribution. Annals Math Stat 1972;43:1481-1490.

47. Greenland S, Poole C. Empirical-Bayes and semi-Bayes approaches to occupational and environmental hazard surveillance. Arch Environ Health 1994;49:9-16.

48. Greenland S. Hierarchical regression for epidemiologic analyses of multiple exposures. Environ Health Perspect 1994;102 Suppl 8:33-9.

49. Greenland S. Principles of multilevel modelling. Int J Epidemiol 2000;29:158-67.

50. Greenland S. Methods for epidemiologic analyses of multiple exposures: a review and comparative study of maximum-likelihood, preliminary-testing, and empirical-Bayes regression. Stat Med 1993;12:717-36.

51. De Roos AJ, Zahm SH, Cantor KP, et al. Integrative assessment of multiple pesticides as risk factors for non-Hodgkin's lymphoma among men. Occup Environ Med 2003;60:E11.

52. Witte JS, Greenland S, Kim LL, Arab L. Multilevel modeling in epidemiology with GLIMMIX. Epidemiology 2000;11:684-8.

53. Abdel-Rahman SZ, El-Zein RA. The 399Gln polymorphism in the DNA repair gene XRCC1 modulates the genotoxic response induced in human lymphocytes by the tobacco-specific nitrosamine NNK. Cancer Lett 2000;159:63-71.

54. Wang Y, Spitz MR, Zhu Y, Dong Q, Shete S, Wu X. From genotype to phenotype: correlating XRCC1 polymorphisms with mutagen sensitivity. DNA Repair (Amst) 2003;2:901-8.

55. Kelsey KT, Park S, Nelson HH, Karagas MR. A population-based case-control study of the XRCC1 Arg399Gln polymorphism and susceptibility to bladder cancer. Cancer Epidemiol Biomarkers Prev 2004;13:1337-41.

56. Patel AV, Calle EE, Pavluck AL, Feigelson HS, Thun MJ, Rodriguez C. A prospective study of XRCC1 (X-ray cross-complementing group 1) polymorphisms and breast cancer risk. Breast Cancer Res 2005;7:R1168-73.

57. Yu HP, Zhang XY, Wang XL, et al. DNA repair gene XRCC1 polymorphisms, smoking, and esophageal cancer risk. Cancer Detect Prev 2004;28:194-9.

58. Pachkowski BF, Winkel S, Kubota Y, Swenberg JA, Millikan RC, Nakamura J. XRCC1 genotype and breast cancer: functional studies and epidemiologic data show interactions between XRCC1 codon 280 His and smoking. Cancer Res 2006;66:2860-8.

59. Yoshihara T, Ishida M, Kinomura A, et al. XRCC3 deficiency results in a defect in recombination and increased endoreduplication in human cells. Embo J 2004;23:670-80.

60. Pritchard JK. Are rare variants responsible for susceptibility to complex diseases? Am J Hum Genet 2001;69:124-37.

61. Greenland S, Robins JM. Empirical-Bayes adjustments for multiple comparisons are sometimes useful. Epidemiology 1991;2:244-51.

62. Witte JS, Greenland S. Simulation study of hierarchical regression. Stat Med 1996;15:1161-70.

B. Manuscript 2: DNA repair genes, cigarette smoking, and intimal medial thickness: The Atherosclerosis Risk in Communities (ARIC) Study

ABSTRACT

Although the association between cigarette smoking and atherosclerosis is well established and consistent across age, sex, racial, and geographic strata, the mechanisms by which smoking initiates vascular disease remain poorly understood. As differences in responses to mutagen exposure have been attributed in part to heritable variation in DNA repair capacity, we examined the degree to which variation in five DNA repair genes (*hOgg1, APEX1, XRCC1*, *XPD* and *XRCC3*) modified the association between ever-smoking and baseline mean intima-medial thickness (IMT) in the Atherosclerosis Risk in Communities (ARIC) cohort. A stratified random sample of 698 Caucasians and 367 African Americans was selected from eligible participants (n=14,255). Analyses were race-stratified and adjusted for age, sex, study center, alcohol intake and physical activity. Baseline mean IMT differences were estimated using hierarchical linear methods that incorporated priors into models including all tagSNPs and models extended to examine gene-by-smoking interaction. When ever-smoking and priors for genetic (within gene) and environmental effects were added to the first-stage model tagSNPs rs31366814 [joint effect$_{XO/XX \text{ vs. } OO}$ = 0.14 (0.03, 0.24)], rs3213282 [joint effect$_{XO \text{ vs. } OO}$ = 0.09 (0, 0.17); joint effect$_{XX \text{ vs. } OO}$ = 0.08 (-0.02, 0.19)], and rs3213245 [joint effect$_{XO \text{ vs. } OO}$ = 0.09 (0.02, 0.17); joint effect$_{XX \text{ vs. } OO}$ = 0.09 (-0.02, 0.20)] were associated with an increase in the estimated effect of ever-smoking on baseline mean IMT. Though imprecise, our results suggest that DNA repair pathway variants might modify the effect of cigarette smoking on subclinical atherosclerosis. Further work examining these pathways is warranted.

1. Introduction

Atherosclerosis is the main pathophysiological process responsible for coronary heart (CHD), cerebrovascular, and peripheral vascular disease(1) and is highly prevalent, as subclinical disease begins early in life (2). Noninvasive imaging techniques provide valid and reproducible means to quantify the burden of subclinical atherosclerosis in vivo, among them B-mode ultrasound of the extracranial carotid arteries (3). Although distal from the epicardial coronary arteries, subclinical atherosclerosis measured in the extracranial carotid beds is associated with prevalent and incident atherothrombotic outcomes, in both the ARIC cohort(4, 5) and in other study populations(6, 7).

The proliferation of smooth muscle cells (SMC) is a fundamental mechanism in the pathophysiology of atherosclerosis(8), although there is disagreement on the exact role SMC proliferation plays. One paradigm, the response to injury or inflammation hypothesis, posits that the joint action of growth factors, proteolytic agents, and extracellular matrix molecules induce SMC migration from the media and their consequent proliferation in the intima(9). Thus, SMC proliferation is only a reactive process.

Another theory, the monoclonal hypothesis, contends that a predisposed SMC population is responsible for the consequent proliferation that typifies atherosclerosis(9). Introduced in 1974(10), the monoclonal hypothesis suggests that plaque initiation of atherosclerotic plaques requires a mutation or viral infection that transforms a SMC into the progenitor of a proliferative clone, analogous to the evolution of neoplastic cells(10). Similarly, increased mutation rates and extensive microsatellite instability have been reported

in human atherosclerotic lesions(11, 12) and SMCs cultured from plaques retain transforming potential throughout many cell generations(13).

Rather than alternatives, the response to injury and monoclonal hypotheses of atherogenesis may be complementary. Initial events leading to plaque formation may reflect the response to injury hypothesis, whereas clone formation and expansion, transforming an inflammatory process into a neoplastic process, requires a longer time. As a corollary, the first stage of atheroma formation may be more readily reversible than the following phases(14). One factor that ties the response to injury and monoclonal hypotheses of atherosclerosis together is exposure to mitotic / proliferative agents, for example compounds found in cigarette smoke. Chronic cigarette smoke exposure has been shown to promote SMC proliferation, a process fundamental to atherogenesis(8) and animal research has demonstrated that the tobacco smoke mutagens benzo(a)pyrene and 1,3-butadiene can induce and stimulate a proliferative vascular SMC phenotype(15, 16).

As differences in responses to mutagen exposure have been attributed in part to heritable variation in DNA repair capacity(17), the identification of DNA repair pathway variants that influence the relationship between cigarette smoke exposure and atherosclerosis could provide new insights into the etiology of this major disease. Using data from the Atherosclerosis Risk in Communities Study (ARIC), we examined the extent to which five DNA repair genes (*hOgg1, APEX1, XRCC1*, *XPD*, and *XRCC3*) from three pathways (base excision repair (BER), nucleotide excision repair (NER), and double-strand break repair (DBS)) modified the relationship between cigarette smoking and mean carotid intimal-medial thickness (IMT), a marker of generalized atherosclerosis.

2. Materials and methods

a. Study population and sources of data

ARIC, the parent population for this study, is a population-based longitudinal study of cardiovascular and pulmonary diseases selected as a probability sample of 15,792 Caucasian and African Americans from four U.S. locales(18). Participants were followed via annual telephone interviews, clinic examinations every three years from 1987-1999, and ongoing hospital and death certificate record abstraction. The study protocol was approved by the Institutional Review Board of each center, and consent was obtained from each participant.

Of the 15,792 ARIC participants, 14,255 met the following eligibility criteria: 1) self-reported race of Caucasian from the NC, MN, or MD field centers or African American from the NC or MS field centers (N= 103 ineligible), and no history of 2) CHD (N=1,102 ineligible), or 3) transient ischemic attack or stroke (N=332 ineligible). A stratified random sample of all eligible ARIC participants (cohort random sample, CRS) was assembled (N = 1,065) at study baseline by stratifying eligible participants into eight mutually exclusive groups based on age (≤ 55 years vs. > 55 years), sex, and race. Sampling fractions varied in an attempt to over-sample higher-risk participants.

Baseline carotid wall thickness was measured by B-mode ultrasound using validated techniques(3), scanning protocols common to each study center, and standardized central readings(19). The far walls of the left and right common carotid, carotid bifurcation, and the internal carotid arteries (six sites total) were measured at designated 1-cm lengths and averaged across as many 1-mm-apart intima-to-media (IMT) distances as were available. As only 13% of study participants had full data for all six sites, missing data were imputed for

participants with at least one IMT measurement using sex- and race-specific models adjusting for age, body mass index, and arterial depth (BMDP 5V). Baseline mean IMT was then defined as the weighted IMT average at the six carotid artery sites after adjustment for measurement drift and reader differences(20). Estimated site-specific reliability coefficients obtained from 36 ARIC participants scanned at three visits 7-14 days apart were 0.77, 0.73, and 0.70 for mean wall thickness at the carotid bifurcation, internal, and common carotid arteries, respectively(21). A natural log transformation was applied to correct for non-normality.

Exposure to cigarette smoke was ascertained at baseline using an interviewer-administered questionnaire. Ever-smoking was defined as greater than 400 cigarettes in a lifetime and used to classify study participants as ever- or never-smokers for these analyses. We did not attempt to capture all genetic variability within the *XRCC1, APEX1, hOGG1, XPD,* and *XRCC3* genes. Instead, selection of the 36 SNPs was informed by functional data, minor allele frequency (MAF, > 0.05), SNP type with preference for non-synonymous SNPs, association studies in the cancer literature, and patterns of pair-wise linkage disequilibrium (LD) reported in the CEPH population by the Hapmap project (http://www.hapmap.org/). Genotyping was performed in two stages. We used matrix-assisted laser desorption/ ionization time-of-flight mass spectrometry(22) for the first stage and the BeadStation system (Illumina, Inc., San Diego, CA) with a custom oligonucleotide pool(23) for the second. Agreement across replicate samples exceeded 0.99 for every valid sample. The data were also compared to five randomly chosen SNPs previously genotyped in the ARIC cohort and the mismatch proportion ranged from 0.005 to 0.02.

b. Statistical methods

All analyses were race-specific. Consistency of SNP genotypes with Hardy-Weinberg equilibrium (HWE) was evaluated among races by chi-square analysis and tagSNPs were identified using a pair-wise $r^2 \geq 0.80$ (24). A general genetic model assuming no mode of inheritance was used when ever-smoking-tagSNP cell frequencies exceeded ten; otherwise an autosomal dominant model was assigned. Missing genotype data were imputed by race using fastPHASE(25). Inferred genotypes were used for analyses if the posterior probability estimate exceeded 0.90.

Confounders of the relationship between cigarette smoking and atherosclerosis were identified from a directed acyclic graph (26) that considered the following variables: age, sex, study center, physical activity, alcohol intake, serum lipoprotein concentrations, body composition, diabetes, and blood pressure. A minimally sufficient adjustment set compromising age, sex, physical activity, field center, and alcohol intake was identified. Physical activity was assessed using three variables that measured leisure, sport, and work-related physical activity (27). Alcohol intake was measured as usual ethanol intake (grams/week). Both variables were collected during the baseline personal interview.

Hierarchical modeling

Genetic analyses typically involve evaluating numerous SNP-disease associations. Standard analytic approaches include: 1) fitting a saturated model containing all variants, 2) reducing the saturated model using a preliminary-testing algorithm, or 3) fitting numerous one-variant-at-a-time models. These approaches are infeasible if parameters are nonestimable, ignore correlation between SNPs, and may provide biased and inefficient

effect estimates(28-31). Furthermore, false positive associations, frequently reflecting point estimates that are inflated and/or unstable(32) are not fully addressed by these approaches.

Instead, hierarchical regression models may be implemented. In hierarchical models, estimates with smaller total error are produced by shrinking unstable estimates towards the mean of the ensemble of variants. The degree of shrinkage is proportional to the precision of each estimate and a prespecified prior variance that represents the range of effects remaining after the first- and second-level effects are estimated(32).

Hierarchical models require two stages. In the first stage, mean IMT is regressed on all variants and covariates (32). tagSNP beta coefficients are then regressed in a second-stage linear model as a function of prior covariates (i.e. tagSNPs s are considered random observations around the second-stage prior covariates) and a pre-specified prior variance. The second stage prior covariates represent categories of exchangeability, which are added to improve the accuracy in parameter estimation and contain variables believed to determine the magnitude of, or explain some variability between, the individual SNP estimates (32). Exchangeability is presumed if a group of tagSNPs are thought to arise from a common distribution with an unknown mean and is a weaker assumption than one presuming all effects are equal (33). We evaluated three categories of exchangeability: all SNPs were exchangeable, SNPs within a given gene were exchangeable, and SNPs within a given pathway were exchangeable.

The prespecified prior variance τ^2 is estimated either from the study data (the empirical-Bayes (EB) approach) or defined by investigators using prior information (the semi-Bayes (SB) approach) (34, 35). τ^2 values for SB analyses are chosen such that at least

95% of the true associations would be captured by the interval $(2(1.96)\sqrt{\tau^2})$. $\tau^2 = 0.00574$ presumes that 95% of all true mean IMT differences lie within a 0.3 range around the prior mean, assuming normality. $\tau^2 = \infty$ yields ordinary maximum-likelihood estimates (36). While our sample size did not permit us to evaluate modification of the estimated mean IMT-ever-smoking relationship by all tagSNPs simultaneously, individual and joint effects of smoking and each variant were assessed individually in models including the main effects of all other variants and confounders (37). Hierarchical models were fit using PROC MIXED (SAS, Cary, NC).

3. Results

Baseline characteristics by race are presented in Table 1. The CRS had a higher proportion of males and older ARIC participants due to the sampling strategy. HWE *P* -values and minor allele frequency (MAF) estimates for 36 DNA repair variants are presented in Table 2. Genetic variation was captured by 20 tagSNPs among the Caucasian stratum and 22 tagSNP among African Americans; only these SNPs will be considered further. The smaller sample size (N = 365, 65% with full tagSNP data) limited the power to detect effects among the African American stratum.

Maximum likelihood (i.e. non-hierarchical) models that included all tagSNPs, an indicator for ever-smoking status, and product term(s) for the tagSNP and ever-smoking status are presented in supplemental Figures 1 (Caucasians) and 2 (African Americans). The estimated main effects of ever-smoking status were relatively precise among Caucasians (95% confidence limit difference (CLD) range = 0.091 – 0.175) and suggested an increase in estimated baseline mean IMT among ever-smokers (range of estimated differences = 0.039 – 0.121, 100% of differences above 0). The estimated tagSNP main effects were considerably imprecise (e.g. the estimated joint effect of rs3213282 and ever-smoking: difference in baseline mean $IMT_{XX\ vs.\ OO}$ = -0.04, 95% CLD = 0.44; the estimated main effect of rs1052133: difference in baseline mean $IMT_{XX\ vs.\ OO}$ = 0.124, 95% CLD = 0.39). While the later estimate suggests considerable effects of tagSNP rs1052133 on differences in baseline mean IMT, the marked imprecision makes the finding unpersuasive. Results in the African American stratum were more variable and difficult to interpret.

The co-occurrence of multiple elevated effect estimates and wide variation in the estimated precision complicated interpretation of the entire panel of results presented in supplemental Figures 1 and 2. Thus, the analyses were extended by examining three categories of exchangeability (all tagSNPs are exchangeable, tagSNPs within each gene are exchangeable, and tagSNPs within each DNA repair pathway are exchangeable) and two prior variance specifications ($\tau^2 = 0.00574$ and 0.0026, corresponding to residual effect ranges of 0.3 and 0.2, respectively). The EB method was evaluated, but this approach appeared to over-shrink estimates, as is typical with this approach when the number of parameters is large relative to the sample size (35, 36, 38).

Main and joint estimated differences in baseline mean IMT obtained from hierarchical models specifying that tagSNPs within each gene were exchangeable and $\tau^2 = 0.00574$ are presented in Figures 1 (Caucasians) and 2 (African Americans). This approach resulted in the estimation of six second-stage fixed effects; five prior means that corresponded to each DNA repair gene (e.g. the estimated main and joint effects of rs1799782 were shrunk towards the estimated *XRCC1* fixed effect) and a sixth representing the estimated effect of ever-smoking (e.g. the main effect of ever-smoking and the joint effect of rs1799782 and ever-smoking were shrunk to the ever-smoking fixed effect). Compared to the estimated differences in baseline mean IMT obtained from the maximum likelihood approach, incorporating a prior mean and variance improved the precision of unstable estimates (e.g. joint effect of the ever-smoking – rs3213245 XX vs. OO contrast , maximum likelihood approach: baseline difference in mean IMT = 0.102, 95% CLD = 0.44; hierarchical approach: baseline difference in mean IMT = 0.09, 95% CLD = 0.22) while producing little change in already stable estimates (e.g. the estimated effect of ever-smoking

for tagSNP rs25489, maximum likelihood approach: baseline difference in mean IMT = 0.09, 95% CLD = 0.19; hierarchical approach: baseline difference in mean IMT = 0.09, 95% CLD = 0.18), a pattern even more apparent within the African American stratum.

As in the maximum likelihood approach, the estimated main effects of ever-smoking status was associated with a greater baseline mean IMT among Caucasians, with estimated differences ranging from 0.047 - 0.12 (100% of estimated differences above 0). The estimated tagSNP main effects among never-smokers were generally null. When evaluating the degree to which DNA repair variants modified the relationship between baseline mean IMT and ever-smoking among Caucasians, tagSNPs rs3213282 (*XRCC1*), rs3213245(*XRCC1*), rs3212024 (*XRCC3*), and rs3136814 (*APEX1*) increased the estimated effect of ever-smoking on differences in baseline mean IMT while tagSNPs rs3136817 (*APEX1*) and rs1799794 (*XRCC3*) decreased the estimated effect of ever-smoking (Figure 1). Although other estimates were elevated, they were difficult to reconcile. A decrease in the estimated effect of ever-smoking was suggested for rs1799793 heterozygotes, but the XX vs. OO stratum was associated with an increase in the estimated effect of ever-smoking. The rs3136814 minor allele also was associated with an increase in the estimated effect of ever-smoking, but the marked imprecision (estimated difference in baseline mean IMT = 0.14, 95% CLD = 0.27) makes interpretation uncertain. The marked imprecision makes the findings unpersuasive. Results among African Americans were highly variable and difficult to interpret.

Many of the estimates suggest consistent effects among heterozygote and homozygote minor allele strata; thus a dominant genetic model that would increase the precision of unstable estimates may be a reasonable alternative. This approach had little effect on stable

estimates (results not shown). However, the unstable estimates demonstrated considerable reduction in CL widths when a dominant model was assigned (e.g. rs3212024, codominant model, joint effect: baseline difference in mean $IMT_{XX \text{ vs. } OO}$ = 0.102 (-0.032, 0.236); dominant model, joint effect: baseline difference in mean $IMT_{XO, XX \text{ vs. } OO}$ = 0.094 (0.017, 0.172)).

Table 28. (MS 2: Table 1) Selected baseline characteristics by race and case status for 14,255 ARIC participants.

		Caucasians		African Americans	
		CRS (N=698) (This study)	All eligible ARIC Participants (N=10,428)	CRS (N=367) (This study)	All eligible ARIC Participants (N=3,827)
Median age at baseline (IQR)		57 (51, 60)	54 (49, 59)	55 (50, 59)	53 (48, 58)
Male (%)		384 (55.0)	4,741 (45.5)	194 (52.9)	1416 (37.0)
Ever smokers (%)		423 (60.7)	6,142 (58.9)	209 (57.1)	2010 (52.5)
Current smokers (%)		154 (22.1)	2,552 (24.5)	113 (30.9)	1120 (29.3)
Former smokers (%)		269 (38.6)	3,590 (34.4)	96 (26.2)	887 (23.2)
Median pack-years of smoking (IQR)		7.5 (0, 28.25)	6.0 (0, 29.0)	3.4 (0, 20.3)	0.5 (0, 17)
Median alcohol intake, grams/week (IQR)		0 (0, 60.4)	0 (0, 52.8)	0 (0, 39.6)	0 (0, 13.2)
Median exercise (IQR)	**Leisure**	2.50 (2.00, 2.75)	2.00 (1.75, 2.25)	2.00 (1.75, 2.25)	2.00 (1.75, 2.50)
	Sport	2.50 (2.00, 3.00)	2.00 (1.75, 2.50)	2.00 (1.75, 2.50)	2.00 (1.75, 2.50)
	Work	2.00 (0, 3.00)	2.50 (1.00, 3.00)	2.50 (1.00, 3.00)	2.50 (1.00, 3.00)

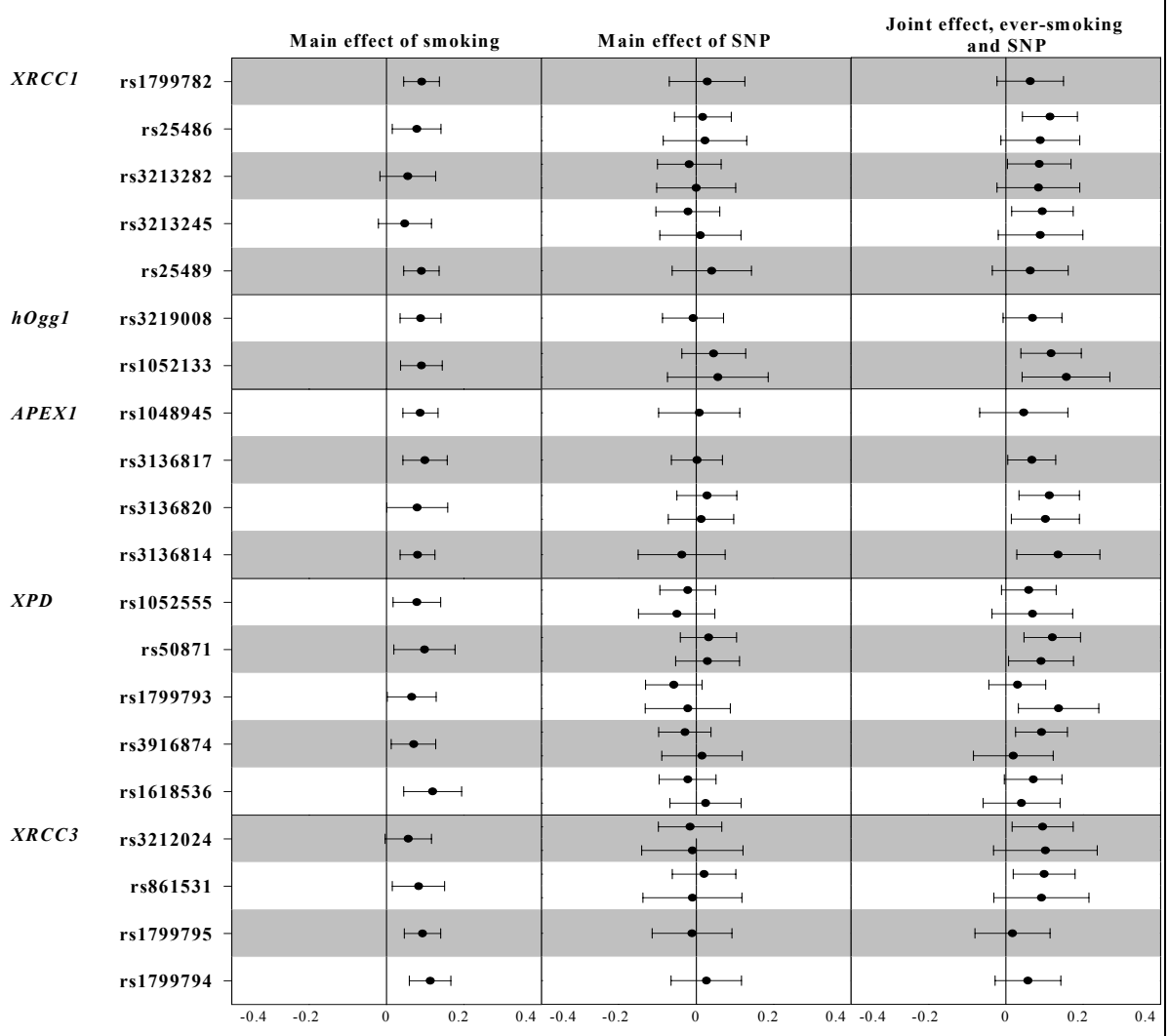
All cells report a count unless otherwise indicated; ARIC, Atherosclerosis Risk in Communities Study; CRS, cohort random sample; IQR, Interquartile range

Table 29. (MS2: Table 2) Hardy-Weinberg Equilibrium *P* - values and minor allele frequency estimates for 36 DNA repair variants in 698 Caucasian and 367 African American ARIC participants.

				Caucasians		African Americans	
Gene	**Cytogenic location**	**Variant**	**SNP function**	**HWE *P*-value**	**MAF**	**HWE *P*-value**	**MAF**
Base Excision Repair (BER)							
XRCC1	19q13.2	rs1799782	Trp194Arg	0.08	0.07†	1.0	0.05†
		rs25489	His280Arg	0.19	0.04†	0.34	0.02¥
		rs25486	Intron	0.31	0.36†	0.54	0.23†
		rs3213282	Intron	0.68	0.44†	0.92	0.43†
		rs3213245	UTR	0.55	0.41†	0.92	0.41†
		rs1475933	Intron	0.93	0.41‡	0.17	0.34†
		rs1799778	Intron	0.98	0.37‡	0.33	0.23‡
		rs25487	Arg339Gln	0.49	0.37‡	0.79	0.15†
		rs915927	Pro206Pro	1.0	0.42‡	0.21	0.40‡
		rs2228487	His107Arg	.	0¥	.	0¥
		rs2307187	UTR	.	0§	.	0¥
		rs2307189	Thr42Thr	.	0¥	.	0¥
		rs25474	Leu514Pro	.	0¥	1.0	0.002¥
		rs25496	Ala72Val	.	0§	.	0¥
hOGG1	3p26.2	rs1052133	Cys326Ser	0.81	0.24†	0.36	0.18†
		rs3219008	Intron	0.11	0.21†	0.11	0.42†
		rs1805373	Gln229Arg	.	0¥	0.90	0.08†
		rs2072668	Intron	0.63	0.24‡	1.0	0.28†
APEX1	14q11.2-q12	rs1048945	His51Gln	1.0	0.03†	1.0	0.009¥
		rs3136820	Glu148Asp	0.12	0.48†	1.0	0.36†
		rs3136817	Intron	0.53	0.23†	0.96	0.15†
		rs3136814	UTR	1.0	0.03†	0.08	0.14†
Nucleotide Excision Repair (NER)							
XPD	19q13.3	rs1052555	Asp711Asp	0.32	0.35†	1.0	0.11‡
		rs1799793	Asn312Asp	0.87	0.36†	0.87	0.12†
		rs1618536	Intron	0.26	0.45†	1.0	0.13†
		rs3916874	Intron	0.38	0.26†	0.0002	0.07†
		rs50871	Intron	0.94	0.46†	0.87	0.09†
		rs1052559	Gln751Lys	0.46	0.40‡	0.08	0.24†
Double-Strand Break Repair (DSB)							
XRCC3	14q32.3	rs861531	Intron	0.57	0.40†	0.21	0.30†
		rs1799795	Intron	0.69	0.01†	1.0	0.02¥
		rs1799794	UTR	0.33	0.19†	0.67	0.20†
		rs3212024	UTR	0.60	0.30†	0.87	0.16†
		rs861539	Thr241Met	0.69	0.39‡	0.69	0.24†
		rs1799796	Intron	0.82	0.31‡	0.65	0.15‡
		rs3212038	UTR	1*10-38	0.38§	.	0§
		rs3212057	UTR	1.0	0.001¥	1.0	0.02¥

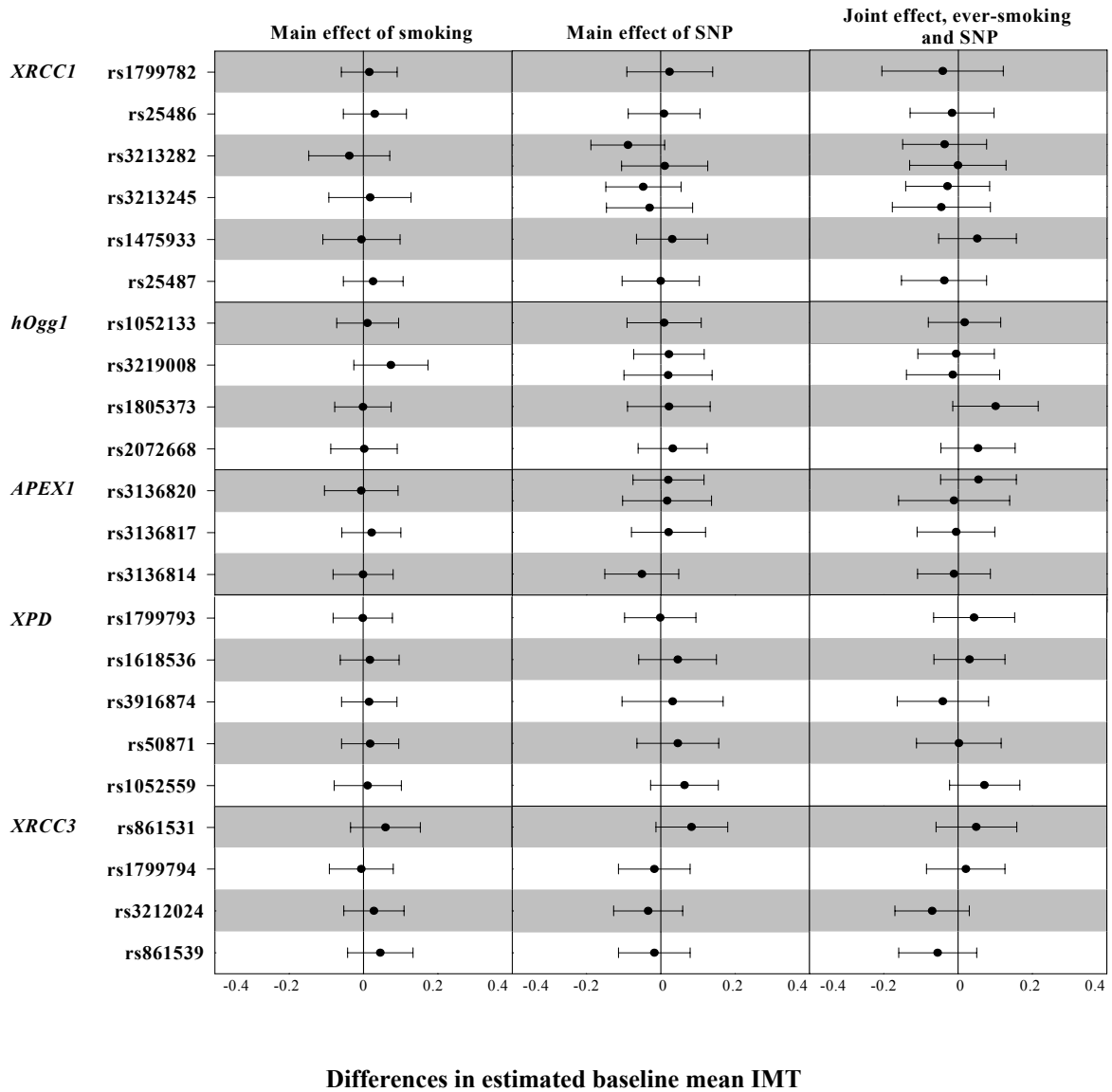
†tagSNP; ‡non-tag SNP, not analyzed further; §Poor quality score, not analyzed; ¥MAF too low, not analyzed; HWE, Hardy-Weinberg Equilibrium, MAF, minor allele frequency; ARIC, Atherosclerosis Risk in Communities Study; *hOGG1*, 8 – hydroxy-2' – deoxyguanosine-glycosylase/apurinic lyase; *APEX1,* apurinic/apyrimidinic endonuclease; *XRCC1,* X-ray repair cross complementing, group 1; *XPD,* xeroderma pigmentosum D; *XRCC3,* X-ray repair complementing defective repair in Chinese hamster cells 3; SNP, single nucleotide polymorphism

Figure 17. (MS2: Figure 1) Main and joint estimated effects of 20 DNA repair tagSNPs and ever-smoking on mean IMT in 470 Caucasian ARIC participants specifying tagSNPs within each gene as exchangeable and a 0.3 residual effect range ($\tau^2 = 0.00574$).



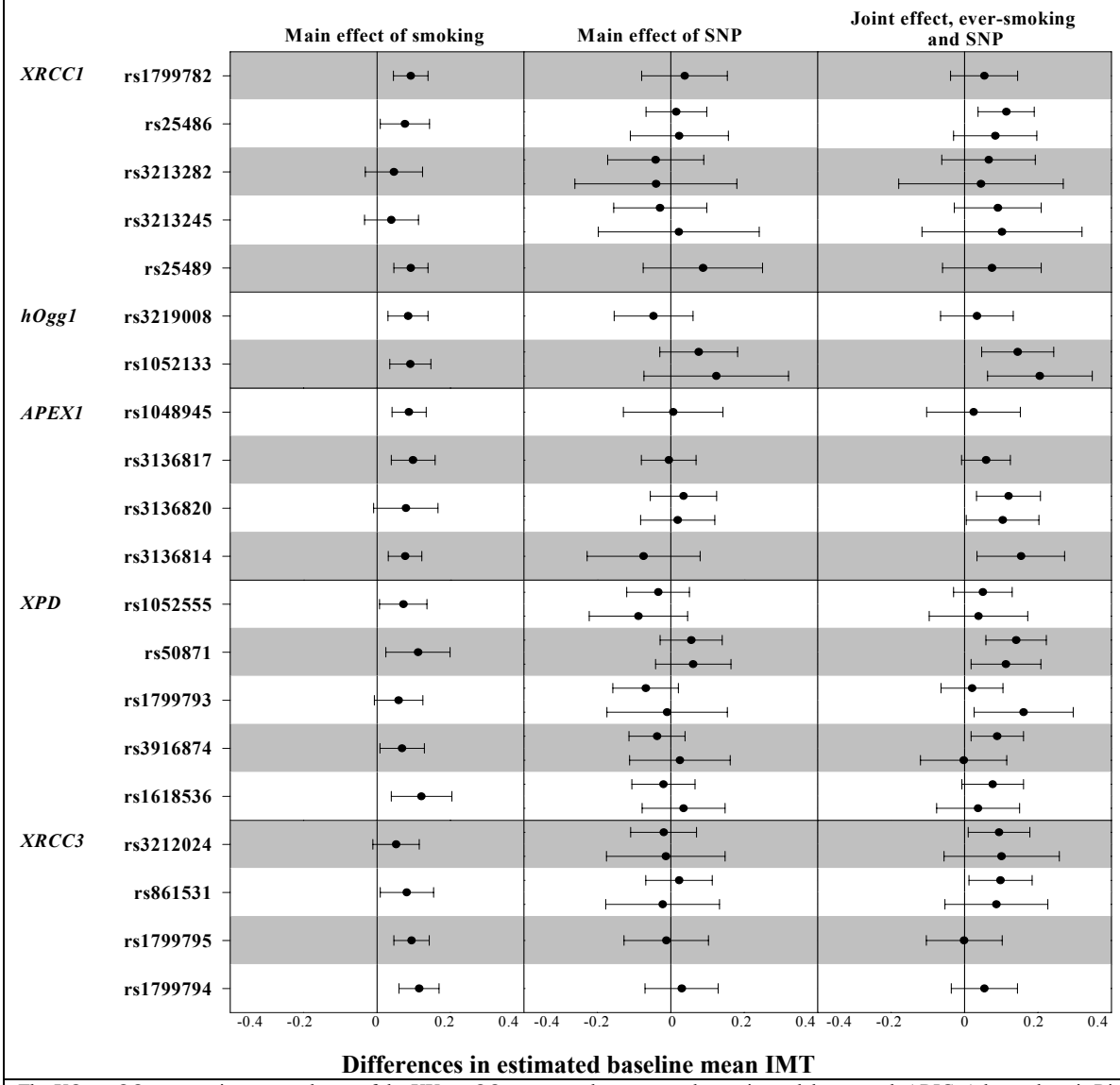
The XO vs. OO contrast is presented atop of the XX vs. OO contrast when a general genetic model was used; ARIC, Atherosclerosis Risk in Communities Study; SNP, single nucleotide polymorphism; IMT, intimal –medial thickness; *hOGG1*, 8 – hydroxy-2' – deoxyguanosine-glycosylase/apurinic lyase; *APEX1,* apurinic/apyrimidinic endonuclease; *XRCC1,* X-ray repair cross complementing, group 1; *XPD,* xeroderma pigmentosum D; *XRCC3,* X-ray repair complementing defective repair in Chinese hamster cells 3

Figure 18. (MS2: Figure 2) Main and joint estimated effects of 22 DNA repair tagSNPs and ever-smoking on mean IMT for 194 African American ARIC participants specifying tagSNPs within each gene as exchangeable and a 0.3 residual effect range ($\tau^2 = 0.00574$).



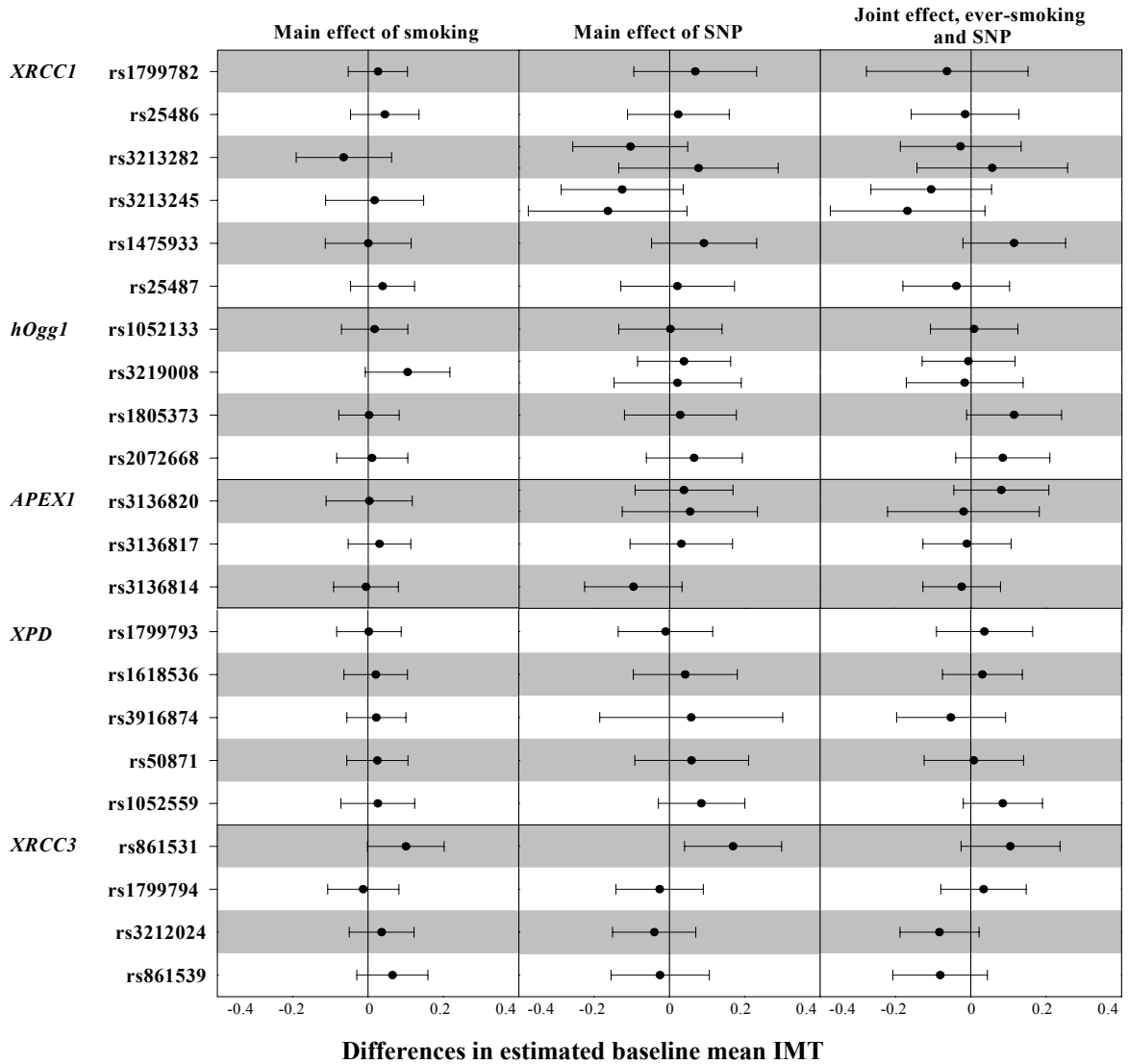
The XO vs. OO contrast is presented atop of the XX vs. OO contrast when a general genetic model was used; ARIC, Atherosclerosis Risk in Communities Study; SNP, single nucleotide polymorphism; IMT, intimal –medial thickness; *hOGG1*, 8 – hydroxy-2' – deoxyguanosine-glycosylase/apurinic lyase; *APEX1,* apurinic/apyrimidinic endonuclease; *XRCC1,* X-ray repair cross complementing, group 1; *XPD,* xeroderma pigmentosum D; *XRCC3,* X-ray repair complementing defective repair in Chinese hamster cells 3

Figure 19. (MS2: Figure S1) Main and joint estimated effects of 20 DNA repair tagSNPs and ever-smoking on mean IMT in 470 Caucasian ARIC participants including a gene-smoking product term for one SNP and the main effects of all others.



The XO vs. OO contrast is presented atop of the XX vs. OO contrast when a general genetic model was used; ARIC, Atherosclerosis Risk in Communities Study; SNP, single nucleotide polymorphism; IMT, intimal –medial thickness; *hOGG1*, 8 – hydroxy-2' – deoxyguanosine-glycosylase/apurinic lyase; *APEX1,* apurinic/apyrimidinic endonuclease; *XRCC1,* X-ray repair cross complementing, group 1; *XPD,* xeroderma pigmentosum D; *XRCC3,* X-ray repair complementing defective repair in Chinese hamster cells 3

Figure 20. (MS2: Figure S2) Main and joint estimated effects of 22 DNA repair tagSNPs and ever-smoking on mean IMT in 194 African American ARIC participants including a gene-smoking product term for one SNP and the main effects of all others.



The XO vs. OO contrast is presented atop of the XX vs. OO contrast when a general genetic model was used; ARIC, Atherosclerosis Risk in Communities Study; SNP, single nucleotide polymorphism; IMT, intimal –medial thickness; *hOGG1*, 8 – hydroxy-2' – deoxyguanosine-glycosylase/apurinic lyase; *APEX1,* apurinic/apyrimidinic endonuclease; *XRCC1,* X-ray repair cross complementing, group 1; *XPD,* xeroderma pigmentosum D; *XRCC3,* X-ray repair complementing defective repair in Chinese hamster cells 3

4. Discussion

Although numerous studies have examined the association between cigarette smoking and subclinical measures of atherosclerosis, studies evaluate this relationship within the context of DNA repair variation are lacking. We confirm the established relationship between cigarette smoking and carotid thickness and show that effect modification by DNA repair genes is a potentially informative hypothesis that warrants further investigation.

This work extends our previous efforts that quantified the association between ever-smoking, DNA repair variation, and incident CHD. Here we focused on subclinical disease measures, as they can provide information not captured by studies of incident events. For example, studies of atherothrombotic events typically focus on factors related to later disease stages and may overlook or underestimate effects of exposures that act earlier(39). If the pattern of somatic DNA damage present in atherosclerotic lesions reflects the mutagenicity of tobacco smoke constituents, one might expect populations who smoke cigarettes and have a reduced DNA repair capacity to have increased carotid thickening. Increased carotid thickening is a marker of the systemic burden of atherosclerosis, a necessary but not sufficient cause of CHD. Indeed, increased IMT is associated with prevalent and incident atherothrombotic outcomes, in both the ARIC cohort(4, 5) and in other study populations(6, 7).

TagSNPs rs3213282 (*XRCC1*), rs1799782 (*XRCC1*), rs3212024 (*XRCC3*), and rs1799794 (*XRCC3*) demonstrated consistent direction of effects for incident CHD (CL Avery, unpublished) and for carotid IMT: tagSNPs rs3213282 and rs3212024 were associated with an increase in the estimated effect of ever-smoking whereas tagSNPs

rs1799782 and rs1799794 were associated with a decrease in the estimated effect of ever-smoking. While there is limited functional data, mutation assays examining rs1799782 (Trp194Arg) suggested that cells with the Trp allele had lower numbers of chromosomal breaks(40), which is consistent with our analysis showing that the Trp allele was associated with a decrease in the estimated effect of ever-smoking on IMT. Population based studies examining the relationship between rs1799782 and cancers and related traits have been contradictory, possibly reflecting the fact that published studies were often underpowered and examined prevalent disease. However, a meta-analysis of 16 published studies of tobacco-related cancers (lung, upper aerodigestive tract, and bladder) estimated a summary odds ratio (OR) (95% confidence interval (CI)) of 0.86 (0.77, 0.95) for the 194Trp contrast and a case-only interaction OR for tobacco smoking, the 194Trp contrast, and tobacco-related cancers of 0.80 (0.56, 1.16) (41), both of which were consistent with our results.

The functions of rs3213282, rs3212024 and rs1799794 have yet to be studied, although rs3213282 and rs1799794 were genotyped by the hapmap project and tag three and two intronic SNPs, respectively. Although the association between smoking status, rs1799794, and incident bladder cancer was examined in 634 Italian males, the marked imprecision precluded comment (42). Consistent with our results, the rs3212024 minor allele was associated with an increased odds of incident follicular lymphoma in 1,035 Scandinavian males and females and increased the effect of cigarette smoking status on the odds of incident follicular lymphoma (43).

Results for tagSNPs rs50871 (*XPD*), rs861531 (*XRCC3*), rs3136814 (*APEX1*), rs3213245 (*XRCC1*), and rs3136817 (*APEX1*) were inconsistent with our incident CHD analyses. For example, rs3213245 was associated with an increase in the estimated effect of

smoking when mean IMT was examined, but suggested a decrease in the estimated effect of ever-smoking when incident CHD was considered, although wide interval estimates were noted. While rs3136814 was associated with an increase in the estimated effect of ever-smoking when mean IMT was examined, results were null when considering incident CHD. These results are difficult to explain, although they do not necessarily represent one or more false positive associations. CHD and IMT are distinct manifestations of a complex disease process (e.g. associations observed for incident CHD might reflect exposures with roles not only in atherosclerosis, but also plaque instability and/or rupture). As variation in DNA repair pathway genes could influence atherogenesis at both its origin and progression, the discrepant results might simply reflect different stages in the natural history of disease.

The generally null estimated main effects in the presence of interaction with ever-smoking status reported for most tagSNPs underscore the necessity of considering genetic effects within the context of biologically plausible environmental exposures. For example, an analysis limited to examining the main effects of DNA repair variants would conclude that none of the variants are associated with subclinical atherosclerosis. However, by incorporating ever-smoking status, an admittedly inexact measure of cigarette smoke exposure, we identified polymorphisms in two genes (*XRCC1* and *XRCC3*) that were associated with variation in baseline mean IMT and incident CHD (CL Avery et al., unpublished).

XRCC1 is a single-strand binding protein and was the first mammalian gene implicated in cellular sensitivity to ionizing radiation(44). While *XRCC1* has no known catalytic activity, it recognizes and binds single-strand DNA breaks(45) and is thought to complex with other BER components during short-patch DNA repair via its role as a

chaperone or central scaffolding protein(46). Animal models of atherosclerosis also associated *XRCC1* upregulation with induced atherosclerotic plaques(47) and Rossi et al., (2004) demonstrated increased *XRCC1* expression in tissue from stable angina plaques(48). *XRCC1* transcription levels were also elevated in diabetic patients when compared to non-diabetic patients, suggesting that metabolites produced under the hyperglycemic state may be mediated by *XRCC1* expression(49).

XRCC3 is involved in the repair of double strand DNA breaks by homologous recombination (HR), the 'non error-prone' DSB repair mechanism(50, 51). Hamster and human cell lines containing *XRCC3* mutations showed a 25-fold decrease in HR, while constitutive wild-type *XRCC3* expression conferred resistance to DNA-damaging agents(52). *XRCC3* mRNA and protein levels were also elevated in malignant prostate cells when compared to normal epithelial cells. Despite the increased *XRCC3* expression, the malignant cells exhibited a defective DNA DSB repair phenotype, suggesting that prostate tumorgenesis may reflect aberrant DNA repair capacity(53). Studies examining the relationship between *XRCC3* variants and atherosclerotic disease are lacking.

Some of our estimates were imprecise, however our analyses are conservative, reflecting our *a priori* preference for general genetic models, as this parameterization does not presume that the heterozygote is intermediate to the two homozygous phenotypes and has been shown to have correct Type I error rates while losing very little power relative to the true genetic model (54). However, many of the heterozygote and homozygote minor allele estimates appeared consistent, and while a dominant model would have little effect on stable estimates, it might be a reasonable alternative for future analyses considering the rarer variants. While we did not account for genotype uncertainty using a weighted analysis, the

posterior probability estimates for the inferred genotypes above the 0.90 criterion consistently exceeded 0.99, thus a weighted analysis would have little effect on the results. Point estimates using a non-imputed data set were comparable (results not shown).

Specifying $\tau^2 = 0.0026$ also increased the precision of our estimates (results not shown), however little is known about the association between DNA repair genes, cigarette smoking, and atherogenesis; thus we chose to present results obtained when $\tau^2 = 0.00574$ was considered. Hierarchical regression simulation studies have demonstrated that coverage proportions of central 95% posterior probability intervals obtained from a correct or over-specified τ^2 approached or exceeded the nominal level, whereas underspecifying τ^2 resulted in subnominal coverage(55). Greenland (1993) also cautioned against underspecifying τ^2 when either the sample size or the number of parameters is not small(35).

We measured cigarette smoke exposure using the ever-smoking metric, although other smoking measures were available including intensity, duration, age at initiation, second hand smoke exposure, and smoking status. Although ever-smoking considers all participants who reported smoking > 400 cigarettes at study baseline as a homogeneous group, 90% of Caucasian and African American participants classified as ever-smokers reported ≥ 10 years of cigarette smoking. While there is sure to be some misclassification of exposure to cigarette smoke, the distribution of smoking duration and intensity indices in ARIC suggest that the majority of participants reporting ever-smoking actually experienced long-term exposure. Practical constraints also limited our analytic options, as power would be reduced considerably if we considered a three-level categorization of cigarette smoke exposure and continuous parameterizations would also be infeasible given our hypothesis of modification by DNA repair variants.

Although the variants we examined were carefully selected, our analysis was limited to 36 polymorphisms (six of which were monomorphic) from five genes. As the BER, NER, and DSB pathways contain over 130 genes, work to further evaluate the role of DNA repair genes in the pathology of smoking-induced atherogenesis is clearly needed. We also used a composite carotid wall thickness measure, although some studies have suggested that diffuse wall thickening resulting from SMC proliferation is best captured by the common carotid IMT metric(56). However, earlier work by ARIC investigators demonstrated that increased IMT at one carotid bed correlates with an increase in IMT at other sites(57). The combined IMT outcome also allowed us to include an additional 39 participants who would not be captured if common carotid IMT was the sole outcome measure.

Because these data are cross-sectional, temporality in the relationship between cigarette smoking and carotid thickness is assumed. However, the relationship between cigarette smoking and carotid thickness has been reported in animal studies and in varied population-based studies (reviewed in (39)). While IMT was also measured during follow-up visits, IMT progression in this and other populations is estimated as 0.01 mm/year in the average(58). Use of repeated measures would improve the precision of the characterization of an individual's IMT, but would not qualitatively add to these analyses.

The use of hierarchical regression methods without the careful consideration of model assumptions can produce estimates that are more biased than those obtained from traditional methods(59). Attempts to improve accuracy could increase bias if the estimated second-state fixed effects are a poor measure of the true mean. However, the three categories of exchangeability we considered provided consistent estimates, suggesting the results are fairly robust to model specifications. While the prior information was somewhat general, a

simplified second stage can outperform maximum likelihood methods(55). Drawbacks of this method include the complete case analysis requirement (i.e. participants have full data on all genetic factors), which may be problematic, especially as researchers assay larger and larger regions of the genome.

As no *a priori* evidence suggestive of an association between specific DNA repair variants, cigarette smoking, and IMT existed, all tagSNPs were examined. We did not account for testing multiple hypotheses by adjusting alpha, as we focused upon describing the magnitude and precision of the estimates, rather than significance testing. However, we exercised awareness of the potential for random error in the interpretation of results.

Cigarette smoking is a major threat to public health and has established atherogenic effects. While imprecise, particularly for African Americans and variants with low MAFs, our results suggest that additional work examining these pathways is warranted. In addition to further characterizing *XRCC1, hOGG1, APEX1, XPD,* and *XRCC3* genetic variation, examining other promising DNA repair pathway candidate genes and extending the study sample to improve statistical power and increase the flexibility for measuring cigarette smoke exposure would allow us to more thoroughly evaluate the role of DNA repair variation in the context of cigarette smoking and atherogenesis. Future studies in varied populations will undoubtedly be required to validate our conclusions. Yet, our results, based on a comprehensive analysis of the role of DNA repair genes in the context of cigarette smoke exposure and subclincal atherosclerosis measures, highlight the importance of incorporating gene-environment interaction when investigating the etiology of complex diseases such as atherosclerosis.

Acknowledgements

This publication was made possible in part by grant HL074377 awarded to Dr. North, training grant HL007055, American Heart Association predoctoral award 0615310U (Ms. Avery), and by a grant from the National Institute of Environmental Health Sciences (P30ES10126, Dr. Olshan). Dr. Poole received partial support from National Institute of Environmental Health Sciences grant PE0ES10126. The ARIC study is carried out as a collaborative effort supported by NHLBI contracts N01-HC-55015, N01-HC-55016, N01-HC-55018, N01-HC-55019, N01-HC-55020, N01-HC-55021, and N01-HC-55022. The authors thank the staff and participants of the ARIC study for their important contributions. These contents are solely the responsibility of the authors and do not necessarily represent the official views of the funding agencies.

5. References

1. Ross R. The pathogenesis of atherosclerosis--an update. N Engl J Med 1986;314:488-500.

2. Stary HC. Evolution and progression of atherosclerotic lesions in coronary arteries of children and young adults. Arteriosclerosis 1989;9:I19-32.

3. Pignoli P, Tremoli E, Poli A, Oreste P, Paoletti R. Intimal plus medial thickness of the arterial wall: a direct measurement with ultrasound imaging. Circulation 1986;74:1399-406.

4. Chambless LE, Heiss G, Folsom AR, et al. Association of coronary heart disease incidence with carotid arterial wall thickness and major risk factors: the Atherosclerosis Risk in Communities (ARIC) Study, 1987-1993. Am J Epidemiol 1997;146:483-94.

5. Chambless LE, Folsom AR, Clegg LX, et al. Carotid wall thickness is predictive of incident clinical stroke: the Atherosclerosis Risk in Communities (ARIC) study. Am J Epidemiol 2000;151:478-87.

6. Salonen JT, Salonen R. Ultrasonographically assessed carotid morphology and the risk of coronary heart disease. Arterioscler Thromb 1991;11:1245-9.

7. Bots ML, Hoes AW, Koudstaal PJ, Hofman A, Grobbee DE. Common carotid intima-media thickness and risk of stroke and myocardial infarction: the Rotterdam Study. Circulation 1997;96:1432-7.

8. Lavezzi AM, Ottaviani G, Matturri L. Biology of the smooth muscle cells in human atherosclerosis. Apmis 2005;113:112-21.

9. Ross R. Atherosclerosis--an inflammatory disease. N Engl J Med 1999;340:115-26.

10. Benditt EP. Evidence for a monoclonal origin of human atherosclerotic plaques and some implications. Circulation 1974;50:650-2.

11. McCaffrey TA, Du B, Consigli S, et al. Genomic instability in the type II TGF-beta1 receptor gene in atherosclerotic and restenotic vascular cells. J Clin Invest 1997;100:2182-8.

12. Hatzistamou J, Kiaris H, Ergazaki M, Spandidos DA. Loss of heterozygosity and microsatellite instability in human atherosclerotic plaques. Biochem Biophys Res Commun 1996;225:186-90.

13. Parkes JL, Cardell RR, Hubbard FC, Jr., Hubbard D, Meltzer A, Penn A. Cultured human atherosclerotic plaque smooth muscle cells retain transforming potential and

display enhanced expression of the myc protooncogene. Am J Pathol 1991;138:765-75.

14. Andreassi MG, Botto N, Colombo MG, Biagini A, Clerico A. Genetic instability and atherosclerosis: can somatic mutations account for the development of cardiovascular diseases? Environ Mol Mutagen 2000;35:265-9.

15. Zhang Y, Ramos KS. The induction of proliferative vascular smooth muscle cell phenotypes by benzo[a]pyrene does not involve mutational activation of ras genes. Mutat Res 1997;373:285-92.

16. Penn A, Snyder CA. 1,3 Butadiene, a vapor phase component of environmental tobacco smoke, accelerates arteriosclerotic plaque development. Circulation 1996;93:552-7.

17. Anonymous. Cancer. In: Stratton K, Shetty P, Wallace R, Bondurant S, eds. Clearing the Smoke. Assessing the Science Base for Tobacco Harm Reduction. Washington, D.C.: National Academy Press, 2001:367-469.

18. The Atherosclerosis Risk in Communities (ARIC) Study: design and objectives. The ARIC investigators. Am J Epidemiol 1989;129:687-702.

19. Group TAS. High-resolution B-mode ultrasound scanning methods in the Atherosclerosis Risk in Communities Study (ARIC). The ARIC Study Group. J Neuroimaging 1991;1:68-73.

20. Tilling K, Smith GD, Chambless L, et al. The relation between birth weight and intima-media thickness in middle-aged adults. Epidemiology 2004;15:557-64.

21. Chambless LE, Zhong MM, Arnett D, Folsom AR, Riley WA, Heiss G. Variability in B-mode ultrasound measurements in the atherosclerosis risk in communities (ARIC) study. Ultrasound Med Biol 1996;22:545-54.

22. Bray MS, Boerwinkle E, Doris PA. High-throughput multiplex SNP genotyping with MALDI-TOF mass spectrometry: practice, problems and promise. Hum Mutat 2001;17:296-304.

23. Shen R, Fan JB, Campbell D, et al. High-throughput SNP genotyping on universal bead arrays. Mutat Res 2005;573:70-82.

24. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics 2005;21:263-5.

25. Scheet P, Stephens M. A fast and flexible statistical model for large-scale population genotype data: applications to inferring missing genotypes and haplotypic phase. Am J Hum Genet 2006;78:629-44.

26. Greenland S, Brumback B. An overview of relations among causal modelling methods. Int J Epidemiol 2002;31:1030-7.

27. Baecke JA, Burema J, Frijters JE. A short questionnaire for the measurement of habitual physical activity in epidemiological studies. Am J Clin Nutr 1982;36:936-42.

28. Witte JS. Genetic analysis with hierarchical models. Genet Epidemiol 1997;14:1137-42.

29. Greenland S. Comment: cautions in the use of preliminary test estimators. Stat Med 1989;8:669-673.

30. Greenland S. Modeling and variable selection in epidemiologic analysis. Am J Public Health 1989;79:340-9.

31. Sclove SL, Morris CN, Radhakrishna R. Non-optimality of preliminary-test estimators for the mean of a multivariate normal distribution. Annals Math Stat 1972;43:1481-1490.

32. Greenland S. Hierarchical regression for epidemiologic analyses of multiple exposures. Environ Health Perspect 1994;102 Suppl 8:33-9.

33. Greenland S. Principles of multilevel modelling. Int J Epidemiol 2000;29:158-67.

34. Greenland S, Poole C. Empirical-Bayes and semi-Bayes approaches to occupational and environmental hazard surveillance. Arch Environ Health 1994;49:9-16.

35. Greenland S. Methods for epidemiologic analyses of multiple exposures: a review and comparative study of maximum-likelihood, preliminary-testing, and empirical-Bayes regression. Stat Med 1993;12:717-36.

36. Greenland S. A semi-Bayes approach to the analysis of correlated multiple associations, with an application to an occupational cancer-mortality study. Stat Med 1992;11:219-30.

37. De Roos AJ, Zahm SH, Cantor KP, et al. Integrative assessment of multiple pesticides as risk factors for non-Hodgkin's lymphoma among men. Occup Environ Med 2003;60:E11.

38. Witte JS, Greenland S, Haile RW, Bird CL. Hierarchical regression analysis applied to a study of multiple dietary exposures and breast cancer. Epidemiology 1994;5:612-21.

39. Anonymous. Cardiovascular Diseases. The health consequences of smoking: a report of the Surgeon General. Atlanta, GA: Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health, 2004:361-419.

40. Wang Y, Spitz MR, Zhu Y, Dong Q, Shete S, Wu X. From genotype to phenotype: correlating XRCC1 polymorphisms with mutagen sensitivity. DNA Repair (Amst) 2003;2:901-8.

41. Hung RJ, Hall J, Brennan P, Boffetta P. Genetic polymorphisms in the base excision repair pathway and cancer risk: a HuGE review. Am J Epidemiol 2005;162:925-42.

42. Matullo G, Guarrera S, Sacerdote C, et al. Polymorphisms/haplotypes in DNA repair genes and smoking: a bladder cancer case-control study. Cancer Epidemiol Biomarkers Prev 2005;14:2569-78.

43. Smedby KE, Lindgren CM, Hjalgrim H, et al. Variation in DNA repair genes ERCC2, XRCC1, and XRCC3 and risk of follicular lymphoma. Cancer Epidemiol Biomarkers Prev 2006;15:258-65.

44. Thompson LH, Brookman KW, Jones NJ, Allen SA, Carrano AV. Molecular cloning of the human XRCC1 gene, which corrects defective DNA strand break repair and sister chromatid exchange. Mol Cell Biol 1990;10:6160-71.

45. Marintchev A, Mullen MA, Maciejewski MW, Pan B, Gryk MR, Mullen GP. Solution structure of the single-strand break repair protein XRCC1 N-terminal domain. Nat Struct Biol 1999;6:884-93.

46. Nash RA, Caldecott KW, Barnes DE, Lindahl T. XRCC1 protein interacts with one of two distinct forms of DNA ligase III. Biochemistry 1997;36:5207-11.

47. Martinet W, Knaapen MW, De Meyer GR, Herman AG, Kockx MM. Oxidative DNA damage and repair in experimental atherosclerosis are reversed by dietary lipid lowering. Circ Res 2001;88:733-9.

48. Rossi ML, Marziliano N, Merlini PA, et al. Different quantitative apoptotic traits in coronary atherosclerotic plaques from patients with stable angina pectoris and acute coronary syndromes. Circulation 2004;110:1767-73.

49. Kapalla M, Yeghiazaryan K, Hricova M, et al. Combined analysis of biochemical parameters in serum and differential gene expression in circulating leukocytes may serve as an ex vivo monitoring system to estimate risk factors for complications in Diabetes mellitus. Amino Acids 2005;28:221-7.

50. Liu N, Lamerdin JE, Tebbs RS, et al. XRCC2 and XRCC3, new human Rad51-family members, promote chromosome stability and protect against DNA cross-links and other damages. Mol Cell 1998;1:783-93.

51. Johnson RD, Jasin M. Double-strand-break-induced homologous recombination in mammalian cells. Biochem Soc Trans 2001;29:196-201.

52. Pierce AJ, Johnson RD, Thompson LH, Jasin M. XRCC3 promotes homology-directed repair of DNA damage in mammalian cells. Genes Dev 1999;13:2633-8.

53. Fan R, Kumaravel TS, Jalali F, Marrano P, Squire JA, Bristow RG. Defective DNA strand break repair after DNA damage in prostate cancer cells: implications for genetic instability and prostate cancer progression. Cancer Res 2004;64:8526-33.

54. Lettre G, Lange C, Hirschhorn JN. Genetic model testing and statistical power in population-based association studies of quantitative traits. Genet Epidemiol 2007;31:358-62.

55. Witte JS, Greenland S. Simulation study of hierarchical regression. Stat Med 1996;15:1161-70.

56. O'Leary DH, Polak JF, Kronmal RA, et al. Thickening of the carotid wall. A marker for atherosclerosis in the elderly? Cardiovascular Health Study Collaborative Research Group. Stroke 1996;27:224-31.

57. Howard G, Burke GL, Evans GW, et al. Relations of intimal-medial thickness among sites within the carotid artery as evaluated by B-mode ultrasound. ARIC Investigators. Atherosclerosis Risk in Communities. Stroke 1994;25:1581-7.

58. Chambless LE, Folsom AR, Davis V, et al. Risk factors for progression of common carotid atherosclerosis: the Atherosclerosis Risk in Communities Study, 1987-1998. Am J Epidemiol 2002;155:38-47.

59. Greenland S, Robins JM. Empirical-Bayes adjustments for multiple comparisons are sometimes useful. Epidemiology 1991;2:244-51.

CHAPTER VI

CONCLUCIONS

A. Recapitulation of overall study aims, results, and degree to which the goals of the doctoral research have been met

1. Overall study aims

The goal of this project was to evaluate how DNA repair pathway variants modify the relationship between cigarette smoking and two CVD measures: incident CHD and subclinical atherosclerosis, quantified using IMT measures of the carotid arteries. Manuscript 1 addresses Aims 1 - 3, while and Manuscript 2 addresses Aims 4 - 6.

AIM 1: To estimate the association between polymorphisms of the DNA repair genes *XRCC1, XRCC3, APEX1, hOgg1,* and *XPD* and incident CHD.

Research question: Are DNA repair pathway variants associated with incident CHD?

AIM 2: Do polymorphisms of the DNA repair genes *XRCC1, XRCC3, APEX1, hOgg1,* and *XPD* modify the association between cigarette smoking and incident CHD.

Research question: To what extent do polymorphisms of the DNA repair genes *XRCC1, XRCC3, APEX1, hOgg1,* and *XPD* modify the association between cigarette smoking and incident CHD?

AIM 3: To incorporate information from multiple genes and cigarette smoke exposure as higher level priors into analyses investigating the relationship between DNA repair variants, cigarette smoking, and incident CHD.

Research question: How does the incorporation of prior probabilities influence the extent to

which polymorphisms of the DNA repair genes *XRCC1, XRCC3, APEX1, hOgg1,* and *XPD* modify the association between cigarette smoking and incident CHD?

AIM 4: To estimate the association between polymorphisms of the DNA repair genes *XRCC1, XRCC3, APEX1, hOgg1,* and *XPD* and subclinical atherosclerosis (quantified as baseline mean IMT).

Research question: Are DNA repair pathway variants associated with subclinical atherosclerosis (quantified as baseline mean IMT)?

AIM 5: To estimate the extent to which polymorphisms of the DNA repair genes *XRCC1, XRCC3, APEX1, hOgg1,* and *XPD* modify the association between cigarette smoking and subclinical atherosclerosis (quantified as baseline mean IMT).

Research question: To what extent do polymorphisms of the DNA repair genes *XRCC1, XRCC3, APEX1, hOgg1,* and *XPD* modify the association between cigarette smoking and subclinical atherosclerosis (quantified as baseline mean IMT)?

AIM 6: To incorporate information from multiple genes and cigarette smoke exposure as higher level priors into analyses investigating the relationship between DNA repair variants, cigarette smoking, and subclinical atherosclerosis (quantified as baseline mean IMT).

Research question: How does the incorporation of prior probabilities influence the extent to which polymorphisms of the DNA repair genes *XRCC1, XRCC3, APEX1, hOgg1,* and *XPD* modify the association between cigarette smoking and subclinical atherosclerosis (quantified as baseline mean IMT)?

2. Results

Results from Manuscript 1 suggested that tagSNPs rs3213282 (*XRCC1*), rs50871 (*XPD*), and rs3212024 (*XRCC3*) were associated with increases in the estimated effect of ever-smoking on incident CHD while tagSNPs rs1799782 (*XRCC1*) and rs861531 (*XRCC3*) were associated with decreases. With regards to Manuscript 2, tagSNPs rs3213282 (*XRCC1*), rs3213245(*XRCC1*), rs3212024 (*XRCC3*), and rs3136814 (*APEX1*) increased the estimated effect of ever-smoking on differences in baseline mean IMT while tagSNPs rs3136817 (*APEX1*) and rs1799794 (*XRCC3*) decreased the estimated effect of ever-smoking. Although we investigated the same study questions among African Americans, the small sample sizes resulted in highly variable estimates that precluded comment.

Evaluating two related, yet distinct phenotypes allowed us to consider different stages in the natural history of atherosclerosis. We extended our study of incident events (Manuscript 1) by evaluating subclinical disease measures, as they can provide information not captured by studies of incident events. For example, studies of atherothrombotic events typically focus on factors related to later disease stages and may overlook or underestimate effects of exposures that act earlier[75]. If the pattern of somatic DNA damage present in atherosclerotic lesions reflects the mutagenicity of tobacco smoke constituents, one might expect populations who smoke cigarettes and have a reduced DNA repair capacity to have increased carotid thickening. Increased carotid thickening is a marker of the systemic burden of atherosclerosis, a necessary but not sufficient cause of CHD. Indeed, increased IMT is associated with prevalent and incident atherothrombotic outcomes, in both the ARIC cohort[469, 470] and in other study populations[471, 472].

A focus on two different stages in the natural history of atherosclerosis may explain why results for tagSNPs rs50871 (*XPD*), rs861531 (*XRCC3*), rs3136814 (*APEX1*), rs3213245 (*XRCC1*), and rs3136817 (*APEX1*) were inconsistent. For example, rs3213245 was associated with an increase in the estimated effect of smoking when mean IMT was examined, but suggested a decrease in the estimated effect of ever-smoking when incident CHD was considered, although wide confidence interval estimates were noted. While rs3136814 was associated with an increase in the estimated effect of ever-smoking when mean IMT was examined, results were null when considering incident CHD. These results are difficult to explain, although they do not necessarily represent one or more false positive associations. CHD and IMT are distinct manifestations of a complex disease process (e.g. associations observed for incident CHD might reflect exposures with roles not only in atherosclerosis, but also plaque instability and/or rupture). As variation in DNA repair pathway genes could influence atherogenesis at both its origin and progression, the discrepant results might simply reflect different stages in the natural history of disease.

We also highlight the advantage of considering gene-by-environment interactions when evaluating complex chronic diseases like atherosclerosis. tagSNP main effects were generally null for both outcomes; thus an analysis limited to examining the main effects of DNA repair variants would conclude that none of the variants are associated with subclinical atherosclerosis or incident CHD. However, by incorporating ever-smoking status, an admittedly inexact measure of cigarette smoke exposure, we identified polymorphisms in two genes (*XRCC1* and *XRCC3*) that were associated with variation in baseline mean IMT and incident CHD.

XRCC1 is a single-strand binding protein and was the first mammalian gene implicated in cellular sensitivity to ionizing radiation[289]. While *XRCC1* has no known catalytic activity, it recognizes and binds single-strand DNA breaks[290] and is thought to complex with other BER components during short-patch DNA repair via its role as a chaperone or central scaffolding protein[292]. Animal models of atherosclerosis also associated *XRCC1* upregulation with induced atherosclerotic plaques[211] and Rossi et al., (2004) demonstrated increased *XRCC1* expression in tissue from stable angina plaques[301]. *XRCC1* transcription levels were also elevated in diabetic patients when compared to non-diabetic patients, suggesting that metabolites produced under the hyperglycemic state may be mediated by *XRCC1* expression[302].

XRCC3 is involved in the repair of double strand DNA breaks by homologous recombination (HR), the 'non error-prone' DSB repair mechanism[419, 420]. Hamster and human cell lines containing *XRCC3* mutations showed a 25-fold decrease in HR, while constitutive wild-type *XRCC3* expression conferred resistance to DNA-damaging agents[422]. *XRCC3* mRNA and protein levels were also elevated in malignant prostate cells when compared to normal epithelial cells. Despite the increased *XRCC3* expression, the malignant cells exhibited a defective DNA DSB repair phenotype, suggesting that prostate tumorgenesis may reflect aberrant DNA repair capacity[300]. Studies examining the relationship between *XRCC3* variants and atherosclerotic disease are lacking.

3. Meeting the goals of doctoral research

A dissertation must be of appropriate scope and considerable rigor to fulfill the goals of doctoral research. I appreciate that it is the committee's responsibility to judge whether I have met said goals. I have been involved in this work since 2003, when I assisted with SNP selection. My role has expanded to the point that I, under Kari North's guidance, am the lead investigator on the design, analysis, consultation, and writing for the two manuscripts presented above. My work has benefited greatly from verbal and written input provided by the Chair and committee members, as well as through consultations with other co-authors. At the dissertation interim committee meeting, all members reached consensus that the scope of the research was appropriate.

I believe that the proposal defense and the preparation, submission for publication, and defense of this dissertation adequately address the four specific goals described in the Epidemiology Academic Policies Manual: originality, depth, scholarship, and writing skills. Originality is acheived through the application of hierarchical regression methods that are described in the genetic epidemiology literature, but are rarely applied. With regards to depth, I have investigated two outcomes related to atherosclerosis: incident CHD and IMT. Thirty-six DNA repair pathway variants and numerous cigarette smoke indices were also considered. I believe that the requirements of scholarship and writing skills are addressed by the careful consideration and thoroughness reflected in this dissertation. Although I received substantive and editorial comments from all committee members, I formulated the initial manuscript structures including the best way to present study results, and how to frame the introduction, materials and methods, results, and discussion sections. Thus, I believe that this dissertation demonstrates my ability to integrate the analytic, organizational, methodological,

and theoretical concepts inherent in the University of North Carolina's Epidemiology curriculum into my current and future research endeavors.

B. Strengths

Results from this study can provide important contributions to public health research. This research could inform investigators on potential mechanism linking cigarette smoking and atherosclerosis (both subclinical disease and atherogenic endpoints), as few studies have evaluated the role of DNA repair genes with regards to incident CHD or subclinical atherosclerosis. This work also underscores the necessity of considering genetic effects within the context of biologically plausible environmental exposures. For example, an analysis limited to examining the main effects of DNA repair variants would conclude that none of the variants are associated with subclinical atherosclerosis. However, by incorporating ever-smoking status, we identified polymorphisms in two genes (*XRCC1* and *XRCC3*) that were associated with variation in baseline mean IMT and incident CHD.

A priori implementation of prior probability distributions through hierarchical analysis adjusted implausible estimates and enhanced precision, thus facilitating the interpretation of the entire panel of results. This represents a likely improvement upon traditional analytic methods. The candidate genes were also carefully selected on the basis of their role in DNA repair pathways and results from functional studies, when available.

C. Limitations

While the study sample is sufficient for the estimation of the main effects of DNA repair variants among the Caucasian and African American strata, power to assess modification for genes with low MAF, especially within the African American stratum was

limited. However, the study is adequately powered to address the main aims, and is an important contribution to the understanding of this major disease.

Although the polymorphisms we examined were carefully selected, our analysis was limited to 36 variants from five genes (six of which were monomorphic). As the pathways we examined contain over 130 genes[196], additional work to evaluate the role of DNA repair pathway variants is needed. In addition, while the indirect candidate association approach we used is a powerful method, it assumes little allelic heterogeneity within loci and the common disease/common variant paradigm. This strategy would be unsuccessful if the genetic component of atherothrombosis involves numerous rare variants at many loci[465]. Because the IMT data are cross-sectional, temporality in the relationship between cigarette smoking and carotid thickness was assumed. However, the relationship between cigarette smoking and carotid thickness has been reported in animal studies and in varied population-based studies (reviewed in 466).

The use of hierarchical regression methods without the careful consideration of model assumptions can produce estimates that are more biased than those obtained from traditional methods[467]. Attempts to improve accuracy could increase bias if the estimated second-state fixed effects are a poor measure of the true mean. However, the three categories of exchangeability we considered provided consistent estimates for both outcomes, suggesting the results are fairly robust to model specifications. While the prior information was somewhat general, a simplified second stage can outperform maximum likelihood methods[468]. Drawbacks of this method include the complete case analysis requirement (i.e. participants have full data on all genetic factors), which may be problematic, especially as researchers assay larger and larger regions of the genome.

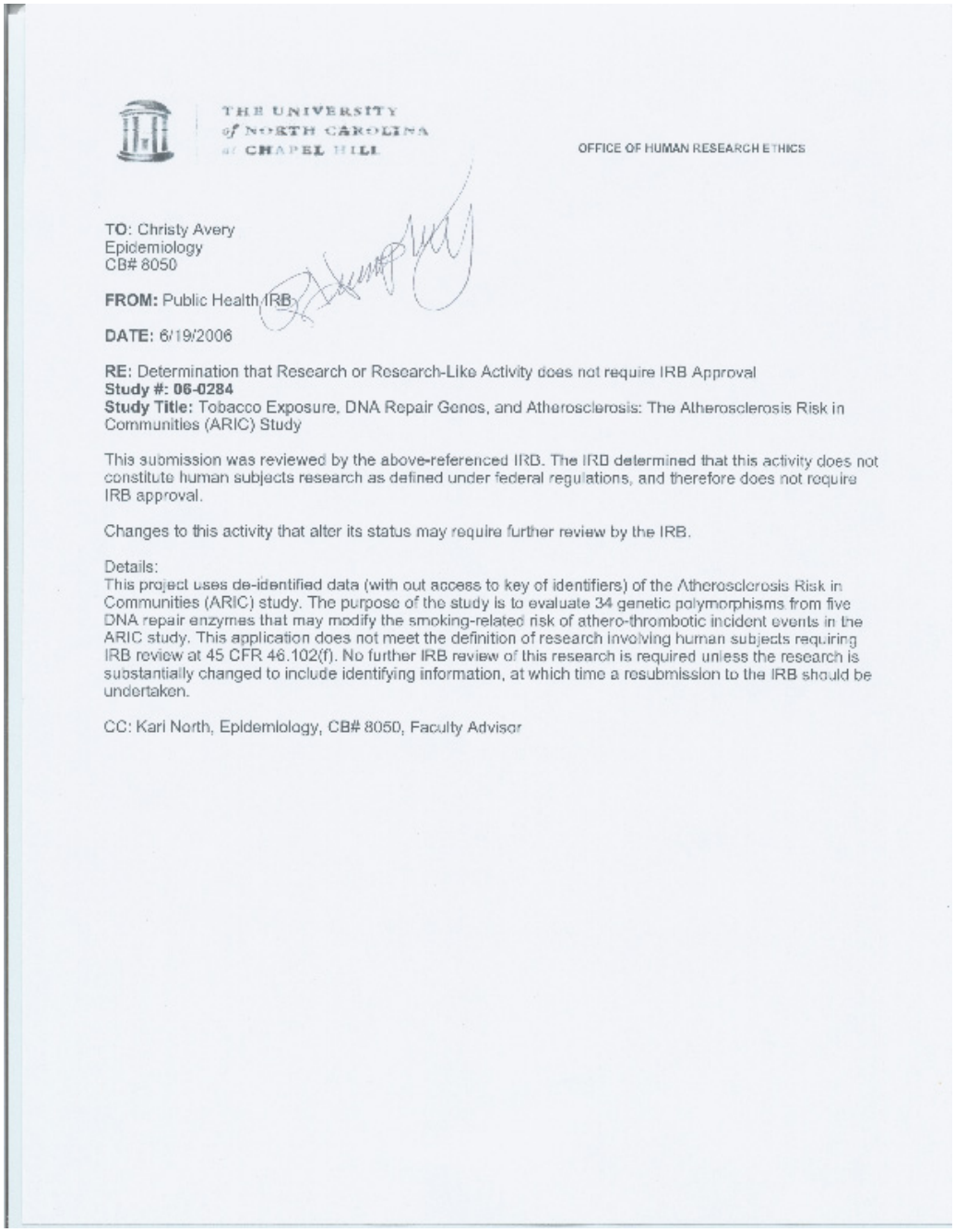
APPENDICES

A. IRB certification

B. Supplemental results, Manuscript 1

C. Supplemental results, Manuscript 2

A. IRB certification



THE UNIVERSITY of NORTH CAROLINA at CHAPEL HILL

OFFICE OF HUMAN RESEARCH ETHICS

TO: Christy Avery
Epidemiology
CB# 8050

FROM: Public Health IRB

DATE: 6/19/2006

RE: Determination that Research or Research-Like Activity does not require IRB Approval
Study #: 06-0284
Study Title: Tobacco Exposure, DNA Repair Genes, and Atherosclerosis: The Atherosclerosis Risk in Communities (ARIC) Study

This submission was reviewed by the above-referenced IRB. The IRB determined that this activity does not constitute human subjects research as defined under federal regulations, and therefore does not require IRB approval.

Changes to this activity that alter its status may require further review by the IRB.

Details:
This project uses de-identified data (with out access to key of identifiers) of the Atherosclerosis Risk in Communities (ARIC) study. The purpose of the study is to evaluate 34 genetic polymorphisms from five DNA repair enzymes that may modify the smoking-related risk of athero-thrombotic incident events in the ARIC study. This application does not meet the definition of research involving human subjects requiring IRB review at 45 CFR 46.102(f). No further IRB review of this research is required unless the research is substantially changed to include identifying information, at which time a resubmission to the IRB should be undertaken.

CC: Kari North, Epidemiology, CB# 8050, Faculty Advisor

I have also completed CITI training and obtained ARIC study approval for this work.

B. Supplemental results, Manuscript 1

Table 30. (MS 1 supplemental results) Number of imputed tagSNP genotype data points by incident CHD status in 1,529 Caucasian ARIC participants.

		Incident CHD (N = 831)				CRS (N = 698)			
Gene	**Variant**	**00**	**X0**	**XX**	**% imputed**	**00**	**X0**	**XX**	**% imputed**
XRCC1	rs1799782	21	0	0	2.7	17	0	0	2.6
	rs25486	1	11	0	1.5	1	6	0	1.0
	rs3213282	8	25	6	5.0	13	25	11	7.4
	rs3213245	0	6	0	0.8	2	1	0	0.4
	rs25489	52	2	0	6.6	32	1	0	4.7
hOGG1	rs3219008	33	0	0	4.3	21	0	0	3.2
	rs1052133	19	4	2	3.2	43	36	5	12.4
APEX1	rs1048945	56	0	0	6.8	46	0	0	6.6
	rs3136817	5	0	0	0.7	3	0	0	0.5
	rs3136820	0	0	0	0	5	0	0	0.9
	rs3136814	49	0	0	6.0	33	0	0	4.7
XPD	rs1052555	8	9	2	2.4	7	3	3	1.9
	rs50871	0	0	0	0	1	0	0	0.2
	rs1799793	1	3	1	0.6	0	0	1	0.2
	rs3916874	1	0	0	0.1	0	0	0	0
	rs1618536	2	0	1	0.4	1	1	0	0.3
XRCC3	rs3212024	4	1	0	0.6	1	4	0	0.7
	rs861531	21	26	11	7.4	21	10	2	5.0
	rs1799795	14	0	0	1.8	13	0	0	2.0
	rs1799794	3	0	0	0.4	8	0	0	1.2

ARIC, Atherosclerosis Risk in Communities Study; CHD, coronary heart disease; CRS, cohort random sample

Table 31. (MS 1 supplemental results) tagSNP genotype frequencies by ever-smoking status in 1,529 Caucasian ARIC participants.

		Ever-smokers (N = 1,023)*				Never-smokers (N = 505)*			
Gene	**Variant**	**00**	**X0**	**XX**	**MAF**	**00**	**X0**	**XX**	**MAF**
XRCC1	rs1799782	840	116	3	0.06	416	60	2	0.07
	rs25486	402	449	115	0.35	217	216	56	0.34
	rs3213282	265	506	182	0.46	142	238	105	0.46
	rs3213245	309	498	157	0.42	164	234	88	0.42
	rs25489	935	79	2	0.04	455	47	2	0.05
hOGG1	rs3219008	601	305	35	0.20	305	153	15	0.19
	rs1052133	582	335	55	0.23	298	165	28	0.23
APEX1	rs1048945	957	62	0	0.03	471	32	0	0.03
	rs3136817	531	347	58	0.25	283	170	22	0.23
	rs3136820	252	442	203	0.47	115	213	114	0.50
	rs3136814	944	67	3	0.04	473	29	0	0.03
XPD	rs1052555	417	443	104	0.34	221	208	58	0.33
	rs50871	250	443	218	0.48	128	224	106	0.48
	rs1799793	394	448	114	0.35	203	220	63	0.36
	rs3916874	526	339	93	0.27	254	191	41	0.28
	rs1618536	289	513	157	0.43	144	237	105	0.46
XRCC3	rs3212024	472	395	98	0.31	224	213	53	0.33
	rs861531	352	459	153	0.40	177	247	59	0.38
	rs1799795	839	120	2	0.06	431	54	1	0.06
	rs1799794	635	289	41	0.19	316	155	19	0.20

*Genotype frequencies based on imputed data; ARIC, Atherosclerosis Risk in Communities Study; MAF, minor allele frequency

Table 32. (MS 1 supplemental results) Imputed tagSNP genotype data by incident CHD status in 623 African American ARIC participants.

		Incident CHD (N = 255)				CRS (N =367)			
Gene	**Variant**	**00**	**X0**	**XX**	**% imputed**	**00**	**X0**	**XX**	**% imputed**
XRCC1	rs1799782	3	0	0	1.4	2	0	0	0.6
	rs25486	0	0	0	0.0	0	3	0	0.9
	rs3213282	1	3	1	2.3	7	10	3	6.1
	rs3213245	0	4	1	2.2	0	1	0	0.3
	rs1475933	12	0	0	5.7	3	1	0	1.2
	rs25487	12	0	0	5.9	5	0	0	1.5
hOGG1	rs3219008	1	0	0	0.5	0	0	0	0.0
	rs1052133	0	0	4	1.8	0	0	0	0.0
	rs1805373	9	0	0	4.2	7	0	0	2.2
	rs2072668	1	0	0	0.5	1	0	4	1.5
APEX1	rs3136817	0	0	4	1.8	0	0	2	0.6
	rs3136820	0	0	0	0.0	0	0	0	0.0
	rs3136814	0	0	0	0.0	0	0	0	0.0
XPD	rs50871	0	0	12	5.5	4	0	4	2.4
	rs1799793	0	1	5	2.7	0	0	4	1.2
	rs3916874	8	0	0	3.5	14	0	0	4.2
	rs1618536	1	0	0	0.4	0	0	1	0.3
	rs1052559	0	0	1	0.5	1	0	0	0.3
XRCC3	rs3212024	0	1	0	0.4	0	0	0	0.0
	rs861531	7	3	3	6.4	3	0	1	1.2
	rs861539	5	0	0	2.3	23	0	0	7.4
	rs1799794	0	0	1	0.4	0	0	1	0.3

ARIC, Atherosclerosis Risk in Communities Study; CHD, coronary heart disease; CRS, cohort random sample

Table 33. (MS 1 supplemental results) tagSNP genotype frequencies by ever-smoking status in 623 African American ARIC participants.

		Ever-smokers (N = 372)*				Never-smokers (N =249)*			
Gene	**Variant**	**00**	**X0**	**XX**	**MAF**	**00**	**X0**	**XX**	**MAF**
XRCC1	rs1799782	292	28	0	0.04	190	28	2	0.07
	rs25486	187	132	18	0.25	133	83	9	0.22
	rs3213282	105	165	56	0.42	72	101	39	0.42
	rs3213245	113	170	53	0.41	78	112	36	0.41
	rs1475933	152	137	39	0.33	96	93	26	0.34
	rs25487	227	87	5	0.15	161	48	3	0.13
hOGG1	rs3219008	96	159	59	0.44	65	112	37	0.43
	rs1052133	223	93	17	0.19	154	63	6	0.17
	rs1805373	267	53	2	0.09	181	34	0	0.08
	rs2072668	169	126	29	0.28	116	83	18	0.27
APEX1	rs3136817	231	93	6	0.16	166	52	5	0.14
	rs3136820	124	157	42	0.37	99	89	33	0.35
	rs3136814	240	82	6	0.14	169	48	0	0.11
XPD	rs50871	268	52	5	0.10	175	44	2	0.11
	rs1799793	266	67	3	0.11	174	46	4	0.12
	rs3916874	301	36	5	0.07	204	16	3	0.05
	rs1618536	256	74	4	0.12	175	42	1	0.10
	rs1052559	186	121	23	0.25	135	76	9	0.21
XRCC3	rs3212024	237	91	9	0.16	166	54	5	0.14
	rs861531	165	131	22	0.28	105	91	19	0.30
	rs861539	206	96	18	0.21	123	63	17	0.24
	rs1799794	210	109	16	0.21	141	66	13	0.21

*Genotype frequencies based on imputed data; ARIC, Atherosclerosis Risk in Communities Study; MAF, minor allele frequency

Table 34. (MS 1 supplemental results) IRR point and interval estimates for the association between 20 DNA repair tagSNPs, ever-smoking and incident CHD in 1,160 Caucasian ARIC participants.

		Conventional analysis (Gene-smoking interaction for one SNP and main effects of all others)			
		Main effect of ever-smoking	Main effect of SNP	Joint effect, SNP and ever-smoking	ICR
XRCC1					
rs1799782	XX, XO vs OO	2.11 (1.44, 3.09)	1.19 (0.54, 2.63)	1.21 (0.63, 2.36)	-1.08 (-2.37, 0.21)
rs25486	XO vs. OO	1.94 (1.18, 3.21)	0.79 (0.42, 1.46)	1.61 (0.93, 2.81)	-0.12 (-1.01, 0.77)
	XX vs. OO	1.94 (1.18, 3.21)	0.47 (0.17, 1.35)	0.67 (0.28, 1.6)	-0.74 (-1.78, 0.29)
rs3213282	XO vs. OO	1.62 (0.85, 3.08)	1.97 (0.65, 6.01)	3.99 (1.32, 12.01)	1.4 (-0.84, 3.64)
	XX vs. OO	1.62 (0.85, 3.08)	3.48 (0.53, 22.65)	7.45 (1, 55.4)	3.35 (-5.49, 12.2)
rs3213245	XO vs. OO	2.37 (1.27, 4.42)	0.48 (0.17, 1.3)	0.83 (0.33, 2.07)	-1.01 (-2.44, 0.41)
	XX vs. OO	2.37 (1.27, 4.42)	0.22 (0.04, 1.33)	0.41 (0.07, 2.45)	-1.17 (-2.7, 0.35)
rs25489	XX, XO vs OO	1.93 (1.33, 2.78)	0.58 (0.14, 2.32)	1.19 (0.45, 3.13)	-0.32 (-1.37, 0.73)
hOGG1					
rs3219008	XX, XO vs OO	2.07 (1.36, 3.17)	1.22 (0.54, 2.79)	2.07 (0.98, 4.38)	-0.22 (-1.37, 0.93)
rs1052133	XO vs. OO	2.24 (1.44, 3.46)	1.11 (0.5, 2.46)	1.44 (0.71, 2.95)	-0.9 (-1.99, 0.2)
	XX vs. OO	2.24 (1.44, 3.46)	1.07 (0.21, 5.37)	3.76 (1.26, 11.22)	1.45 (-2.32, 5.22)
APEX1					
rs1048945	XX, XO vs OO	1.92 (1.32, 2.79)	1.57 (0.58, 4.24)	3.33 (1.45, 7.66)	0.84 (-2.01, 3.69)
rs3136817	XX, XO vs OO	1.6 (1.04, 2.48)	0.68 (0.38, 1.2)	1.76 (1.07, 2.9)	0.48 (-0.31, 1.28)
rs3136820	XO vs. OO	1.58 (0.83, 3.01)	1.17 (0.59, 2.34)	2.28 (1.21, 4.31)	0.53 (-0.52, 1.58)
	XX vs. OO	1.58 (0.83, 3.01	0.93 (0.41, 2.1)	2.29 (1.11, 4.74)	0.78 (-0.41, 1.98)
rs3136814	XX, XO vs OO	1.95 (1.34, 2.83)	1.53 (0.52, 4.48)	2.71 (1.25, 5.87)	0.23 (-2.28, 2.74)
XPD					
rs1052555	XO vs. OO	2.13 (1.28, 3.55)	1.52 (0.79, 2.91)	2.6 (1.36, 4.97)	-0.05 (-1.3, 1.19)
	XX vs. OO	2.13 (1.28, 3.55)	0.62 (0.21, 1.86)	1.44 (0.49, 4.21)	-0.3 (-1.69, 1.09)
rs50871	XO vs. OO	0.85 (0.45, 1.59)	0.56 (0.3, 1.06)	1.42 (0.78, 2.58)	1 (0.44, 1.56)
	XX vs. OO	0.85 (0.45, 1.59)	0.81 (0.37, 1.76)	2.32 (1.14, 4.71)	1.66 (0.44, 2.89)
rs1799793	XO vs. OO	2.29 (1.34, 3.92)	1.21 (0.61, 2.41)	2.29 (1.17, 4.46)	-0.22 (-1.44,1)
	XX vs. OO	2.29 (1.34, 3.92)	2.31 (0.67,8)	2.8 (1.02, 7.7)	-0.81 (-3.33, 1.71)
rs3916874	XO vs. OO	1.69 (1.06, 2.69)	0.95 (0.53, 1.7)	1.79 (1.08, 2.96)	0.15 (-0.74, 1.03)
	XX vs. OO	1.69 (1.06, 2.69)	0.69 (0.26, 1.86)	3.71 (1.61, 8.53)	2.33 (-0.55, 5.2)
rs1618536	XO vs. OO	1.71 (0.91, 3.18)	0.79 (0.41, 1.51)	1.73 (0.91, 3.3)	0.24 (-0.71, 1.19)
	XX vs. OO	1.71 (0.91, 3.18)	0.94 (0.4, 2.21)	1.59 (0.69, 3.69)	-0.05 (-1.33, 1.23)
XRCC3					
rs3212024	XO vs. OO	1.35 (0.85, 2.15)	0.81 (0.42, 1.56)	2 (1.08, 3.69)	0.83 (-0.05, 1.71)
	XX vs. OO	1.35 (0.85, 2.15)	0.7 (0.22, 2.22)	2.56 (0.83, 7.91)	1.51 (-0.9, 3.91)
rs861531	XO vs. OO	3.66 (2.05, 6.53)	2.39 (1.15, 4.95)	3.22 (1.61, 6.43)	-1.83 (-3.92, 0.26)
	XX vs. OO	3.66 (2.05, 6.53)	1.75 (0.55, 5.55)	2.61 (0.89, 7.64)	-1.79 (-4.18, 0.6)
rs1799795	XX, XO vs OO	1.78 (1.22, 2.61)	0.52 (0.2, 1.35)	1.98 (0.94, 4.19)	0.68 (-0.67, 2.03)
rs1799794	XX, XO vs OO	2.11 (1.38, 3.21)	1.25 (0.59, 2.64)	2.08 (1.07, 4.04)	-0.28 (-1.35, 0.8)

Piecewise exponential model adjusted for sampling strategy, alcohol intake, physical activity, and study center; ICR, interaction contrast ratio; CHD, coronary heart disease; ARIC, Atherosclerosis Risk in Communities Study

Table 35. (MS 1 supplemental results) IRR point and interval estimates for the association between 20 DNA repair tagSNPs, ever-smoking and incident CHD in 1160 Caucasian ARIC participants.

		tagSNPs within Each Gene Considered Exchangeable			
		Main effect of ever-smoking	**Main effect of SNP**	**Joint effect, SNP and ever-smoking**	**ICR**
XRCC1					
rs1799782	XX, XO vs OO	1.86 (1.29, 2.68)	1.11 (0.56, 2.21)	1.23 (0.66, 2.3)	-0.74 (-1.79, 0.31)
rs25486	XO vs. OO	1.8 (1.13, 2.88)	0.87 (0.5, 1.5)	1.58 (0.94, 2.63)	-0.09 (-0.9, 0.71)
	XX vs. OO	1.8 (1.13, 2.88)	0.64 (0.28, 1.47)	0.84 (0.39, 1.81)	-0.6 (-1.53, 0.33)
rs3213282	XO vs. OO	1.4 (0.79, 2.48)	1.04 (0.54, 2)	2.01 (1.04, 3.86)	0.57 (-0.34, 1.48)
	XX vs. OO	1.4 (0.79, 2.48)	1.05 (0.47, 2.32)	1.97 (0.85, 4.6)	0.52 (-0.75, 1.79)
rs3213245	XO vs. OO	2.02 (1.14, 3.55)	0.9 (0.48, 1.71)	1.45 (0.79, 2.65)	-0.47 (-1.55, 0.6)
	XX vs. OO	2.02 (1.14, 3.55)	0.81 (0.36, 1.83)	1.49 (0.65, 3.4)	-0.34 (-1.61, 0.94)
rs25489	XX, XO vs OO	1.77 (1.24, 2.53)	1.15 (0.52, 2.51)	1.9 (0.95, 3.79)	-0.02 (-1.39, 1.35)
hOGG1					
rs3219008	XX, XO vs OO	1.97 (1.31, 2.97)	1.34 (0.73, 2.45)	1.92 (1.06, 3.45)	-0.4 (-1.42, 0.63)
rs1052133	XO vs. OO	2.12 (1.4, 3.21)	1.13 (0.6, 2.12)	1.32 (0.74, 2.35)	-0.93 (-1.92, 0.06)
	XX vs. OO	2.12 (1.4, 3.21)	1.1 (0.39, 3.11)	2.77 (1.15, 6.65)	0.55 (-1.77, 2.87)
APEX1					
rs1048945	XX, XO vs OO	1.73 (1.22, 2.47)	1.34 (0.61, 2.95)	2.93 (1.36, 6.33)	0.86 (-1.33, 3.05)
rs3136817	XX, XO vs OO	1.51 (1, 2.29)	0.74 (0.44, 1.23)	1.63 (1.03, 2.57)	0.38 (-0.34, 1.09)
rs3136820	XO vs. OO	1.57 (0.88, 2.82)	1.25 (0.69, 2.26)	2.17 (1.23, 3.85)	0.35 (-0.63, 1.33)
	XX vs. OO	1.57 (0.88, 2.82)	1.05 (0.53, 2.05)	2.16 (1.12, 4.14)	0.54 (-0.56, 1.63)
rs3136814	XX, XO vs OO	1.76 (1.24, 2.52)	1.28 (0.56, 2.91)	2.23 (1.1, 4.52)	0.18 (-1.58, 1.94)
XPD					
rs1052555	XO vs. OO	1.92 (1.19, 3.1)	1.46 (0.85, 2.53)	2.32 (1.33, 4.06)	-0.06 (-1.12, 1)
	XX vs. OO	1.92 (1.19, 3.1)	0.78 (0.35, 1.71)	1.65 (0.73, 3.75)	-0.04 (-1.3, 1.21)
rs50871	XO vs. OO	0.96 (0.55, 1.69)	0.68 (0.4, 1.18)	1.48 (0.86, 2.53)	0.83 (0.25, 1.41)
	XX vs. OO	0.96 (0.55, 1.69)	0.92 (0.49, 1.73)	2.08 (1.11, 3.9)	1.2 (0.19, 2.21)
rs1799793	XO vs. OO	1.99 (1.2, 3.28)	1.13 (0.64, 1.98)	1.91 (1.07, 3.4)	-0.2 (-1.2, 0.79)
	XX vs. OO	1.99 (1.2, 3.28)	1.5 (0.65, 3.47)	2.06 (0.94, 4.5)	-0.43 (-2, 1.15)
rs3916874	XO vs. OO	1.6 (1.03, 2.47)	1.05 (0.62, 1.76)	1.78 (1.1, 2.86)	0.14 (-0.68, 0.95)
	XX vs. OO	1.6 (1.03, 2.47)	0.97 (0.45, 2.09)	3.08 (1.45, 6.53)	1.52 (-0.62, 3.66)
rs1618536	XO vs. OO	1.67 (0.95, 2.93)	0.87 (0.5, 1.52)	1.68 (0.95, 2.96)	0.14 (-0.73, 1)
	XX vs. OO	1.67 (0.95, 2.93)	1.03 (0.51, 2.09)	1.57 (0.74, 3.31)	-0.13 (-1.28, 1.01)
XRCC3					
rs3212024	XO vs. OO	1.35 (0.87, 2.09)	0.89 (0.49, 1.61)	1.87 (1.05, 3.31)	0.63 (-0.15, 1.41)
	XX vs. OO	1.35 (0.87, 2.09)	0.84 (0.32, 2.15)	2.27 (0.86, 6.02)	1.09 (-0.7, 2.87)
rs861531	XO vs. OO	2.87 (1.68, 4.88)	2 (1.05, 3.83)	2.65 (1.42, 4.95)	-1.22 (-2.75, 0.31)
	XX vs. OO	2.87 (1.68, 4.88)	1.42 (0.55, 3.72)	2.18 (0.85, 5.56)	-1.11 (-2.82, 0.59)
rs1799795	XX, XO vs OO	1.66 (1.16, 2.39)	0.6 (0.27, 1.37)	1.75 (0.87, 3.5)	0.49 (-0.6, 1.58)
rs1799794	XX, XO vs OO	1.93 (1.29, 2.89)	1.23 (0.62, 2.43)	1.84 (0.99, 3.4)	-0.33 (-1.26, 0.61)

Piecewise exponential model adjusted for sampling strategy, alcohol intake, physical activity, and study center; ICR, interaction contrast ratio; CHD, coronary heart disease; ARIC, Atherosclerosis Risk in Communities Study; τ^2 = 0.35, corresponds to a 10-fold residual effect range around the prior mean

Table 36. (MS 1 supplemental results) IRR point and interval estimates for the association between 20 DNA repair tagSNPs, ever-smoking and incident CHD in 345 African American ARIC participants.

		Conventional analysis (Gene-smoking interaction for one SNP and main effects of all others)			
		Main effect of ever-smoking	**Main effect of SNP**	**Joint effect, SNP and ever-smoking**	**ICR**
XRCC1					
rs1799782	XX, XO vs OO	1.51 (0.66, 3.43)	3.1 (0.68, 14.21)	5.15 (0.81, 32.76)	1.54 (-7.42, 10.49)
rs25486	XX, XO vs OO	1.1 (0.41, 2.99)	1.59 (0.29, 8.69)	4.19 (0.82, 21.31)	2.5 (-2.43, 7.43)
rs3213282	XO vs. OO	0.92 (0.24, 3.49)	0.61 (0.11, 3.51)	1.09 (0.14, 8.49)	0.56 (-0.94, 2.05)
	XX vs. OO	0.92 (0.24, 3.49)	3.06 (0.21, 44.22)	6.38 (0.51, 79.9)	3.4 (-6.72, 13.51)
rs3213245	XO vs. OO	1.21 (0.28, 5.24)	0.65 (0.12, 3.45)	1.15 (0.19, 6.92)	0.29 (-1.47, 2.05)
	XX vs. OO	1.21 (0.28, 5.24)	0.17 (0.01, 2.17)	0.25 (0.02, 3.67)	-0.13 (-1.92, 1.66)
rs1475933	XX, XO vs OO	1.8 (0.5, 6.55)	1.74 (0.35, 8.64)	2.43 (0.48, 12.41)	-0.12 (-2.92, 2.68)
rs25487	XX, XO vs OO	1.41 (0.55, 3.6)	0.59 (0.11, 3.01)	1.15 (0.22, 5.95)	0.16 (-1.87, 2.19)
hOGG1					
rs1052133	XX, XO vs OO	2.37 (0.91, 6.22)	2.11 (0.48, 9.24)	1.07 (0.25, 4.59)	-2.42 (-6.16, 1.33)
rs3219008	XO vs. OO	0.93 (0.25, 3.55)	0.65 (0.15, 2.79)	2.08 (0.47, 9.16)	1.5 (-1.05, 4.05)
	XX vs. OO	0.93 (0.25, 3.55)	5.34 (0.78, 36.77)	2.9 (0.44, 19.17)	-2.38 (-10.7, 5.95)
rs1805373	XX, XO vs OO	1.24 (0.55, 2.79)	1.36 (0.32, 5.69)	5.37 (1.57, 18.33)	3.78 (-2.48, 10.03)
rs2072668	XX, XO vs OO	3.9 (1.1, 13.77)	1.81 (0.43, 7.72)	0.91 (0.2, 4.11)	-3.8 (-9.54, 1.94)
APEX1					
rs3136820	XO vs. OO	1.09 (0.33, 3.66)	1.37 (0.33, 5.73)	2.3 (0.55, 9.59)	0.84 (-1.31, 2.99)
	XX vs. OO	1.09 (0.33, 3.66)	2.09 (0.37, 11.7)	6.34 (0.73, 54.77)	4.16 (-7.84, 16.17)
rs3136817	XX, XO vs OO	1.3 (0.55, 3.11)	0.47 (0.11, 1.99)	1.21 (0.37, 3.91)	0.43 (-0.92, 1.79)
rs3136814	XX, XO vs OO	1.38 (0.56, 3.42)	0.47 (0.08, 2.71)	1.02 (0.31, 3.34)	0.17 (-1.33, 1.68)
XPD					
rs1799793	XX, XO vs OO	1.14 (0.44, 2.93)	0.4 (0.1, 1.6)	1.35 (0.35, 5.24)	0.82 (-0.84, 2.48)
rs1618536	XX, XO vs OO	2.11 (0.85, 5.27)	1.69 (0.36, 8.02)	0.47 (0.14, 1.6)	-2.33 (-6.01, 1.35)
rs3916874	XX, XO vs OO	1.57 (0.69, 3.59)	0.84 (0.1, 7.45)	0.88 (0.2, 3.82)	-0.54 (-2.86, 1.78)
rs50871	XX, XO vs OO	1.68 (0.7, 4.07)	2.48 (0.62, 9.83)	2.71 (0.69, 10.67)	-0.45 (-4.64, 3.74)
rs1052559	XX, XO vs OO	1.57 (0.54, 4.57)	0.78 (0.25, 2.49)	1.16 (0.39, 3.43)	-0.19 (-2.05, 1.67)
XRCC3					
rs861531	XX, XO vs OO	1.62 (0.55, 4.72)	0.67 (0.15, 3)	0.96 (0.2, 4.66)	-0.32 (-1.94, 1.29)
rs1799794	XX, XO vs OO	1.72 (0.55, 5.38)	1.59 (0.48, 5.31)	2.13 (0.61, 7.48)	-0.18 (-2.8, 2.44)
rs3212024	XX, XO vs OO	1.63 (0.67, 3.96)	1.24 (0.4, 3.79)	1.55 (0.46, 5.2)	-0.31 (-2.46, 1.83)
rs861539	XX, XO vs OO	1.57 (0.56, 4.37)	1.97 (0.39, 10.05)	2.9 (0.62, 13.66)	0.36 (-2.67, 3.4)

Piecewise exponential model adjusted for sampling strategy, alcohol intake, physical activity, and study center; ICR, interaction contrast ratio; CHD, coronary heart disease; ARIC, Atherosclerosis Risk in Communities Study

Table 37. (MS 1 supplemental results) IRR point and interval estimates for the association between 20 DNA repair tagSNPs, ever-smoking and incident CHD in 345 African American ARIC participants.

		tagSNPs within Each Gene Considered Exchangeable			
		Main effect of ever-smoking	**Main effect of SNP**	**Joint effect, SNP and ever-smoking**	**ICR**
XRCC1					
rs1799782	XX, XO vs OO	1.23 (0.59, 2.57)	1.47 (0.56, 3.84)	1.88 (0.48, 7.35)	0.18 (-2.12, 2.48)
rs25486	XX, XO vs OO	1.04 (0.45, 2.4)	1.29 (0.53, 3.13)	2.14 (0.77, 5.98)	0.82 (-0.97, 2.6)
rs3213282	XO vs. OO	1.03 (0.37, 2.88)	0.86 (0.35, 2.07)	1.08 (0.36, 3.23)	0.2 (-0.95, 1.35)
	XX vs. OO	1.03 (0.37, 2.88)	1.2 (0.44, 3.29)	1.62 (0.5, 5.29)	0.39 (-1.29, 2.07)
rs3213245	XO vs. OO	1.16 (0.41, 3.29)	1.23 (0.51, 2.93)	1.44 (0.49, 4.26)	0.06 (-1.4, 1.51)
	XX vs. OO	1.16 (0.41, 3.29)	0.86 (0.31, 2.38)	1.29 (0.37, 4.57)	0.28 (-1.23, 1.79)
rs1475933	XX, XO vs OO	1.34 (0.48, 3.78)	1.31 (0.56, 3.05)	1.56 (0.58, 4.23)	-0.09 (-1.58, 1.39)
rs25487	XX, XO vs OO	1.17 (0.53, 2.6)	1.02 (0.42, 2.45)	1.4 (0.46, 4.19)	0.21 (-1.34, 1.76)
hOGG1					
rs1052133	XX, XO vs OO	1.56 (0.71, 3.45)	1.43 (0.59, 3.45)	1.17 (0.42, 3.27)	-0.82 (-2.45, 0.8)
rs3219008	XO vs. OO	1.07 (0.39, 2.93)	0.82 (0.33, 2.06)	1.27 (0.46, 3.51)	0.38 (-0.88, 1.64)
	XX vs. OO	1.07 (0.39, 2.93)	1.47 (0.54, 4.04)	1.41 (0.42, 4.77)	-0.13 (-1.9, 1.64)
rs1805373	XX, XO vs OO	1.12 (0.53, 2.34)	1.34 (0.52, 3.49)	2.36 (0.86, 6.46)	0.9 (-1.36, 3.16)
rs2072668	XX, XO vs OO	1.65 (0.63, 4.33)	1.21 (0.51, 2.84)	1.13 (0.41, 3.09)	-0.73 (-2.53, 1.08)
APEX1					
rs3136820	XO vs. OO	1.04 (0.39, 2.75)	0.72 (0.31, 1.7)	1.03 (0.4, 2.66)	0.26 (-0.73, 1.25)
	XX vs. OO	1.04 (0.39, 2.75)	1 (0.37, 2.69)	1.41 (0.38, 5.25)	0.37 (-1.35, 2.09)
rs3136817	XX, XO vs OO	1.12 (0.51, 2.45)	0.84 (0.34, 2.04)	1.33 (0.53, 3.35)	0.37 (-0.82, 1.56)
rs3136814	XX, XO vs OO	1.24 (0.57, 2.68)	0.79 (0.3, 2.07)	0.96 (0.35, 2.64)	-0.07 (-1.18, 1.05)
XPD					
rs1799793	XX, XO vs OO	1.14 (0.51, 2.55)	0.89 (0.37, 2.13)	1.47 (0.52, 4.12)	0.44 (-0.95, 1.82)
rs1618536	XX, XO vs OO	1.36 (0.63, 2.93)	0.85 (0.34, 2.13)	0.77 (0.3, 2)	-0.44 (-1.7, 0.83)
rs3916874	XX, XO vs OO	1.29 (0.62, 2.69)	1.2 (0.42, 3.43)	1.01 (0.33, 3.1)	-0.48 (-2.04, 1.08)
rs50871	XX, XO vs OO	1.29 (0.60 2.78)	1.49 (0.6, 3.73)	1.58 (0.58, 4.27)	-0.2 (-2.05, 1.64)
rs1052559	XX, XO vs OO	1.33 (0.55, 3.24)	1.02 (0.46, 2.28)	1.16 (0.47, 2.86)	-0.2 (-1.52, 1.13)
XRCC3					
rs861531	XX, XO vs OO	1.39 (0.56, 3.46)	0.9 (0.37, 2.17)	0.99 (0.35, 2.79)	-0.29 (-1.53, 0.94)
rs1799794	XX, XO vs OO	1.21 (0.49, 2.99)	0.99 (0.41, 2.41)	1.23 (0.45, 3.35)	0.02 (-1.2, 1.25)
rs3212024	XX, XO vs OO	1.25 (0.57, 2.75)	0.91 (0.4, 2.08)	1.13 (0.43, 3.02)	-0.03 (-1.23, 1.18)
rs861539	XX, XO vs OO	1.34 (0.56, 3.22)	1.09 (0.44, 2.7)	1.21 (0.44, 3.31)	-0.22 (-1.55, 1.1)

Piecewise exponential model adjusted for sampling strategy, alcohol intake, physical activity, and study center; ICR, interaction contrast ratio; CHD, coronary heart disease; ARIC, Atherosclerosis Risk in Communities Study; $\tau^2 = 0.35$, corresponds to a 10-fold residual effect range around the prior mean

C. Supplemental results, Manuscript 2

Table 38. (MS 2 supplemental results) Point and interval estimates for the association between 20 DNA repair tagSNPs, ever-smoking, and baseline mean IMT in 470 Caucasian ARIC participants

		Conventional analysis (Gene-smoking interaction for one SNP and main effects of all others)		
SNP		**Main effect of ever-smoking**	**Main effect of SNP**	**Joint effect, SNP and ever-smoking**
XRCC1				
rs1799782	XX, XO vs OO	0.092 (0.045, 0.139)	0.038 (-0.079, 0.154)	0.054 (-0.038, 0.145)
rs25486	XO vs. OO	0.076 (0.009, 0.143)	0.015 (-0.067, 0.098)	0.114 (0.037, 0.19)
	XX vs. OO	0.076 (0.009, 0.143)	0.023 (-0.11, 0.157)	0.084 (-0.03, 0.197)
rs3213282	XO vs. OO	0.046 (-0.032, 0.124)	-0.041 (-0.172, 0.09)	0.066 (-0.062, 0.193)
	XX vs. OO	0.046 (-0.032, 0.124)	-0.04 (-0.261, 0.18)	0.045 (-0.179, 0.269)
rs3213245	XO vs. OO	0.039 (-0.034, 0.113)	-0.029 (-0.155, 0.098)	0.091 (-0.028, 0.209)
	XX vs. OO	0.039 (-0.034, 0.113)	0.022 (-0.197, 0.241)	0.102 (-0.116, 0.32)
rs25489	XX, XO vs OO	0.092 (0.046, 0.139)	0.088 (-0.075, 0.25)	0.075 (-0.06, 0.209)
hOGG1				
rs3219008	XX, XO vs OO	0.085 (0.03, 0.139)	-0.047 (-0.154, 0.061)	0.034 (-0.065, 0.133)
rs1052133	XO vs. OO	0.091 (0.035, 0.147)	0.076 (-0.03, 0.182)	0.145 (0.047, 0.243)
	XX vs. OO	0.091 (0.035, 0.147)	0.124 (-0.073, 0.321)	0.205 (0.063, 0.348)
APEX1				
rs1048945	XX, XO vs OO	0.087 (0.041, 0.134)	0.007 (-0.129, 0.142)	0.025 (-0.103, 0.152)
rs3136817	XX, XO vs OO	0.098 (0.039, 0.158)	-0.005 (-0.08, 0.069)	0.059 (-0.008, 0.125)
rs3136820	XO vs. OO	0.079 (-0.009, 0.166)	0.035 (-0.056, 0.125)	0.12 (0.033, 0.207)
	XX vs. OO	0.079 (-0.009, 0.166)	0.019 (-0.082, 0.12)	0.104 (0.005, 0.203)
rs3136814	XX, XO vs OO	0.077 (0.031, 0.122)	-0.074 (-0.228, 0.08)	0.154 (0.034, 0.273)
XPD				
rs1052555	XO vs. OO	0.072 (0.007, 0.136)	-0.034 (-0.12, 0.051)	0.05 (-0.03, 0.13)
	XX vs. OO	0.072 (0.007, 0.136)	-0.088 (-0.222, 0.046)	0.038 (-0.096, 0.172)
rs50871	XO vs. OO	0.112 (0.024, 0.199)	0.056 (-0.029, 0.14)	0.141 (0.058, 0.223)
	XX vs. OO	0.112 (0.024, 0.199)	0.061 (-0.041, 0.164)	0.113 (0.018, 0.208)
rs1799793	XO vs. OO	0.059 (-0.007, 0.125)	-0.068 (-0.158, 0.021)	0.021 (-0.064, 0.105)
	XX vs. OO	0.059 (-0.007, 0.125)	-0.01 (-0.174, 0.154)	0.161 (0.026, 0.296)
rs3916874	XO vs. OO	0.068 (0.008, 0.129)	-0.037 (-0.114, 0.039)	0.089 (0.018, 0.161)
	XX vs. OO	0.068 (0.008, 0.129)	0.025 (-0.112, 0.162)	-0.002 (-0.12, 0.115)
rs1618536	XO vs. OO	0.121 (0.039, 0.204)	-0.02 (-0.106, 0.066)	0.077 (-0.007, 0.161)
	XX vs. OO	0.121 (0.039, 0.204)	0.035 (-0.078, 0.148)	0.037 (-0.076, 0.15)
XRCC3				
rs3212024	XO vs. OO	0.052 (-0.011, 0.115)	-0.019 (-0.109, 0.07)	0.094 (0.01, 0.178)
	XX vs. OO	0.052 (-0.011, 0.115)	-0.013 (-0.175, 0.148)	0.101 (-0.056, 0.258)
rs861531	XO vs. OO	0.081 (0.009, 0.154)	0.023 (-0.068, 0.113)	0.098 (0.012, 0.184)
	XX vs. OO	0.081 (0.009, 0.154)	-0.022 (-0.177, 0.133)	0.087 (-0.053, 0.227)
rs1799795	XX, XO vs OO	0.094 (0.046, 0.142)	-0.012 (-0.127, 0.103)	-0.001 (-0.104, 0.103)
rs1799794	XX, XO vs OO	0.115 (0.06, 0.169)	0.03 (-0.07, 0.129)	0.054 (-0.036, 0.144)

General linear mixed model adjusted for sampling strategy and study center; ARIC, Atherosclerosis Risk in Communities Study; IMT, intima-medial thickness; SNP, single nucleotide polymorphism; *hOGG1*, 8 – hydroxy-2' – deoxyguanosine-glycosylase/apurinic lyase; *APEX1,* apurinic/apyrimidinic endonuclease; *XRCC1,* X-ray repair cross complementing, group 1; *XPD,* xeroderma pigmentosum D; *XRCC3,* X-ray repair complementing defective repair in Chinese hamster cells 3

Table 39. (MS 2 supplemental results) Point and interval estimates for the association between 22 DNA repair tagSNPs, ever-smoking, and baseline mean IMT in 194 African American ARIC participants.

		Conventional analysis (Gene-smoking interaction for one SNP and main effects of all others)		
SNP		**Main effect of ever-smoking**	**Main effect of SNP**	**Joint effect, SNP and ever-smoking**
XRCC1				
rs1799782	XX, XO vs OO	0.026 (-0.053, 0.104)	0.068 (-0.095, 0.231)	-0.064 (-0.278, 0.151)
rs25486	XX, XO vs OO	0.044 (-0.047, 0.135)	0.023 (-0.112, 0.158)	-0.016 (-0.159, 0.127)
rs3213282	XO vs. OO	-0.065 (-0.191, 0.062)	-0.104 (-0.257, 0.048)	-0.028 (-0.188, 0.133)
	XX vs. OO	-0.065 (-0.191, 0.062)	0.077 (-0.135, 0.288)	0.056 (-0.144, 0.256)
rs3213245	XO vs. OO	0.017 (-0.113, 0.147)	-0.126 (-0.288, 0.036)	-0.106 (-0.266, 0.054)
	XX vs. OO	0.017 (-0.113, 0.147)	-0.164 (-0.375, 0.046)	-0.169 (-0.374, 0.037)
rs1475933	XX, XO vs OO	0 (-0.114, 0.114)	0.091 (-0.048, 0.231)	0.114 (-0.022, 0.251)
rs25487	XX, XO vs OO	0.038 (-0.047, 0.123)	0.021 (-0.13, 0.172)	-0.039 (-0.181, 0.102)
hOGG1				
rs1052133	XX, XO vs OO	0.017 (-0.071, 0.105)	0.002 (-0.135, 0.139)	0.008 (-0.108, 0.124)
rs3219008	XO vs. OO	0.105 (-0.008, 0.217)	0.038 (-0.085, 0.162)	-0.007 (-0.13, 0.117)
	XX vs. OO	0.105 (-0.008, 0.217)	0.021 (-0.148, 0.19)	-0.017 (-0.172, 0.138)
rs1805373	XX, XO vs OO	0.002 (-0.078, 0.082)	0.028 (-0.12, 0.177)	0.114 (-0.012, 0.24)
rs2072668	XX, XO vs OO	0.01 (-0.084, 0.105)	0.065 (-0.062, 0.193)	0.084 (-0.041, 0.209)
APEX1				
rs3136820	XO vs. OO	0.003 (-0.112, 0.117)	0.038 (-0.092, 0.168)	0.08 (-0.046, 0.206)
	XX vs. OO	0.003 (-0.112, 0.117)	0.054 (-0.126, 0.233)	-0.02 (-0.222, 0.181)
rs3136817	XX, XO vs OO	0.03 (-0.053, 0.113)	0.031 (-0.105, 0.167)	-0.011 (-0.128, 0.106)
rs3136814	XX, XO vs OO	-0.006 (-0.092, 0.08)	-0.096 (-0.226, 0.033)	-0.025 (-0.128, 0.078)
XPD				
rs1799793	XX, XO vs OO	0.002 (-0.083, 0.088)	-0.01 (-0.136, 0.115)	0.036 (-0.091, 0.164)
rs1618536	XX, XO vs OO	0.021 (-0.064, 0.105)	0.042 (-0.096, 0.18)	0.031 (-0.075, 0.137)
rs3916874	XX, XO vs OO	0.022 (-0.056, 0.101)	0.058 (-0.185, 0.301)	-0.053 (-0.197, 0.092)
rs50871	XX, XO vs OO	0.025 (-0.056, 0.106)	0.059 (-0.092, 0.21)	0.008 (-0.124, 0.14)
rs1052559	XX, XO vs OO	0.026 (-0.072, 0.124)	0.085 (-0.029, 0.2)	0.085 (-0.02, 0.19)
XRCC3				
rs861531	XX, XO vs OO	0.101 (-0.001, 0.202)	0.169 (0.04, 0.298)	0.105 (-0.026, 0.237)
rs1799794	XX, XO vs OO	-0.013 (-0.107, 0.082)	-0.026 (-0.142, 0.09)	0.034 (-0.08, 0.147)
rs3212024	XX, XO vs OO	0.036 (-0.05, 0.122)	-0.04 (-0.151, 0.07)	-0.083 (-0.188, 0.022)
rs861539	XX, XO vs OO	0.065 (-0.03, 0.159)	-0.025 (-0.155, 0.106)	-0.081 (-0.207, 0.044)

General linear mixed model adjusted for sampling strategy and study center; ARIC, Atherosclerosis Risk in Communities Study; IMT, intima-medial thickness; SNP, single nucleotide polymorphism; *hOGG1*, 8 – hydroxy-2' – deoxyguanosine-glycosylase/apurinic lyase; *APEX1,* apurinic/apyrimidinic endonuclease; *XRCC1,* X-ray repair cross complementing, group 1; *XPD,* xeroderma pigmentosum D; *XRCC3,* X-ray repair complementing defective repair in Chinese hamster cells 3

Table 40. (MS 2 supplemental results) Point and interval estimates for the association between 20 DNA repair tagSNPs, ever-smoking, and baseline mean IMT in 470 Caucasian ARIC participants specifying tagSNPs within each gene as exchangeable and a 0.3 residual effect range ($\tau^2 = 0.00574$).

		tagSNPs within a Given Gene Considered Exchangeable		
SNP		**Main effect of ever-smoking**	**Main effect of SNP**	**Joint effect, SNP and ever-smoking**
XRCC1				
rs1799782	XX, XO vs OO	0.091 (0.044, 0.137)	0.029 (-0.069, 0.126)	0.063 (-0.023, 0.149)
rs25486	XO vs. OO	0.078 (0.015, 0.141)	0.017 (-0.056, 0.091)	0.114 (0.043, 0.185)
	XX vs. OO	0.078 (0.015, 0.141)	0.023 (-0.085, 0.131)	0.089 (-0.013, 0.191)
rs3213282	XO vs. OO	0.055 (-0.017, 0.127)	-0.018 (-0.1, 0.065)	0.086 (0.004, 0.168)
	XX vs. OO	0.055 (-0.017, 0.127)	0 (-0.102, 0.102)	0.084 (-0.023, 0.191)
rs3213245	XO vs. OO	0.047 (-0.021, 0.116)	-0.021 (-0.104, 0.061)	0.094 (0.015, 0.174)
	XX vs. OO	0.047 (-0.021, 0.116)	0.011 (-0.094, 0.116)	0.089 (-0.02, 0.199)
rs25489	XX, XO vs OO	0.09 (0.044, 0.136)	0.04 (-0.062, 0.143)	0.063 (-0.035, 0.161)
hOGG1				
rs3219008	XX, XO vs OO	0.088 (0.035, 0.141)	-0.008 (-0.087, 0.071)	0.069 (-0.007, 0.145)
rs1052133	XO vs. OO	0.09 (0.036, 0.144)	0.045 (-0.037, 0.128)	0.117 (0.039, 0.195)
	XX vs. OO	0.09 (0.036, 0.144)	0.056 (-0.074, 0.186)	0.156 (0.042, 0.269)
APEX1				
rs1048945	XX, XO vs OO	0.087 (0.042, 0.133)	0.008 (-0.097, 0.113)	0.046 (-0.068, 0.16)
rs3136817	XX, XO vs OO	0.099 (0.042, 0.157)	0.002 (-0.064, 0.068)	0.067 (0.005, 0.129)
rs3136820	XO vs. OO	0.079 (0.001, 0.158)	0.028 (-0.05, 0.105)	0.112 (0.034, 0.19)
	XX vs. OO	0.079 (0.001, 0.158)	0.013 (-0.072, 0.097)	0.102 (0.014, 0.19)
rs3136814	XX, XO vs OO	0.08 (0.035, 0.125)	-0.037 (-0.15, 0.075)	0.135 (0.028, 0.243)
XPD				
rs1052555	XO vs. OO	0.078 (0.017, 0.14)	-0.022 (-0.094, 0.05)	0.059 (-0.011, 0.13)
	XX vs. OO	0.078 (0.017, 0.14)	-0.05 (-0.149, 0.048)	0.069 (-0.036, 0.173)
rs50871	XO vs. OO	0.098 (0.019, 0.177)	0.032 (-0.041, 0.104)	0.12 (0.047, 0.193)
	XX vs. OO	0.098 (0.019, 0.177)	0.029 (-0.053, 0.112)	0.091 (0.007, 0.175)
rs1799793	XO vs. OO	0.065 (0.002, 0.128)	-0.058 (-0.131, 0.015)	0.03 (-0.044, 0.103)
	XX vs. OO	0.065 (0.002, 0.128)	-0.022 (-0.132, 0.088)	0.136 (0.032, 0.24)
rs3916874	XO vs. OO	0.07 (0.012, 0.127)	-0.029 (-0.097, 0.038)	0.092 (0.025, 0.159)
	XX vs. OO	0.07 (0.012, 0.127)	0.015 (-0.089, 0.119)	0.019 (-0.084, 0.122)
rs1618536	XO vs. OO	0.119 (0.044, 0.194)	-0.022 (-0.096, 0.051)	0.071 (-0.004, 0.145)
	XX vs. OO	0.119 (0.044, 0.194)	0.024 (-0.068, 0.116)	0.04 (-0.059, 0.14)
XRCC3				
rs3212024	XO vs. OO	0.056 (-0.004, 0.116)	-0.016 (-0.098, 0.066)	0.095 (0.016, 0.174)
	XX vs. OO	0.056 (-0.004, 0.116)	-0.01 (-0.141, 0.121)	0.102 (-0.032, 0.236)
rs861531	XO vs. OO	0.083 (0.015, 0.15)	0.02 (-0.062, 0.102)	0.099 (0.019, 0.179)
	XX vs. OO	0.083 (0.015, 0.15)	-0.01 (-0.138, 0.118)	0.092 (-0.031, 0.215)
rs1799795	XX, XO vs OO	0.093 (0.046, 0.14)	-0.011 (-0.114, 0.092)	0.017 (-0.08, 0.114)
rs1799794	XX, XO vs OO	0.113 (0.059, 0.166)	0.026 (-0.065, 0.117)	0.057 (-0.028, 0.142)

General linear mixed model adjusted for sampling strategy and study center; ARIC, Atherosclerosis Risk in Communities Study; IMT, intima-medial thickness; SNP, single nucleotide polymorphism; *hOGG1*, 8 – hydroxy-2' – deoxyguanosine-glycosylase/apurinic lyase; *APEX1,* apurinic/apyrimidinic endonuclease; *XRCC1,* X-ray repair cross complementing, group 1; *XPD,* xeroderma pigmentosum D; *XRCC3,* X-ray repair complementing defective repair in Chinese hamster cells 3

Table 41. (MS 2 supplemental results) Point and interval estimates for the association between 22 DNA repair tagSNPs, ever-smoking, and baseline mean IMT in 194 African American ARIC participants specifying tagSNPs within each gene as exchangeable and a 0.3 residual effect range ($\tau^2 = 0.00574$).

		tagSNPs within a Given Gene Considered Exchangeable		
SNP		**Main effect of ever-smoking**	**Main effect of SNP**	**Joint effect, SNP and ever-smoking**
XRCC1				
rs1799782	XX, XO vs OO	0.016 (-0.059, 0.091)	0.024 (-0.091, 0.14)	-0.042 (-0.206, 0.121)
rs25486	XX, XO vs OO	0.031 (-0.054, 0.116)	0.009 (-0.087, 0.106)	-0.017 (-0.13, 0.096)
rs3213282	XO vs. OO	-0.038 (-0.147, 0.071)	-0.088 (-0.188, 0.011)	-0.037 (-0.15, 0.076)
	XX vs. OO	-0.038 (-0.147, 0.071)	0.011 (-0.105, 0.127)	-0.001 (-0.131, 0.128)
rs3213245	XO vs. OO	0.018 (-0.093, 0.128)	-0.047 (-0.148, 0.055)	-0.029 (-0.142, 0.084)
	XX vs. OO	0.018 (-0.093, 0.128)	-0.03 (-0.146, 0.086)	-0.046 (-0.178, 0.086)
rs1475933	XX, XO vs OO	-0.005 (-0.109, 0.099)	0.031 (-0.065, 0.126)	0.051 (-0.053, 0.156)
rs25487	XX, XO vs OO	0.026 (-0.054, 0.107)	0 (-0.104, 0.104)	-0.038 (-0.153, 0.076)
hOGG1				
rs1052133	XX, XO vs OO	0.011 (-0.072, 0.095)	0.009 (-0.091, 0.109)	0.017 (-0.081, 0.114)
rs3219008	XO vs. OO	0.074 (-0.026, 0.174)	0.022 (-0.073, 0.117)	-0.006 (-0.109, 0.097)
	XX vs. OO	0.074 (-0.026, 0.174)	0.02 (-0.099, 0.138)	-0.015 (-0.14, 0.111)
rs1805373	XX, XO vs OO	-0.001 (-0.077, 0.075)	0.022 (-0.09, 0.133)	0.1 (-0.015, 0.215)
rs2072668	XX, XO vs OO	0.002 (-0.088, 0.091)	0.032 (-0.061, 0.125)	0.053 (-0.047, 0.153)
APEX1				
rs3136820	XO vs. OO	-0.006 (-0.105, 0.093)	0.02 (-0.075, 0.116)	0.054 (-0.048, 0.156)
	XX vs. OO	-0.006 (-0.105, 0.093)	0.017 (-0.103, 0.136)	-0.012 (-0.161, 0.138)
rs3136817	XX, XO vs OO	0.022 (-0.058, 0.101)	0.021 (-0.079, 0.12)	-0.006 (-0.111, 0.098)
rs3136814	XX, XO vs OO	-0.001 (-0.082, 0.08)	-0.051 (-0.151, 0.048)	-0.012 (-0.11, 0.086)
XPD				
rs1799793	XX, XO vs OO	-0.001 (-0.081, 0.079)	-0.001 (-0.097, 0.095)	0.043 (-0.066, 0.152)
rs1618536	XX, XO vs OO	0.018 (-0.062, 0.097)	0.046 (-0.059, 0.15)	0.031 (-0.065, 0.126)
rs3916874	XX, XO vs OO	0.016 (-0.058, 0.091)	0.032 (-0.104, 0.168)	-0.041 (-0.164, 0.082)
rs50871	XX, XO vs OO	0.019 (-0.058, 0.096)	0.046 (-0.064, 0.156)	0.002 (-0.112, 0.116)
rs1052559	XX, XO vs OO	0.012 (-0.078, 0.103)	0.064 (-0.027, 0.155)	0.071 (-0.023, 0.166)
XRCC3				
rs861531	XX, XO vs OO	0.06 (-0.034, 0.154)	0.083 (-0.013, 0.18)	0.049 (-0.059, 0.158)
rs1799794	XX, XO vs OO	-0.006 (-0.092, 0.08)	-0.017 (-0.114, 0.079)	0.021 (-0.085, 0.126)
rs3212024	XX, XO vs OO	0.028 (-0.053, 0.109)	-0.034 (-0.127, 0.059)	-0.07 (-0.17, 0.03)
rs861539	XX, XO vs OO	0.045 (-0.043, 0.133)	-0.017 (-0.114, 0.079)	-0.055 (-0.16, 0.05)

General linear mixed model adjusted for sampling strategy and study center; ARIC, Atherosclerosis Risk in Communities Study; IMT, intima-medial thickness; SNP, single nucleotide polymorphism; *hOGG1*, 8 – hydroxy-2' – deoxyguanosine-glycosylase/apurinic lyase; *APEX1,* apurinic/apyrimidinic endonuclease; *XRCC1,* X-ray repair cross complementing, group 1; *XPD,* xeroderma pigmentosum D; *XRCC3,* X-ray repair complementing defective repair in Chinese hamster cells 3

REFERENCES

1. Doll R, Hill AB. Lung cancer and other causes of death in relation to smoking; a second report on the mortality of British doctors. *Br Med J* 1956;12:1071-81.

2. Hammond EC, Horn D. Smoking and death rates; report on forty-four monghs of follow-up of 187,783 men. II. Death rates by cause. *J Am Med Assoc* 1958;166:1294-308.

3. Doll R, Peto R. Mortality in relation to smoking: 20 years' observations on male British doctors. *Br Med J* 1976;2:1525-36.

4. Doll R, Gray R, Hafner B, Peto R. Mortality in relation to smoking: 22 years' observations on female British doctors. *Br Med J* 1980;280:967-71.

5. Willett WC, Green A, Stampfer MJ et al. Relative and absolute excess risks of coronary heart disease among women who smoke cigarettes. *N Engl J Med* 1987;317:1303-9.

6. Price JF, Mowbray PI, Lee AJ, Rumley A, Lowe GD, Fowkes FG. Relationship between smoking and cardiovascular risk factors in the development of peripheral arterial disease and coronary artery disease: Edinburgh Artery Study. *Eur Heart J* 1999;20:344-53.

7. Jonas MA, Oates JA, Ockene JK, Hennekens CH. Statement on smoking and cardiovascular disease for health care professionals. American Heart Association. *Circulation* 1992;86:1664-9.

8. Anonymous. Annual smoking-attributable mortality, years of potential life lost, and productivity losses--United States, 1997-2001. *MMWR Morb Mortal Wkly Rep* 2005;54:625-8.

9. Green CR, Rodgman A. The Tobacco Chemists' Research Conference: A half-century of advances in analytical methodology of tobacco and its products. *Recent Adv. Tob. Sci.* 1996;22:131-304.

10. Van Schooten FJ, Hirvonen A, Maas LM et al. Putative susceptibility markers of coronary artery disease: association between VDR genotype, smoking, and aromatic DNA adduct levels in human right atrial tissue. *Faseb J* 1998;12:1409-17.

11. De Flora S, Izzotti A, Walsh D, Degan P, Petrilli GL, Lewtas J. Molecular epidemiology of atherosclerosis. *Faseb J* 1997;11:1021-31.

12. Zhang Y, Ramos KS. The induction of proliferative vascular smooth muscle cell phenotypes by benzo[a]pyrene does not involve mutational activation of ras genes. *Mutat Res* 1997;373:285-92.

13. Penn A, Snyder CA. 1,3 Butadiene, a vapor phase component of environmental tobacco smoke, accelerates arteriosclerotic plaque development. *Circulation* 1996;93:552-7.

14. Anonymous. Cancer. In: Stratton K, Shetty P, Wallace R, Bondurant S, eds. *Clearing the Smoke. Assessing the Science Base for Tobacco Harm Reduction*. Washington, D.C.: National Academy Press; 2001:367-469.

15. Ross R. The pathogenesis of atherosclerosis--an update. *N Engl J Med* 1986;314:488-500.

16. Leibowitz J. The History of Coronary Heart Disease. Berkeley: University of California Press; 1970.

17. Faxon DP, Fuster V, Libby P et al. Atherosclerotic Vascular Disease Conference: Writing Group III: pathophysiology. *Circulation* 2004;109:2617-25.

18. Schoen FJ, Cotran RS. Blood Vessels. In: Cotran RS, Kumar V, Collins T, eds. *Pathologic Basis of Disease*. 6th ed. Philadelphia, PA: W.B. Saunders Co.; 1999:493-542.

19. McGill HC, Jr. George Lyman Duff memorial lecture. Persistent problems in the pathogenesis of atherosclerosis. *Arteriosclerosis* 1984;4:443-51.

20. Stary HC. Evolution and progression of atherosclerotic lesions in coronary arteries of children and young adults. *Arteriosclerosis* 1989;9:I19-32.

21. Libby P. The Vascular Biology of Atherosclerosis. In: Braunwald E, Zipes D, Libby P, eds. *Heart Disease: a Textbook of Cardiovascular Medicine*. Philadelphia: W.B. Saunders; 2001:995-1009.

22. Luepker RV, Apple FS, Christenson RH et al. Case definitions for acute coronary heart disease in epidemiology and clinical research studies: a statement from the AHA Council on Epidemiology and Prevention; AHA Statistics Committee; World Heart Federation Council on Epidemiology and Prevention; the European Society of Cardiology Working Group on Epidemiology and Prevention; Centers for Disease Control and Prevention; and the National Heart, Lung, and Blood Institute. *Circulation* 2003;108:2543-9.

23. Hessler JR, Robertson AL, Jr., Chisolm GM, 3rd. LDL-induced cytotoxicity and its inhibition by HDL in human vascular smooth muscle and endothelial cells in culture. *Atherosclerosis* 1979;32:213-29.

24. Kume N, Kita T. Apoptosis of vascular cells by oxidized LDL: involvement of caspases and LOX-1 and its implication in atherosclerotic plaque rupture. *Circ Res* 2004;94:269-70.

25. Libby P. Inflammation in atherosclerosis. *Nature* 2002;420:868-74.

26. Gimbrone MA, Jr., Cybulsky MI, Kume N, Collins T, Resnick N. Vascular endothelium. An integrator of pathophysiological stimuli in atherogenesis. *Ann N Y Acad Sci* 1995;748:122-31; discussion 131-2.

27. Wight TN. The extracellular matrix and atherosclerosis. *Curr Opin Lipidol* 1995;6:326-34.

28. O'Brien ER, Garvin MR, Dev R et al. Angiogenesis in human coronary atherosclerotic plaques. *Am J Pathol* 1994;145:883-94.

29. Hiatt WR. Medical treatment of peripheral arterial disease and claudication. *N Engl J Med* 2001;344:1608-21.

30. Sacco RL. Clinical practice. Extracranial carotid stenosis. *N Engl J Med* 2001;345:1113-8.

31. Pasternak RC, Abrams J, Greenland P, Smaha LA, Wilson PW, Houston-Miller N. 34th Bethesda Conference: Task force #1--Identification of coronary heart disease risk: is there a detection gap? *J Am Coll Cardiol* 2003;41:1863-74.

32. Castelli WP. Lipids, risk factors and ischaemic heart disease. *Atherosclerosis* 1996;124 Suppl:S1-9.

33. Marmot MG. In: Elliott P, ed. *Coronary Heart Disease Epidemiology*; 2005.

34. Dzielak DJ. The immune system and hypertension. *Hypertension* 1992;19:I36-44.

35. Bataillard A, Renaudin C, Sassard J. Silica attenuates hypertension in Lyon hypertensive rats. *J Hypertens* 1995;13:1581-4.

36. McCarron RM, Wang L, Siren AL, Spatz M, Hallenbeck JM. Monocyte adhesion to cerebromicrovascular endothelial cells derived from hypertensive and normotensive rats. *Am J Physiol* 1994;267:H2491-7.

37. Schmid-Schonbein GW, Seiffge D, DeLano FA, Shen K, Zweifach BW. Leukocyte counts and activation in spontaneously hypertensive and normotensive rats. *Hypertension* 1991;17:323-30.

38. Fuster V, Gotto AM, Libby P, Loscalzo J, McGill HC. 27th Bethesda Conference: matching the intensity of risk factor management with the hazard for coronary disease events. Task Force 1. Pathogenesis of coronary disease: the biologic role of risk factors. *J Am Coll Cardiol* 1996;27:964-76.

39. Malek AM, Alper SL, Izumo S. Hemodynamic shear stress and its role in atherosclerosis. *Jama* 1999;282:2035-42.

40. Kannel WB, Wilson PW. An update on coronary risk factors. *Med Clin North Am* 1995;79:951.

41. AHA. Heart Disease and Stroke Statistics - 2005 Update. Dallas, Texas: American Heart Association; 2005.

42. Fusegawa Y, Goto S, Handa S, Kawada T, Ando Y. Platelet spontaneous aggregation in platelet-rich plasma is increased in habitual smokers. *Thromb Res* 1999;93:271-8.

43. Rival J, Riddle JM, Stein PD. Effects of chronic smoking on platelet function. *Thromb Res* 1987;45:75-85.

44. Smith FB, Lee AJ, Fowkes FG, Price JF, Rumley A, Lowe GD. Hemostatic factors as predictors of ischemic heart disease and stroke in the Edinburgh Artery Study. *Arterioscler Thromb Vasc Biol* 1997;17:3321-5.

45. Kannel WB, D'Agostino RB, Belanger AJ. Fibrinogen, cigarette smoking, and risk of cardiovascular disease: insights from the Framingham Study. *Am Heart J* 1987;113:1006-10.

46. Barua RS, Ambrose JA, Saha DC, Eales-Reynolds LJ. Smoking is associated with altered endothelial-derived fibrinolytic and antithrombotic factors: an in vitro demonstration. *Circulation* 2002;106:905-8.

47. Newby DE, Wright RA, Labinjoh C et al. Endothelial dysfunction, impaired endogenous fibrinolysis, and cigarette smoking: a mechanism for arterial thrombosis and myocardial infarction. *Circulation* 1999;99:1411-5.

48. Pretorius M, Rosenbaum DA, Lefebvre J, Vaughan DE, Brown NJ. Smoking impairs bradykinin-stimulated t-PA release. *Hypertension* 2002;39:767-71.

49. De Vriese AS, Verbeuren TJ, Van de Voorde J, Lameire NH, Vanhoutte PM. Endothelial dysfunction in diabetes. *Br J Pharmacol* 2000;130:963-74.

50. Verma S, Anderson TJ. The ten most commonly asked questions about endothelial function in cardiology. *Cardiol Rev* 2001;9:250-2.

51. Sarkar R, Meinberg EG, Stanley JC, Gordon D, Webb RC. Nitric oxide reversibly inhibits the migration of cultured vascular smooth muscle cells. *Circ Res* 1996;78:225-30.

52. Kubes P, Suzuki M, Granger DN. Nitric oxide: an endogenous modulator of leukocyte adhesion. *Proc Natl Acad Sci U S A* 1991;88:4651-5.

53. Steinberg HO, Baron AD. Vascular function, insulin resistance and fatty acids. *Diabetologia* 2002;45:623-34.

54. Kim JA, Montagnani M, Koh KK, Quon MJ. Reciprocal relationships between insulin resistance and endothelial dysfunction: molecular and pathophysiological mechanisms. *Circulation* 2006;113:1888-904.

55. Furukawa S, Fujita T, Shimabukuro M et al. Increased oxidative stress in obesity and its impact on metabolic syndrome. *J Clin Invest* 2004;114:1752-61.

56. Awtry EH, Loscalzo J. Coronary Heart Disease. In: Carpenter CCJ, Griggs RC, Loscalzo J, eds. *Cecil Essentials of Medicine*. Philadelphia, PA: W.B. Saunders Co.; 2001:79-99.

57. Paffenbarger RS, Hyde RT. Exercise in the prevension of coronary heart disease. *Preventive Medicine* 1984;13:3-22.

58. Kramsch DM, Aspen AJ, Abramowitz BM, Kreimendahl T, Hood WB. Reduction of coronary atherosclerosis by moderate conditioning exercise in monkeys on an atherogenic diet. *N Engl J Med* 1981;305:1483-1489.

59. Yano K, Rhoads GG, Kagan A. Coffee, alcohol and risk of coronary heart disease among Japanese men living in Hawaii. *N Engl J Med* 1977;297:405-9.

60. Marmot MG, Rose G, Shipley MJ, Thomas BJ. Alcohol and mortality: a U-shaped curve. *Lancet* 1981;1:580-3.

61. Klatsky AL, Friedman GD, Siegelaub AB. Alcohol use and cardiovascular disease: the Kaiser-Permanente experience. *Circulation* 1981;64:III 32-41.

62. Kozarevic D, Demirovic J, Gordon T, Kaelber CT, McGee D, Zukel WJ. Drinking habits and coronary heart disease: the Yugoslavia cardiovascular disease study. *Am J Epidemiol* 1982;116:748-58.

63. Rimm E. Alcohol and cardiovascular disease. *Curr Atheroscler Rep* 2000;2:529-35.

64. Nora JJ, Berg K, Nora AH. Cardiovascular diseases: genetics, epidemiology, and prevention. In: Motulsky AG, Harper PS, Bobrow M, Scriver C, eds. *Oxford monographs on medical genetics, number 22*. New York: Oxford University Press; 1991:4-5.

65. Jousilahti P, Puska P, Vartiainen E, Pekkanen J, Tuomilehto J. Parental history of premature coronary heart disease: an independent risk factor of myocardial infarction. *J Clin Epidemiol* 1996;49:497-503.

66. Grech ED, Ramsdale DR, Bray CL, Faragher EB. Family history as an independent risk factor of coronary artery disease. *Eur Heart J* 1992;13:1311-5.

67. Myers RH, Kiely DK, Cupples LA, Kannel WB. Parental history is an independent risk factor for coronary artery disease: the Framingham Study. *Am Heart J* 1990;120:963-9.

68. Hopkins PN, Williams RR, Kuida H et al. Family history as an independent risk factor for incident coronary artery disease in a high-risk cohort in Utah. *Am J Cardiol* 1988;62:703-7.

69. Perkins KA. Family history of coronary heart disease: is it an independent risk factor? *Am J Epidemiol* 1986;124:182-94.

70. Shea S, Ottman R, Gabrieli C, Stein Z, Nichols A. Family history as an independent risk factor for coronary artery disease. *J Am Coll Cardiol* 1984;4:793-801.

71. Sesso HD, Lee IM, Gaziano JM, Rexrode KM, Glynn RJ, Buring JE. Maternal and paternal history of myocardial infarction and risk of cardiovascular disease in men and women. *Circulation* 2001;104:393-8.

72. Keavney B. Genetic epidemiological studies of coronary heart disease. *Int J Epidemiol* 2002;31:730-6.

73. Schoen FJ. The Heart. In: Collins T, ed. *Pathologic Basis of Disease*. 6th ed. Philadelphia, PA: W.B. Saunders Co.; 1999:543-599.

74. Awtry EH, Loscalzo J. Cardiac Arrhythmias. In: Carpenter CCJ, Griggs RC, Loscalzo J, eds. *Cecil Essentials of Medicine*. Philadelphia, PA: W.B. Saunders Co.; 2001:100-126.

75. Anonymous. Cardiovascular Diseases. *The health consequences of smoking: a report of the Surgeon General*. Atlanta, GA: Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health; 2004:361-419.

76. Alpert JS, Thygesen K, Antman E, Bassand JP. Myocardial infarction redefined--a consensus document of The Joint European Society of Cardiology/American College of Cardiology Committee for the redefinition of myocardial infarction. *J Am Coll Cardiol* 2000;36:959-69.

77. Newby LK, Alpert JS, Ohman EM, Thygesen K, Califf RM. Changing the diagnosis of acute myocardial infarction: implications for practice and clinical investigations. *Am Heart J* 2002;144:957-80.

78. Weinstein BJ, Epstein FH. Comparability of Criteria and Methods in the Epidemiology of Cardiovascular Disease. Report of a Survey. *Circulation* 1964;30:643-53.

79. World Health Organization Regional Office for Europe WGitEoIHDR. Report of the Fifth Working Group, Copenhagen, 26-29 April 1971. Copenhagen, Denmark: Regional Office for Europe, World Health Organization; 1971.

80. Tunstall-Pedoe H, Kuulasmaa K, Amouyel P, Arveiler D, Rajakangas AM, Pajak A. Myocardial infarction and coronary deaths in the World Health Organization MONICA Project. Registration procedures, event rates, and case-fatality rates in 38 populations from 21 countries in four continents. *Circulation* 1994;90:583-612.

81. Marshall GD. Overview of recruitment to the coronary primary prevention trial. *Circulation* 1982;66:IV5-9.

82. Murray CJ, Lopez AD. Alternative projections of mortality and disability by cause 1990-2020: Global Burden of Disease Study. *Lancet* 1997;349:1498-504.

83. Lloyd-Jones DM, Larson MG, Beiser A, Levy D. Lifetime risk of developing coronary heart disease. *Lancet* 1999;353:89-92.

84. Jones DW, Chambless LE, Folsom AR et al. Risk factors for coronary heart disease in African Americans: the atherosclerosis risk in communities study, 1987-1997. *Arch Intern Med* 2002;162:2565-71.

85. Howard BV, Lee ET, Cowan LD et al. Rising tide of cardiovascular disease in American Indians. The Strong Heart Study. *Circulation* 1999;99:2389-95.

86. Benfante R. Studies of cardiovascular disease and cause-specific mortality trends in Japanese-American men living in Hawaii and risk factor comparisons with other Japanese populations in the Pacific region: a review. *Hum Biol* 1992;64:791-805.

87. Lavezzi AM, Ottaviani G, Matturri L. Biology of the smooth muscle cells in human atherosclerosis. *Apmis* 2005;113:112-21.

88. Ross R. Atherosclerosis--an inflammatory disease. *N Engl J Med* 1999;340:115-26.

89. Libby P, Hansson GK. Involvement of the immune system in human atherogenesis: current knowledge and unanswered questions. *Lab Invest* 1991;64:5-15.

90. Benditt EP. Evidence for a monoclonal origin of human atherosclerotic plaques and some implications. *Circulation* 1974;50:650-2.

91. Thomas WA, Kim DN. Biology of disease. Atherosclerosis as a hyperplastic and/or neoplastic process. *Lab Invest* 1983;48:245-55.

92. Pearson TA, Dillman JM, Solez K, Heptinstall RH. Clonal characteristics in layers of human atherosclerotic plaques. A study of the selection hypothesis of monoclonality. *Am J Pathol* 1978;93:93-116.

93. Pearson TA, Dillman JM, Solex K, Heptinstall RH. Clonal markers in the study of the origin and growth of human atherosclerotic lesions. *Circ Res* 1978;43:10-8.

94. Hatzistamou J, Kiaris H, Ergazaki M, Spandidos DA. Loss of heterozygosity and microsatellite instability in human atherosclerotic plaques. *Biochem Biophys Res Commun* 1996;225:186-90.

95. McCaffrey TA, Du B, Consigli S et al. Genomic instability in the type II TGF-beta1 receptor gene in atherosclerotic and restenotic vascular cells. *J Clin Invest* 1997;100:2182-8.

96. Penn A, Garte SJ, Warren L, Nesta D, Mindich B. Transforming gene in human atherosclerotic plaque DNA. *Proc Natl Acad Sci U S A* 1986;83:7951-5.

97. Parkes JL, Cardell RR, Hubbard FC, Jr., Hubbard D, Meltzer A, Penn A. Cultured human atherosclerotic plaque smooth muscle cells retain transforming potential and display enhanced expression of the myc protooncogene. *Am J Pathol* 1991;138:765-75.

98. Andreassi MG, Botto N, Colombo MG, Biagini A, Clerico A. Genetic instability and atherosclerosis: can somatic mutations account for the development of cardiovascular diseases? *Environ Mol Mutagen* 2000;35:265-9.

99. Black HR. Smoking and cardiovascular disease. In: Laragh JH, Brenner BM, eds. *Hypertension: Pathophysiology, Diagnosis and Management*. 2nd ed. New York, NY: Raven Press, Ltd.; 1995:2621-47.

100. Church DF, Pryor WA. The oxidative stress placed on the lung by cigarette smoke. In: Crystal RG, West JB, Barres PJ, Cherniack NS, Weibel ER, eds. *The Lung*. 2 vol. New York, NY: Raven Press; 1991:1975-1979.

101. Halliwell B, Gutteridge JM. Free Radicals in Biology and Medicine. 2nd ed. Oxford: Clarendon Press; 1989.

102. Orrenius S, McConkey DJ, Bellomo G, Nicotera P. Role of Ca2+ in toxic cell killing. *Trends Pharmacol Sci* 1989;10:281-5.

103. Halliwell B. Oxidants and human disease: some new concepts. *Faseb J* 1987;1:358-64.

104. Pryor WA, Stone K. Oxidants in cigarette smoke. Radicals, hydrogen peroxide, peroxynitrate, and peroxynitrite. *Ann N Y Acad Sci* 1993;686:12-27; discussion 27-8.

105. Pryor WA, Prier DG, Church DF. Electron-spin resonance study of mainstream and sidestream cigarette smoke: nature of the free radicals in gas-phase smoke and in cigarette tar. *Environ Health Perspect* 1983;47:345-55.

106. Church DF, Pryor WA. Free-radical chemistry of cigarette smoke and its toxicological implications. *Environ Health Perspect* 1985;64:111-26.

107. Pryor WA, Dooley MM, Church DF. Mechanisms of cigarette smoke toxicity: the inactivation of human alpha-1-proteinase inhibitor by nitric oxide/isoprene mixtures in air. *Chem Biol Interact* 1985;54:171-83.

108. Pryor WA, Dooley MM, Church DF. Inactivation of human alpha-1-proteinase inhibitor by gas-phase cigarette smoke. *Biochem Biophys Res Commun* 1984;122:676-81.

109. Jaimes EA, DeMaster EG, Tian RX, Raij L. Stable compounds of cigarette smoke induce endothelial superoxide anion production via NADPH oxidase activation. *Arterioscler Thromb Vasc Biol* 2004;24:1031-6.

110. Cooper KO, Witz G, Witmer C. The effects of alpha, beta-unsaturated aldehydes on hepatic thiols and thiol-containing enzymes. *Fundam Appl Toxicol* 1992;19:343-9.

111. Corte ED, Stirpe F. The regulation of rat liver xanthine oxidase. Involvement of thiol groups in the conversion of the enzyme activity from dehydrogenase (type D) into oxidase (type O) and purification of the enzyme. *Biochem J* 1972;126:739-45.

112. Inanami O, Johnson JL, Babior BM. The leukocyte NADPH oxidase subunit p47PHOX: the role of the cysteine residues. *Arch Biochem Biophys* 1998;350:36-40.

113. Borish ET, Pryor WA, Venugopal S, Deutsch WA. DNA synthesis is blocked by cigarette tar-induced DNA single-strand breaks. *Carcinogenesis* 1987;8:1517-20.

114. Cross CE, Halliwell B, Borish ET et al. Oxygen radicals and human disease. *Ann Intern Med* 1987;107:526-45.

115. Ambrose JA, Barua RS. The pathophysiology of cigarette smoking and cardiovascular disease. *J Am Coll Cardiol* 2004;43:1731-1737.

116. Clarkson TB, Weingand KW, Kaplan JR, Adams MR. Mechanisms of atherogenesis. *Circulation* 1987;76:I20-8.

117. Ijzerman RG, Serne EH, van Weissenbruch MM, de Jongh RT, Stehouwer CD. Cigarette smoking is associated with an acute impairment of microvascular function in humans. *Clin Sci (Lond)* 2003;104:247-52.

118. Celermajer DS, Sorensen KE, Georgakopoulos D et al. Cigarette smoking is associated with dose-related and potentially reversible impairment of endothelium-dependent dilation in healthy young adults. *Circulation* 1993;88:2149-55.

119. Kugiyama K, Yasue H, Ohgushi M et al. Deficiency in nitric oxide bioactivity in epicardial coronary arteries of cigarette smokers. *J Am Coll Cardiol* 1996;28:1161-7.

120. Sumida H, Watanabe H, Kugiyama K, Ohgushi M, Matsumura T, Yasue H. Does passive smoking impair endothelium-dependent coronary artery dilation in women? *J Am Coll Cardiol* 1998;31:811-5.

121. Barua RS, Ambrose JA, Eales-Reynolds LJ, DeVoe MC, Zervas JG, Saha DC. Dysfunctional endothelial nitric oxide biosynthesis in healthy smokers with impaired endothelium-dependent vasodilatation. *Circulation* 2001;104:1905-10.

122. Mayhan WG, Sharpe GM. Effect of cigarette smoke extract on arteriolar dilatation in vivo. *J Appl Physiol* 1996;81:1996-2003.

123. Mayhan WG, Patel KP. Effect of nicotine on endothelium-dependent arteriolar dilatation in vivo. *Am J Physiol* 1997;272:H2337-42.

124. Celermajer DS, Adams MR, Clarkson P et al. Passive smoking and impaired endothelium-dependent arterial dilatation in healthy young adults. *N Engl J Med* 1996;334:150-4.

125. Ota Y, Kugiyama K, Sugiyama S et al. Impairment of endothelium-dependent relaxation of rabbit aortas by cigarette smoke extract--role of free radicals and attenuation by captopril. *Atherosclerosis* 1997;131:195-202.

126. Napoli C, de Nigris F, Williams-Ignarro S, Pignalosa O, Sica V, Ignarro LJ. Nitric oxide and atherosclerosis: An update. *Nitric Oxide* 2006.

127. Napoli C, Ignarro LJ. Nitric oxide and atherosclerosis. *Nitric Oxide* 2001;5:88-97.

128. Tracy RP, Psaty BM, Macy E et al. Lifetime smoking exposure affects the association of C-reactive protein with cardiovascular disease risk factors and subclinical disease in healthy elderly subjects. *Arterioscler Thromb Vasc Biol* 1997;17:2167-76.

129. Bermudez EA, Rifai N, Buring JE, Manson JE, Ridker PM. Relation between markers of systemic vascular inflammation and smoking in women. *Am J Cardiol* 2002;89:1117-9.

130. Mendall MA, Patel P, Asante M et al. Relation of serum cytokine concentrations to cardiovascular risk factors and coronary heart disease. *Heart* 1997;78:273-7.

131. Tappia PS, Troughton KL, Langley-Evans SC, Grimble RF. Cigarette smoking influences cytokine production and antioxidant defences. *Clin Sci (Lond)* 1995;88:485-9.

132. Craig WY, Palomaki GE, Haddow JE. Cigarette smoking and serum lipid and lipoprotein concentrations: an analysis of published data. *Bmj* 1989;298:784-8.

133. Binkova B, Strejc P, Boubelik O, Stavkova Z, Chvatalova I, Sram RJ. DNA adducts and human atherosclerotic lesions. *Int J Hyg Environ Health* 2001;204:49-54.

134. Wannamethee SG, Shaper AG, Whincup PH, Walker M. Role of risk factors for major coronary heart disease events with increasing length of follow up. *Heart* 1999;81:374-9.

135. Goldberg RJ, Burchfiel CM, Benfante R, Chiu D, Reed DM, Yano K. Lifestyle and biologic factors associated with atherosclerotic disease in middle-aged men. 20-year findings from the Honolulu Heart Program. *Arch Intern Med* 1995;155:686-94.

136. Kawachi I, Colditz GA, Stampfer MJ et al. Smoking cessation and decreased risks of total mortality, stroke, and coronary heart disease incidence among women: a prospective cohort study. In: Burns DM, Garfinkel L, Samet JM, eds. *Changes in Cigarette-*

Related Disease Risks and Their Implication for Prevention and Control. Bethesda, MD: NIH publications; 1997:531-565.

137. Baba S, Iso H, Mannami T, Sasaki S, Okada K, Konishi M. Cigarette smoking and risk of coronary heart disease incidence among middle-aged Japanese men and women: the JPHC Study Cohort I. *Eur J Cardiovasc Prev Rehabil* 2006;13:207-13.

138. Hoffmann D, Hoffmann I, El-Bayoumy K. The less harmful cigarette: a controversial issue. a tribute to Ernst L. Wynder. *Chem Res Toxicol* 2001;14:767-90.

139. Kensler CJ, Battista SP. Components of Cigarette Smoke with Ciliary-Depressant Activity. Their Selective Removal by Filters Containing Activated Charcoal Granules. *N Engl J Med* 1963;269:1161-6.

140. Dalhamn T, Rylander R. Cigarette smoke and ciliastasis. Effect of varying composition of smoke. *Arch Environ Health* 1966;13:47-50.

141. Fisher B. Filtering new technology. *Tobacco Reporter* 2000;127:46-47.

142. John AL. Japan: always something new. *Tobacco International*; 1996.

143. Howard G, Wagenknecht LE, Burke GL et al. Cigarette smoking and progression of atherosclerosis: The Atherosclerosis Risk in Communities (ARIC) Study. *Jama* 1998;279:119-24.

144. Hrubec Z, McLaughlin JK. Former cigarette smoking and mortality among US veterans: a 26-year follow-up, 1954-1980. In: Burns DM, Garfinkel L, Samet JM, eds. *Changes in cigarette-related disease risks and their implication for prevention and control. Smoking and Tobacco Control Monograph 8. National Institutes of Health, National Cancer Institute*. Bethesda (MD): U.S. Government Printing Office; 1997:501-30.

145. Lee J, Heng D, Chia KS, Chew SK, Tan BY, Hughes K. Risk factors and incident coronary heart disease in Chinese, Malay and Asian Indian males: the Singapore Cardiovascular Cohort Study. *Int J Epidemiol* 2001;30:983-988.

146. Wilson PW, Anderson KM, Castelli WP. Twelve-year incidence of coronary heart disease in middle-aged adults during the era of hypertensive therapy: the Framingham offspring study. *Am J Med* 1991;90:11-6.

147. Neaton JD, Wentworth D. Serum cholesterol, blood pressure, cigarette smoking, and death from coronary heart disease. Overall findings and differences by age for 316,099 white men. Multiple Risk Factor Intervention Trial Research Group. *Arch Intern Med* 1992;152:56-64.

148. Kleinman JC, Feldman JJ, Monk MA. Trends in Smoking and Ischemic Heart Disease Mortality. In: Havlik RJ, Feinleib M, eds. *Proceedings of the Conference on the Decline in Coronary Heart Disease Mortality*. Bethesda, MD: NIH; 1978:196-211.

149. Kannel WB. Update on the role of cigarette smoking in coronary artery disease. *Am Heart J* 1981;101:319-328.

150. Harris JE. Cigarette smoking and the decline in coronary heart disease mortality. In: Havlik RJ, Feinleib M, eds. *Proceedings of the Conference on the Decline in Coronary Heart Disease Mortality*. Bethesda, MD: NIH; 1976:212-214.

151. Seltzer CC. Framingham study data and "established wisdom" about cigarette smoking and coronary heart disease. *J Clin Epidemiol* 1989;42:743-50.

152. Scott DA, Palmer RM, Stapleton JA. Dose-years as an improved index of cumulative tobacco smoke exposure. *Med Hypotheses* 2001;56:735-6.

153. Joseph AM, Hecht SS, Murphy SE et al. Relationships between cigarette consumption and biomarkers of tobacco toxin exposure. *Cancer Epidemiol Biomarkers Prev* 2005;14:2963-8.

154. Doll R, Peto R, Boreham J, Sutherland I. Mortality from cancer in relation to smoking: 50 years observations on British doctors. *Br J Cancer* 2005;92:426-9.

155. Cook DG, Shaper AG, Pocock SJ, Kussick SJ. Giving up smoking and the risk of heart attacks. A report from The British Regional Heart Study. *Lancet* 1986;2:1376-80.

156. Critchley JA, Capewell S. Mortality risk reduction associated with smoking cessation in patients with coronary heart disease: a systematic review. *Jama* 2003;290:86-97.

157. Wei Q, Cheng L, Hong WK, Spitz MR. Reduced DNA repair capacity in lung cancer patients. *Cancer Res* 1996;56:4103-7.

158. Helzlsouer KJ, Harris EL, Parshad R, Perry HR, Price FM, Sanford KK. DNA repair proficiency: potential susceptiblity factor for breast cancer. *J Natl Cancer Inst* 1996;88:754-5.

159. Mohrenweiser HW, Jones IM. Variation in DNA repair is a factor in cancer susceptibility: a paradigm for the promises and perils of individual and population risk estimation? *Mutat Res* 1998;400:15-24.

160. Lu T, Pan Y, Kao SY et al. Gene regulation and DNA damage in the ageing human brain. *Nature* 2004;429:883-91.

161. Suh Y, Vijg J. Maintaining genetic integrity in aging: a zero sum game. *Antioxid Redox Signal* 2006;8:559-71.

162. Ames BN, Shigenaga MK, Hagen TM. Oxidants, antioxidants, and the degenerative diseases of aging. *Proc Natl Acad Sci U S A* 1993;90:7915-22.

163. Cadenas E. Biochemistry of oxygen toxicity. *Annu Rev Biochem* 1989;58:79-110.

164. DiGuiseppi J, Fridovich I. The toxicology of molecular oxygen. *Crit Rev Toxicol* 1984;12:315-42.

165. Esterbauer H, Eckl P, Ortner A. Possible mutagens derived from lipids and lipid precursors. *Mutat Res* 1990;238:223-33.

166. Joenje H. Genetic toxicology of oxygen. *Mutat Res* 1989;219:193-208.

167. Smart DJ, Chipman JK, Hodges NJ. Activity of OGG1 variants in the repair of pro-oxidant-induced 8-oxo-2'-deoxyguanosine. *DNA Repair (Amst)* 2006.

168. Frei B, England L, Ames BN. Ascorbate is an outstanding antioxidant in human blood plasma. *Proc Natl Acad Sci U S A* 1989;86:6377-81.

169. Halliwell B, Aruoma OI. DNA damage by oxygen-derived species. Its mechanism and measurement in mammalian systems. *FEBS Lett* 1991;281:9-19.

170. Halliwell B, Gutteridge JM. The antioxidants of human extracellular fluids. *Arch Biochem Biophys* 1990;280:1-8.

171. Hayes JD, Pulford DJ. The glutathione S-transferase supergene family: regulation of GST and the contribution of the isoenzymes to cancer chemoprotection and drug resistance. *Crit Rev Biochem Mol Biol* 1995;30:445-600.

172. Cooke MS, Evans MD, Dizdaroglu M, Lunec J. Oxidative DNA damage: mechanisms, mutation, and disease. *Faseb J* 2003;17:1195-214.

173. Frosina G. Overexpression of enzymes that repair endogenous damage to DNA. *Eur J Biochem* 2000;267:2135-49.

174. Kopplin A, Eberle-Adamkiewicz G, Glusenkamp KH, Nehls P, Kirstein U. Urinary excretion of 3-methyladenine and 3-ethyladenine after controlled exposure to tobacco smoke. *Carcinogenesis* 1995;16:2637-41.

175. Lindahl T. Instability and decay of the primary structure of DNA. *Nature* 1993;362:709-15.

176. Nakamura J, Swenberg JA. Endogenous apurinic/apyrimidinic sites in genomic DNA of mammalian tissues. *Cancer Res* 1999;59:2522-6.

177. Schuller HM. Mechanisms of smoking-related lung and pancreatic adenocarcinoma development. *Nat Rev Cancer* 2002;2:455-63.

178. Hecht SS. Biochemistry, biology, and carcinogenicity of tobacco-specific N-nitrosamines. *Chem Res Toxicol* 1998;11:559-603.

179. Jin Z, May WS, Gao F, Flagg T, Deng X. Bcl2 suppresses DNA repair by enhancing c-Myc transcriptional activity. *J Biol Chem* 2006;281:14446-56.

180. Yu Z, Chen J, Ford BN, Brackley ME, Glickman BW. Human DNA repair systems: an overview. *Environ Mol Mutagen* 1999;33:3-20.

181. Shigenaga MK, Gimeno CJ, Ames BN. Urinary 8-hydroxy-2'-deoxyguanosine as a biological marker of in vivo oxidative DNA damage. *Proc Natl Acad Sci U S A* 1989;86:9697-701.

182. Kuchino Y, Mori F, Kasai H et al. Misreading of DNA templates containing 8-hydroxydeoxyguanosine at the modified base and at adjacent residues. *Nature* 1987;327:77-9.

183. Floyd RA. The role of 8-hydroxyguanine in carcinogenesis. *Carcinogenesis* 1990;11:1447-50.

184. Goode EL, Ulrich CM, Potter JD. Polymorphisms in DNA repair genes and associations with cancer risk. *Cancer Epidemiol Biomarkers Prev* 2002;11:1513-30.

185. Gupta RC, Reddy MV, Randerath K. 32P-postlabeling analysis of non-radioactive aromatic carcinogen--DNA adducts. *Carcinogenesis* 1982;3:1081-92.

186. Loft S, Astrup A, Buemann B, Poulsen HE. Oxidative DNA damage correlates with oxygen consumption in humans. *Faseb J* 1994;8:534-7.

187. Loft S, Vistisen K, Ewertz M, Tjonneland A, Overvad K, Poulsen HE. Oxidative DNA damage estimated by 8-hydroxydeoxyguanosine excretion in humans: influence of smoking, gender and body mass index. *Carcinogenesis* 1992;13:2241-7.

188. Tagesson C, Kallberg M, Klintenberg C, Starkhammar H. Determination of urinary 8-hydroxydeoxyguanosine by automated coupled-column high performance liquid chromatography: a powerful technique for assaying in vivo oxidative DNA damage in cancer patients. *Eur J Cancer* 1995;31A:934-40.

189. Squire JA, Whitmore GF, Phillips RA. A Genetic Basis of Cancer. In: Tannock IF, Hill RP, eds. *The Basic Science of Oncology*. 4 ed: McGraw-Hill Companies, Inc.; 1998:48-78.

190. Neumann AS, Sturgis EM, Wei Q. Nucleotide excision repair as a marker for susceptibility to tobacco-related cancers: a review of molecular epidemiological studies. *Mol Carcinog* 2005;42:65-92.

191. Chen RH, Maher VM, Brouwer J, van de Putte P, McCormick JJ. Preferential repair and strand-specific repair of benzo[a]pyrene diol epoxide adducts in the HPRT gene of diploid human fibroblasts. *Proc Natl Acad Sci U S A* 1992;89:5413-7.

192. Xiong P, Bondy ML, Li D et al. Sensitivity to benzo(a)pyrene diol-epoxide associated with risk of breast cancer in young women and modulation by glutathione S-transferase polymorphisms: a case-control study. *Cancer Res* 2001;61:8465-9.

193. Kuschel B, Auranen A, McBride S et al. Variants in DNA double-strand break repair genes and breast cancer susceptibility. *Hum Mol Genet* 2002;11:1399-407.

194. Khanna KK, Jackson SP. DNA double-strand breaks: signaling, repair and the cancer connection. *Nat Genet* 2001;27:247-54.

195. Luo LZ, Werner KM, Gollin SM, Saunders WS. Cigarette smoke induces anaphase bridges and genomic imbalances in normal cells. *Mutat Res* 2004;554:375-85.

196. Wood RD, Mitchell M, Sgouros J, Lindahl T. Human DNA repair genes. *Science* 2001;291:1284-9.

197. Wu X, Gu J, Amos CI, Jiang H, Hong WK, Spitz MR. A parallel study of in vitro sensitivity to benzo[a]pyrene diol epoxide and bleomycin in lung carcinoma cases and controls. *Cancer* 1998;83:1118-27.

198. Wu X, Gu J, Hong WK et al. Benzo[a]pyrene diol epoxide and bleomycin sensitivity and susceptibility to cancer of upper aerodigestive tract. *J Natl Cancer Inst* 1998;90:1393-9.

199. Cloos J, Nieuwenhuis EJ, Boomsma DI et al. Inherited susceptibility to bleomycin-induced chromatid breaks in cultured peripheral blood lymphocytes. *J Natl Cancer Inst* 1999;91:1125-30.

200. Miller RD, Kwok PY. The birth and death of human single-nucleotide polymorphisms: new experimental evidence and implications for human history and medicine. *Hum Mol Genet* 2001;10:2195-8.

201. Taylor JG, Choi EH, Foster CB, Chanock SJ. Using genetic variation to study human disease. *Trends Mol Med* 2001;7:507-12.

202. Gray IC, Campbell DA, Spurr NK. Single nucleotide polymorphisms as tools in human genetics. *Hum Mol Genet* 2000;9:2403-8.

203. Shastry BS. SNP alleles in human disease and evolution. *J Hum Genet* 2002;47:561-6.

204. Wu X, Zhao H, Suk R, Christiani DC. Genetic susceptibility to tobacco-related cancer. *Oncogene* 2004;23:6500-23.

205. Eisen JA, Hanawalt PC. A phylogenomic study of DNA repair genes, proteins, and processes. *Mutat Res* 1999;435:171-213.

206. Lu AL, Li X, Gu Y, Wright PM, Chang DY. Repair of oxidative DNA damage: mechanisms and functions. *Cell Biochem Biophys* 2001;35:141-70.

207. Lindahl T, Wood RD. Quality control by DNA repair. *Science* 1999;286:1897-905.

208. Krokan HE, Nilsen H, Skorpen F, Otterlei M, Slupphaug G. Base excision repair of DNA in mammalian cells. *FEBS Lett* 2000;476:73-7.

209. Kunsch C, Medford RM. Oxidative stress as a regulator of gene expression in the vasculature. *Circ Res* 1999;85:753-66.

210. Irani K. Oxidant signaling in vascular cell growth, death, and survival : a review of the roles of reactive oxygen species in smooth muscle and endothelial cell mitogenic and apoptotic signaling. *Circ Res* 2000;87:179-83.

211. Martinet W, Knaapen MW, De Meyer GR, Herman AG, Kockx MM. Elevated levels of oxidative DNA damage and DNA repair enzymes in human atherosclerotic plaques. *Circulation* 2002;106:927-32.

212. Boiteux S, Radicella JP. The human OGG1 gene: structure, functions, and its implication in the process of carcinogenesis. *Arch Biochem Biophys* 2000;377:1-8.

213. Shinmura K, Kohno T, Takeuchi-Sasaki M et al. Expression of the OGG1-type 1a (nuclear form) protein in cancerous and non-cancerous human cells. *Int J Oncol* 2000;16:701-7.

214. Nishioka K, Ohtsubo T, Oda H et al. Expression and differential intracellular localization of two major forms of human 8-oxoguanine DNA glycosylase encoded by alternatively spliced OGG1 mRNAs. *Mol Biol Cell* 1999;10:1637-52.

215. Shinmura K, Yokota J. The OGG1 gene encodes a repair enzyme for oxidatively damaged DNA and is involved in human carcinogenesis. *Antioxid Redox Signal* 2001;3:597-609.

216. Nishimura S. Involvement of mammalian OGG1(MMH) in excision of the 8-hydroxyguanine residue in DNA. *Free Radic Biol Med* 2002;32:813-21.

217. Grollman AP, Moriya M. Mutagenesis by 8-oxoguanine: an enemy within. *Trends Genet* 1993;9:246-9.

218. Fearon ER. Human cancer syndromes: clues to the origin and nature of cancer. *Science* 1997;278:1043-50.

219. Kohno T, Shinmura K, Tosaka M et al. Genetic polymorphisms and alternative splicing of the hOGG1 gene, that is involved in the repair of 8-hydroxyguanine in damaged DNA. *Oncogene* 1998;16:3219-25.

220. Michaels ML, Miller JH. The GO system protects organisms from the mutagenic effect of the spontaneous lesion 8-hydroxyguanine (7,8-dihydro-8-oxoguanine). *J Bacteriol* 1992;174:6321-5.

221. Thomas D, Scot AD, Barbey R, Padula M, Boiteux S. Inactivation of OGG1 increases the incidence of G . C-->T . A transversions in Saccharomyces cerevisiae: evidence for endogenous oxidative damage to DNA in eukaryotic cells. *Mol Gen Genet* 1997;254:171-8.

222. Inoue M, Osaki T, Noguchi M, Hirohashi S, Yasumoto K, Kasai H. Lung cancer patients have increased 8-hydroxydeoxyguanosine levels in peripheral lung tissue DNA. *Jpn J Cancer Res* 1998;89:691-5.

223. Jaiswal M, LaRusso NF, Nishioka N, Nakabeppu Y, Gores GJ. Human Ogg1, a protein involved in the repair of 8-oxoguanine, is inhibited by nitric oxide. *Cancer Res* 2001;61:6388-93.

224. Beckman JS, Beckman TW, Chen J, Marshall PA, Freeman BA. Apparent hydroxyl radical production by peroxynitrite: implications for endothelial injury from nitric oxide and superoxide. *Proc Natl Acad Sci U S A* 1990;87:1620-4.

225. Janssen K, Schlink K, Gotte W, Hippler B, Kaina B, Oesch F. DNA repair activity of 8-oxoguanine DNA glycosylase 1 (OGG1) in human lymphocytes is not dependent on genetic polymorphism Ser326/Cys326. *Mutat Res* 2001;486:207-16.

226. Chevillard S, Radicella JP, Levalois C et al. Mutations in OGG1, a gene involved in the repair of oxidative DNA damage, are found in human lung and kidney tumours. *Oncogene* 1998;16:3083-6.

227. Sunaga N, Kohno T, Yanagitani N et al. Contribution of the NQO1 and GSTT1 polymorphisms to lung adenocarcinoma susceptibility. *Cancer Epidemiol Biomarkers Prev* 2002;11:730-8.

228. Bruce CR, Carey AL, Hawley JA, Febbraio MA. Intramuscular heat shock protein 72 and heme oxygenase-1 mRNA are reduced in patients with type 2 diabetes: evidence that insulin resistance is associated with a disturbed antioxidant defense mechanism. *Diabetes* 2003;52:2338-45.

229. Ceriello A, Quagliaro L, Piconi L et al. Effect of postprandial hypertriglyceridemia and hyperglycemia on circulating adhesion molecules and oxidative stress generation and the possible role of simvastatin treatment. *Diabetes* 2004;53:701-10.

230. Evans JL, Goldfine ID, Maddux BA, Grodsky GM. Are oxidative stress-activated signaling pathways mediators of insulin resistance and beta-cell dysfunction? *Diabetes* 2003;52:1-8.

231. Talior I, Yarkoni M, Bashan N, Eldar-Finkelman H. Increased glucose uptake promotes oxidative stress and PKC-delta activation in adipocytes of obese, insulin-resistant mice. *Am J Physiol Endocrinol Metab* 2003;285:E295-302.

232. Wang CL, Hsieh MC, Hsin SC et al. The hOGG1 Ser326Cys gene polymorphism is associated with decreased insulin sensitivity in subjects with normal glucose tolerance. *J Hum Genet* 2006;51:124-8.

233. Chen L, Elahi A, Pow-Sang J, Lazarus P, Park J. Association between polymorphism of human oxoguanine glycosylase 1 and risk of prostate cancer. *J Urol* 2003;170:2471-4.

234. Cai Q, Shu XO, Wen W et al. Functional Ser326Cys polymorphism in the hOGG1 gene is not associated with breast cancer risk. *Cancer Epidemiol Biomarkers Prev* 2006;15:403-4.

235. Choi JY, Hamajima N, Tajima K et al. hOGG1 Ser326Cys polymorphism and breast cancer risk among Asian women. *Breast Cancer Res Treat* 2003;79:59-62.

236. Vogel U, Nexo BA, Olsen A et al. No association between OGG1 Ser326Cys polymorphism and breast cancer risk. *Cancer Epidemiol Biomarkers Prev* 2003;12:170-1.

237. Kim JI, Park YJ, Kim KH et al. hOGG1 Ser326Cys polymorphism modifies the significance of the environmental risk factor for colon cancer. *World J Gastroenterol* 2003;9:956-60.

238. Le Marchand L, Donlon T, Lum-Jones A, Seifried A, Wilkens LR. Association of the hOGG1 Ser326Cys polymorphism with lung cancer risk. *Cancer Epidemiol Biomarkers Prev* 2002;11:409-12.

239. Liang G, Pu Y, Yin L. Rapid detection of single nucleotide polymorphisms related with lung cancer susceptibility of Chinese population. *Cancer Lett* 2005;223:265-74.

240. Sugimura H, Kohno T, Wakai K et al. hOGG1 Ser326Cys polymorphism and lung cancer susceptibility. *Cancer Epidemiol Biomarkers Prev* 1999;8:669-74.

241. Hanaoka T, Sugimura H, Nagura K et al. hOGG1 exon7 polymorphism and gastric cancer in case-control studies of Japanese Brazilians and non-Japanese Brazilians. *Cancer Lett* 2001;170:53-61.

242. Takezaki T, Gao CM, Wu JZ et al. hOGG1 Ser(326)Cys polymorphism and modification by environmental factors of stomach cancer risk in Chinese. *Int J Cancer* 2002;99:624-7.

243. Cho EY, Hildesheim A, Chen CJ et al. Nasopharyngeal carcinoma and genetic polymorphisms of DNA repair enzymes XRCC1 and hOGG1. *Cancer Epidemiol Biomarkers Prev* 2003;12:1100-4.

244. Elahi A, Zheng Z, Park J, Eyring K, McCaffrey T, Lazarus P. The human OGG1 DNA repair enzyme and its association with orolaryngeal cancer risk. *Carcinogenesis* 2002;23:1229-34.

245. Xing DY, Tan W, Song N, Lin DX. Ser326Cys polymorphism in hOGG1 gene and risk of esophageal cancer in a Chinese population. *Int J Cancer* 2001;95:140-3.

246. Zhang Z, Shi Q, Wang LE et al. No Association between hOGG1 Ser326Cys polymorphism and risk of squamous cell carcinoma of the head and neck. *Cancer Epidemiol Biomarkers Prev* 2004;13:1081-3.

247. Nam RK, Zhang WW, Jewett MA et al. The use of genetic markers to determine risk for prostate cancer at prostate biopsy. *Clin Cancer Res* 2005;11:8391-7.

248. Xu J, Zheng SL, Turner A et al. Associations between hOGG1 sequence variants and prostate cancer susceptibility. *Cancer Res* 2002;62:2253-7.

249. Kau HC, Tsai CC, Hsu WM, Liu JH, Wei YH. Genetic polymorphism of hOGG1 and risk of pterygium in Chinese. *Eye* 2004;18:635-9.

250. Olshan AF, Shaw GM, Millikan RC, Laurent C, Finnell RH. Polymorphisms in DNA repair genes as risk factors for spina bifida and orofacial clefts. *Am J Med Genet A* 2005;135:268-73.

251. Wikman H, Risch A, Klimek F et al. hOGG1 polymorphism and loss of heterozygosity (LOH): significance for lung cancer susceptibility in a caucasian population. *Int J Cancer* 2000;88:932-7.

252. Parsons JL, Dianova, II, Dianov GL. APE1 is the major 3'-phosphoglycolate activity in human cell extracts. *Nucleic Acids Res* 2004;32:3531-6.

253. Chaudhry MA, Dedon PC, Wilson DM, 3rd, Demple B, Weinfeld M. Removal by human apurinic/apyrimidinic endonuclease 1 (Ape 1) and Escherichia coli exonuclease III of 3'-phosphoglycolates from DNA treated with neocarzinostatin, calicheamicin, and gamma-radiation. *Biochem Pharmacol* 1999;57:531-8.

254. Demple B, Harrison L. Repair of oxidative damage to DNA: enzymology and biology. *Annu Rev Biochem* 1994;63:915-48.

255. Xanthoudakis S, Miao GG, Curran T. The redox and DNA-repair activities of Ref-1 are encoded by nonoverlapping domains. *Proc Natl Acad Sci U S A* 1994;91:23-7.

256. Walker LJ, Robson CN, Black E, Gillespie D, Hickson ID. Identification of residues in the human DNA repair enzyme HAP1 (Ref-1) that are essential for redox regulation of Jun DNA binding. *Mol Cell Biol* 1993;13:5370-6.

257. Wiederhold L, Leppard JB, Kedar P et al. AP endonuclease-independent DNA base excision repair in human cells. *Mol Cell* 2004;15:209-20.

258. Xanthoudakis S, Miao G, Wang F, Pan YC, Curran T. Redox activation of Fos-Jun DNA binding activity is mediated by a DNA repair enzyme. *Embo J* 1992;11:3323-35.

259. Abate C, Patel L, Rauscher FJ, 3rd, Curran T. Redox regulation of fos and jun DNA-binding activity in vitro. *Science* 1990;249:1157-61.

260. Demple B, Halbrook J, Linn S. Escherichia coli xth mutants are hypersensitive to hydrogen peroxide. *J Bacteriol* 1983;153:1079-82.

261. Sammartano LJ, Tuveson RW. Escherichia coli xthA mutants are sensitive to inactivation by broad-spectrum near-UV (300- to 400-nm) radiation. *J Bacteriol* 1983;156:904-6.

262. Ramotar D, Popoff SC, Gralla EB, Demple B. Cellular role of yeast Apn1 apurinic endonuclease/3'-diesterase: repair of oxidative and alkylation DNA damage and control of spontaneous mutation. *Mol Cell Biol* 1991;11:4537-44.

263. Walker LJ, Craig RB, Harris AL, Hickson ID. A role for the human DNA repair enzyme HAP1 in cellular protection against DNA damaging agents and hypoxic stress. *Nucleic Acids Res* 1994;22:4884-9.

264. Grosch S, Fritz G, Kaina B. Apurinic endonuclease (Ref-1) is induced in mammalian cells by oxidative stress and involved in clastogenic adaptation. *Cancer Res* 1998;58:4410-6.

265. Fung H, Demple B. A vital role for Ape1/Ref1 protein in repairing spontaneous DNA damage in human cells. *Mol Cell* 2005;17:463-70.

266. Izumi T, Brown DB, Naidu CV et al. Two essential but distinct functions of the mammalian abasic endonuclease. *Proc Natl Acad Sci U S A* 2005;102:5739-43.

267. Ramana CV, Boldogh I, Izumi T, Mitra S. Activation of apurinic/apyrimidinic endonuclease in human cells by reactive oxygen species and its correlation with their adaptive response to genotoxicity of free radicals. *Proc Natl Acad Sci U S A* 1998;95:5061-6.

268. Ludwig DL, MacInnes MA, Takiguchi Y et al. A murine AP-endonuclease gene-targeted deficiency with post-implantation embryonic progression and ionizing radiation sensitivity. *Mutat Res* 1998;409:17-29.

269. Meira LB, Devaraj S, Kisby GE et al. Heterozygosity for the mouse Apex gene results in phenotypes associated with oxidative stress. *Cancer Res* 2001;61:5552-7.

270. Xanthoudakis S, Smeyne RJ, Wallace JD, Curran T. The redox/DNA repair protein, Ref-1, is essential for early embryonic development in mice. *Proc Natl Acad Sci U S A* 1996;93:8919-23.

271. Vidal AE, Boiteux S, Hickson ID, Radicella JP. XRCC1 coordinates the initial and late stages of DNA abasic site repair through protein-protein interactions. *Embo J* 2001;20:6530-9.

272. Sossou M, Flohr-Beckhaus C, Schulz I, Daboussi F, Epe B, Radicella JP. APE1 overexpression in XRCC1-deficient cells complements the defective repair of oxidative single strand breaks but increases genomic instability. *Nucleic Acids Res* 2005;33:298-306.

273. Martinet W, Knaapen MW, De Meyer GR, Herman AG, Kockx MM. Oxidative DNA damage and repair in experimental atherosclerosis are reversed by dietary lipid lowering. *Circ Res* 2001;88:733-9.

274. Jeon BH, Gupta G, Park YC et al. Apurinic/apyrimidinic endonuclease 1 regulates endothelial NO production and vascular tone. *Circ Res* 2004;95:902-10.

275. Izumi T, Mitra S. Deletion analysis of human AP-endonuclease: minimum sequence required for the endonuclease activity. *Carcinogenesis* 1998;19:525-7.

276. Hu JJ, Smith TR, Miller MS, Mohrenweiser HW, Golden A, Case LD. Amino acid substitution variants of APE1 and XRCC1 genes associated with ionizing radiation sensitivity. *Carcinogenesis* 2001;22:917-22.

277. Hadi MZ, Coleman MA, Fidelis K, Mohrenweiser HW, Wilson DM, 3rd. Functional characterization of Ape1 variants identified in the human population. *Nucleic Acids Res* 2000;28:3871-9.

278. Hsieh MM, Hegde V, Kelley MR, Deutsch WA. Activation of APE/Ref-1 redox activity is mediated by reactive oxygen species and PKC phosphorylation. *Nucleic Acids Res* 2001;29:3116-22.

279. Misra RR, Ratnasinghe D, Tangrea JA et al. Polymorphisms in the DNA repair genes XPD, XRCC1, XRCC3, and APE/ref-1, and the risk of lung cancer among male smokers in Finland. *Cancer Lett* 2003;191:171-8.

280. Ito H, Matsuo K, Hamajima N et al. Gene-environment interactions between the smoking habit and polymorphisms in the DNA repair genes, APE1 Asp148Glu and XRCC1 Arg399Gln, in Japanese lung cancer risk. *Carcinogenesis* 2004;25:1395-401.

281. Popanda O, Schattenberg T, Phong CT et al. Specific combinations of DNA repair gene variants and increased risk for non-small cell lung cancer. *Carcinogenesis* 2004;25:2433-41.

282. Zienolddiny S, Campa D, Lind H et al. Polymorphisms of DNA repair genes and risk of non-small cell lung cancer. *Carcinogenesis* 2006;27:560-7.

283. Hao B, Wang H, Zhou K et al. Identification of genetic variants in base excision repair pathway and their associations with risk of esophageal squamous cell carcinoma. *Cancer Res* 2004;64:4378-84.

284. Zhang Y, Newcomb PA, Egan KM et al. Genetic polymorphisms in base-excision repair pathway genes and risk of breast cancer. *Cancer Epidemiol Biomarkers Prev* 2006;15:353-8.

285. Jiao L, Bondy ML, Hassan MM et al. Selected polymorphisms of DNA repair genes and risk of pancreatic cancer. *Cancer Detect Prev* 2006;30:284-91.

286. Matullo G, Dunning AM, Guarrera S et al. DNA repair polymorphisms and cancer risk in non-smokers in a cohort study. *Carcinogenesis* 2006;27:997-1007.

287. Terry PD, Umbach DM, Taylor JA. APE1 genotype and risk of bladder cancer: evidence for effect modification by smoking. *Int J Cancer* 2006;118:3170-3.

288. Cappelli E, Taylor R, Cevasco M, Abbondandolo A, Caldecott K, Frosina G. Involvement of XRCC1 and DNA ligase III gene products in DNA base excision repair. *J Biol Chem* 1997;272:23970-5.

289. Thompson LH, Brookman KW, Jones NJ, Allen SA, Carrano AV. Molecular cloning of the human XRCC1 gene, which corrects defective DNA strand break repair and sister chromatid exchange. *Mol Cell Biol* 1990;10:6160-71.

290. Marintchev A, Mullen MA, Maciejewski MW, Pan B, Gryk MR, Mullen GP. Solution structure of the single-strand break repair protein XRCC1 N-terminal domain. *Nat Struct Biol* 1999;6:884-93.

291. Caldecott KW, Aoufouchi S, Johnson P, Shall S. XRCC1 polypeptide interacts with DNA polymerase beta and possibly poly (ADP-ribose) polymerase, and DNA ligase III is a novel molecular 'nick-sensor' in vitro. *Nucleic Acids Res* 1996;24:4387-94.

292. Nash RA, Caldecott KW, Barnes DE, Lindahl T. XRCC1 protein interacts with one of two distinct forms of DNA ligase III. *Biochemistry* 1997;36:5207-11.

293. Masson M, Niedergang C, Schreiber V, Muller S, Menissier-de Murcia J, de Murcia G. XRCC1 is specifically associated with poly(ADP-ribose) polymerase and negatively regulates its activity following DNA damage. *Mol Cell Biol* 1998;18:3563-71.

294. Caldecott KW. XRCC1 and DNA strand break repair. *DNA Repair (Amst)* 2003;2:955-69.

295. Thompson LH, West MG. XRCC1 keeps DNA from getting stranded. *Mutat Res* 2000;459:1-18.

296. Moore DJ, Taylor RM, Clements P, Caldecott KW. Mutation of a BRCT domain selectively disrupts DNA single-strand break repair in noncycling Chinese hamster ovary cells. *Proc Natl Acad Sci U S A* 2000;97:13649-54.

297. Lee JM, Lee YC, Yang SY et al. Genetic polymorphisms of XRCC1 and risk of the esophageal cancer. *Int J Cancer* 2001;95:240-6.

298. Thompson LH, Bachinski LL, Stallings RL et al. Complementation of repair gene mutations on the hemizygous chromosome 9 in CHO: a third repair gene on human chromosome 19. *Genomics* 1989;5:670-9.

299. Dillehay LE, Thompson LH, Carrano AV. DNA-strand breaks associated with halogenated pyrimidine incorporation. *Mutat Res* 1984;131:129-36.

300. Fan R, Kumaravel TS, Jalali F, Marrano P, Squire JA, Bristow RG. Defective DNA strand break repair after DNA damage in prostate cancer cells: implications for genetic instability and prostate cancer progression. *Cancer Res* 2004;64:8526-33.

301. Rossi ML, Marziliano N, Merlini PA et al. Different quantitative apoptotic traits in coronary atherosclerotic plaques from patients with stable angina pectoris and acute coronary syndromes. *Circulation* 2004;110:1767-73.

302. Kapalla M, Yeghiazaryan K, Hricova M et al. Combined analysis of biochemical parameters in serum and differential gene expression in circulating leukocytes may serve as an ex vivo monitoring system to estimate risk factors for complications in Diabetes mellitus. *Amino Acids* 2005;28:221-7.

303. Qu T, Morii E, Oboki K, Lu Y, Morimoto K. Micronuclei in EM9 cells expressing polymorphic forms of human XRCC1. *Cancer Lett* 2005;221:91-5.

304. Savas S, Kim DY, Ahmad MF, Shariff M, Ozcelik H. Identifying functional genetic variants in DNA repair pathway using protein conservation analysis. *Cancer Epidemiol Biomarkers Prev* 2004;13:801-7.

305. Takanami T, Nakamura J, Kubota Y, Horiuchi S. The Arg280His polymorphism in X-ray repair cross-complementing gene 1 impairs DNA repair ability. *Mutat Res* 2005;582:135-45.

306. Abdel-Rahman SZ, El-Zein RA. The 399Gln polymorphism in the DNA repair gene XRCC1 modulates the genotoxic response induced in human lymphocytes by the tobacco-specific nitrosamine NNK. *Cancer Lett* 2000;159:63-71.

307. Cornetta T, Festa F, Testa A, Cozzi R. DNA damage repair and genetic polymorphisms: Assessment of individual sensitivity and repair capacity. *Int J Radiat Oncol Biol Phys* 2006;66:537-45.

308. Wang Y, Spitz MR, Zhu Y, Dong Q, Shete S, Wu X. From genotype to phenotype: correlating XRCC1 polymorphisms with mutagen sensitivity. *DNA Repair (Amst)* 2003;2:901-8.

309. Pachkowski BF, Winkel S, Kubota Y, Swenberg JA, Millikan RC, Nakamura J. XRCC1 genotype and breast cancer: functional studies and epidemiologic data show interactions between XRCC1 codon 280 His and smoking. *Cancer Res* 2006;66:2860-8.

310. Ladiges WC. Mouse models of XRCC1 DNA repair polymorphisms and cancer. *Oncogene* 2006;25:1612-9.

311. Duell EJ, Wiencke JK, Cheng TJ et al. Polymorphisms in the DNA repair genes XRCC1 and ERCC2 and biomarkers of DNA damage in human blood mononuclear cells. *Carcinogenesis* 2000;21:965-71.

312. Lei YC, Hwang SJ, Chang CC et al. Effects on sister chromatid exchange frequency of polymorphisms in DNA repair gene XRCC1 in smokers. *Mutat Res* 2002;519:93-101.

313. Hung RJ, Hall J, Brennan P, Boffetta P. Genetic Polymorphisms in the Base Excision Repair Pathway and Cancer Risk: A HuGE Review. *Am J Epidemiol* 2005.

314. Shen M, Berndt SI, Rothman N et al. Polymorphisms in the DNA nucleotide excision repair genes and lung cancer risk in Xuan Wei, China. *Int J Cancer* 2005;116:768-73.

315. Kim SU, Park SK, Yoo KY et al. XRCC1 genetic polymorphism and breast cancer risk. *Pharmacogenetics* 2002;12:335-8.

316. Hung RJ, Brennan P, Canzian F et al. Large-scale investigation of base excision repair genetic polymorphisms and lung cancer risk in a multicenter study. *J Natl Cancer Inst* 2005;97:567-76.

317. Matullo G, Guarrera S, Sacerdote C et al. Polymorphisms/haplotypes in DNA repair genes and smoking: a bladder cancer case-control study. *Cancer Epidemiol Biomarkers Prev* 2005;14:2569-78.

318. Stern MC, Umbach DM, van Gils CH, Lunn RM, Taylor JA. DNA repair gene XRCC1 polymorphisms, smoking, and bladder cancer risk. *Cancer Epidemiol Biomarkers Prev* 2001;10:125-31.

319. Chacko P, Rajan B, Joseph T, Mathew BS, Pillai MR. Polymorphisms in DNA repair gene XRCC1 and increased genetic susceptibility to breast cancer. *Breast Cancer Res Treat* 2005;89:15-21.

320. Patel AV, Calle EE, Pavluck AL, Feigelson HS, Thun MJ, Rodriguez C. A prospective study of XRCC1 (X-ray cross-complementing group 1) polymorphisms and breast cancer risk. *Breast Cancer Res* 2005;7:R1168-73.

321. Shen J, Gammon MD, Terry MB et al. Polymorphisms in XRCC1 modify the association between polycyclic aromatic hydrocarbon-DNA adducts, cigarette smoking, dietary antioxidants, and breast cancer risk. *Cancer Epidemiol Biomarkers Prev* 2005;14:336-42.

322. Smith TR, Levine EA, Perrier ND et al. DNA-repair genetic polymorphisms and breast cancer risk. *Cancer Epidemiol Biomarkers Prev* 2003;12:1200-4.

323. Thyagarajan B, Anderson KE, Folsom AR et al. No association between XRCC1 and XRCC3 gene polymorphisms and breast cancer risk: Iowa Women's Health Study. *Cancer Detect Prev* 2006;30:313-21.

324. Skjelbred CF, Saebo M, Wallin H et al. Polymorphisms of the XRCC1, XRCC3 and XPD genes and risk of colorectal adenoma and carcinoma, in a Norwegian cohort: a case control study. *BMC Cancer* 2006;6:67.

325. Stern MC, Siegmund KD, Corral R, Haile RW. XRCC1 and XRCC3 polymorphisms and their role as effect modifiers of unsaturated fatty acids and antioxidant intake on colorectal adenomas risk. *Cancer Epidemiol Biomarkers Prev* 2005;14:609-15.

326. Ratnasinghe LD, Abnet C, Qiao YL et al. Polymorphisms of XRCC1 and risk of esophageal and gastric cardia cancer. *Cancer Lett* 2004;216:157-64.

327. Xing D, Qi J, Miao X, Lu W, Tan W, Lin D. Polymorphisms of DNA repair genes XRCC1 and XPD and their associations with risk of esophageal squamous cell carcinoma in a Chinese population. *Int J Cancer* 2002;100:600-5.

328. Yu HP, Zhang XY, Wang XL et al. DNA repair gene XRCC1 polymorphisms, smoking, and esophageal cancer risk. *Cancer Detect Prev* 2004;28:194-9.

329. Lee SG, Kim B, Choi J, Kim C, Lee I, Song K. Genetic polymorphisms of XRCC1 and risk of gastric cancer. *Cancer Lett* 2002;187:53-60.

330. Shen H, Xu Y, Qian Y et al. Polymorphisms of the DNA repair gene XRCC1 and risk of gastric cancer in a Chinese population. *Int J Cancer* 2000;88:601-6.

331. Chen S, Tang D, Xue K et al. DNA repair gene XRCC1 and XPD polymorphisms and risk of lung cancer in a Chinese population. *Carcinogenesis* 2002;23:1321-5.

332. Ratnasinghe D, Yao SX, Tangrea JA et al. Polymorphisms of the DNA repair gene XRCC1 and lung cancer risk. *Cancer Epidemiol Biomarkers Prev* 2001;10:119-23.

333. Han J, Hankinson SE, Colditz GA, Hunter DJ. Genetic variation in XRCC1, sun exposure, and risk of skin cancer. *Br J Cancer* 2004;91:1604-9.

334. Cao Y, Miao XP, Huang MY et al. Polymorphisms of XRCC1 genes and risk of nasopharyngeal carcinoma in the Cantonese population. *BMC Cancer* 2006;6:167.

335. Hirata H, Hinoda Y, Matsuyama H et al. Polymorphisms of DNA repair genes are associated with renal cell carcinoma. *Biochem Biophys Res Commun* 2006;342:1058-62.

336. Demokan S, Demir D, Suoglu Y, Kiyak E, Akar U, Dalay N. Polymorphisms of the XRCC1 DNA repair gene in head and neck cancer. *Pathol Oncol Res* 2005;11:22-5.

337. Olshan AF, Watson MA, Weissler MC, Bell DA. XRCC1 polymorphisms and head and neck cancer. *Cancer Lett* 2002;178:181-6.

338. Sturgis EM, Castillo EJ, Li L et al. Polymorphisms of DNA repair gene XRCC1 in squamous cell carcinoma of the head and neck. *Carcinogenesis* 1999;20:2125-9.

339. Metsola K, Kataja V, Sillanpaa P et al. XRCC1 and XPD genetic polymorphisms, smoking and breast cancer risk in a Finnish case-control study. *Breast Cancer Res* 2005;7:R987-97.

340. Vogel U, Nexo BA, Wallin H, Overvad K, Tjonneland A, Raaschou-Nielsen O. No association between base excision repair gene polymorphisms and risk of lung cancer. *Biochem Genet* 2004;42:453-60.

341. Kelsey KT, Park S, Nelson HH, Karagas MR. A population-based case-control study of the XRCC1 Arg399Gln polymorphism and susceptibility to bladder cancer. *Cancer Epidemiol Biomarkers Prev* 2004;13:1337-41.

342. Sanyal S, Festa F, Sakano S et al. Polymorphisms in DNA repair and metabolic genes in bladder cancer. *Carcinogenesis* 2004;25:729-34.

343. Shen M, Hung RJ, Brennan P et al. Polymorphisms of the DNA repair genes XRCC1, XRCC3, XPD, interaction with environmental exposures, and bladder cancer risk in a case-control study in northern Italy. *Cancer Epidemiol Biomarkers Prev* 2003;12:1234-40.

344. Figueiredo JC, Knight JA, Briollais L, Andrulis IL, Ozcelik H. Polymorphisms XRCC1-R399Q and XRCC3-T241M and the risk of breast cancer at the Ontario site of the Breast Cancer Family Registry. *Cancer Epidemiol Biomarkers Prev* 2004;13:583-91.

345. Shu XO, Cai Q, Gao YT, Wen W, Jin F, Zheng W. A population-based case-control study of the Arg399Gln polymorphism in DNA repair gene XRCC1 and risk of breast cancer. *Cancer Epidemiol Biomarkers Prev* 2003;12:1462-7.

346. Cai L, You NC, Lu H et al. Dietary selenium intake, aldehyde dehydrogenase-2 and X-ray repair cross-complementing 1 genetic polymorphisms, and the risk of esophageal squamous cell carcinoma. *Cancer* 2006;106:2345-54.

347. Huang WY, Chow WH, Rothman N et al. Selected DNA repair polymorphisms and gastric cancer in Poland. *Carcinogenesis* 2005;26:1354-9.

348. Park JY, Park SH, Choi JE et al. Polymorphisms of the DNA repair gene xeroderma pigmentosum group A and risk of primary lung cancer. *Cancer Epidemiol Biomarkers Prev* 2002;11:993-7.

349. Zhang X, Miao X, Liang G et al. Polymorphisms in DNA base excision repair genes ADPRT and XRCC1 and risk of lung cancer. *Cancer Res* 2005;65:722-6.

350. Ritchey JD, Huang WY, Chokkalingam AP et al. Genetic variants of DNA repair genes and prostate cancer: a population-based study. *Cancer Epidemiol Biomarkers Prev* 2005;14:1703-9.

351. Cleaver JE. Defective repair replication of DNA in xeroderma pigmentosum. *Nature* 1968;218:652-6.

352. Kraemer KH, Lee MM, Scotto J. Xeroderma pigmentosum. Cutaneous, ocular, and neurologic abnormalities in 830 published cases. *Arch Dermatol* 1987;123:241-50.

353. Itin PH, Sarasin A, Pittelkow MR. Trichothiodystrophy: update on the sulfur-deficient brittle hair syndromes. *J Am Acad Dermatol* 2001;44:891-920; quiz 921-4.

354. Nakura J, Ye L, Morishima A, Kohara K, Miki T. Helicases and aging. *Cell Mol Life Sci* 2000;57:716-30.

355. Nance MA, Berry SA. Cockayne syndrome: review of 140 cases. *Am J Med Genet* 1992;42:68-84.

356. Rapin I, Lindenbaum Y, Dickson DW, Kraemer KH, Robbins JH. Cockayne syndrome and xeroderma pigmentosum. *Neurology* 2000;55:1442-9.

357. Friedberg EC. How nucleotide excision repair protects against cancer. *Nat Rev Cancer* 2001;1:22-33.

358. Hanawalt PC. Transcription-coupled repair and human disease. *Science* 1994;266:1957-8.

359. Queille S, Drougard C, Sarasin A, Daya-Grosjean L. Effects of XPD mutations on ultraviolet-induced apoptosis in relation to skin cancer-proneness in repair-deficient syndromes. *J Invest Dermatol* 2001;117:1162-70.

360. Egly JM. The 14th Datta Lecture. TFIIH: from transcription to clinic. *FEBS Lett* 2001;498:124-8.

361. Giglia-Mari G, Coin F, Ranish JA et al. A new, tenth subunit of TFIIH is responsible for the DNA repair syndrome trichothiodystrophy group A. *Nat Genet* 2004;36:714-9.

362. Weeda G, Eveno E, Donker I et al. A mutation in the XPB/ERCC3 DNA repair transcription gene, associated with trichothiodystrophy. *Am J Hum Genet* 1997;60:320-9.

363. Benhamou S, Sarasin A. ERCC2 /XPD gene polymorphisms and lung cancer: a HuGE review. *Am J Epidemiol* 2005;161:1-14.

364. Drapkin R, Reardon JT, Ansari A et al. Dual role of TFIIH in DNA excision repair and in transcription by RNA polymerase II. *Nature* 1994;368:769-72.

365. Tirode F, Busso D, Coin F, Egly JM. Reconstitution of the transcription factor TFIIH: assignment of functions for the three enzymatic subunits, XPB, XPD, and cdk7. *Mol Cell* 1999;3:87-95.

366. de Boer J, Donker I, de Wit J, Hoeijmakers JH, Weeda G. Disruption of the mouse xeroderma pigmentosum group D DNA repair/basal transcription gene results in preimplantation lethality. *Cancer Res* 1998;58:89-94.

367. de Boer J, Andressoo JO, de Wit J et al. Premature aging in mice deficient in DNA repair and transcription. *Science* 2002;296:1276-9.

368. de Boer J, van Steeg H, Berg RJ et al. Mouse model for the DNA repair/basal transcription disorder trichothiodystrophy reveals cancer predisposition. *Cancer Res* 1999;59:3489-94.

369. Dolle ME, Busuttil RA, Garcia AM et al. Increased genomic instability is not a prerequisite for shortened lifespan in DNA repair deficient mice. *Mutat Res* 2006;596:22-35.

370. Merkel P, Khoury N, Bertolotto C, Perfetti R. Insulin and glucose regulate the expression of the DNA repair enzyme XPD. *Mol Cell Endocrinol* 2003;201:75-85.

371. Lunn RM, Helzlsouer KJ, Parshad R et al. XPD polymorphisms: effects on DNA repair proficiency. *Carcinogenesis* 2000;21:551-5.

372. Vodicka P, Kumar R, Stetina R et al. Genetic polymorphisms in DNA repair genes and possible links with DNA repair rates, chromosomal aberrations and single-strand breaks in DNA. *Carcinogenesis* 2004;25:757-63.

373. Seker H, Butkiewicz D, Bowman ED et al. Functional significance of XPD polymorphic variants: attenuated apoptosis in human lymphoblastoid cells with the XPD 312 Asp/Asp genotype. *Cancer Res* 2001;61:7430-4.

374. Shen MR, Jones IM, Mohrenweiser H. Nonconservative amino acid substitution variants exist at polymorphic frequency in DNA repair genes in healthy humans. *Cancer Res* 1998;58:604-8.

375. Coin F, Bergmann E, Tremeau-Bravard A, Egly JM. Mutations in XPB and XPD helicases found in xeroderma pigmentosum patients impair the transcription function of TFIIH. *Embo J* 1999;18:1357-66.

376. Butkiewicz D, Rusin M, Enewold L, Shields PG, Chorazy M, Harris CC. Genetic polymorphisms in DNA repair genes and risk of lung cancer. *Carcinogenesis* 2001;22:593-7.

377. Pavanello S, Pulliero A, Siwinska E, Mielzynska D, Clonfero E. Reduced nucleotide excision repair and GSTM1-null genotypes influence anti-B[a]PDE-DNA adduct levels in mononuclear white blood cells of highly PAH-exposed coke oven workers. *Carcinogenesis* 2005;26:169-75.

378. Zhou W, Liu G, Miller DP et al. Gene-environment interaction for the ERCC2 polymorphisms and cumulative cigarette smoking exposure in lung cancer. *Cancer Res* 2002;62:1377-81.

379. Liang G, Xing D, Miao X et al. Sequence variations in the DNA repair gene XPD and risk of lung cancer in a Chinese population. *Int J Cancer* 2003;105:669-73.

380. Schabath MB, Delclos GL, Grossman HB et al. Polymorphisms in XPD exons 10 and 23 and bladder cancer risk. *Cancer Epidemiol Biomarkers Prev* 2005;14:878-84.

381. Garcia-Closas M, Malats N, Real FX et al. Genetic variation in the nucleotide excision repair pathway and bladder cancer risk. *Cancer Epidemiol Biomarkers Prev* 2006;15:536-42.

382. Costa S, Pinto D, Pereira D et al. Importance of xeroderma pigmentosum group D polymorphisms in susceptibility to ovarian cancer. *Cancer Lett* 2006.

383. Yin J, Li J, Ma Y, Guo L, Wang H, Vogel U. The DNA repair gene ERCC2/XPD polymorphism Arg 156Arg (A22541C) and risk of lung cancer in a Chinese population. *Cancer Lett* 2005;223:219-26.

384. Jorgensen TJ, Visvanathan K, Ruczinski I, Thuita L, Hoffman S, Helzlsouer KJ. Breast Cancer Risk is not Associated with Polymorphic Forms of Xeroderma Pigmentosum Genes in a Cohort of Women from Washington County, Maryland. *Breast Cancer Res Treat* 2006.

385. Justenhoven C, Hamann U, Pesch B et al. ERCC2 genotypes and a corresponding haplotype are linked with breast cancer risk in a German population. *Cancer Epidemiol Biomarkers Prev* 2004;13:2059-64.

386. Lee SA, Lee KM, Park WY et al. Obesity and genetic polymorphism of ERCC2 and ERCC4 as modifiers of risk of breast cancer. *Exp Mol Med* 2005;37:86-90.

387. Zhang L, Zhang Z, Yan W. Single nucleotide polymorphisms for DNA repair genes in breast cancer patients. *Clin Chim Acta* 2005;359:150-5.

388. Ye W, Kumar R, Bacova G, Lagergren J, Hemminki K, Nyren O. The XPD 751Gln allele is associated with an increased risk for esophageal adenocarcinoma: a population-based case-control study in Sweden. *Carcinogenesis* 2006;27:1835-41.

389. Hu Z, Xu L, Shao M et al. Polymorphisms in the two helicases ERCC2/XPD and ERCC3/XPB of the transcription factor IIH complex and risk of lung cancer: a case-control analysis in a Chinese population. *Cancer Epidemiol Biomarkers Prev* 2006;15:1336-40.

390. Spitz MR, Wu X, Wang Y et al. Modulation of nucleotide excision repair capacity by XPD polymorphisms in lung cancer patients. *Cancer Res* 2001;61:1354-7.

391. Millikan RC, Hummer A, Begg C et al. Polymorphisms in nucleotide excision repair genes and risk of multiple primary melanoma: the Genes Environment and Melanoma Study. *Carcinogenesis* 2006;27:610-8.

392. Jiao L, Hassan MM, Bondy ML, Abbruzzese JL, Evans DB, Li D. The XPD Asp(312)Asn and Lys(751)Gln polymorphisms, corresponding haplotype, and pancreatic cancer risk. *Cancer Lett* 2006.

393. Buch S, Zhu B, Davis AG et al. Association of polymorphisms in the cyclin D1 and XPD genes and susceptibility to cancers of the upper aero-digestive tract. *Mol Carcinog* 2005;42:222-8.

394. Yu HP, Wang XL, Sun X et al. Polymorphisms in the DNA repair gene XPD and susceptibility to esophageal squamous cell carcinoma. *Cancer Genet Cytogenet* 2004;154:10-5.

395. Stern MC, Johnson LR, Bell DA, Taylor JA. XPD codon 751 polymorphism, metabolism genes, smoking, and bladder cancer risk. *Cancer Epidemiol Biomarkers Prev* 2002;11:1004-11.

396. Brewster AM, Jorgensen TJ, Ruczinski I et al. Polymorphisms of the DNA repair genes XPD (Lys751Gln) and XRCC1 (Arg399Gln and Arg194Trp): relationship to breast cancer risk and familial predisposition to breast cancer. *Breast Cancer Res Treat* 2006;95:73-80.

397. Huang WY, Berndt SI, Kang D et al. Nucleotide excision repair gene polymorphisms and risk of advanced colorectal adenoma: XPC polymorphisms modify smoking-related risk. *Cancer Epidemiol Biomarkers Prev* 2006;15:306-11.

398. David-Beabes GL, Lunn RM, London SJ. No association between the XPD (Lys751G1n) polymorphism or the XRCC3 (Thr241Met) polymorphism and lung cancer risk. *Cancer Epidemiol Biomarkers Prev* 2001;10:911-2.

399. Yin J, Vogel U, Ma Y, Guo L, Wang H, Qi R. Polymorphism of the DNA repair gene ERCC2 Lys751Gln and risk of lung cancer in a northeastern Chinese population. *Cancer Genet Cytogenet* 2006;169:27-32.

400. Pfeiffer P, Goedecke W, Kuhfittig-Kulle S, Obe G. Pathways of DNA double-strand break repair and their impact on the prevention and formation of chromosomal aberrations. *Cytogenet Genome Res* 2004;104:7-13.

401. Lengauer C, Kinzler KW, Vogelstein B. Genetic instabilities in human cancers. *Nature* 1998;396:643-9.

402. Rich T, Allen RL, Wyllie AH. Defying death after DNA damage. *Nature* 2000;407:777-83.

403. Allen C, Kurimasa A, Brenneman MA, Chen DJ, Nickoloff JA. DNA-dependent protein kinase suppresses double-strand break-induced and spontaneous homologous recombination. *Proc Natl Acad Sci U S A* 2002;99:3758-63.

404. Clikeman JA, Khalsa GJ, Barton SL, Nickoloff JA. Homologous recombinational repair of double-strand breaks in yeast is enhanced by MAT heterozygosity through yKU-dependent and -independent mechanisms. *Genetics* 2001;157:579-89.

405. Fukushima T, Takata M, Morrison C et al. Genetic analysis of the DNA-dependent protein kinase reveals an inhibitory role of Ku in late S-G2 phase DNA double-strand break repair. *J Biol Chem* 2001;276:44413-8.

406. Liang F, Han M, Romanienko PJ, Jasin M. Homology-directed repair is a major double-strand break repair pathway in mammalian cells. *Proc Natl Acad Sci U S A* 1998;95:5172-7.

407. Ferguson DO, Alt FW. DNA double strand break repair and chromosomal translocation: lessons from animal models. *Oncogene* 2001;20:5572-9.

408. van Gent DC, Hoeijmakers JH, Kanaar R. Chromosomal stability and the DNA double-stranded break connection. *Nat Rev Genet* 2001;2:196-206.

409. Jasin M. Chromosome breaks and genomic instability. *Cancer Invest* 2000;18:78-86.

410. Bay JO, Udar N, Bignon YJ, Gatti RA. [Ataxia telangiectasia and genetic predisposition to cancer]. *Bull Cancer* 1996;83:171-5.

411. Zhang H, Tombline G, Weber BL. BRCA1, BRCA2, and DNA damage response: collision or collusion? *Cell* 1998;92:433-6.

412. Brenneman MA, Wagener BM, Miller CA, Allen C, Nickoloff JA. XRCC3 controls the fidelity of homologous recombination: roles for XRCC3 in late stages of recombination. *Mol Cell* 2002;10:387-95.

413. Nickoloff JA, Brenneman MA. Recombination. In: Creighton TE, ed. *The Encyclopedia of Molecular Medicine*. New York: Wiley; 2001:2736-2741.

414. Richardson C, Jasin M. Frequent chromosomal translocations induced by DNA double-strand breaks. *Nature* 2000;405:697-700.

415. Richardson C, Moynahan ME, Jasin M. Double-strand break repair by interchromosomal recombination: suppression of chromosomal translocations. *Genes Dev* 1998;12:3831-42.

416. Featherstone C, Jackson SP. DNA repair: the Nijmegen breakage syndrome protein. *Curr Biol* 1998;8:R622-5.

417. Chen L, Trujillo K, Sung P, Tomkinson AE. Interactions of the DNA ligase IV-XRCC4 complex with DNA ends and the DNA-dependent protein kinase. *J Biol Chem* 2000;275:26196-205.

418. Nick McElhinny SA, Snowden CM, McCarville J, Ramsden DA. Ku recruits the XRCC4-ligase IV complex to DNA ends. *Mol Cell Biol* 2000;20:2996-3003.

419. Liu N, Lamerdin JE, Tebbs RS et al. XRCC2 and XRCC3, new human Rad51-family members, promote chromosome stability and protect against DNA cross-links and other damages. *Mol Cell* 1998;1:783-93.

420. Johnson RD, Jasin M. Double-strand-break-induced homologous recombination in mammalian cells. *Biochem Soc Trans* 2001;29:196-201.

421. Griffin CS, Simpson PJ, Wilson CR, Thacker J. Mammalian recombination-repair genes XRCC2 and XRCC3 promote correct chromosome segregation. *Nat Cell Biol* 2000;2:757-61.

422. Pierce AJ, Johnson RD, Thompson LH, Jasin M. XRCC3 promotes homology-directed repair of DNA damage in mammalian cells. *Genes Dev* 1999;13:2633-8.

423. Yoshihara T, Ishida M, Kinomura A et al. XRCC3 deficiency results in a defect in recombination and increased endoreduplication in human cells. *Embo J* 2004;23:670-80.

424. Smith TR, Miller MS, Lohman K et al. Polymorphisms of XRCC1 and XRCC3 genes and susceptibility to breast cancer. *Cancer Lett* 2003;190:183-90.

425. Jacobsen NR, Raaschou-Nielsen O, Nexo B et al. XRCC3 polymorphisms and risk of lung cancer. *Cancer Lett* 2004;213:67-72.

426. Han J, Hankinson SE, Ranu H, De Vivo I, Hunter DJ. Polymorphisms in DNA double-strand break repair genes and breast cancer risk in the Nurses' Health Study. *Carcinogenesis* 2004;25:189-95.

427. Shen H, Sturgis EM, Dahlstrom KR, Zheng Y, Spitz MR, Wei Q. A variant of the DNA repair gene XRCC3 and risk of squamous cell carcinoma of the head and neck: a case-control analysis. *Int J Cancer* 2002;99:869-72.

428. Jacobsen NR, Nexo BA, Olsen A et al. No association between the DNA repair gene XRCC3 T241M polymorphism and risk of skin cancer and breast cancer. *Cancer Epidemiol Biomarkers Prev* 2003;12:584-5.

429. Shen H, Wang X, Hu Z et al. Polymorphisms of DNA repair gene XRCC3 Thr241Met and risk of gastric cancer in a Chinese population. *Cancer Lett* 2004;206:51-8.

430. Bray MS, Boerwinkle E, Doris PA. High-throughput multiplex SNP genotyping with MALDI-TOF mass spectrometry: practice, problems and promise. *Hum Mutat* 2001;17:296-304.

431. Shen R, Fan JB, Campbell D et al. High-throughput SNP genotyping on universal bead arrays. *Mutat Res* 2005;573:70-82.

432. Hosking L, Lumsden S, Lewis K et al. Detection of genotyping errors by Hardy-Weinberg equilibrium testing. *Eur J Hum Genet* 2004;12:395-9.

433. Deng HW, Chen WM, Recker RR. Population admixture: detection by Hardy-Weinberg test and its quantitative effects on linkage-disequilibrium methods for localizing genes underlying complex traits. *Genetics* 2001;157:885-97.

434. Volovics A, Van Den Brandt PA. Methods for the analyses of case-cohort studies. Biometrical journal 1997;39:195-214.

435. Schouten EG, Dekker JM, Kok FJ, et al. Risk ratio and rate ratio estimation in case-cohort designs: hypertension and cardiovascular mortality. Stat Med 1993;12:1733-45.

436. Carlson CS, Eberle MA, Rieder MJ, Yi Q, Kruglyak L, Nickerson DA. Selecting a maximally informative set of single-nucleotide polymorphisms for association analyses using linkage disequilibrium. *Am J Hum Genet* 2004;74:106-20.

437. Yen Y-C, Kraft P. Model selection in genetic associatin studies. *Genet Epidemiol* 2005;29:234-398.

438. Greenland S, Brumback B. An overview of relations among causal modelling methods. *Int J Epidemiol* 2002;31:1030-7.

439. Rothman KJ. Epidemiology: An Introduction. New York: Oxford University Press; 2002.

440. Akey J, Jin L, Xiong M. Haplotypes vs single marker linkage disequilibrium tests: what do we gain? *Eur J Hum Genet* 2001;9:291-300.

441. Morris RW, Kaplan NL. On the advantage of haplotype analysis in the presence of multiple disease susceptibility alleles. *Genet Epidemiol* 2002;23:221-33.

442. Schaid DJ. Evaluating associations of haplotypes with traits. *Genet Epidemiol* 2004;27:348-64.

443. Schaid DJ, Rowland CM, Tines DE, Jacobson RM, Poland GA. Score tests for association between traits and haplotypes when linkage phase is ambiguous. *Am J Hum Genet* 2002;70:425-34.

444. Stephens M, Smith NJ, Donnelly P. A new statistical method for haplotype reconstruction from population data. *Am J Hum Genet* 2001;68:978-89.

445. Witte JS. Genetic analysis with hierarchical models. *Genet Epidemiol* 1997;14:1137-42.

446. Greenland S. Modeling and variable selection in epidemiologic analysis. *Am J Public Health* 1989;79:340-9.

447. Greenland S. Comment: cautions in the use of preliminary test estimators. *Stat Med* 1989;8:669-673.

448. Sclove SL, Morris CN, Radhakrishna R. Non-optimality of preliminary-test estimators for the mean of a multivariate normal distribution. *Annals Math Stat* 1972;43:1481-1490.

449. Morris CN. Parametric empirical Bayes inference: theory and applications (with discussion). *J Am Stat Assoc* 1983;78:47-65.

450. Greenland S, Poole C. Empirical-Bayes and semi-Bayes approaches to occupational and environmental hazard surveillance. *Arch Environ Health* 1994;49:9-16.

451. Efron B, Morris CN. Stein's estimation rule and its competitors: an empirical-Bayes approach. *J Am Stat Assoc* 1973;68:117-30.

452. Efron B, Morris CN. Data analysis using Stein's estimator and its generalizations. *J Am Stat Assoc* 1975;70:311-19.

453. Hung RJ, Brennan P, Malaveille C et al. Using hierarchical modeling in genetic association studies with multiple markers: application to a case-control study of bladder cancer. *Cancer Epidemiol Biomarkers Prev* 2004;13:1013-21.

454. Greenland S. Principles of multilevel modelling. *Int J Epidemiol* 2000;29:158-67.

455. Greenland S. Hierarchical regression for epidemiologic analyses of multiple exposures. *Environ Health Perspect* 1994;102 Suppl 8:33-9.

456. Greenland S. Methods for epidemiologic analyses of multiple exposures: a review and comparative study of maximum-likelihood, preliminary-testing, and empirical-Bayes regression. *Stat Med* 1993;12:717-36.

457. Witte JS, Greenland S, Kim LL, Arab L. Multilevel modeling in epidemiology with GLIMMIX. *Epidemiology* 2000;11:684-8.

458. De Roos AJ, Zahm SH, Cantor KP et al. Integrative assessment of multiple pesticides as risk factors for non-Hodgkin's lymphoma among men. *Occup Environ Med* 2003;60:E11.

459. Snijders T, Bosker R. Multilevel analysis: An Introduction to Basic and Advanced Multilevel Modeling. Thousand Oaks, CA: Sage Publications; 1999.

460. Pignoli P, Tremoli E, Poli A, Oreste P, Paoletti R. Intimal plus medial thickness of the arterial wall: a direct measurement with ultrasound imaging. Circulation 1986;74:1399-406.

461. Group TAS. High-resolution B-mode ultrasound scanning methods in the Atherosclerosis Risk in Communities Study (ARIC). The ARIC Study Group. J Neuroimaging 1991;1:68-73.

462. Tilling K, Smith GD, Chambless L, et al. The relation between birth weight and intima-media thickness in middle-aged adults. Epidemiology 2004;15:557-64.

463. Chambless LE, Zhong MM, Arnett D, Folsom AR, Riley WA, Heiss G. Variability in B-mode ultrasound measurements in the atherosclerosis risk in communities (ARIC) study. Ultrasound Med Biol 1996;22:545-54.

464. Scheet P, Stephens M. A fast and flexible statistical model for large-scale population genotype data: applications to inferring missing genotypes and haplotypic phase. Am J Hum Genet 2006;78:629-44.

465. Pritchard JK. Are rare variants responsible for susceptibility to complex diseases? Am J Hum Genet 2001;69:124-37.

466. Anonymous. Cardiovascular Diseases. The health consequences of smoking: a report of the Surgeon General. Atlanta, GA: Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health, 2004:361-419.

467. Greenland S, Robins JM. Empirical-Bayes adjustments for multiple comparisons are sometimes useful. Epidemiology 1991;2:244-51.

468. Witte JS, Greenland S. Simulation study of hierarchical regression. Stat Med 1996;15:1161-70.

469. Chambless LE, Heiss G, Folsom AR, et al. Association of coronary heart disease incidence with carotid arterial wall thickness and major risk factors: the Atherosclerosis Risk in Communities (ARIC) Study, 1987-1993. Am J Epidemiol 1997;146:483-94.

470. Chambless LE, Folsom AR, Clegg LX, et al. Carotid wall thickness is predictive of incident clinical stroke: the Atherosclerosis Risk in Communities (ARIC) study. Am J Epidemiol 2000;151:478-87.

471. Salonen JT, Salonen R. Ultrasonographically assessed carotid morphology and the risk of coronary heart disease. Arterioscler Thromb 1991;11:1245-9.

472. Bots ML, Hoes AW, Koudstaal PJ, Hofman A, Grobbee DE. Common carotid intima-media thickness and risk of stroke and myocardial infarction: the Rotterdam Study. Circulation 1997;96:1432-7.