Genetic Susceptibility to Arsenic Exposure and Arsenical Skin Lesion Prevalence in Bangladesh

Maria Argos

Submitted in partial fulfillment of the requirements for

the degree of Doctor of Philosophy

under the Executive Committee of the Graduate School of

Arts and Sciences

COLUMBIA UNIVERSITY

2011

ABSTRACT

Genetic Susceptibility to Arsenic Exposure and Arsenical Skin Lesion Prevalence in Bangladesh

Maria Argos

Elevated concentrations of arsenic in groundwater pose a public health threat to millions of people worldwide. While arsenic is an established human carcinogen, a mode of action has yet to be determined for arsenic carcinogenesis. However, the oxidative stress and DNA repair pathways have been implicated in arsenic toxicity and have been hypothesized to underlie arsenic carcinogenesis. To date, few epidemiologic studies have evaluated genetic susceptibility to arsenical skin lesions based on single nucleotide polymorphisms (SNPs) in antioxidant enzyme or DNA repair genes. Utilizing cross-sectional data from the 2000-2002 survey of the Health Effects of Arsenic Longitudinal Study (HEALS) for 610 prevalent arsenical skin lesion cases and 1,079 randomly selected controls, I evaluated the associations of SNPs in genes encoding antioxidant enzymes and DNA repair enzymes on skin lesion prevalence. I also evaluated potential interactions between the SNPS as well as SNP-environment interactions in determining skin lesion prevalence.

In the first study of this dissertation (Chapter 2), I assessed the relationship between SNPs in antioxidant enzyme genes and skin lesion prevalence, as well as possible interactions of these associations on the additive scale by various environmental factors. There were no statistically significant associations between these SNPs (*SOD2,* rs4880; *CAT,* rs1001179; *GPX1,* rs1050450; and *MPO*, rs2333227) and skin lesion prevalence. Additionally, there was no evidence of additive interaction by arsenic

exposure levels, body mass index, smoking status, or fruit and vegetable intake with the SNPs in relation to skin lesion prevalence. However, there was marginal evidence that skin lesion prevalence was increased among individuals who carried 4 or more risk alleles compared to individuals carrying 0-3 risk alleles in these SNPs. Additionally, I observed a significant departure from additivity for the risk allele score and primary methylation index on skin lesion prevalence.

In the second study of this dissertation (Chapter 3), I assessed the relationship between SNPs in DNA repair genes (*OGG1,* rs1052133; *XRCC1,* rs25487 and rs1799782; *XRCC3,* rs861539; *ERCC2,* rs1052559; *ERCC5,* rs17655; and *LIG4,* rs1805388) and skin lesion prevalence, as well as possible interactions of these associations on the additive scale by various environmental factors. In logistic regression models controlling for sex, age, and well water arsenic concentration, no associations were observed between measured SNPs and skin lesion prevalence. The results did not vary by arsenic exposure levels, body mass index, or smoking status. However, I did observe a significant inverse association of total fruit and vegetable consumption with skin lesion prevalence, and its additive interaction with the polymorphism in *ERCC5*.

In the third study of this dissertation (Chapter 4), I utilized a multi-analytic approach to explore gene-gene, gene-environment, and higher-order interactions among 10 SNPs related to the oxidative stress and DNA repair pathways by MDR, CART, and logistic regression models. As shown in Chapters 2 and 3, none of these SNPs were associated with skin lesion prevalence, however, were evaluated for potential SNP-SNP interactions. MDR and CART modeling approaches were utilized for the selection of potential gene-gene and gene-environment interactions. Considerable overlap of the

interactions detected by both these methods was observed, which were further evaluated by logistic regression. Results from logistic regression modeling, provided some evidence of these statistical interactions; however, their biological interpretation was limited.

In summary, there was marginal evidence that skin lesion prevalence was increased among individuals who carried 4 or more risk alleles in genotyped SNPs related to the oxidative stress pathway compared to individuals carrying 0-3 risk alleles in these SNPs and, a significant departure from additivity was observed for the risk allele score and primary methylation index on skin lesion prevalence. Additionally, a significant inverse association of total fruit and vegetable consumption with skin lesion prevalence was observed and, a significant interaction between the polymorphism in *ERCC5* and total fruit and vegetable intake was observed in relation to skin lesion prevalence on the additive scale. However, these finding require replication in other studies.

TABLE OF CONTENTS

LIST OF TABLES AND FIGURES

Chapter 1

Chapter 2

ACKNOWLEDGEMENTS

I especially thank my dissertation sponsor, Dr. Habibul Ahsan, for his mentorship over the years. My most valuable training in epidemiology has come through the research I have had the opportunity to participate in with him, from whom I have learned a great deal. I would like to thank my committee members Dr. Alfredo Morabia, Dr. Andrew Rundle, Dr. Yu Chen, and Dr. Regina Santella for their critical read of this work and their generous feedback. I also would like to thank Liliane Zaretsky for her persistent encouragement and support, humor, and guidance through the program. A special thank you also goes to Dr. Sharon Schwartz who challenged me to think critically, and encouraged my enthusiasm for epidemiology, teaching, and cycling. My classmates Teresa Janevic, Danella Hafeman, Tamarra James, Matt Lamb, and Kellee White have been wonderful study partners and friends. I additionally thank all my professors and colleagues at the Mailman School of Public Health who made my years as a graduate student a memorable and rewarding experience. I express special thanks to Faruque Parvez who has been an excellent research colleague and friend. I also would like to thank Paul Rathouz and Brandon Pierce who provided tremendous insight to me regarding research methods and continuously engaged me in thoughtful debates. Lastly, I would like to thank all my family and friends who were supportive of me through this process. Bridget Wadzuk and Kalpana Ramakrishnan for their friendship and always listening ears; Nicholas Mader for making my dissertation writing enjoyable through his company and enthusiasm to work late nights, ability to defuse dissertation “crises”, and continual encouragement; and, my parents for their tremendous support and love.

DEDICATION

To my parents,

who made all of this possible,

for their endless encouragement and patience.

Chapter 1

Background and Introduction

ARSENIC IN DRINKING WATER IN BANGLADESH

Arsenic is a naturally occurring metalloid, ubiquitously present in the environment. Through reduction-oxidation reactions, arsenic can be released from soil and rock into the surrounding aquifers. Elevated concentrations of arsenic in groundwater were first realized in Bangladesh in the 1990s with the appearance of skin lesion epidemics in Bangladeshi villages, which accessed drinking water by tubewells that tap into the arsenic-enriched aquifers (1). The tubewells were installed through Bangladesh governmental initiatives supported by UNICEF beginning in the 1970s to provide safe drinking water to the population through the consumption of groundwater. This was an effort to reduce mortality and morbidity from cholera and other waterborne diseases that had plagued the population, and proved to be effective towards this end with the subsequent reduction of infant mortality.

The permissible level of arsenic in drinking water regulated by the Government of Bangladesh is 50 μg/L. In the Health Effects of Arsenic Longitudinal Study (HEALS)—a prospective cohort study to examine the health effects of arsenic exposure in a Bangladeshi adult population established in 2000 through funding from the National Institute of Environmental Health Sciences (NIEHS)-sponsored Columbia University Superfund Research Program (Project PI: Habibul Ahsan)—we have the ability to characterize exposure distributions and health outcomes within a representative rural population in Bangladesh. Within the HEALS study sample alone, 55.3% of the cohort has been exposed to arsenic concentrations greater than the Bangladeshi national drinking water standard. In 1993, the World Health Organization revised its guideline from 50 μg/L to 10 μg/L. By this new standard, 76.6% of the study cohort is considered at risk for

arsenic toxicity. Estimates from a 1998 British Geological Survey of tubewells in Bangladesh indicate that 46% of samples had water arsenic concentrations greater than 10 μg/L and 27% of samples had water arsenic concentrations greater than 50 μg/L; based on population estimates from 1999, it is believed that approximately 28-35 million in Bangladesh were exposed to arsenic concentrations greater than 50 μg/L in their drinking water and 46-57 million exposed to arsenic concentrations greater than 10 μg/L. There has not been reliable data to quantify the global burden of arsenic in drinking water worldwide; however, arsenic in drinking water has been detected at concentrations greater than 10 μg/L or the prevailing national standard in several countries including Argentina, Australia, Bangladesh, Chile, China, Hungary, India, Mexico, Peru, Thailand, and the United States of America (2).

Chronic exposure to arsenic in drinking water, particularly at concentrations >100 μg/L, has been associated with a multitude of health effects including cancers of the skin, lung, liver, bladder and kidney, cardiovascular disease, peripheral neuropathy, obstetric outcomes and respiratory diseases (3-16). Additionally, the International Agency for Research on Cancer has classified arsenic as a class I human carcinogen (17). Health effects at the lower dose range (≤100 μg/L) remain in question due to limited power of studies to observe dose-dependent associations in this range.

ARSENIC AND SKIN LESIONS

Skin is the major target organ of arsenic, with skin lesions a hallmark characteristic of chronic arsenic exposure and an early manifestation of arsenic toxicity (18). These lesions are chiefly subclassified as (listed in order of increasing severity):

melanosis, leucomelanosis, and keratosis. Melanosis, increased melanin deposition in the skin present as diffuse or spotted hyperpigmentation over the trunk and extremities, usually is the first effect observed with chronic arsenic ingestion. Melanosis often coincides with or progresses to leucomelanosis, which is characterized by diffuse hypopigmentation of the skin over the trunk and extremities. The most severe type of arsenical skin lesion is keratosis. Arsenical keratoses are thought to be precancerous lesions with the potential to become squamous cell or basal cell carcinomas. Classically, they are located on the palms and soles but may be found elsewhere on the body, with the most characteristic form as punctate papules, yellow in color, with numerous small, horny, corn-like elevations, usually 2 to 10 mm in diameter (19, 20). Arsenical keratoses may also present as plaques with slightly elevated, erythematous, scaly or pigmented features (20). It is estimated that the latency for the appearance of arsenical keratoses is extremely variable, but averages around 9 to 19 years (21, 22). The appearance of melanosis and leucomenalosis may appear with a shorter latency period following chronic arsenic exposure; however, this has not been well documented. Depending on the amount of exposure to arsenic, hyperpigmentation can be observed within 6 months.

The association between arsenic exposure and arsenical skin lesion prevalence has been well-established in various ethnic populations and a dose-response trend has been clearly demonstrated. Yoshida et al. (23) published a recent review of the major epidemiologic studies examining the association between arsenic and skin lesions. A striking aspect of this systematic summary is that a dose-response relationship with arsenic exposure was nearly consistently seen across various ethnic populations, despite varying definitions of arsenic exposure and skin lesions. This suggests that even with

variability in the measurement constructs for exposure and disease classification, the effects are probably an underestimate of the true dose-response and the effects are plausible. The association between arsenic exposure (as measured by various exposure constructs) and prevalent skin lesions (defined as presence of at least 1 type of skin lesion) has been previously evaluated in the HEALS cohort and a dose-response trend was demonstrated based on quintiles of increasing water arsenic exposure (prevalence odds ratios=1.0, 1.91, 3.03, 3.71, and 5.39) (24). In recent prospective analyses from the HEALS cohort, we found multivariate adjusted HRs for incident skin lesions comparing 10.1–50.0, 50.1–100.0, 100.1–200.0, and ≥200.1 μg/L to ≤10 μg/L of well water arsenic exposure to be 1.17 (95% CI: 0.92, 1.49), 1.69 (95% CI: 1.33, 2.14), 1.97 (95% CI: 1.58, 2.46), and 2.98 (95% CI: 2.40, 3.71), respectively (P for trend=0.0001) (25).

Evidence has suggested that arsenic exposure itself fails to fully explain the presence of arsenical skin lesions in an exposed population and that genetic susceptibility may play an important role in determining sub-populations at higher risk of developing the disease at similarly exposed levels (26). Epidemiologic studies have shown interaction with sex, age, body mass index (24), smoking (27), socioeconomic status (28), nutritional status (29), and genetic polymorphisms (as reviewed in (30, 31)).

ARSENIC METABOLITES AND SKIN LESIONS

Arsenic is primarily present in the inorganic form (arsenate and arsenite) in drinking water (32). Once internalized, it goes through a series of reduction and oxidative methylation steps (33). While methylation of an exogenous compound is typically considered to be a detoxification process, there is mounting evidence that the methylation

of arsenic increases it's toxicity *in vivo* particularly with the trivalent methylated arsenic species that are more toxic than the inorganic and pentavalent methylated arsenic species (34-37). The proposed pathway for arsenic methylation (38), with *S*-adenosyl methionine (SAM) serving as the methyl donor (CH^{3+}) is:

Arsenate + 2 e^- → Arsenite + CH^{3+} → Methylarsonic acid + 2 e^- → Methylarsonous acid + CH^{3+} → Dimethylarsinic acid + 2 e^- → Dimethylarsinous acid.

The chemical structures of these arsenic compounds are shown in **Figure 1**. Typically, ingested inorganic arsenic is excreted as 10–20% inorganic arsenic, 10–15% monomethylated arsenic (MMA), and 60–75% dimethylated arsenic (DMA) (39). However, there is known inter-individual variability in the methylation capacity of arsenic (as reviewed in (40)), which has been hypothesized to partly explain the variability in susceptibility to arsenic toxicity. Recent *in vitro* evidence by Kojima et al. showed that arsenic methylation enhanced arsenic-induced oxidative DNA damage (41).

In a cross-sectional study in the Lagunera region of Mexico, Del Razo et al. showed that arsenic-exposed individual with cutaneous symptoms of arsenic toxicity compared to exposed individuals without skin lesions had significantly elevated levels of %MMA in urine (14.3% versus 9.5%, respectively), higher primary methylation index—ratio of MMA to inorganic arsenic—(0.5 versus 0.3, respectively), and a lower secondary methylation index—ratio of DMA to MMA—(3.5 versus 6.0, respectively) (42). In a population-based study conducted by Valenzuela et al. among residents of Zimapan, Mexico—an area known to have high arsenic exposure in groundwater caused by mining activity in the region—showed that arsenic-exposed individuals with manifest skin lesions had a significantly larger mean percentage of MMA^{III} in their urine as compared

to arsenic-exposed individuals with no visible skin lesions (7.7% versus 5.9%, respectively), while there was no difference in the percent of MMA^V (43). In a Taiwanese population, Yu et al. showed in a case-control study that individuals with skin lesions (selected with non-melanoma skin cancers, Bowen's disease, melanosis and hyperkeratosis) had significantly larger mean percentage of MMA in urine, lower mean percentage of DMA, and higher ratio of MMA to DMA than control subjects (44). They showed that individuals with high %MMA (>15.5%) had 5.5 (95% CI=1.2-24.8) times the risk of skin lesions compared to individuals with low MMA and, individuals with low %DMA (<72.2%) had 3.2 (95% CI=1.1-10.0) times the risk of skin lesions compared to individuals with high DMA (44). Finally, in a cross-sectional study in a Bangladeshi population, Ahsan et al. previously reported increased prevalence of skin lesions with increasing urinary %MMA and decreasing secondary methylation index (45). In summary, based on these cross-sectional studies of prevalent skin lesion cases, there is good epidemiologic evidence to suggest an association between higher %MMA and lower %DMA in urine and increased skin lesion prevalence.

LIFESTYLE FACTORS AND SKIN LESIONS

There have been several lifestyle factors that have been associated with skin lesion prevalence. The associations between body mass index (24), cigarette smoking (27), and dietary factors (29, 46) with skin lesions have been previously demonstrated in the HEALS cohort. These lifestyle factors all share in common their implication in the oxidative stress pathway. There is substantial evidence to suggest that body mass index (47, 48), cigarette smoking (49), and fruit/vegetable consumption (49-51) are related to

oxidative stress in humans. Additionally, there has been some evidence to suggest that these lifestyle factors also play a role in DNA repair capacity (52, 53).

Body Mass Index

In a hospital-based cross-sectional study in Bangladesh, it was observed that body mass index was inversely associated with skin lesion prevalence and duration (54). In a cross-sectional analysis from HEALS, Ahsan et al. showed a dose-response trend for skin lesion prevalence based on quintiles of increasing body mass index (prevalence odds ratios=1.0, 0.94, 1.01, 0.82, and 0.76), adjusted for potential confounders including well water arsenic concentration (24).

Cigarette Smoking

In a previous cross-sectional study from HEALS, Ahsan et al. showed marginal increased risk of skin lesion prevalence with tobacco smoking, adjusted for potential confounders including well water arsenic concentration (24). Additionally, significant additive interaction was observed between arsenic exposure and tobacco use in males on skin lesion prevalence (27). In recent prospective analyses from HEALS, we observed that former smokers (HR=5.10, 95% CI=4.19, 6.21) and current smokers (HR=3.39, 95% CI=2.93, 3.92) were at increased risk of incident skin lesions compared to never smokers (25). Additionally, we also observed significant additive interaction between arsenic exposure and tobacco use in males on skin lesion incidence (55).

Dietary Factors

Several analyses from the HEALS cohort have evaluated the association of dietary factors with skin lesion prevalence and incidence. Zablotska et al. observed riboflavin, pyridoxine, folic acid, vitamin A, vitamin C, and vitamin E were all inversely associated with skin lesion prevalence (29). In a recent analysis, Pierce et al. observed that dietary patterns that were related to increased gourd, root vegetable, and vegetable intake were inversely associated with skin lesion incidence (56). In a nested case-control study from HEALS, folate deficiency was found to be associated with increased skin lesion incidence (57). Additionally, in a prospective analysis, Chen et al. found blood selenium to be inversely associated with skin lesion incidence (46).

MECHANISMS OF ARSENIC TOXICITY

Arsenic is a well-established human carcinogen (17) and dose-response associations with skin lesion prevalence (24) and incidence (25) have been observed—precursor conditions to non-melanoma skin cancers. However, the mechanisms by which arsenic acts remain in question, which has primarily been attributed to the absence of a suitable animal model to study arsenic toxicity (58, 59). Yet, several mechanisms of arsenic toxicity have been proposed, including genotoxicity (60), increased cell proliferation through the activation of signal transduction pathways (61), induction of oxidative stress (62-64), altered DNA methylation and gene expression (65), impairment of DNA repair (58), increased tumor promotion (66-68), and as a co-carcinogen (69).

ARSENIC EXPOSURE AND OXIDATIVE STRESS

One of the proposed pathways that may be implicated in the association between arsenic and skin lesion risk is oxidative stress (62, 63). Arsenic has been shown to induce reactive oxygen species (ROS) (70). Oxidative stress is hypothesized to play a role in initiation, promotion, and progression within the framework of a multistage carcinogenesis model (71). Matsui et al. conducted a study of 8-hydroxy-2'-deoxyguanosine (8-OHdG), a well established marker of oxidative stress, in skin tissue and found 8-OHdG to be significantly higher in arsenic-related Bowen's disease (squamous cell carcinoma in situ) as compared to arsenic-unrelated Bowen's disease (72). Similar findings were seen by An et al., who additionally showed that 8-OHdG was elevated in the normal and keratotic skin tissue of arsenic-exposed individuals (73).

GENES ENCODING OXIDATIVE STRESS RELATED ENZYMES

There are several well-established antioxidant enzyme genes that are related to the oxidative stress pathway: superoxide dismutase 2, mitochondrial (*SOD2*); catalase (*CAT*); glutathione peroxidase 1 (*GPX1*); and myeloperoxidase (MPO).

SOD2

SOD2 protein plays a major role in maintaining oxidative balance by converting superoxide (O_2^-)—a precursor molecule for all other ROS—into hydrogen peroxide and oxygen (74). Additionally, prior *in vitro* experiments have shown SOD2 to play a role in mitigating arsenic-induced DNA damage from oxidative stress (75, 76). This gene is located in chromosome region 6q25.3. A nonsynonymous SNP (C>T, Ala16Val, rs4880)

in *SOD2* has been extensively studied (77). The C allele retains the alpha helical structure of the protein for normal activity of the enzyme (78). This particular SNP in *SOD2* has been previously shown to be associated with other diseases caused by oxidative stress, including lung cancer (79), prostate cancer (80), diabetic neuropathy (81), Alzheimer's disease (82), radiation injury in patients treated with radiation therapy for prostate adenocarcinoma (83), alcoholic cirrhosis (84), urolithiasis (85), and acoustic neuroma (86). While contrary to the biological direction that would be hypothesized based on the current literature, Hsueh et al. in a study of arsenic-exposed individuals in Taiwan found the C allele of this SNP to be associated with increased risk (OR=2.0; 95% CI=1.0-3.9) of hypertension, adjusted for arsenic exposure; there was no evidence of statistical interaction between arsenic and SOD2 (87).

CAT

CAT protein plays a major role at times of severe oxidative stress by converting hydrogen peroxide (H_2O_2) into water and oxygen (88). Additionally, prior *in vitro* experiments have shown CAT to play a role in mitigating arsenic-induced DNA damage from oxidative stress (70, 75, 89). This gene is located in chromosome region 11p13. A synonymous SNP (T>C, rs1001179) in *CAT* has been investigated with regard to arsenic toxicity in the epidemiologic literature. The T allele has been associated with higher erythrocyte CAT levels (90). Hsueh et al. examined the association of this SNP with hypertension and saw no significant independent risk of the SNP among arsenic-exposed individuals in Taiwan (87). In pilot work, Ahsan et al. examined this SNP in relation to

arsenical keratosis and found a nonsignificant increased risk (OR=1.9; 95% CI=0.8-4.7) associated with the T allele, adjusted for arsenic exposure (91).

GPX1

GPX1 is the main antioxidant enzyme in normal conditions that functions in the detoxification of hydrogen peroxide by using it to oxidize glutathione (92). Additionally, prior *in vitro* experiments have shown GPx to play a role in mitigating arsenic-induced DNA damage from oxidative stress (93, 94). This gene is located in chromosome region 3p21.3. A nonsynonymous SNP (C>T, Pro200Leu, rs1050450) in *GPX1* has been associated with oxidative stress-related diseases. There is some evidence to suggest that the T allele is associated with decreased enzyme activity levels (95, 96), particularly among men (97). This particular SNP in *GPX1* has been previously shown to be associated with other diseases caused by oxidative stress, including lung cancer (98, 99), atherosclerosis in diabetics (100), aging and longevity (101), and interaction with SOD2 for breast cancer risk (102).

MPO

MPO has been implicated in oxidative stress because of its utilization of hydrogen peroxide to produce the oxidant hypochlorite, which may lead to oxidative damage of biological tissues (103). Additionally, prior *in vitro* experiments have shown MPO to play a role in mitigating arsenic-induced DNA damage from oxidative stress (76). This gene is located in chromosome region 17q23.1. A synonymous SNP (G>A, rs2333227) in *MPO* has been previously examined with respect to arsenic toxicity. The A allele has

been shown to have lower enzyme activity (104) and lower transcriptional activity (105). The A allele has been associated with lower risk of aerodigestive tract cancer (106) and lung cancer (107) and increased risk of prostate cancer among a subset of men with aggressive disease (108). In pilot work, Ahsan et al. examined this SNP in relation to arsenical keratosis and found a nonsignificant increased risk (OR=2.1; 95% CI=0.7-6.2) associated with the GG genotype, adjusted for arsenic exposure (91). Additionally, Huang et al. examined this SNP in relation to urinary arsenic concentrations and found marginally increased urinary total arsenic concentrations among smokers with the GG genotype compared to the GA/AA genotype (109).

ARSENIC AND DNA REPAIR

DNA repair has been an additionally proposed pathway that may be implicated in the association between arsenic and skin lesion risk (110). In a cross-sectional study, arsenic concentrations in drinking water and toenail clippings were positively associated with ERCC1 mRNA expression levels suggesting arsenic exposure may induce a DNA repair response (111); although, an inverse association was found in another study (112). In vitro experiments have shown that arsenic induces DNA strand breaks in a concentration dependent manner (113). Arsenic has been shown to inhibit nucleotide excision repair (114, 115). DNA repair deficiency has been associated with increased micronuclei frequency (116). Reduced DNA repair capacity has been associated with increased arsenical skin lesion risk (117).

GENES ENCODING DNA REPAIR PROTEINS

There are several genes that are known to encode DNA repair proteins including: 8-oxoguanine DNA glycosylase (*OGG1)*; X-ray repair complementing defective repair in Chinese hamster cells 1 (*XRCC1*); X-ray repair complementing defective repair in Chinese hamster cells 3 (*XRCC3*); excision repair cross-complementing rodent repair deficiency, complementation group 2 (*ERCC2*); excision repair cross-complementing rodent repair deficiency, complementation group 5 (*ERCC5*); and ligase IV, DNA, ATP-dependent (*LIG4*).

OGG1

OGG1 is a DNA glycosylase involved in base excision repair (BER) of 8-OHdG, an adduct formed from oxidative stress (118). During BER, OGG1 removes damaged bases by cleaving N-glycosylic bonds (118). Prior *in vitro* (119) and animal (120, 121) studies have shown OGG1 to play a role in mitigating arsenic-induced oxidative damage. This gene is located in chromosome region 3p26.2. A nonsynonymous SNP (C>G, Ser326Cys, rs1052133) in *OGG1* has been previously associated with cancer (122). Proteins encoded with the G allele exhibit reduced 8-OHdG repair activity (123). This specific polymorphism in *OGG1* has been shown to be associated with multiple cancers (124-129). Mo et al. showed *OGG1* expression to be significantly correlated with water arsenic concentration among a Mongolian adult population and associated with hyperkeratosis in males (130).

XRCC1

XRCC1 is a DNA repair enzyme that interacts with polynucleotide kinase (PNK), DNA polymerase-beta (POLB) and DNA ligase III (LIG3) as part of a complex to repair single-strand breaks and functions in BER to repair damage caused by agents such as ROS (131). This gene is located in chromosome region 19q13.2.

There is a well characterized nonsynonymous SNP (G>A, Arg399Gln, rs25487) in this gene. Phenotype studies suggest that the A allele is associated with reduced DNA repair (132-135) and ionizing radiation sensitivity (136, 137). The A allele has been associated with several cancers, including increased risk of colorectal (138), lung (139), and breast cancers (140, 141). Although, carriers of the A allele had a reduced risk of non-melanoma skin cancers (142), squamous cell carcinoma of the head and neck (143), bladder cancer (144), and lung cancer (145).

Another known nonsynonymous SNP (C>T, Arg194Trp, rs1799782) in this gene has been previously studied in relation to arsenic. The T allele has been associated with deficient DNA repair (146). A case-control study of lung cancer found homozygous variants (TT genotype) to have a 3-fold increased risk (147); while other case-control studies have shown a reduced risk for lung cancer among carriers of the T allele (145, 148). The T allele has been associated with borderline increased risk of colorectal cancer (138) and squamous cell carcinoma of the head and neck (143), but lower risk of bladder cancer (144). Additionally, increased risk of esophageal squamous cell carcinoma was observed among homozygous variants (149). The presence of both variants in rs1799782 (T allele) and rs25487 (A allele) showed an increased risk of gastric cancer (150). Breton

et al. demonstrated that the rs1799782 SNP modified the association between arsenic and skin lesions, with a three-fold larger odds ratio for skin lesions among individuals in the highest tertile of arsenic exposure with the CC genotype compared to TT in the same exposure category (151).

XRCC3

XRCC3 encodes a protein that forms a complex with RAD51 and RAD51C to repair double-strand DNA breaks through the homologous recombination pathway (152-155). Moreover, studies from our research group have demonstrated arsenic's ability to induce breaks in DNA strands of mammalian cells (156, 157); therefore, reduced ability to repair double-strand breaks due to polymorphisms in the *XRCC3* gene may be relevant for arsenic carcinogenesis. This gene is located in chromosome region 14q32.3. A nonsynonymous SNP (C>T, Thr241Met, rs861539) in *XRCC3* has been characterized. A recent study suggests that the T allele is associated with decreased repair capacity (158). The T allele has been associated with increased risk of melanocytic skin cancer (159), breast cancer (160), and bladder cancer (161). However, the T allele has also been associated with decreased basal cell carcinoma risk (162). Moreover, in a recent case-control study in West Bengal, India, Kundu et al. observed a significant decreased risk of arsenical skin lesions associated with T allele of this polymorphism (163).

ERCC2

ERCC2 is a DNA helicase involved in transcription and nucleotide excision repair (NER) (164). This gene is located in chromosome region 19q13.2–13.3. A

nonsynonymous SNP (A>C, Lys751Gln, rs1052559) in *ERCC2* has been previously established as a risk factor for various cancers and diseases. ERCC2 deficiency is associated with the disease xeroderma pigmentosum—characterized by a deficiency in NER and a greatly increased predisposition to skin cancer (165)—trichothiodystrophy and Cockayne syndrome (166). A phenotype study showed that carriers of the C allele have sub-optimal DNA repair (167). The C allele has been associated with increased risk of lung cancer (147), basal cell carcinoma (168), chronic lymphocytic leukemia (169), squamous cell carcinomas of the head and neck (170), acute myeloid leukemia (171), and breast cancer (170). In pilot work, Ahsan et al. showed a synergistic effect of the AA genotype of this SNP and arsenic exposure on increased risk of arsenical keratosis (172). However, in a US arsenic-exposed population no effects of this SNP were seen with squamous or basal cell carcinoma (173).

ERCC5

ERCC5 is a DNA repair protein involved in the NER pathway of UV-induced damage (174). While there was no current literature on the association of this gene with arsenic, there is evidence of a potential interaction between arsenic and UV (69, 175). This gene is located in chromosome region 13q33. A nonsynonymous SNP (G>C, Asp1104His, rs17655) in *ERCC5* has been examined by prior studies. Case-control studies have found the G allele to be associated with a decreased risk of lung cancer and squamous cell carcinomas of the oropharynx, larynx and esophagus (176) and of lung cancer risk in African Americans (177). Additionally, multifactor dimensionality reduction analysis found this SNP to be predictive of lung cancer risk among Hispanics

(177). Additionally, the C allele was associated with marginally increased breast cancer risk (178).

LIG4

LIG4 is a DNA double-strand break repair enzyme whose two main roles are the ligation step of the non-homologous end-joining pathway and V(D)J recombination (179). Prior *in vitro* studies have shown DNA ligases to play a role in mitigating arsenic-induced oxidative damage (180, 181). This gene is located in chromosome region 13q33-q34. A nonsynonymous SNP (C>T, Thr9Ile, rs1805388) in *LIG4* has been previously examined. This polymorphism is not believed to significantly alter the structural conformation of the protein (182). The T allele of this polymorphism has been shown to be associated with a reduced risk of multiple myeloma (183) as well as glioma (184) and non-Hodgkin lymphoma (185).

GENETIC SUSCEPTIBILITY TO ARSENIC-RELATED SKIN LESIONS

Genetic susceptibility to arsenic-related skin lesions has been evaluated by several candidate gene studies in the last decade. To our knowledge, 18 genes have been evaluated in studies examining polymorphisms in relation to prevalent skin lesions, which are summarized in **Table 1**. For most of the candidate genes, only a single SNP in each gene has been evaluated and few SNPs or genes have been replicated in a second study. The vast majority of these studies suffer from small sample sizes and, significant main effects of the SNPs in most instances were not observed. Although some promising

results have been detected for various SNPs, the few replication studies that have been conducted have found divergent results (i.e., *GSTT1*, *GSTM1*, *ERCC2*).

There are only a few pathways that have been examined to date in relation to arsenical skin lesions. Several genes related to arsenic or xenobiotic metabolism (e.g., *PNP*, *AS3MT*, *GSTO1*, *GSTO2*, *GSTP1*, *GSTTI*, *GSTM1*, *MTHFR*) have been examined with mixed results. Additionally, several candidate SNPs have been evaluated in DNA repair genes (e.g., *ERCC2*, *APEX1*, *XRCC1*, *XRCC3*, *OGG1*) as well as oxidative stress-related genes (e.g., *MPO*, *CAT*), inflammation (e.g., *TNF*, *IL10*), and tumor suppression (e.g., *TP53*).

The limited scope of work that has been done to evaluate genetic susceptibility to arsenical skin lesions clearly warrants further investigation. There has been inconsistency in case definitions (i.e., severity of skin lesions), differences in arsenic exposure distributions across populations as well as other covariates, and under-powered studies to be able to synthesize genetic risk factors from the current literature.

DISSERTATION AIMS

This dissertation utilized cross-sectional data from the 2000-2002 survey of the Health Effects of Arsenic Longitudinal Study (HEALS), with 610 prevalent arsenical skin lesion cases and 1,079 controls selected for the present project, which has been adequately powered to evaluate SNP and SNP-environment associations (**Appendix A**). Data was utilized on SNPs that were genotyped in the following genes encoding antioxidant enzymes: *SOD2, GPX1, CAT,* and *MPO*; as well as the following genes encoding DNA repair enzymes: *OGG1, XRCC1, XRCC3, ERCC2, ERCC5,* and *LIG4.*

Data on arsenic exposure levels as well as clinical, sociodemographic and food frequency questionnaire data were also utilized.

Specifically, this dissertation aims:

Chapter 2: Arsenic Exposure, Antioxidant Enzymes, and Skin Lesion Prevalence in an Adult Bangladeshi Population.

Aim: To evaluate whether SNPs in genes encoding antioxidant enzymes (*SOD2, GPX1, CAT, MPO*) are associated with prevalent arsenical skin lesion status in 610 cases and 1,079 controls from the HEALS cohort.

Hypotheses: Specifically, I tested the following hypotheses:

a. SNPs in these candidate genes are associated with arsenical skin lesions.
b. SNPs in these candidate genes modify the association between arsenic exposure (as measured by well water arsenic and urinary total arsenic concentrations) and arsenical skin lesions.
c. SNPs in these candidate genes modify the association between arsenic methylation capacity (as measured by urinary arsenic species) and arsenical skin lesions.
d. SNPs in these candidate genes modify the association between lifestyle factors (body mass index, cigarette smoking status, and fruit/vegetable consumption) and arsenical skin lesions.

Chapter 3: Arsenic Exposure, DNA Repair Genes, and Skin Lesion Prevalence in an Adult Bangladeshi Population.

Aim: To evaluate whether SNPs in genes encoding DNA repair proteins (*OGG1, XRCC1, XRCC3, ERCC2, ERCC5*, and *LIG4*) are associated with prevalent arsenical skin lesion status in 610 cases and 1,079 controls from the HEALS cohort.

Hypotheses: Specifically, I tested the following hypotheses:

a. SNPs in these candidate genes are associated with arsenical skin lesions.
b. SNPs in these candidate genes modify the association between arsenic exposure (as measured by well water arsenic and urinary total arsenic concentrations) and arsenical skin lesions.
c. SNPs in these candidate genes modify the association between arsenic methylation capacity (as measured by urinary arsenic species) and arsenical skin lesions.
d. SNPs in these candidate genes modify the association between lifestyle factors (body mass index, cigarette smoking status, and fruit/vegetable consumption) and arsenical skin lesions.

Chapter 4: Gene-Gene, Gene-Environment, and Higher Order Interactions in Relation to Arsenic-related Skin Lesions in an Adult Bangladeshi Population.

Aim: To explore SNP-SNP and higher order associations to predict arsenical skin lesion prevalence.

Hypotheses: Specifically, I tested the following hypotheses:

a. Explore interactions among SNPs using multifactor dimensionality reduction (MDR) and classification and regression trees (CART).

b. Explore interactions among SNPs and arsenic exposure (as measured by well water arsenic, creatinine-adjusted urinary total arsenic, primary methylation index, and secondary methylation index) using MDR and CART.
c. Model interaction predictions consistent between MDR and CART using logistic regression models to estimate interactions on the multiplicative scale.
d. Evaluate the findings of these modeling scenarios together to propose interactions for confirmation in future analyses.

REFERENCES

1. Dhar R, Biswas B, Samanta G, et al. Groundwater arsenic calamity in Bangladesh. *Current Science.* 1997;73(1):48-59.

2. BGS, DPHE. Arsenic contamination of groundwater in Bangladesh. Kinniburgh D, Smedley P, eds. (BGS Technical Report WC/00/19, Volume 2). Keyworth: British Geological Survey, 2001.

3. Smith AH, Goycolea M, Haque R, et al. Marked increase in bladder and lung cancer mortality in a region of Northern Chile due to arsenic in drinking water. *Am J Epidemiol.* 1998;147(7):660-669.

4. Chen CJ, Chen CW, Wu MM, et al. Cancer potential in liver, lung, bladder and kidney due to ingested inorganic arsenic in drinking water. *British journal of cancer.* 1992;66(5):888-892.

5. Hopenhayn-Rich C, Biggs ML, Fuchs A, et al. Bladder cancer mortality associated with arsenic in drinking water in Argentina. *Epidemiology.* 1996;7(2):117-124.

6. Hopenhayn-Rich C, Biggs ML, Smith AH. Lung and kidney cancer mortality associated with arsenic in drinking water in Cordoba, Argentina. *Int J Epidemiol.* 1998;27(4):561-569.

7. Tseng WP. Effects and dose--response relationships of skin cancer and blackfoot disease with arsenic. *Environ Health Perspect.* 1977;19109-119.

8. Brouwer OF, Onkenhout W, Edelbroek PM, et al. Increased neurotoxicity of arsenic in methylenetetrahydrofolate reductase deficiency. *Clinical neurology and neurosurgery.* 1992;94(4):307-310.

9. Navas-Acien A, Sharrett AR, Silbergeld EK, et al. Arsenic exposure and cardiovascular disease: a systematic review of the epidemiologic evidence. *Am J Epidemiol.* 2005;162(11):1037-1049.

10. States JC, Srivastava S, Chen Y, et al. Arsenic and cardiovascular disease. *Toxicol Sci.* 2009;107(2):312-323.

11. Tseng CH. Cardiovascular disease in arsenic-exposed subjects living in the arseniasis-hyperendemic areas in Taiwan. *Atherosclerosis.* 2008;199(1):12-18.

12. Ahmad SA, Sayed MH, Barua S, et al. Arsenic in drinking water and pregnancy outcomes. *Environ Health Perspect.* 2001;109(6):629-631.

13. Milton AH, Smith W, Rahman B, et al. Chronic arsenic exposure and adverse pregnancy outcomes in bangladesh. *Epidemiology.* 2005;16(1):82-86.

14. von Ehrenstein OS, Guha Mazumder DN, Hira-Smith M, et al. Pregnancy outcomes, infant mortality, and arsenic in drinking water in West Bengal, India. *Am J Epidemiol.* 2006;163(7):662-669.

15. Milton AH, Rahman M. Respiratory effects and arsenic contaminated well water in Bangladesh. *International journal of environmental health research.* 2002;12(2):175-179.

16. Parvez F, Chen Y, Brandt-Rauf PW, et al. Nonmalignant respiratory effects of chronic arsenic exposure from drinking water among never-smokers in Bangladesh. *Environ Health Perspect.* 2008;116(2):190-195.

17. IARC Working Group on the Evaluation of Carcinogenic Risks to Humans. Some Drinking-water Disinfectants and Contaminants, including Arsenic. Volume 84. Lyon, France, 2004.

18. Byrd DM, Roegner ML, Griffiths JC, et al. Carcinogenic risks of inorganic arsenic in perspective. *International archives of occupational and environmental health.* 1996;68(6):484-494.

19. Jackson R, Grainge JW. Arsenic and cancer. *Canadian Medical Association journal.* 1975;113(5):396-401.

20. Duncan K, Geisse J, Leffell D. Chapter 113. Epithelial Precancerous Lesions (Chapter). Wolff K, Goldsmith L, Katz S, et al., Eds. Fitzpatrick's Dermatology in General Medicine, 7e: The McGraw-Hill Companies, Inc., 2008.

21. Wong SS, Tan KC, Goh CL. Cutaneous manifestations of chronic arsenicism: review of seventeen cases. *Journal of the American Academy of Dermatology.* 1998;38(2 Pt 1):179-185.

22. Haque R, Mazumder DN, Samanta S, et al. Arsenic in drinking water and skin lesions: dose-response data from West Bengal, India. *Epidemiology.* 2003;14(2):174-182.

23. Yoshida T, Yamauchi H, Fan Sun G. Chronic health effects in people exposed to arsenic via the drinking water: dose-response relationships in review. *Toxicol Appl Pharmacol.* 2004;198(3):243-252.

24. Ahsan H, Chen Y, Parvez F, et al. Arsenic exposure from drinking water and risk of premalignant skin lesions in Bangladesh: baseline results from the Health Effects of Arsenic Longitudinal Study. *Am J Epidemiol.* 2006;163(12):1138-1148.

25. Argos M, Kalra T, Pierce B, et al. A prospective study of arsenic exposure from drinking water and incidence of skin lesions in Bangladesh. *American Journal of Epidemiology.* In press.

26. Concha G, Vogler G, Nermell B, et al. Intra-individual variation in the metabolism of inorganic arsenic. *International archives of occupational and environmental health.* 2002;75(8):576-580.

27. Chen Y, Graziano JH, Parvez F, et al. Modification of risk of arsenic-induced skin lesions by sunlight exposure, smoking, and occupational exposures in Bangladesh. *Epidemiology.* 2006;17(4):459-467.

28. Argos M, Parvez F, Chen Y, et al. Socioeconomic status and risk for arsenic-related skin lesions in Bangladesh. *Am J Public Health.* 2007;97(5):825-831.

29. Zablotska LB, Chen Y, Graziano JH, et al. Protective effects of B vitamins and antioxidants on the risk of arsenic-related skin lesions in Bangladesh. *Environ Health Perspect.* 2008;116(8):1056-1062.

30. Ghosh P, Banerjee M, Giri AK, et al. Toxicogenomics of arsenic: classical ideas and recent advances. *Mutat Res.* 2008;659(3):293-301.

31. Hernandez A, Marcos R. Genetic variations associated with interindividual sensitivity in the response to arsenic exposure. *Pharmacogenomics.* 2008;9(8):1113-1132.

32. Andreae MO. Determination of arsenic species in natural waters. *Analytical chemistry.* 1977;49(6):820-823.

33. Aposhian HV. Enzymatic methylation of arsenic species and other new approaches to arsenic toxicity. *Annu Rev Pharmacol Toxicol.* 1997;37397-419.

34. Le XC, Ma M, Cullen WR, et al. Determination of monomethylarsonous acid, a key arsenic methylation intermediate, in human urine. *Environ Health Perspect.* 2000;108(11):1015-1018.

35. Del Razo LM, Styblo M, Cullen WR, et al. Determination of trivalent methylated arsenicals in biological matrices. *Toxicol Appl Pharmacol.* 2001;174(3):282-293.

36. Ahmad S, Anderson WL, Kitchin KT. Dimethylarsinic acid effects on DNA damage and oxidative stress related biochemical parameters in B6C3F1 mice. *Cancer Lett.* 1999;139(2):129-135.

37. Styblo M, Drobna Z, Jaspers I, et al. The role of biomethylation in toxicity and carcinogenicity of arsenic: a research update. *Environ Health Perspect.* 2002;110 Suppl 5767-771.

38. Challenger F. Biological methylation. *Chem Rev.* 1945;36(3):315-361.

39. Hopenhayn-Rich C, Smith AH, Goeden HM. Human studies do not support the methylation threshold hypothesis for the toxicity of inorganic arsenic. *Environmental research.* 1993;60(2):161-177.

40. Tseng CH. A review on environmental factors regulating arsenic methylation in humans. *Toxicol Appl Pharmacol.* 2009;235(3):338-350.

41. Kojima C, Ramirez DC, Tokar EJ, et al. Requirement of arsenic biomethylation for oxidative DNA damage. *Journal of the National Cancer Institute.* 2009;101(24):1670-1681.

42. Del Razo LM, Garcia-Vargas GG, Vargas H, et al. Altered profile of urinary arsenic metabolites in adults with chronic arsenicism. A pilot study. *Archives of toxicology.* 1997;71(4):211-217.

43. Valenzuela OL, Borja-Aburto VH, Garcia-Vargas GG, et al. Urinary trivalent methylated arsenic species in a population chronically exposed to inorganic arsenic. *Environ Health Perspect.* 2005;113(3):250-254.

44. Yu RC, Hsu KH, Chen CJ, et al. Arsenic methylation capacity and skin cancer. *Cancer Epidemiol Biomarkers Prev.* 2000;9(11):1259-1262.

45. Ahsan H, Chen Y, Kibriya MG, et al. Arsenic metabolism, genetic susceptibility, and risk of premalignant skin lesions in Bangladesh. *Cancer Epidemiol Biomarkers Prev.* 2007;16(6):1270-1278.

46. Chen Y, Hall M, Graziano JH, et al. A prospective study of blood selenium levels and the risk of arsenic-related premalignant skin lesions. *Cancer Epidemiol Biomarkers Prev.* 2007;16(2):207-213.

47. Furukawa S, Fujita T, Shimabukuro M, et al. Increased oxidative stress in obesity and its impact on metabolic syndrome. *The Journal of clinical investigation.* 2004;114(12):1752-1761.

48. Keaney JF, Jr., Larson MG, Vasan RS, et al. Obesity and systemic oxidative stress: clinical correlates of oxidative stress in the Framingham Study. *Arteriosclerosis, thrombosis, and vascular biology.* 2003;23(3):434-439.

49. Lesgards JF, Durand P, Lassarre M, et al. Assessment of lifestyle effects on the overall antioxidant capacity of healthy subjects. *Environ Health Perspect.* 2002;110(5):479-486.

50. Talegawkar SA, Beretta G, Yeum KJ, et al. Total antioxidant performance is associated with diet and serum antioxidants in participants of the diet and physical activity substudy of the Jackson Heart Study. *The Journal of nutrition.* 2009;139(10):1964-1971.

51. Staruchova M, Volkova K, Lajdova A, et al. Importance of diet in protection against oxidative damage. *Neuro endocrinology letters.* 2006;27 Suppl 2112-115.

52. Tyson J, Caple F, Spiers A, et al. Inter-individual variation in nucleotide excision repair in young adults: effects of age, adiposity, micronutrient supplementation and genotype. *The British journal of nutrition.* 2009;101(9):1316-1323.

53. Raji NS, Surekha A, Rao KS. Improved DNA-repair parameters in PHA-stimulated peripheral blood lymphocytes of human subjects with low body mass index. *Mechanisms of ageing and development.* 1998;104(2):133-148.

54. Mitra AK, Bose BK, Kabir H, et al. Arsenic-related health problems among hospital patients in southern Bangladesh. *J Health Popul Nutr.* 2002;20(3):198-204.

55. Melkonian S, Argos M, Pierce BL, et al. A prospective study of the synergistic effects of arsenic exposure and smoking, sun exposure, fertilizer use, and pesticide use on risk of premalignant skin lesions in Bangladeshi men. *Am J Epidemiol.* 2011;173(2):183-191.

56. Pierce BL, Argos M, Chen Y, et al. Arsenic exposure, dietary patterns, and skin lesion risk in bangladesh: a prospective study. *Am J Epidemiol.* 2011;173(3):345-354.

57. Pilsner JR, Liu X, Ahsan H, et al. Folate deficiency, hyperhomocysteinemia, low urinary creatinine, and hypomethylation of leukocyte DNA are risk factors for arsenic-induced skin lesions. *Environ Health Perspect.* 2009;117(2):254-260.

58. Rossman TG. Mechanism of arsenic carcinogenesis: an integrated approach. *Mutat Res.* 2003;533(1-2):37-65.

59. Tokar EJ, Benbrahim-Tallaa L, Ward JM, et al. Cancer in experimental animals exposed to arsenic and arsenic compounds. *Critical reviews in toxicology.*40(10):912-927.

60. Colognato R, Coppede F, Ponti J, et al. Genotoxicity induced by arsenic compounds in peripheral human lymphocytes analysed by cytokinesis-block micronucleus assay. *Mutagenesis.* 2007;22(4):255-261.

61. Simeonova PP, Wang S, Toriuma W, et al. Arsenic mediates cell proliferation and gene expression in the bladder epithelium: association with activating protein-1 transactivation. *Cancer Res.* 2000;60(13):3445-3453.

62. Shi H, Shi X, Liu KJ. Oxidative mechanism of arsenic toxicity and carcinogenesis. *Mol Cell Biochem.* 2004;255(1-2):67-78.

63. Valko M, Rhodes CJ, Moncol J, et al. Free radicals, metals and antioxidants in oxidative stress-induced cancer. *Chem Biol Interact.* 2006;160(1):1-40.

64. Kitchin KT, Conolly R. Arsenic-induced carcinogenesis--oxidative stress as a possible mode of action and future research needs for more biologically based risk assessment. *Chem Res Toxicol.* 2010;23(2):327-335.

65. Huang C, Ke Q, Costa M, et al. Molecular mechanisms of arsenic carcinogenesis. *Mol Cell Biochem.* 2004;255(1-2):57-66.

66. Yamamoto S, Konishi Y, Matsuda T, et al. Cancer induction by an organic arsenic compound, dimethylarsinic acid (cacodylic acid), in F344/DuCrj rats after pretreatment with five carcinogens. *Cancer Res.* 1995;55(6):1271-1276.

67. Wanibuchi H, Yamamoto S, Chen H, et al. Promoting effects of dimethylarsinic acid on N-butyl-N-(4-hydroxybutyl)nitrosamine-induced urinary bladder carcinogenesis in rats. *Carcinogenesis.* 1996;17(11):2435-2439.

68. Yamanaka K, Ohtsubo K, Hasegawa A, et al. Exposure to dimethylarsinic acid, a main metabolite of inorganic arsenics, strongly promotes tumorigenesis initiated by 4-nitroquinoline 1-oxide in the lungs of mice. *Carcinogenesis.* 1996;17(4):767-770.

69. Rossman TG, Uddin AN, Burns FJ. Evidence that arsenite acts as a cocarcinogen in skin cancer. *Toxicol Appl Pharmacol.* 2004;198(3):394-404.

70. Kessel M, Liu SX, Xu A, et al. Arsenic induces oxidative DNA damage in mammalian cells. *Mol Cell Biochem.* 2002;234-235(1-2):301-308.

71. Klaunig JE, Kamendulis LM. The role of oxidative stress in carcinogenesis. *Annu Rev Pharmacol Toxicol.* 2004;44239-267.

72. Matsui M, Nishigori C, Toyokuni S, et al. The role of oxidative DNA damage in human arsenic carcinogenesis: detection of 8-hydroxy-2'-deoxyguanosine in arsenic-related Bowen's disease. *J Invest Dermatol.* 1999;113(1):26-31.

73. An Y, Gao Z, Wang Z, et al. Immunohistochemical analysis of oxidative DNA damage in arsenic-related human skin samples from arsenic-contaminated area of China. *Cancer Lett.* 2004;214(1):11-18.

74. Skrzycki M, Majewska M, Podsiad M, et al. Expression and activity of superoxide dismutase isoenzymes in colorectal cancer. *Acta biochimica Polonica.* 2009;56(4):663-670.

75. Nordenson I, Beckman L. Is the genotoxic effect of arsenic mediated by oxygen free radicals? *Hum Hered.* 1991;41(1):71-73.

76. Wang TS, Hsu TY, Chung CH, et al. Arsenite induces oxidative DNA adducts and DNA-protein cross-links in mammalian cells. *Free Radic Biol Med.* 2001;31(3):321-330.

77. Rosenblum JS, Gilula NB, Lerner RA. On signal sequence polymorphisms and diseases of distribution. *Proc Natl Acad Sci U S A.* 1996;93(9):4471-4473.

78. Shimoda-Matsubayashi S, Matsumine H, Kobayashi T, et al. Structural dimorphism in the mitochondrial targeting sequence in the human manganese superoxide dismutase gene. A predictive evidence for conformational change to influence mitochondrial transport and a study of allelic association in Parkinson's disease. *Biochem Biophys Res Commun.* 1996;226(2):561-565.

79. Zejnilovic J, Akev N, Yilmaz H, et al. Association between manganese superoxide dismutase polymorphism and risk of lung cancer. *Cancer Genet Cytogenet.* 2009;189(1):1-4.

80. Kang D, Lee KM, Park SK, et al. Functional variant of manganese superoxide dismutase (SOD2 V16A) polymorphism is associated with prostate cancer risk in the prostate, lung, colorectal, and ovarian cancer study. *Cancer Epidemiol Biomarkers Prev.* 2007;16(8):1581-1586.

81. Mollsten A, Marklund SL, Wessman M, et al. A functional polymorphism in the manganese superoxide dismutase gene and diabetic nephropathy. *Diabetes.* 2007;56(1):265-269.

82. Wiener HW, Perry RT, Chen Z, et al. A polymorphism in SOD2 is associated with development of Alzheimer's disease. *Genes Brain Behav.* 2007;6(8):770-775.

83. Burri RJ, Stock RG, Cesaretti JA, et al. Association of single nucleotide polymorphisms in SOD2, XRCC1 and XRCC3 with susceptibility for the

development of adverse effects resulting from radiotherapy for prostate cancer. *Radiat Res.* 2008;170(1):49-59.

84. Nahon P, Sutton A, Pessayre D, et al. Genetic dimorphism in superoxide dismutase and susceptibility to alcoholic cirrhosis, hepatocellular carcinoma, and death. *Clin Gastroenterol Hepatol.* 2005;3(3):292-298.

85. Tugcu V, Ozbek E, Aras B, et al. Manganese superoxide dismutase (Mn-SOD) gene polymorphisms in urolithiasis. *Urol Res.* 2007;35(5):219-224.

86. Rajaraman P, Hutchinson A, Rothman N, et al. Oxidative response gene polymorphisms and risk of adult brain tumors. *Neuro Oncol.* 2008;10(5):709-715.

87. Hsueh YM, Lin P, Chen HW, et al. Genetic polymorphisms of oxidative and antioxidant enzymes and arsenic-related hypertension. *J Toxicol Environ Health A.* 2005;68(17-18):1471-1484.

88. Kinnula VL, Everitt JI, Mangum JB, et al. Antioxidant defense mechanisms in cultured pleural mesothelial cells. *American journal of respiratory cell and molecular biology.* 1992;7(1):95-103.

89. Wang TS, Huang H. Active oxygen species are involved in the induction of micronuclei by arsenite in XRS-5 cells. *Mutagenesis.* 1994;9(3):253-257.

90. Forsberg L, Lyrenas L, de Faire U, et al. A common functional C-T substitution polymorphism in the promoter region of the human catalase gene influences transcription factor binding, reporter gene transcription and is correlated to blood catalase levels. *Free Radic Biol Med.* 2001;30(5):500-505.

91. Ahsan H, Chen Y, Kibriya MG, et al. Susceptibility to arsenic-induced hyperkeratosis and oxidative stress genes myeloperoxidase and catalase. *Cancer Lett.* 2003;201(1):57-65.

92. Halliwell B. Free radicals and metal ions in health and disease. *The Proceedings of the Nutrition Society.* 1987;46(1):13-26.

93. Wang TS, Shu YF, Liu YC, et al. Glutathione peroxidase and catalase modulate the genotoxicity of arsenite. *Toxicology.* 1997;121(3):229-237.

94. Lee TC, Ho IC. Modulation of cellular antioxidant defense activities by sodium arsenite in human fibroblasts. *Archives of toxicology.* 1995;69(7):498-504.

95. Hamanishi T, Furuta H, Kato H, et al. Functional variants in the glutathione peroxidase-1 (GPx-1) gene are associated with increased intima-media thickness of carotid arteries and risk of macrovascular diseases in japanese type 2 diabetic patients. *Diabetes.* 2004;53(9):2455-2460.

96. Hu YJ, Diamond AM. Role of glutathione peroxidase 1 in breast cancer: loss of heterozygosity and allelic differences in the response to selenium. *Cancer Res.* 2003;63(12):3347-3351.

97. Bastaki M, Huen K, Manzanillo P, et al. Genotype-activity relationship for Mn-superoxide dismutase, glutathione peroxidase 1 and catalase in humans. *Pharmacogenet Genomics.* 2006;16(4):279-286.

98. Rosenberger A, Illig T, Korb K, et al. Do genetic factors protect for early onset lung cancer? A case control study before the age of 50 years. *BMC cancer.* 2008;860.

99. Raaschou-Nielsen O, Sorensen M, Hansen RD, et al. GPX1 Pro198Leu polymorphism, interactions with smoking and alcohol consumption, and risk for lung cancer. *Cancer Lett.* 2007;247(2):293-300.

100. Nemoto M, Nishimura R, Sasaki T, et al. Genetic association of glutathione peroxidase-1 with coronary artery calcification in type 2 diabetes: a case control study with multi-slice computed tomography. *Cardiovascular diabetology.* 2007;623.

101. Soerensen M, Christensen K, Stevnsner T, et al. The Mn-superoxide dismutase single nucleotide polymorphism rs4880 and the glutathione peroxidase 1 single nucleotide polymorphism rs1050450 are associated with aging and longevity in the oldest old. *Mechanisms of ageing and development.* 2009;130(5):308-314.

102. Cox DG, Tamimi RM, Hunter DJ. Gene x Gene interaction between MnSOD and GPX-1 and breast cancer risk: a nested case-control study. *BMC cancer.* 2006;6217.

103. Babior BM. Phagocytes and oxidative stress. *The American journal of medicine.* 2000;109(1):33-44.

104. Van Schooten FJ, Boots AW, Knaapen AM, et al. Myeloperoxidase (MPO) -463G->A reduces MPO activity and DNA adduct levels in bronchoalveolar lavages of smokers. *Cancer Epidemiol Biomarkers Prev.* 2004;13(5):828-833.

105. Piedrafita FJ, Molander RB, Vansant G, et al. An Alu element in the myeloperoxidase promoter contains a composite SP1-thyroid hormone-retinoic acid response element. *J Biol Chem.* 1996;271(24):14412-14420.

106. Cascorbi I, Henning S, Brockmoller J, et al. Substantially reduced risk of cancer of the aerodigestive tract in subjects with variant--463A of the myeloperoxidase gene. *Cancer Res.* 2000;60(3):644-649.

107. London SJ, Lehman TA, Taylor JA. Myeloperoxidase genetic polymorphism and lung cancer risk. *Cancer Res.* 1997;57(22):5001-5003.

108. Choi JY, Neuhouser ML, Barnett MJ, et al. Iron intake, oxidative stress-related genes (MnSOD and MPO) and prostate cancer risk in CARET cohort. *Carcinogenesis.* 2008;29(5):964-970.

109. Huang YK, Huang YL, Hsueh YM, et al. Arsenic exposure, urinary arsenic speciation, and the incidence of urothelial carcinoma: a twelve-year follow-up study. *Cancer Causes Control.* 2008;19(8):829-839.

110. Kitchin KT. Recent advances in arsenic carcinogenesis: modes of action, animal model systems, and methylated arsenic metabolites. *Toxicol Appl Pharmacol.* 2001;172(3):249-261.

111. Mo J, Xia Y, Ning Z, et al. Elevated ERCC1 gene expression in blood cells associated with exposure to arsenic from drinking water in Inner Mongolia. *Anticancer research.* 2009;29(8):3253-3259.

112. Andrew AS, Karagas MR, Hamilton JW. Decreased DNA repair gene expression among individuals exposed to arsenic in United States drinking water. *Int J Cancer.* 2003;104(3):263-268.

113. Dong JT, Luo XM. Arsenic-induced DNA-strand breaks associated with DNA-protein crosslinks in human fetal lung fibroblasts. *Mutat Res.* 1993;302(2):97-102.

114. Bau DT, Wang TS, Chung CH, et al. Oxidative DNA adducts and DNA-protein cross-links are the major DNA lesions induced by arsenite. *Environ Health Perspect.* 2002;110 Suppl 5753-756.

115. Schwerdtle T, Walter I, Mackiw I, et al. Induction of oxidative DNA damage by arsenite and its trivalent and pentavalent methylated metabolites in cultured human cells and isolated DNA. *Carcinogenesis.* 2003;24(5):967-974.

116. Mateuca RA, Roelants M, Iarmarcovai G, et al. hOGG1(326), XRCC1(399) and XRCC3(241) polymorphisms influence micronucleus frequencies in human lymphocytes in vivo. *Mutagenesis.* 2008;23(1):35-41.

117. Banerjee M, Sarma N, Biswas R, et al. DNA repair deficiency leads to susceptibility to develop arsenic-induced premalignant skin lesions. *Int J Cancer.* 2008;123(2):283-287.

118. Lu AL, Li X, Gu Y, et al. Repair of oxidative DNA damage: mechanisms and functions. *Cell Biochem Biophys.* 2001;35(2):141-170.

119. Pu YS, Jan KY, Wang TC, et al. 8-Oxoguanine DNA glycosylase and MutY homolog are involved in the incision of arsenite-induced DNA adducts. *Toxicol Sci.* 2007;95(2):376-382.

120. Wanibuchi H, Salim EI, Kinoshita A, et al. Understanding arsenic carcinogenicity by the use of animal models. *Toxicol Appl Pharmacol.* 2004;198(3):366-376.

121. Kinoshita A, Wanibuchi H, Morimura K, et al. Carcinogenicity of dimethylarsinic acid in Ogg1-deficient mice. *Cancer science.* 2007;98(6):803-814.

122. Ishida T, Hippo Y, Nakahori Y, et al. Structure and chromosome location of human OGG1. *Cytogenet Cell Genet.* 1999;85(3-4):232-236.

123. 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(25):3219-3225.

124. Sugimura H, Kohno T, Wakai K, et al. hOGG1 Ser326Cys polymorphism and lung cancer susceptibility. *Cancer Epidemiol Biomarkers Prev.* 1999;8(8):669-674.

125. 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(6):932-937.

126. Xing DY, Tan W, Song N, et al. Ser326Cys polymorphism in hOGG1 gene and risk of esophageal cancer in a Chinese population. *Int J Cancer.* 2001;95(3):140-143.

127. Goode EL, Ulrich CM, Potter JD. Polymorphisms in DNA repair genes and associations with cancer risk. *Cancer Epidemiol Biomarkers Prev.* 2002;11(12):1513-1530.

128. Le Marchand L, Donlon T, Lum-Jones A, et al. Association of the hOGG1 Ser326Cys polymorphism with lung cancer risk. *Cancer Epidemiol Biomarkers Prev.* 2002;11(4):409-412.

129. Xu J, Zheng SL, Turner A, et al. Associations between hOGG1 sequence variants and prostate cancer susceptibility. *Cancer Res.* 2002;62(8):2253-2257.

130. Mo J, Xia Y, Wade TJ, et al. Chronic arsenic exposure and oxidative stress: OGG1 expression and arsenic exposure, nail selenium, and skin hyperkeratosis in Inner Mongolia. *Environ Health Perspect.* 2006;114(6):835-841.

131. Whitehouse CJ, Taylor RM, Thistlethwaite A, et al. XRCC1 stimulates human polynucleotide kinase activity at damaged DNA termini and accelerates DNA single-strand break repair. *Cell.* 2001;104(1):107-117.

132. 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(1):63-71.

133. 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(5):965-971.

134. Lei Y, Hwang S, Chang C, et al. Effects on sister chromatid exchange frequency of polymorphisms in DNA repair gene XRCC1 in smokers. *Mutat Res.* 2002;519(1-2):93.

135. Lunn RM, Langlois RG, Hsieh LL, et al. XRCC1 polymorphisms: effects on aflatoxin B1-DNA adducts and glycophorin A variant frequency. *Cancer Res.* 1999;59(11):2557-2561.

136. Hu JJ, Smith TR, Miller MS, et al. Amino acid substitution variants of APE1 and XRCC1 genes associated with ionizing radiation sensitivity. *Carcinogenesis.* 2001;22(6):917-922.

137. Hu JJ, Smith TR, Miller MS, et al. Genetic regulation of ionizing radiation sensitivity and breast cancer risk. *Environ Mol Mutagen.* 2002;39(2-3):208-215.

138. Abdel-Rahman SZ, Soliman AS, Bondy ML, et al. Inheritance of the 194Trp and the 399Gln variant alleles of the DNA repair gene XRCC1 are associated with increased risk of early-onset colorectal carcinoma in Egypt. *Cancer Lett.* 2000;159(1):79-86.

139. Divine KK, Gilliland FD, Crowell RE, et al. The XRCC1 399 glutamine allele is a risk factor for adenocarcinoma of the lung. *Mutat Res.* 2001;461(4):273-278.

140. Duell EJ, Millikan RC, Pittman GS, et al. Polymorphisms in the DNA repair gene XRCC1 and breast cancer. *Cancer Epidemiol Biomarkers Prev.* 2001;10(3):217-222.

141. Kim SU, Park SK, Yoo KY, et al. XRCC1 genetic polymorphism and breast cancer risk. *Pharmacogenetics.* 2002;12(4):335-338.

142. Nelson HH, Kelsey KT, Mott LA, et al. The XRCC1 Arg399Gln polymorphism, sunburn, and non-melanoma skin cancer: evidence of gene-environment interaction. *Cancer Res.* 2002;62(1):152-155.

143. Olshan AF, Watson MA, Weissler MC, et al. XRCC1 polymorphisms and head and neck cancer. *Cancer Lett.* 2002;178(2):181-186.

144. Stern MC, Umbach DM, van Gils CH, et al. DNA repair gene XRCC1 polymorphisms, smoking, and bladder cancer risk. *Cancer Epidemiol Biomarkers Prev.* 2001;10(2):125-131.

145. David-Beabes GL, London SJ. Genetic polymorphism of XRCC1 and lung cancer risk among African-Americans and Caucasians. *Lung Cancer.* 2001;34(3):333-339.

146. 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(4):604-608.

147. 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(8):1321-1325.

148. 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(2):119-123.

149. Xing D, Qi J, Miao X, et al. 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(5):600-605.

150. 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(4):601-606.

151. Breton CV, Zhou W, Kile ML, et al. Susceptibility to arsenic-induced skin lesions from polymorphisms in base excision repair genes. *Carcinogenesis.* 2007;28(7):1520-1525.

152. 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(6):783-793.

153. Liu N, Schild D, Thelen MP, et al. Involvement of Rad51C in two distinct protein complexes of Rad51 paralogs in human cells. *Nucleic Acids Res.* 2002;30(4):1009-1015.

154. Pierce AJ, Johnson RD, Thompson LH, et al. XRCC3 promotes homology-directed repair of DNA damage in mammalian cells. *Genes Dev.* 1999;13(20):2633-2638.

155. Brenneman MA, Weiss AE, Nickoloff JA, et al. XRCC3 is required for efficient repair of chromosome breaks by homologous recombination. *Mutat Res.* 2000;459(2):89-97.

156. Liu F, Jan KY. DNA damage in arsenite- and cadmium-treated bovine aortic endothelial cells. *Free Radic Biol Med.* 2000;28(1):55-63.

157. Liu SX, Athar M, Lippai I, et al. Induction of oxyradicals by arsenic: implication for mechanism of genotoxicity. *Proc Natl Acad Sci U S A.* 2001;98(4):1643-1648.

158. Matullo G, Palli D, Peluso M, et al. XRCC1, XRCC3, XPD gene polymorphisms, smoking and (32)P-DNA adducts in a sample of healthy subjects. *Carcinogenesis.* 2001;22(9):1437-1445.

159. Winsey SL, Haldar NA, Marsh HP, et al. A variant within the DNA repair gene XRCC3 is associated with the development of melanoma skin cancer. *Cancer Res.* 2000;60(20):5612-5616.

160. 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(12):1399-1407.

161. Matullo G, Guarrera S, Carturan S, et al. DNA repair gene polymorphisms, bulky DNA adducts in white blood cells and bladder cancer in a case-control study. *Int J Cancer.* 2001;92(4):562-567.

162. Thirumaran RK, Bermejo JL, Rudnai P, et al. Single nucleotide polymorphisms in DNA repair genes and basal cell carcinoma of skin. *Carcinogenesis.* 2006;27(8):1676-1681.

163. Kundu M, Ghosh P, Mitra S, et al. Precancerous and non-cancer disease endpoints of chronic arsenic exposure: the level of chromosomal damage and XRCC3 T241M polymorphism. *Mutat Res.* 2011;706(1-2):7-12.

164. Chen J, Suter B. Xpd, a structural bridge and a functional link. *Cell cycle.* 2003;2(6):503-506.

165. Tomescu D, Kavanagh G, Ha T, et al. Nucleotide excision repair gene XPD polymorphisms and genetic predisposition to melanoma. *Carcinogenesis.* 2001;22(3):403-408.

166. Cleaver JE, Thompson LH, Richardson AS, et al. A summary of mutations in the UV-sensitive disorders: xeroderma pigmentosum, Cockayne syndrome, and trichothiodystrophy. *Hum Mutat.* 1999;14(1):9-22.

167. Lunn RM, Helzlsouer KJ, Parshad R, et al. XPD polymorphisms: effects on DNA repair proficiency. *Carcinogenesis.* 2000;21(4):551-555.

168. Dybdahl M, Vogel U, Frentz G, et al. Polymorphisms in the DNA repair gene XPD: correlations with risk and age at onset of basal cell carcinoma. *Cancer Epidemiol Biomarkers Prev.* 1999;8(1):77-81.

169. Ganster C, Neesen J, Zehetmayer S, et al. DNA repair polymorphisms associated with cytogenetic subgroups in B-cell chronic lymphocytic leukemia. *Genes, chromosomes & cancer.* 2009;48(9):760-767.

170. Mitra AK, Singh N, Garg VK, et al. Statistically significant association of the single nucleotide polymorphism (SNP) rs13181 (ERCC2) with predisposition to Squamous Cell Carcinomas of the Head and Neck (SCCHN) and Breast cancer in the north Indian population. *J Exp Clin Cancer Res.* 2009;28104.

171. Shi JY, Ren ZH, Jiao B, et al. Genetic variations of DNA repair genes and their prognostic significance in patients with acute myeloid leukemia. *Int J Cancer.* 2011;128(1):233-238.

172. Ahsan H, Chen Y, Wang Q, et al. DNA repair gene XPD and susceptibility to arsenic-induced hyperkeratosis. *Toxicol Lett.* 2003;143(2):123-131.

173. Applebaum KM, Karagas MR, Hunter DJ, et al. Polymorphisms in nucleotide excision repair genes, arsenic exposure, and non-melanoma skin cancer in New Hampshire. *Environ Health Perspect.* 2007;115(8):1231-1236.

174. Thompson LH, Carrano AV, Sato K, et al. Identification of nucleotide-excision-repair genes on human chromosomes 2 and 13 by functional complementation in hamster-human hybrids. *Somatic cell and molecular genetics.* 1987;13(5):539-551.

175. Rossman TG. Enhancement of UV-mutagenesis by low concentrations of arsenite in E. coli. *Mutat Res.* 1981;91(3):207-211.

176. Cui Y, Morgenstern H, Greenland S, et al. Polymorphism of Xeroderma Pigmentosum group G and the risk of lung cancer and squamous cell carcinomas of the oropharynx, larynx and esophagus. *Int J Cancer.* 2006;118(3):714-720.

177. Chang JS, Wrensch MR, Hansen HM, et al. Nucleotide excision repair genes and risk of lung cancer among San Francisco Bay Area Latinos and African Americans. *Int J Cancer.* 2008;123(9):2095-2104.

178. Rajaraman P, Bhatti P, Doody MM, et al. Nucleotide excision repair polymorphisms may modify ionizing radiation-related breast cancer risk in US radiologic technologists. *Int J Cancer.* 2008;123(11):2713-2716.

179. Tomkinson AE, Mackey ZB. Structure and function of mammalian DNA ligases. *Mutat Res.* 1998;407(1):1-9.

180. Lynn S, Lai HT, Gurr JR, et al. Arsenite retards DNA break rejoining by inhibiting DNA ligation. *Mutagenesis.* 1997;12(5):353-358.

181. Li JH, Rossman TG. Inhibition of DNA ligase activity by arsenite: a possible mechanism of its comutagenesis. *Molecular toxicology.* 1989;2(1):1-9.

182. O'Driscoll M, Cerosaletti KM, Girard PM, et al. DNA ligase IV mutations identified in patients exhibiting developmental delay and immunodeficiency. *Molecular cell.* 2001;8(6):1175-1185.

183. Roddam PL, Rollinson S, O'Driscoll M, et al. Genetic variants of NHEJ DNA ligase IV can affect the risk of developing multiple myeloma, a tumour characterised by aberrant class switch recombination. *Journal of medical genetics.* 2002;39(12):900-905.

184. Liu Y, Zhou K, Zhang H, et al. Polymorphisms of LIG4 and XRCC4 involved in the NHEJ pathway interact to modify risk of glioma. *Hum Mutat.* 2008;29(3):381-389.

185. Hill DA, Wang SS, Cerhan JR, et al. Risk of non-Hodgkin lymphoma (NHL) in relation to germline variation in DNA repair and related genes. *Blood.* 2006;108(9):3161-3167.

186. De Chaudhuri S, Ghosh P, Sarma N, et al. Genetic variants associated with arsenic susceptibility: study of purine nucleoside phosphorylase, arsenic (+3) methyltransferase, and glutathione S-transferase omega genes. *Environ Health Perspect.* 2008;116(4):501-505.

187. Valenzuela OL, Drobna Z, Hernandez-Castellanos E, et al. Association of AS3MT polymorphisms and the risk of premalignant arsenic skin lesions. *Toxicol Appl Pharmacol.* 2009;239(2):200-207.

188. Ghosh P, Basu A, Mahata J, et al. Cytogenetic damage and genetic variants in the individuals susceptible to arsenic-induced cancer through drinking water. *Int J Cancer.* 2006;118(10):2470-2478.

189. McCarty KM, Ryan L, Houseman EA, et al. A case-control study of GST polymorphisms and arsenic related skin lesions. *Environ Health.* 2007;65.

190. Banerjee N, Nandy S, Kearns JK, et al. Polymorphisms in the TNF-α and IL10-gene promoters and risk of arsenic-induced skin lesions and other non-dermatological health effects. *Toxicol Sci.* 2011;In press.

191. De Chaudhuri S, Mahata J, Das JK, et al. Association of specific p53 polymorphisms with keratosis in individuals exposed to arsenic through drinking water in West Bengal, India. *Mutat Res.* 2006;601(1-2):102-112.

192. Banerjee M, Sarkar J, Das JK, et al. Polymorphism in the ERCC2 codon 751 is associated with arsenic-induced premalignant hyperkeratosis and significant chromosome aberrations. *Carcinogenesis.* 2007;28(3):672-676.

193. McCarty KM, Smith TJ, Zhou W, et al. Polymorphisms in XPD (Asp312Asn and Lys751Gln) genes, sunburn and arsenic-related skin lesions. *Carcinogenesis.* 2007;28(8):1697-1702.

TABLE 1. Studies evaluating genetic susceptibility assessed by single nucleotide polymorphisms to arsenical skin lesions

First author (year)	Design	Location	Endpoint Assessment	SNPs Assessed	Allele Frequency	No. of Cases	No. of Controls	Measure of Effect	95% CI
PNP									
De Chaudhuri (2008) (186)	Case-control	West Bengal, India	Presence of more than 1 characteristic skin lesion	rs1049562	T=0.13	229	199	OR=1.69 (CT+TT vs CC)	1.08, 2.66
				rs1049564	A=0.12			OR=1.66 (GA+AA vs GG)	1.04, 2.64
				rs1130650	T=0.13			OR=1.67 (CT+TT vs CC)	1.05, 2.66
AS3MT									
De Chaudhuri (2008) (186)	Case-control	West Bengal, India	Presence of more than 1 characteristic skin lesion	rs11191439	C=0.05	229	199	OR=1.02 (TC+CC vs TT)	0.53, 1.98
Valenzuela (2009) (187)	Case-control	Mexico	Presence of at least 1 characteristic skin lesion	rs7085104	G=0.38	71	51	OR=1.60 (AA+AG vs GG)	0.6, 4.3
				rs11191439	C=0.066			OR=4.28 (TC+CC vs TT)	1.0, 18.5
				rs11191453	C=0.21			OR=0.76 (TT vs TC)	0.3, 1.9
GSTO1									
Ahsan (2007) (45)	Case-control	Bangladesh	Presence of at least 1 characteristic skin lesion	rs4925	A=0.18	594	1041	OR=0.98 (CA vs CC) OR=1.73 (AA vs CC)	0.74, 1.29 0.91, 3.30
				rs11509438	T=0.10			OR=0.85 (AT vs AA) OR=0.45 (TT vs AA)	0.60, 1.19 0.14, 1.44
				rs11509437	A=0.10			OR=0.88 (GA vs GG) OR=0.52 (AA vs GG)	0.63, 1.24 0.18, 1.52
De Chaudhuri (2008) (186)	Case-control	West Bengal, India	Presence of more than 1 characteristic skin lesion	rs4925	A=0.13	229	199	OR=1.34 (CA+AA vs CC)	0.84, 2.13
				rs11509437	−=0.03			OR=1.44 (−/AGG vs AGG/AGG)	0.67, 3.17
				rs11509438	A=0.11			OR=0.72 (GA+AA vs AA)	0.45, 1.16

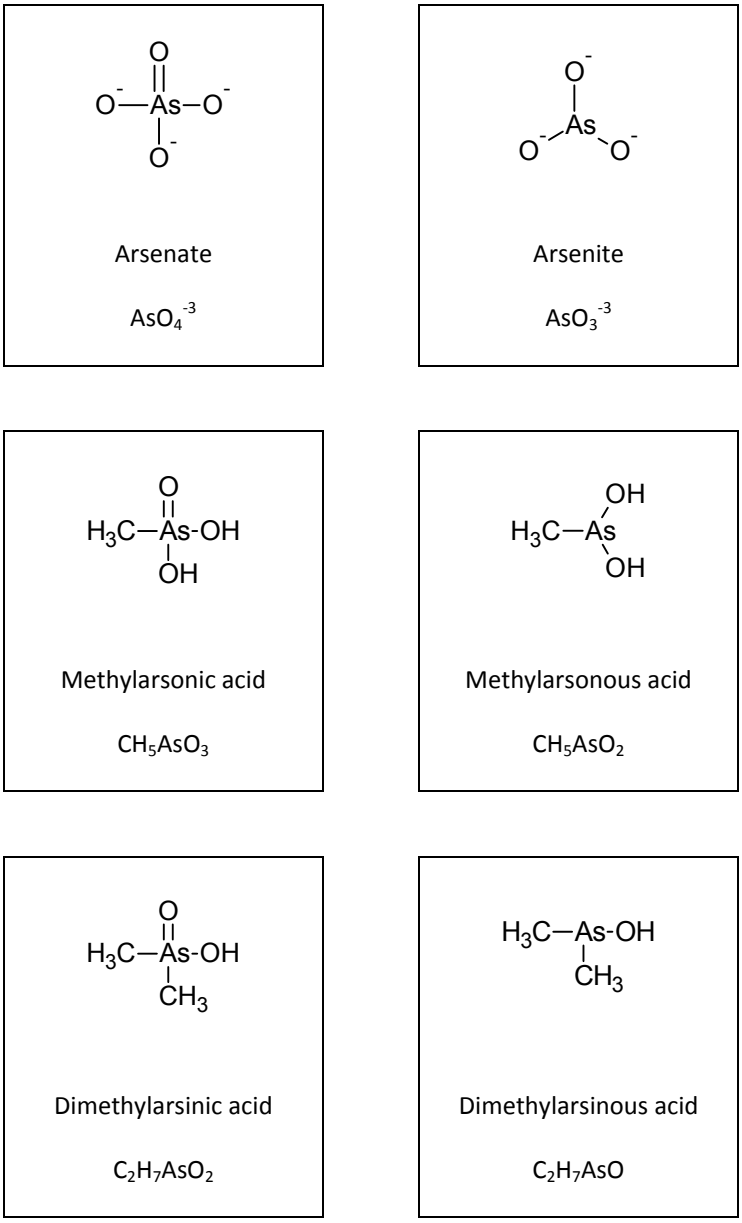
GSTO2									
De Chaudhuri (2008) (186)	Case-control	West Bengal, India	Presence of more than 1 characteristic skin lesion	rs156697	G=0.28	229	199	OR=0.88 (AG+GG vs AA)	0.60, 1.29
GSTP1									
Ghosh (2006) (188)	Case-control	West Bengal, India	Presence of at least 1 characteristic skin lesion	rs1695	G=0.14	86	110	OR=0.91 (AG vs AA) OR=1.26 (GG vs AA)	0.46, 1.80 0.25, 6.40
				rs1138272	T=0.04			OR=2.09 (CT vs CC) OR=N.E. (TT vs CC)	0.67, 6.53 –
McCarty (2007) (189)	Case-control	Bangladesh	Presence of at least 1 characteristic skin lesion	rs1695	G=0.27	592	597	OR=0.89 (AG vs AA) OR=1.86 (GG vs AA)	0.68, 1.17 1.15, 3.00
GSTT1									
Ghosh (2006) (188)	Case-control	West Bengal, India	Presence of at least 1 characteristic skin lesion	GSTT1 null	null=0.13	244	178	OR=0.91 (–/+, +/+ vs –/–)	0.33, 1.47
McCarty (2007) (189)	Case-control	Bangladesh	Presence of at least 1 characteristic skin lesion	GSTT1 null	null=0.18	592	597	OR=1.56 (wt vs null)	1.10, 2.19
GSTM1									
Ghosh (2006) (188)	Case-control	West Bengal, India	Presence of at least 1 characteristic skin lesion	GSTM1 null	null=0.18	244	178	OR=1.73 (–/+, +/+ vs –/–)	1.24, 2.22
McCarty (2007) (189)	Case-control	Bangladesh	Presence of at least 1 characteristic skin lesion	GSTM1 null	null=0.41	592	597	OR=0.99 (wt vs null)	0.77, 1.28

MTHFR									
Ahsan (2007) (45)	Case-control	Bangladesh	Presence of at least 1 characteristic skin lesion	rs1801133	T=0.12	594	1041	OR=1.14 (CT vs CC) OR=1.56 (TT vs CC)	0.84, 1.54 0.57, 4.30
				rs1801131	C=0.37			OR=1.11 (CA vs CC) OR=1.37 (AA vs CC)	0.75, 1.66 0.91, 2.06
TNF									
Banerjee (2011) (190)	Case-control	West Bengal, India	Presence of at least 1 characteristic skin lesion	rs1800629	N.E.[1]	207	190	OR=3.04 (GA/AA vs GG)	1.78, 5.21
IL10									
Banerjee (2011) (190)	Case-control	West Bengal, India	Presence of at least 1 characteristic skin lesion	rs1800890	N.E.[1]	207	190	OR=2.03 (TA/AA vs TT)	1.26, 3.28
TP53									
De Chaudhuri (2006) (191)	Case-control	West Bengal, India	Keratosis	rs17878362[2]	16bp dup=0.13	177	189	OR=2.09 (–/– vs –/ ACCTGGAGGGCTGGGG+ ACCTGGAGGGCTGGGG/ ACCTGGAGGGCTGGGG)	1.26, 3.46
				rs1042522	G=0.43			OR=2.09 (CC vs GG+GC)	1.32, 3.30
				rs1625895	A=0.15			OR=5.31 (AA vs AG+GG)	0.60, 46.96
ERCC2									
Ahsan (2003) (172)	Case-control	Bangladesh	Hyperkeratosis	rs13181	N.E.[1]	29	105	OR=1.7 (AA vs AC/CC)	0.7, 4.3
Banerjee (2007) (192)	Case-control	West Bengal, India	Hyperkeratosis	rs13181	A=0.4	165	153	OR=4.77 (AA vs AC/CC)	2.75, 8.23

[1] N.E., Not estimateable from reported data.
[2] Deletion insertion polymorphism

McCarty (2007) (193)	Case-control	Bangladesh	Presence of at least 1 characteristic skin lesion	rs13181	A=0.4	555	560	OR=0.92 (AC vs AA) OR=0.98 (CC vs AA)	0.69, 1.23 0.66, 1.45
				rs1799793	A=0.3			OR=0.87 (AG vs GG) OR=0.76 (AA vs GG)	0.65, 1.15 0.50, 1.15
APEX1									
Breton (2007) (151)	Case-control	Bangladesh	Presence of at least 1 characteristic skin lesion	rs3136820 (rs1130409)	G=0.26	792	792	OR=0.96 (TG vs TT) OR=1.93 (GG vs TT)	0.75, 1.22 1.15, 3.19
XRCC1									
Breton (2007) (151)	Case-control	Bangladesh	Presence of at least 1 characteristic skin lesion	rs1799782	T=0.11	792	792	OR=1.01 (TC vs CC) OR=0.52 (TT vs CC)	0.76, 1.35 0.17, 1.66
				rs25487	A=0.37			OR=1.17 (GA vs GG) OR=1.39 (AA vs GG)	0.91, 1.52 0.96, 2.03
XRCC3									
Kundu (2011) (163)	Case-control	West Bengal, India	Presence of at least 1 characteristic skin lesion	rs861539	T=0.28	206	215	OR=0.45 (CT/TT vs CC)	0.30, 0.67
OGG1									
Breton (2007) (151)	Case-control	Bangladesh	Presence of at least 1 characteristic skin lesion	rs1052133	G=0.32	792	792	OR=1.04 (GC vs CC) OR=0.87 (GG vs CC)	0.81, 1.32 0.58, 1.32
MPO									
Ahsan (2003) (91)	Case-control	Bangladesh	Hyperkeratosis	rs2333227	A=0.14	30	104	OR=2.1 (GG vs GA+AA)	0.7, 6.2
CAT									
Ahsan (2003) (91)	Case-control	Bangladesh	Hyperkeratosis	rs1001179	T=0.19	30	104	OR=1.9 (CT+TT vs CC)	0.8, 4.7

Figure 1. Chemical structure of inorganic and organic arsenic species



Chapter 2

Arsenic Exposure, Antioxidant Enzymes, and Skin Lesion Prevalence in an Adult Bangladeshi Population

ABSTRACT

One of the proposed pathways that may be related to arsenic toxicity is oxidative stress and therefore, may play a role in the association between arsenic and skin lesion risk. The objective of this analysis was to assess the relationship between SNPs in antioxidant enzyme genes and skin lesion prevalence among 610 population-based prevalent skin lesion cases and 1,079 controls enrolled in the HEALS cohort. There were no statistically significant associations between these SNPs (*SOD2,* rs4880; *CAT,* rs1001179; *GPX1,* rs1050450; and *MPO*, rs2333227) and skin lesion prevalence. The results did not vary by arsenic exposure levels (as measured by well water arsenic concentration, urinary total arsenic concentration, primary methylation index, and secondary methylation index), body mass index, smoking status, or fruit and vegetable intake. However, there was marginal evidence that skin lesion prevalence was increased among individuals who carried 4 or more risk alleles compared to individuals carrying 0-3 risk alleles in these SNPs. Additionally, we observed a significant departure from additivity for the risk allele score and primary methylation index on skin lesion prevalence. There was no evidence that these SNPs were associated with skin lesion severity, such as pigmentation changes and keratosis. In summary, there is some evidence of a cumulative effect of these antioxidant SNPs that should be explored further, particularly in conjunction with arsenic methylation ability.

INTRODUCTION

While arsenic is a well-established human carcinogen (1) and has additionally been associated with an array of chronic diseases (2-15), the underlying mechanism of arsenic toxicity has not yet been determined. Dose-response associations between arsenic exposure with both prevalent and incident skin lesions have been observed (16, 17). One of the proposed pathways that may modify the association between arsenic and skin lesion risk is oxidative stress (18, 19). Arsenic has been shown to induce reactive oxygen species (ROS) (20). Oxidative stress is hypothesized to play a role in initiation, promotion, and progression within the framework of a multistage carcinogenesis model (21). Matsui et al. conducted a study of 8-hydroxy-2'-deoxyguanosine (8-OHdG), a well established marker of oxidative stress, in skin tissue and found 8-OHdG to be significantly higher in arsenic-related Bowen's disease (squamous cell carcinoma in situ) as compared to arsenic-unrelated Bowen's disease (22). Similar findings were seen by An et al., who additionally showed that 8-OHdG was elevated in the normal and keratotic skin tissue of arsenic-exposed individuals (23). The antioxidant defense system is comprised of a cell's enzymatic and non-enzymatic antioxidants (18). The construct measured for the purposes of this analysis is the enzymatic antioxidant potential of the individual based on SNPs in genes selected from the oxidative stress pathway.

DNA was genotyped for four SNPs in the following antioxidant enzyme genes: superoxide dismutase 2, mitochondrial (*SOD2;* rs4880); catalase (*CAT;* rs1001179); glutathione peroxidase 1 (*GPX1;* rs1050450); and myeloperoxidase (*MPO*; rs2333227). SOD2 protein plays a major role in maintaining oxidative balance by converting

superoxide (O_2^-)—a precursor molecule for all other ROS—into hydrogen peroxide and oxygen (24). CAT protein plays a major role at times of severe oxidative stress by converting hydrogen peroxide into water and oxygen (25). GPX1 is the main antioxidant enzyme in normal conditions that functions in the detoxification of hydrogen peroxide by using it to oxidize glutathione (26). MPO has been implicated in oxidative stress because of its utilization of hydrogen peroxide to produce the oxidant hypochlorite, which may lead to oxidative damage of biological tissues (27). Additionally, prior *in vitro* experiments have shown SOD2 (28, 29), CAT (20, 28, 30), GPx (31, 32), and MPO (29) to play roles in mitigating arsenic-induced DNA damage from oxidative stress.

In this study, we evaluate whether SNPs in genes encoding antioxidant enzymes (*SOD2, GPX1, CAT, MPO*) are associated with prevalent arsenical skin lesion status in 610 cases and 1,079 controls from the Health Effects of Arsenic Longitudinal Study (HEALS) cohort.

MATERIALS AND METHODS

Study population

HEALS is an ongoing, population-based cohort study examining both the short- and long-term health effects of arsenic exposure. The study was launched in Araihazar, Bangladesh, in 2000. The selection of cohort participants, study design and methods have been described in detail elsewhere (33).

The HEALS cohort served as the study population for the prevalent cases and controls utilized in these analyses. Between October 2000 and May 2002, married

individuals were sampled—an eligibility criteria to minimize loss to follow-up—who were aged 18–75 years and residents of the study area for at least 5 years. From the 12,050 residents who met the eligibility criteria from an enumerated total 65,876 persons in the study area, 11,746 (97.5% response rate) men and women (4,801 married couples and 2,144 married individuals whose spouses did not participate) were enrolled into the HEALS cohort. Among the 11,746 enrolled baseline cohort participants, 11,224 participants (95.6%) provided urine samples. At baseline, 610 cases of skin lesions were identified among the 11,224 cohort members who provided urine samples and were included as prevalent cases in this study. Of the remaining participants with an available urine sample and known to be free of skin lesions at baseline (n=10,614), a ~10% random sample (n=1,079) was selected and included as controls in this study.

Structured baseline questionnaires were administered to participants in Bengali by trained interviewers, who were blind to the water arsenic concentration of the participants' wells. The questionnaires assessed sociodemographic characteristics, current and past tubewell use, typical water consumption patterns, food frequency of thirty-nine items common to the population (34), occupational exposures, and smoking habits. The study physicians also conducted a clinical examination, which included a skin evaluation for the detection of arsenical skin lesions based on a structured protocol (16, 33). Venous blood and spot urine samples were also collected from each participant on the day of the interview.

The study protocol was approved by the Institutional Review Boards of Columbia University, The University of Chicago, and the Bangladesh Medical Research Council. Informed consent was obtained from all participants prior to baseline interview.

Arsenic exposure assessment

Three primary measures of arsenic exposure were estimated for each case-control participant: well water arsenic concentration, creatinine-adjusted urinary total arsenic concentration, and urinary arsenic metabolites.

At the baseline survey, participants were asked to identify the well they primarily used as their source of drinking water, from which we were able to assign the appropriate well water arsenic concentration exposure. Well water arsenic concentrations of all 5,966 tubewells in the study area were measured by graphite furnace atomic absorption spectrometry, with a detection limit of 5 μg/L. Samples below the limit of detection were subsequently reanalyzed by inductively coupled plasma-mass spectrometry (ICP-MS), with a detection limit of 0.1 μg/L (35).

Urinary total arsenic concentration was measured in a spot urine sample collected at baseline by graphite furnace atomic absorption spectrometry, with a detection limit of 2 μg/L (36). Urinary creatinine was measured by a colorimetric Sigma Diagnostics Kit (Sigma, St. Louis, MO), and urinary total arsenic was subsequently divided by creatinine to obtain a creatinine-adjusted urinary total arsenic concentration, expressed as μg/g creatinine (37).

Urinary arsenic metabolites were measured based on the method by Reuter et al. (38) using high-performance liquid chromatography separation of arsenobetaine, arsenocholine, arsenate (As^{V}), arsenite (As^{III}), total monomethyl arsenic (MMA^{III} + MMA^{V}), and total dimethyl arsenic (DMA^{III} + DMA^{V}) followed by detection by ICP-MS-dynamic reaction cell. Total inorganic arsenic (As^{III} + As^{V}) is utilized in these analyses since As^{III} can oxidize to As^{V} during sample transport, storage, and preparation.

For purposes of analysis, well water arsenic was categorized into quartiles based on the distribution in the controls; however, since the first and second quartiles roughly corresponded to the World Health Organization's guideline for arsenic in drinking water (10 μg/L) and the national standard for arsenic in drinking water in Bangladesh (50 μg/L), respectively, we adjusted the cutoff points slightly to correspond to these regulatory levels. Urinary total arsenic concentration and arsenic metabolite metrics were quartiled based on the distribution in the controls. The percentages of MMA, DMA and inorganic arsenic were calculated after exclusion of arsenobetaine and arsenocholine from the total. In addition, two methylation indices were constructed: primary methylation index (PMI)—the ratio of MMA to inorganic arsenic—and secondary methylation index (SMI)—the ratio of DMA to MMA.

Individuals who were missing data on a particular arsenic exposure measure were excluded from the analysis of that specific exposure. Well water arsenic and urinary total arsenic concentrations were available for all cases and controls. Arsenic metabolite data was missing for 18 case subjects and 37 control subjects.

Single nucleotide polymorphism assessment

High-throughput DNA extraction was completed in 96-well format using the QIAmp DNA 96 DNA Blood kit (Qiagen, Valencia, CA). Replica plates were made with 12.5 ng DNA in 2.5 μL per well. In the first step, the genomic DNA was amplified by PCR using appropriate primers. After PCR amplification, the primers and deoxynucleotide triphosphates in 10 μL PCR product were digested with the 10 μL shrimp alkaline phosphatase cocktail containing 1.0 μL (1 unit/μL) of shrimp alkaline phosphatase, 0.1 μL of Escherichia coli exonuclease I (10 units/μL; U.S. Biochemical, Cleveland, OH), 1.0 μL of 10x shrimp alkaline phosphatase buffer, and 7.9 μL of DNase and RNase-free water for 45 min at 37°C followed by heating at 95°C for 15 min for enzyme deactivation. Then, single nucleotide extension was carried out in the presence of the appropriate allele-specific dideoxynucleotide triphosphates fluorescence labeled with either R110 or TAMRA (PerkinElmer, Waltham, MA). For single nucleotide extension reactions, both the forward and reverse probes were initially tested to select the better probe based on clear signal clustering. Reaction mixture (13 μL/well) containing 0.025 μL AcycloPrime enzyme, 0.5 μL terminator dye, 1 μL reaction buffer, 0.25 μL extension probe (10 pmol/μL), and 11.225 μL water was added to 7 μL of digested PCR product to make 20 μL reaction volume. Thermocycling was done at 95°C for 3 min followed by optimum number of cycles of 95°C for 15 s and 55°C for 30 s. Finally, the fluorescence was measured with Wallac 1420 Multilabel Counter Victor 3 (PerkinElmer, Waltham, MA). In addition to our assay-specific quality control samples, 10% of the samples were run in duplicate after relabeling to keep laboratory researchers blinded to its identity.

Concordance based on the duplicates was >0.92. Call rates for the SNPs of interest ranged from 95.2% to 97.9%, as shown in **Table 1**.

Skin lesion assessment

Arsenical skin lesion assessment was conducted through skin examination at the baseline interview following a structured protocol by a trained study physician (16, 33). Arsenical skin lesions were categorized as the presence of melanosis on the body surface, leucomelanosis on the body surface, and keratosis on the hands or feet. For the purposes of this study, cases were selected as having the presence of any type of arsenical skin lesion.

The distribution of skin lesion severity among the 610 cases was as follows: 359 cases had melanosis only, 20 had leucomelanosis only, 10 had melanosis and leucomelanosis, 170 had melanosis and keratosis, 40 had leucomelanosis and keratosis, and 11 had melanosis, leucomelanosis, and keratosis.

Individuals without keratotic arsenical lesions are generally thought to have a milder form of skin lesions, manifest as changes in the dermal pigmentation of the skin. This suggests that there is some heterogeneity of the skin lesion definition; therefore, subset analyses were conducted to evaluate the association of SNPs with skin lesion severity. Skin lesion severity was stratified into absence of keratotic skin lesions (n=389) and presence of keratotic skin lesions (n=221).

Covariates

All covariate data was derived from the baseline interview. Demographic factors included sex and reported age at baseline. Self-reported smoking history was ascertained and categorized as ever versus never smoker for the purposes of these analyses. At the time of the baseline interview, height and weight of each participant was measured by the study physician; body mass index was constructed as weight in kilograms divided by height in meters squared and categorized by the World Health Organization cutoff points for underweight, normal weight, and overweight/obese.

Fruit and vegetable consumption as well as total caloric intake was constructed based on the reported intake from the food frequency questionnaire (FFQ) (34). Participants were exclude from this analysis if they did not complete or reported an unlikely total caloric intake value on the FFQ (<500 or >4000 kcal/ day; n=47 missing and n=29 extreme values (total n=76 excluded)). A single variable for total fruit and vegetable intake was constructed by summing the average daily amount of fruit and vegetables consumed that was reported for 26 food items (**Table B.1**). The total fruit and vegetable intake was then tertiled based on the distribution in controls for the purposes of these analyses.

Statistical analysis

Hardy Weinberg Equilibrium (HWE) was calculated for the controls based on Pearson (χ^2) tests and α. Violation of HWE is calculated based on statistical departure

from expected HWE frequencies based on the exact test. α is used to denote the magnitude of departure from HWE.

$$\alpha = \frac{1}{2}\log\left(\frac{4P_{gg}P_{GG}}{P_{Gg}^2}\right)$$

where P_{gg}, P_{Gg}, and P_{GG} are the genotype proportions (39).

The association between each SNP and arsenical skin lesion status was estimated by prevalence odds ratios (PORs) and their 95% confidence intervals (CIs) from logistic regression models employed by the LOGISTIC procedure in SAS. The PORs were adjusted for sex, age, and well water arsenic concentration, primarily because of the strong association of these covariates with the outcome. The genotype-specific PORs were estimated for each SNP using the homozygous major genotype as the reference category. Additionally, the per allele POR was estimated assuming a log-additive model. In these models, a single ordinal variable was included indicating the number of minor alleles (0, 1, 2) and the associated Wald statistic was interpreted as the P for trend.

A summary risk allele count was created for the measured biallelic SNPs by summing the number of risk alleles carried by an individual, based on *a priori* knowledge of the SNP alleles on protein structure and function (SOD2=T allele, GPX1=T allele, CAT=C allele, MPO=A allele). For purposes of analysis, the risk allele count was dichotomized.

We also evaluated interactions of genetic factors with arsenic exposure (as measured by well water arsenic and urinary total arsenic concentrations) as well as arsenic methylation capacity in relation to arsenical skin lesions on the additive scale

using both genotype status and risk allele count in separate analyses. Multivariate adjusted estimates were used to estimate the relative excess risk for interaction (RERI), calculated as

$$RERI = \exp(\beta1+\beta2+\beta3) - \exp(\beta1) - \exp(\beta2) +1.$$

Here β1 is the coefficient of the ordinal arsenic exposure measure, β2 is the coefficient of the dichotomous SNP effect modifier measure, and β3 is the coefficient of the cross-product of the ordinal arsenic exposure and dichotomous SNP measure (40, 41). Bias corrected and accelerated (BCa) 95% CIs of the RERI were estimated via 1000 bootstrap samples. CIs of the RERI were also calculated using the delta method described by Hosmer and Lemeshow with similar results (not shown) (42). We also evaluated effect modification of the associations between lifestyle factors (BMI, smoking status, and fruit and vegetable intake) and skin lesions by each SNP and risk allele count in separate analyses on the additive scale, adjusting for well water arsenic exposure and other covariates. In these analyses, β1 is the coefficient of the ordinal lifestyle measure, β2 is the coefficient of the dichotomous SNP effect modifier measure, and β3 is the coefficient of the cross-product of the ordinal lifestyle measure and dichotomous SNP effect modifier—with the exception of smoking status that was modeled as a dichotomous lifestyle measure.

In exploratory subset analyses, skin lesion severity was evaluated by stratifying arsenical skin lesion status into absence of keratotic skin lesions and presence of keratotic skin lesions. Prevalent skin lesion cases with melanosis and/or leucomelanosis but not keratosis (n=389) were classified as non-keratotic skin lesions and, individuals with

keratosis (n=221) were classified as keratotic skin lesions. Ordered polytomous regression models were employed by the LOGISTIC procedure in SAS, comparing non-keratotic skin lesions (coded=1) and keratotic skin lesions (coded=2) to the controls (coded=0, reference category).

RESULTS

Sample characteristics

Information on the genotyped antioxidant enzyme SNPs is shown in **Table 1**. Based on the α statistic, no SNPs were observed to deviate from Hardy-Weinberg equilibrium indicating no major evidence of genotyping errors for each SNP. The distribution of selected characteristics in the total HEALS cohort and by the 610 prevalent skin lesion cases and 1,079 controls are shown in **Table 2.** Since a random sample of the baseline cohort without skin lesions was selected as control participants for genotyping, the distribution of characteristics in the controls is not appreciably different from the total baseline cohort except for the distribution of age, where we see that control subjects were underrepresented in the 31–40 age range and overrepresented in the 41–50 age range relative to the total cohort. Similar to the previous HEALS cohort analysis for skin lesion prevalence (16), we observed based on the selected cases and controls for this analysis that males, older age, low BMI (<18.5 kg/m^2), current or past cigarette smoking, and no formal education were associated with increased skin lesion prevalence. Additionally, clear dose-response associations were observed with arsenic exposure as

measured by well water arsenic and urinary total arsenic concentrations (**Table 2**), as well as with the distribution of urinary arsenic metabolites (**Table 3**).

Evaluation of associations with single and cumulative SNP effects

No significant associations were observed between each of the antioxidant enzyme SNPs and skin lesion prevalence, based on the genotype and per allele POR estimates (**Table 4**). A risk allele score for the genotyped SNPs was constructed based on *a priori* evidence of the risk allele for each SNP and showed evidence of marginal increased risk when 4 or more risk alleles were present (adjusted POR=1.2, 95% CI=0.9, 1.6; **Table 5**). These results did not vary by skin lesion severity, classified as non-keratotic and keratotic skin lesions (**Table B.2**). Additionally, in exploratory analyses we evaluated these associations in male participants only since their baseline skin lesion prevalence was appreciably higher than females and, we did not observe notable differences in the distribution of SNPs or the associations of the SNPs with skin lesion prevalence in males as compared to the total study sample, precluding evidence of interaction by sex (**Table B.3**).

Additive interaction evaluation

We evaluated whether the associations between antioxidant enzyme SNPs and skin lesion prevalence were modified by various arsenic exposure measures and lifestyle factors on the additive scale. Departure from additivity was evaluated by well water arsenic concentration (**Table 6**), creatinine-adjusted urinary total arsenic concentration

(**Table 7**), primary methylation index (**Table 8**), secondary methylation index (**Table 9**), BMI (**Table 10**), cigarette smoking status (**Table 11**), and total fruit and vegetable intake (**Table 12**). A significant departure from additivity was observed for the risk allele score and primary methylation index on skin lesion prevalence (RERI=0.22, 95% CI=0.03, 0.36; **Table 8**). On the additive scale, we observed that skin lesion prevalence was greater in individuals carrying 4 or more risk alleles with each 1 quartile increase in primary methylation index than would be expected based on the additive independent effects of risk allele score and primary methylation index alone. We did not observed any other evidence of interaction on the additive scale by measures of arsenic exposure, arsenic metabolism, or lifestyle factors as characterized by RERI estimates that were not appreciably different from zero.

DISCUSSION

In this population-based case-control study of arsenical skin lesion prevalence in a Bangladeshi population, we systematically evaluated four candidate SNPs in antioxidant enzymes. There were no statistically significant associations between these SNPs (*SOD2,* rs4880; *CAT,* rs1001179; *GPX1,* rs1050450; and *MPO*, rs2333227) and skin lesion prevalence. However, there was marginal evidence that skin lesion prevalence was increased among individuals who carried 4 or more risk alleles compared to individuals carrying 0-3 risk alleles in these SNPs. Additionally, we observed a significant departure from additivity for the risk allele score and primary methylation index on skin lesion prevalence.

Our observation of significant modification of the association between primary methylation index and skin lesion prevalence by the risk allele score on the additive scale is consistent with evidence from the current literature. Individuals with a higher proportion of MMA to inorganic arsenic in their urine (represented by a larger primary methylation index score) were observed to have increased prevalence of arsenical skin lesions compared to those with a lower relative proportion, which is consistent with other studies examining the association between skin lesion prevalence and urinary arsenic metabolites (16, 43-47) as well as blood arsenic metabolites (48). Notably, a cross-sectional study among arsenic-exposed individuals showed that lower primary methylation index was associated with lower plasma antioxidant capacity (49). Moreover, *in vitro* studies have shown that MMA induces reactive oxygen species, particularly the trivalent form of this organic arsenic species (50, 51). In the recent study by Bailey et al (51), *in vitro* experiments using normal human epidermal keratinocytes showed that MMA^{III} exhibited the greatest potential for skin carcinogenesis through the induction of oxidative stress, increased transcript levels of keratinocyte growth factors, and modulation of MAPK and NF-κB pathways, compared to the trivalent forms of inorganic arsenic and DMA. In a comparative functional genomics analysis in yeast, Jo et al (52) showed that glutathione (related to cellular antioxidant status) was more important in MMA^{III} toxicity than trivalent inorganic arsenic toxicity. Therefore, the fact that an interaction was observed only with the primary methylation index and not other measures of arsenic exposure that do not reflect the distribution of arsenic metabolites but rather capture the total burden of arsenic exposure, is quite consistent with the fact that

oxidative stress may be a mechanism that is associated only with MMA—a construct that is best captured by the primary methylation index. Thus, individuals who do not fully methylate arsenic efficiently and are carriers of several risk allele variants in antioxidant enzyme genes could potentially be at increased risk due to this gene-environment interaction.

In pilot work conducted by Ahsan et al (53) in the same cohort population as this analysis but using a non-overlapping set of prevalent skin lesion cases with keratosis and control subjects, non-significant increased risks were shown for the GG genotype of MPO (rs2333227) and the CT/TT genotype of CAT (rs1001179). Based on the polytomous regression analysis conducted in this current study, stratifying cases by presence of keratosis did not reveal any significant difference in estimates between keratotic versus non-keratotic cases for the SNPs. However, examination of the PORs from this analysis showed a trend toward an increased risk for the T allele of CAT (per allele POR=1.3, 95% CI=1.0,1.7) among keratotic skin lesion cases compared to controls, which was not observed among non-keratotic skin lesion cases compared to controls (POR=0.8, 95% CI=0.8, 1.1). Examination of the PORs for MPO showed less appreciable differences in the risk estimates for the A allele (non-keratotic skin lesions POR=0.9, 95% CI=0.7, 1.2; keratotic skin lesions POR=1.2, 95% CI=0.9, 1.7), with a more notable difference in estimates among individuals with the AA genotype (non-keratotic skin lesions POR=1.1, 95% CI=0.4, 3.0; keratotic skin lesions POR=1.7, 95% CI=0.5, 5.1). However, only 5 keratotic cases had the AA genotype which urges caution for the interpretation of the POR estimate. Additionally, increased prevalence for MPO

would potentially be suggested for the A allele in this study, which is the opposite of what was observed previously for the G allele by Ahsan et al. Since both the pilot work and the stratified analysis in this study were based on a small number of cases (n=30 and n=221, respectively), further investigation of these potential findings is warranted in a larger sample of well-defined keratotic skin lesion cases to draw more definitive conclusions.

The major strengths of this study were the measurement of genetic variants, large size of the study sample, the wide range of arsenic exposure, and the multiple constructs of arsenic exposure. The advantage of measuring genetic variants is because they can be measured with little measurement error and bias. Additionally, SNPs are time-independent measures (54), which strengthens causal inference of the associations evaluated in this analysis between the measured SNPs and prevalent skin lesions. Conversely, the amount of oxidative stress could have been measured but there would have been major limitations in the temporality of these measures with the use of prevalent cases. Additionally, in very early or mild cases of skin lesions (some forms of melanosis) may not be readily visible with clinical examination; therefore, oxidative stress biomarkers could also suffer from reverse causation if it was not ensured that the measures were taken at an appropriate period before disease onset. A major strength of this study is the assessment of arsenic exposure. Arsenic exposure was ascertained based on well water arsenic concentration, as well as urinary arsenic measures including total arsenic concentration and arsenic metabolites, which allowed us to evaluate various

constructs of exposure including total exposure burden as well as exposure methylation ability.

While there are many strengths of this study, there are several limitations that are acknowledged. The antioxidant defense system is comprised of enzymatic and non-enzymatic antioxidants. The focus of this study was to examine the influence of polymorphisms of selected enzymatic antioxidants. The role of non-enzymatic antioxidants and activity of enzymatic antioxidants are unmeasured; therefore, the total antioxidant potential of the cell remains unknown. Additionally, arsenic serves as a proxy measure to the actual measure of interest, which is arsenic-induced oxidative stress. There is evidence from prior studies to suggest that MMA may be the most relevant arsenic species to the oxidative stress pathway, as discussed previously; however, we can only infer from this work that individuals with a higher concentration of MMA experience higher levels of oxidative stress and have not measured it directly. Alternatively, biomarkers of oxidative stress could have been measured. The assessment of the effect of SNPs on prevalent skin lesions is also limited for the evaluation of mechanistic hypotheses and phenotype considerations since the association may be related to disease prevalence or progression and not disease incidence. Therefore, findings from these analyses should be replicated with incident cases. However, we deem the effect of disease prevalence to be minimal since arsenical skin lesions are typically not fatal. Additionally, in the subset analysis we conducted to evaluate the associations of SNPs with arsenical skin lesion severity, it did not appear that these SNPs were related to disease progression. Finally, the SNPs and genes for which there is data available were

selected based on a candidate-gene approach, and only a single SNP in each gene of interest was examined. Future studies should take a more comprehensive genomewide approach for evaluating genetic variants in the antioxidant enzyme pathway in relation to skin lesion status. Findings from this type of "discovery" approach could then be evaluated in candidate-gene studies employing SNPs which comprehensively tag the genes of interest.

The findings of this study have potentially important scientific and public health implications for arsenic in drinking water. Prior epidemiologic research has suggested oxidative stress as a potential mechanism of arsenic toxicity, which observations from this study suggest is a viable mechanism for arsenic toxicity and warrants further investigation. Additionally, individuals deficient in arsenic metabolism (as characterized by the primary methylation index) and carried multiple risk alleles in antioxidant enzymes had increased prevalence of skin lesions, which highlights potential pathways for intervention in individuals with manifest skin lesions.

In conclusion, there was marginal evidence that skin lesion prevalence was increased among individuals who carried 4 or more risk alleles compared to individuals carrying 0-3 risk alleles in antioxidant enzyme SNPs (*SOD2,* rs4880; *CAT,* rs1001179; *GPX1,* rs1050450; and *MPO*, rs2333227). Additionally, we observed a significant departure from additivity for the risk allele score and primary methylation index on skin lesion prevalence, indicating that skin lesion prevalence was greater in individuals carrying 4 or more risk alleles and with poorer arsenic methylation ability than was expected based on the additive independent effects of these characteristics alone. Further

investigation into cumulative effects of variants in antioxidant enzyme genes is warranted, particularly among individuals deficient in arsenic methylation.

REFERENCES

1. IARC Working Group on the Evaluation of Carcinogenic Risks to Humans. Some Drinking-water Disinfectants and Contaminants, including Arsenic. Volume 84. Lyon, France, 2004.

2. Smith AH, Goycolea M, Haque R, et al. Marked increase in bladder and lung cancer mortality in a region of Northern Chile due to arsenic in drinking water. *Am J Epidemiol.* 1998;147(7):660-669.

3. Chen CJ, Chen CW, Wu MM, et al. Cancer potential in liver, lung, bladder and kidney due to ingested inorganic arsenic in drinking water. *British journal of cancer.* 1992;66(5):888-892.

4. Hopenhayn-Rich C, Biggs ML, Fuchs A, et al. Bladder cancer mortality associated with arsenic in drinking water in Argentina. *Epidemiology.* 1996;7(2):117-124.

5. Hopenhayn-Rich C, Biggs ML, Smith AH. Lung and kidney cancer mortality associated with arsenic in drinking water in Cordoba, Argentina. *Int J Epidemiol.* 1998;27(4):561-569.

6. Tseng WP. Effects and dose--response relationships of skin cancer and blackfoot disease with arsenic. *Environ Health Perspect.* 1977;19109-119.

7. Brouwer OF, Onkenhout W, Edelbroek PM, et al. Increased neurotoxicity of arsenic in methylenetetrahydrofolate reductase deficiency. *Clinical neurology and neurosurgery.* 1992;94(4):307-310.

8. Navas-Acien A, Sharrett AR, Silbergeld EK, et al. Arsenic exposure and cardiovascular disease: a systematic review of the epidemiologic evidence. *Am J Epidemiol.* 2005;162(11):1037-1049.

9. States JC, Srivastava S, Chen Y, et al. Arsenic and cardiovascular disease. *Toxicol Sci.* 2009;107(2):312-323.

10. Tseng CH. Cardiovascular disease in arsenic-exposed subjects living in the arseniasis-hyperendemic areas in Taiwan. *Atherosclerosis.* 2008;199(1):12-18.

11. Ahmad SA, Sayed MH, Barua S, et al. Arsenic in drinking water and pregnancy outcomes. *Environ Health Perspect.* 2001;109(6):629-631.

12. Milton AH, Smith W, Rahman B, et al. Chronic arsenic exposure and adverse pregnancy outcomes in bangladesh. *Epidemiology.* 2005;16(1):82-86.

13. von Ehrenstein OS, Guha Mazumder DN, Hira-Smith M, et al. Pregnancy outcomes, infant mortality, and arsenic in drinking water in West Bengal, India. *Am J Epidemiol.* 2006;163(7):662-669.

14. Milton AH, Rahman M. Respiratory effects and arsenic contaminated well water in Bangladesh. *International journal of environmental health research.* 2002;12(2):175-179.

15. Parvez F, Chen Y, Brandt-Rauf PW, et al. Nonmalignant respiratory effects of chronic arsenic exposure from drinking water among never-smokers in Bangladesh. *Environ Health Perspect.* 2008;116(2):190-195.

16. Ahsan H, Chen Y, Parvez F, et al. Arsenic exposure from drinking water and risk of premalignant skin lesions in Bangladesh: baseline results from the Health Effects of Arsenic Longitudinal Study. *Am J Epidemiol.* 2006;163(12):1138-1148.

17. Argos M, Kalra T, Pierce B, et al. A prospective study of arsenic exposure from drinking water and incidence of skin lesions in Bangladesh. *American Journal of Epidemiology.* In press.

18. Valko M, Rhodes CJ, Moncol J, et al. Free radicals, metals and antioxidants in oxidative stress-induced cancer. *Chem Biol Interact.* 2006;160(1):1-40.

19. Shi H, Shi X, Liu KJ. Oxidative mechanism of arsenic toxicity and carcinogenesis. *Mol Cell Biochem.* 2004;255(1-2):67-78.

20. Kessel M, Liu SX, Xu A, et al. Arsenic induces oxidative DNA damage in mammalian cells. *Mol Cell Biochem.* 2002;234-235(1-2):301-308.

21. Klaunig JE, Kamendulis LM. The role of oxidative stress in carcinogenesis. *Annu Rev Pharmacol Toxicol.* 2004;44239-267.

22. Matsui M, Nishigori C, Toyokuni S, et al. The role of oxidative DNA damage in human arsenic carcinogenesis: detection of 8-hydroxy-2'-deoxyguanosine in arsenic-related Bowen's disease. *J Invest Dermatol.* 1999;113(1):26-31.

23. An Y, Gao Z, Wang Z, et al. Immunohistochemical analysis of oxidative DNA damage in arsenic-related human skin samples from arsenic-contaminated area of China. *Cancer Lett.* 2004;214(1):11-18.

24. Skrzycki M, Majewska M, Podsiad M, et al. Expression and activity of superoxide dismutase isoenzymes in colorectal cancer. *Acta biochimica Polonica.* 2009;56(4):663-670.

25. Kinnula VL, Everitt JI, Mangum JB, et al. Antioxidant defense mechanisms in cultured pleural mesothelial cells. *American journal of respiratory cell and molecular biology.* 1992;7(1):95-103.

26. Halliwell B. Free radicals and metal ions in health and disease. *The Proceedings of the Nutrition Society.* 1987;46(1):13-26.

27. Babior BM. Phagocytes and oxidative stress. *The American journal of medicine.* 2000;109(1):33-44.

28. Nordenson I, Beckman L. Is the genotoxic effect of arsenic mediated by oxygen free radicals? *Hum Hered.* 1991;41(1):71-73.

29. Wang TS, Hsu TY, Chung CH, et al. Arsenite induces oxidative DNA adducts and DNA-protein cross-links in mammalian cells. *Free Radic Biol Med.* 2001;31(3):321-330.

30. Wang TS, Huang H. Active oxygen species are involved in the induction of micronuclei by arsenite in XRS-5 cells. *Mutagenesis.* 1994;9(3):253-257.

31. Wang TS, Shu YF, Liu YC, et al. Glutathione peroxidase and catalase modulate the genotoxicity of arsenite. *Toxicology.* 1997;121(3):229-237.

32. Lee TC, Ho IC. Modulation of cellular antioxidant defense activities by sodium arsenite in human fibroblasts. *Archives of toxicology.* 1995;69(7):498-504.

33. Ahsan H, Chen Y, Parvez F, et al. Health Effects of Arsenic Longitudinal Study (HEALS): description of a multidisciplinary epidemiologic investigation. *J Expo Sci Environ Epidemiol.* 2006;16(2):191-205.

34. Chen Y, Ahsan H, Parvez F, et al. Validity of a food-frequency questionnaire for a large prospective cohort study in Bangladesh. *The British journal of nutrition.* 2004;92(5):851-859.

35. Cheng Z, Zheng Y, Mortlock R, et al. Rapid multi-element analysis of groundwater by high-resolution inductively coupled plasma mass spectrometry. *Anal Bioanal Chem.* 2004;379(3):512-518.

36. Nixon DE, Mussmann GV, Eckdahl SJ, et al. Total arsenic in urine: palladium-persulfate vs nickel as a matrix modifier for graphite furnace atomic absorption spectrophotometry. *Clin Chem.* 1991;37(9):1575-1579.

37. Nermell B, Lindberg AL, Rahman M, et al. Urinary arsenic concentration adjustment factors and malnutrition. *Environmental research.* 2008;106(2):212-218.

38. Reuter W, Davidowski L, Neubauer K. Speciation of five arsenic compounds in urine by HPLC/ICP-MS. Perkin Elmer Application Notes: Norwalk, CT: Perkin-Elmer, 2003.

39. Minelli C, Thompson JR, Abrams KR, et al. How should we use information about HWE in the meta-analyses of genetic association studies? *Int J Epidemiol.* 2008;37(1):136-146.

40. Knol MJ, van der Tweel I, Grobbee DE, et al. Estimating interaction on an additive scale between continuous determinants in a logistic regression model. *Int J Epidemiol.* 2007;36(5):1111-1118.

41. Li R, Chambless L. Test for additive interaction in proportional hazards models. *Annals of epidemiology.* 2007;17(3):227-236.

42. Hosmer DW, Lemeshow S. Confidence interval estimation of interaction. *Epidemiology.* 1992;3(5):452-456.

43. Rahman M, Vahter M, Sohel N, et al. Arsenic exposure and age and sex-specific risk for skin lesions: a population-based case-referent study in Bangladesh. *Environmental health perspectives.* 2006;114(12):1847-1852.

44. Lindberg AL, Rahman M, Persson LA, et al. The risk of arsenic induced skin lesions in Bangladeshi men and women is affected by arsenic metabolism and the age at first exposure. *Toxicology and applied pharmacology.* 2008;230(1):9-16.

45. Hopenhayn-Rich C, Biggs ML, Smith AH, et al. Methylation study of a population environmentally exposed to arsenic in drinking water. *Environ Health Perspect.* 1996;104(6):620-628.

46. Tseng CH. Blackfoot disease and arsenic: a never-ending story. *J Environ Sci Health C Environ Carcinog Ecotoxicol Rev.* 2005;23(1):55-74.

47. Hsueh YM, Huang YL, Huang CC, et al. Urinary levels of inorganic and organic arsenic metabolites among residents in an arseniasis-hyperendemic area in Taiwan. *Journal of toxicology and environmental health.* 1998;54(6):431-444.

48. Jackson R, Grainge JW. Arsenic and cancer. *Canadian Medical Association journal.* 1975;113(5):396-401.

49. Wu MM, Chiou HY, Wang TW, et al. Association of blood arsenic levels with increased reactive oxidants and decreased antioxidant capacity in a human population of northeastern Taiwan. *Environ Health Perspect.* 2001;109(10):1011-1017.

50. Nesnow S, Roop BC, Lambert G, et al. DNA damage induced by methylated trivalent arsenicals is mediated by reactive oxygen species. *Chem Res Toxicol.* 2002;15(12):1627-1634.

51. Bailey KA, Hester SD, Knapp GW, et al. Gene expression of normal human epidermal keratinocytes modulated by trivalent arsenicals. *Molecular carcinogenesis.* 2010;49(12):981-998.

52. Jo WJ, Loguinov A, Wintz H, et al. Comparative functional genomic analysis identifies distinct and overlapping sets of genes required for resistance to monomethylarsonous acid (MMAIII) and arsenite (AsIII) in yeast. *Toxicol Sci.* 2009;111(2):424-436.

53. Ahsan H, Chen Y, Kibriya MG, et al. Susceptibility to arsenic-induced hyperkeratosis and oxidative stress genes myeloperoxidase and catalase. *Cancer Lett.* 2003;201(1):57-65.

54. Le Marchand L, Wilkens LR. Design considerations for genomic association studies: importance of gene-environment interactions. *Cancer Epidemiol Biomarkers Prev.* 2008;17(2):263-267.

TABLE 1. Single Nucleotide Polymorphisms

Gene	SNP rs number	Chromosome	SNP	Amino acid substitution	Genotyping success rate (%)	α
SOD2	4880	6	Ex2+24C>T	A16V	96.5	0.03
CAT	1001179	11	-329T>C		96.3	0.01
GPX1	1050450	3	Ex1-226C>T	P200L	96.4	-0.03
MPO	2333227	17	-642G>A		96.8	-0.03

TABLE 2. Selected Characteristics for the Study Sample, Araihazar, Bangladesh, 2000-2002

Characteristic	HEALS Cohort (n=11,746)		Cases (n=610)		Controls (n=1,079)		POR[3] (95% CI)
	N	%	N	%	N	%	
Sex							
Male	5,042	42.9	507	82.6	440	40.8	1.0
Female	6,704	57.1	106	17.4	639	59.2	0.1 (0.1, 0.1)
Age, years							
18-30	3,653	31.1	161	26.4	386	35.8	1.0
31-40	4,186	35.7	224	36.7	237	22.0	2.5 (1.8, 3.4)
41-50	2,730	23.2	60	9.8	355	32.9	5.6 (4.0, 7.8)
51-75	1,176	10.0	165	27.0	100	9.3	9.8 (6.7, 14.1)
BMI[4], kg/m^2							
<18.50	4,555	39.7	305	50.4	402	37.5	1.0
18-50-24.99	6,107	53.3	280	46.3	589	54.9	0.6 (0.5, 0.8)
≥25.00	805	7.0	20	3.3	81	7.6	0.3 (0.2, 0.5)
Cigarette smoking							
Never	7,568	64.5	177	29.0	698	64.7	1.0
Ever	4,173	35.5	433	71.0	381	35.3	4.5 (3.6, 5.6)
Fruit/vegetable intake, g/day							
74.0-415.5	3,868	33.4	232	39.9	344	33.3	1.0
415.6-593.0	3,752	32.4	168	28.9	344	33.3	0.7 (0.6, 0.9)
≥593.1	3,950	34.2	181	31.2	344	33.3	0.8 (0.6, 1.0)
Well water arsenic, μg/L[5]							
0.1-10	2,743	23.4	72	11.8	265	24.6	1.0
10.1-50	2,511	21.4	97	15.9	248	23.0	1.4 (1.0, 2.0)
50.1-150	3,600	30.7	202	33.1	330	30.6	2.2 (1.6, 3.1)
150.1-864	2,889	24.6	239	39.2	236	21.9	3.7 (2.7, 5.1)
Urinary total arsenic, μg/g[6]							
16-103	2,725	24.2	64	10.5	269	24.9	1.0
104-192	2,713	24.2	121	19.8	271	25.1	1.9 (1.3, 2.6)
193-339	2,822	25.1	164	26.9	269	24.9	2.6 (1.8, 3.6)
340-8556	2,964	26.4	261	42.8	270	25.1	4.1 (2.9, 5.6)

[3] POR, Prevalence odds ratio; unadjusted.
[4] BMI cut-point defined by WHO BMI classification for underweight, normal, and overweight/obese.
[5] Water arsenic concentration cut-points roughly correspond to quartiles of the exposure distribution for the controls but have been slightly adjust to reflect policy relevant parameters.
[6] Creatinine-adjusted urinary total arsenic concentration.

TABLE 3. Urinary Arsenic Metabolite Distribution for the Study Sample, Araihazar, Bangladesh, 2000-2002

Characteristic	Cases (n=610)[7]		Controls (n=1,079)[8]		Crude Effect Estimate	
	N	%	N	%	POR	95% CI
% Total MMA						
<8.650	59	10.0	260	25.0	1.0	
8.650-11.765	130	22.0	261	25.0	2.2	1.5, 3.1
11.766-14.903	146	24.7	269	25.0	2.5	1.7, 3.5
14.904-33.415	257	43.4	260	25.0	4.4	3.1, 6.0
% Total DMA						
<63.500	192	32.4	260	25.0	1.0	
63.500-69.215	152	25.7	260	25.0	0.8	0.6, 1.0
69.216-74.480	124	21.0	262	25.1	0.6	0.5, 0.8
74.481-90.828	124	21.0	260	24.9	0.6	0.5, 0.9
% Inorganic arsenic[9]						
<10.81	152	25.7	260	25.0	1.0	
10.81-14.20	145	24.5	261	25.0	0.9	0.7, 1.3
14.21-18.31	141	23.8	262	25.1	0.9	0.7, 1.2
18.32-69.30	154	26.0	259	24.9	1.0	0.7, 1.3
Primary methylation index[10,11]						
<0.580	70	11.8	260	25.0	1.0	
0.580-0.835	132	22.3	259	25.0	1.9	1.3, 2.6
0.836-1.154	171	28.9	260	25.0	2.4	1.8, 3.4
1.155-19.570	218	36.9	260	25.0	3.1	2.3, 4.3
Secondary methylation index[12,13]						
<4.316	236	39.9	259	25.0	1.0	
4.316-5.855	153	25.9	259	25.0	0.6	0.5, 0.8
5.856-8.211	128	21.7	259	25.0	0.5	0.4, 0.7
8.212-32.300	74	12.5	259	25.0	0.3	0.2, 0.4

[7] All arsenic metabolites data missing for 18 case subjects.
[8] All arsenic metabolite data missing for 37 control subjects.
[9] Inorganic arsenic is sum of arsenate and arsenite.
[10] The primary methylation index was calculated as % MMA / % Inorganic arsenic.
[11] Primary methylation index missing for 1 additional case and 3 additional controls subjects due to 0 values.
[12] The secondary methylation index was calculated as % DMA / % MMA.
[13] Secondary methylation index missing for 1 additional case and 6 additional control subjects due to 0 values.

TABLE 4. Prevalence Odds Ratios and 95% CIs for Skin Lesion Prevalence in Relation to Antioxidant Enzyme SNPs, Araihazar, Bangladesh, 2000-2002

SNP	MAF[14]	Cases (n=610)		Controls (n=1,079)		Crude Estimate		*P* for trend	Multivariate Estimate[15]		*P* for trend
		N	%	N	%	POR	95% CI		POR	95% CI	
SOD2	T=0.45										
CC		173	30.6	302	31.5	1.0			1.0		
CT		268	47.3	456	47.5	1.0	0.8, 1.3		1.0	0.8, 1.3	
TT		125	22.1	202	21.0	1.1	0.8, 1.4		1.0	0.7, 1.4	
Per allele						1.0	0.9, 1.2	0.61	1.0	0.9, 1.2	0.91
GPX1	T=0.19										
CC		372	65.4	623	65.2	1.0			1.0		
CT		174	30.6	297	31.1	1.0	0.8, 1.2		1.0	0.7, 1.3	
TT		23	4.0	36	3.8	1.1	0.6, 1.8		1.2	0.6, 2.3	
Per allele						1.0	0.8, 1.2	0.98	1.0	0.8, 1.3	0.87
CAT	T=0.23										
CC		358	63.5	571	59.5	1.0			1.0		
CT		185	32.8	343	35.8	0.9	0.7, 1.1		1.0	0.8, 1.3	
TT		21	3.7	45	4.7	0.7	0.4, 1.3		0.9	0.5, 1.7	
Per allele						0.9	0.7, 1.0	0.11	1.0	0.8, 1.2	0.80
MPO	A=0.14										
GG		423	74.3	709	73.6	1.0			1.0		
GA		134	23.6	237	24.6	0.9	0.7, 1.2		1.0	0.7, 1.3	
AA		12	2.1	17	1.8	1.2	0.6, 2.5		1.3	0.5, 3.1	
Per allele						0.9	0.8, 1.2	0.88	1.0	0.8, 1.3	0.94

[14] Minor allele frequency in the controls.
[15] Adjusted for sex, age, and well water arsenic concentration.

TABLE 5. Prevalence Odds Ratios and 95% CIs for Skin Lesion Prevalence in Relation to Antioxidant Enzyme Risk Allele Count, Araihazar, Bangladesh, 2000-2002

Risk Allele Count	Cases (n=517)		Controls (n=868)		Crude Estimate		*P* for trend	Multivariate Estimate[16]		*P* for trend
	N	%	N	%	POR	95% CI		POR	95% CI	
0–3	309	59.8	559	64.4	1.0			1.0		
4–7	208	40.2	309	35.6	1.2	1.0, 1.6	0.08	1.1	0.9, 1.5	0.29

[16] Adjusted for sex, age, and well water arsenic concentration.

TABLE 6. Prevalence Odds Ratios[17] and 95% CIs for Skin Lesion Prevalence in Relation to Antioxidant Enzyme SNPs by Well Water Arsenic Concentration, Araihazar, Bangladesh, 2000-2002

SNP	Well Water Arsenic Concentration (µg/L)				RERI
	0.1-50	50.1-100	100.1-150	150.1-864	
SOD2					0.01
CC	1.00	1.23 (0.62, 2.45)	1.67 (0.88, 3.16)	3.69 (1.99, 6.86)	(-0.20, 0. 20)
CT/TT	1.00[18]	1.33 (0.74, 2.40)	2.14 (1.24, 3.71)	3.23 (1.84, 5.65)	
GPX1					-0.13
CC	1.00	1.42 (0.87, 2.31)	1.94 (1.24, 3.04)	4.08 (2.61, 6.37)	(-0.42, 0.10)
CT/TT	1.00[19]	1.37 (0.76, 2.45)	2.49 (1.53, 4.08)	2.66 (1.59, 4.45)	
CAT					0.14
TT	1.00	1.52 (0.27, 8.48)	1.76 (0.34, 9.16)	2.86 (0.48, 17.03)	(-0.87, 0.66)
CT/CC	1.00[20]	1.51 (0.43, 5.33)	2.20 (0.63, 7.67)	3.72 (1.06, 12.99)	
MPO					-0.02
GG	1.00	1.71 (1.05, 2.78)	3.23 (2.07, 5.04)	4.09 (2.60, 6.43)	(-0.22, 0.23)
GA/AA	1.00[21]	1.76 (0.92, 3.35)	1.64 (0.91, 2.96)	6.05 (3.42, 10.73)	
Risk allele					0.05
0–3	1.00	1.72 (1.02, 2.91)	2.22 (1.36, 3.64)	3.79 (2.32, 6.18)	(-0.20, 0.27)
4–7	1.00[22]	1.65 (0.89, 3.04)	2.79 (1.66, 4.69)	4.02 (2.35, 6.89)	

[17] Adjusted for sex and age.
[18] POR = 0.82 comparing CT/TT to CC in this lowest exposure quartile.
[19] POR = 1.11 comparing CT/TT to CC in this lowest exposure quartile.
[20] POR = 1.04 comparing CT/CC to TT in this lowest exposure quartile.
[21] POR = 1.51 comparing GA/AA to GG in this lowest exposure quartile.
[22] POR = 1.37 comparing 4-7 to 0-3 in this lowest exposure quartile.

TABLE 7. Prevalence Odds Ratios[23] and 95% CIs for Skin Lesion Prevalence in Relation to Antioxidant Enzyme SNPs by Creatinine-adjusted Urinary Total Arsenic Concentration, Araihazar, Bangladesh, 2000-2002

SNP	Creatinine-adjusted urinary total arsenic concentration (µg/g)				RERI
	16-103	104-192	193-339	340-8556	
SOD2					0.02
CC	1.00	1.58 (0.78, 3.19)	3.14 (1.56, 6.30)	4.71 (2.44, 9.08)	(-0.23, 0.23)
CT/TT	1.00[24]	1.88 (1.01, 3.51)	2.44 (1.32, 4.49)	5.43 (2.97, 9.95)	
GPX1					-0.07
CC	1.00	1.86 (1.14, 3.01)	2.55 (1.58, 4.12)	5.28 (3.36, 8.31)	(-0.32, 0.16)
CT/TT	1.00[25]	1.63 (0.93, 2.86)	2.78 (1.65, 4.70)	4.44 (2.64, 7.46)	
CAT					0.17
TT	1.00	3.33 (0.54, 20.56)	1.41 (0.29, 6.82)	3.73 (0.72, 19.35)	(-0.62, 0.65)
CT/CC	1.00[26]	1.62 (0.52, 4.98)	2.41 (0.78, 7.40)	4.69 (1.53, 14.35)	
MPO					-0.03
GG	1.00	1.87 (1.16, 3.00)	2.73 (1.72, 4.32)	5.37 (3.42, 8.42)	(-0.26, 0.23)
GA/AA	1.00[27]	1.88 (1.02, 3.47)	2.65 (1.43, 4.91)	5.04 (2.91, 8.74)	
Risk allele					0.05
0–3	1.00	2.13 (1.24, 3.64)	2.75 (1.62, 4.68)	5.55 (3.32, 9.30)	(-0.24, 0.30)
4–7	1.00[28]	2.00 (1.10, 3.66)	3.81 (2.13, 6.82)	5.48 (3.15, 9.51)	

[23] Adjusted for sex and age.
[24] POR = 1.03 comparing CT/TT to CC in this lowest exposure quartile.
[25] POR = 1.24 comparing CT/TT to CC in this lowest exposure quartile.
[26] POR = 0.98 comparing CT/CC to TT in this lowest exposure quartile.
[27] POR = 1.09 comparing GA/AA to GG in this lowest exposure quartile.
[28] POR = 1.65 comparing 4-7 to 0-3 in this lowest exposure quartile.

TABLE 8. Prevalence Odds Ratios[29] and 95% CIs for Skin Lesion Prevalence in Relation to Antioxidant Enzyme SNPs by Primary Methylation Index, Araihazar, Bangladesh, 2000-2002

SNP	Primary Methylation Index				RERI
	<0.580	**0.580-0.835**	**0.836-1.154**	**1.155-19.570**	
SOD2					0.10
CC	1.00	1.56 (0.76, 3.20)	1.48 (0.74, 2.98)	1.22 (0.61, 2.44)	(-0.14, 0.26)
CT/TT	1.00[30]	1.37 (0.72, 2.62)	1.37 (0.72, 2.59)	1.51 (0.80, 2.85)	
GPX1					-0.03
CC	1.00	1.44 (0.88, 2.37)	1.29 (0.79, 2.10)	1.59 (0.98, 2.58)	(-0.26, 0.17)
CT/TT	1.00[31]	1.34 (0.73, 2.45)	1.72 (0.99, 2.97)	1.24 (0.72, 2.14)	
CAT					0.24
TT	1.00	0.65 (0.10, 4.22)	0.42 (0.07, 2.40)	0.60 (0.12, 3.04)	(-0.50, 0.58)
CT/CC	1.00[32]	0.84 (0.23, 3.10)	0.82 (0.22, 3.01)	0.83 (0.23, 3.05)	
MPO					-0.03
GG	1.00	1.40 (0.88, 2.20)	1.52 (0.97, 2.37)	1.42 (0.92, 2.20)	(-0.34, 0.15)
GA/AA	1.00[33]	1.38 (0.76, 2.47)	1.20 (0.68, 2.10)	1.33 (0.76, 2.35)	
Risk allele					0.22
0–3	1.00	1.16 (0.69, 1.95)	1.10 (0.67, 1.82)	1.07 (0.65, 1.77)	(0.03, 0.36)
4–7	1.00[34]	1.28 (0.72, 2.28)	1.36 (0.78, 2.37)	1.68 (0.98, 2.90)	

[29] Adjusted for sex and age.
[30] POR = 0.91 comparing CT/TT to CC in this lowest exposure quartile.
[31] POR = 0.94 comparing CT/TT to CC in this lowest exposure quartile.
[32] POR = 0.52 comparing CT/CC to TT in this lowest exposure quartile.
[33] POR = 0.98 comparing GA/AA to GG in this lowest exposure quartile.
[34] POR = 0.63 comparing 4-7 to 0-3 in this lowest exposure quartile.

TABLE 9. Prevalence Odds Ratios[35] and 95% CIs for Skin Lesion Prevalence in Relation to Antioxidant Enzyme SNPs by Secondary Methylation Index, Araihazar, Bangladesh, 2000-2002

SNP	Secondary Methylation Index				RERI
	<4.316	**4.316-5.855**	**5.856-8.211**	**8.212-32.300**	
SOD2					0.004
CC	1.00	1.01 (0.58, 1.75)	1.00 (0.56, 1.76)	0.51 (0.25, 1.04)	(-0.25, 0.17)
CT/TT	1.00[36]	0.87 (0.55, 1.37)	0.80 (0.50, 1.28)	0.70 (0.42, 1.17)	
GPX1					-0.03
CC	1.00	0.70 (0.48, 1.03)	0.69 (0.46, 1.03)	0.54 (0.34, 0.86)	(-0.29, 0.16)
CT/TT	1.00[37]	0.80 (0.51, 1.27)	0.69 (0.42, 1.14)	0.49 (0.27, 0.89)	
CAT					0.03
TT	1.00	0.64 (0.14, 2.97)	0.37 (0.07, 2.02)	0.63 (0.11, 3.54)	(-0.77, 0.45)
CT/CC	1.00[38]	0.83 (0.30, 2.35)	0.80 (0.28, 2.26)	0.57 (0.20, 1.65)	
MPO					0.09
GG	1.00	0.99 (0.69, 1.41)	0.81 (0.56, 1.18)	0.64 (0.41, 1.00)	(-0.08, 0.24)
GA/AA	1.00[39]	0.63 (0.37, 1.08)	0.75 (0.44, 1.29)	0.46 (0.24, 0.88)	
Risk allele					-0.03
0–3	1.00	0.75 (0.50, 1.13)	0.66 (0.43, 1.02)	0.48 (0.29, 0.80)	(-0.38, 0.20)
4–7	1.00[40]	0.95 (0.59, 1.52)	0.85 (0.52, 1.41)	0.59 (0.33, 1.05)	

[35] Adjusted for sex and age.
[36] POR = 1.12 comparing CT/TT to CC in this lowest exposure quartile.
[37] POR = 0.85 comparing CT/TT to CC in this lowest exposure quartile.
[38] POR = 1.09 comparing CT/CC to TT in this lowest exposure quartile.
[39] POR = 1.31 comparing GA/AA to GG in this lowest exposure quartile.
[40] POR = 1.02 comparing 4-7 to 0-3 in this lowest exposure quartile.

TABLE 10. Prevalence Odds Ratios[41] and 95% CIs for Skin Lesion Prevalence in Relation to Antioxidant Enzyme SNPs by Body Mass Index, Araihazar, Bangladesh, 2000-2002

SNP	Body Mass Index (kg/m^2)			RERI
	<18.5	18.5-24.9	>25.0	
SOD2				0.13
CC	1.00	0.98 (0.63, 1.53)	0.51 (0.18, 1.41)	(-0.27, 0.38)
CT/TT	1.00[42]	0.86 (0.58, 1.27)	0.58 (0.27, 1.25)	
GPX1				-0.22
CC	1.00	0.69 (0.51, 0.94)	0.42 (0.20, 0.91)	(-0.97, 0.15)
CT/TT	1.00[43]	0.81 (0.56, 1.15)	0.45 (0.18, 1.11)	
CAT				0.28
TT	1.00	0.98 (0.27, 3.62)	1.06 (0.12, 9.29)	(-1.03, 0.75)
CT/CC	1.00[44]	0.99 (0.35, 2.82)	0.53 (0.16, 1.74)	
MPO				0.10
GG	1.00	0.83 (0.62, 1.10)	0.48 (0.24, 0.97)	(-0.41, 0.41)
GA/AA	1.00[45]	0.82 (0.56, 1.21)	0.33 (0.10, 1.09)	
Risk allele				0.02
0–3	1.00	0.73 (0.52, 1.02)	0.54 (0.24, 1.22)	(-0.63, 0.40)
4–7	1.00[46]	0.93 (0.64, 1.35)	0.45 (0.17, 1.16)	

[41] Adjusted for sex, age, and well water arsenic concentration.
[42] POR = 1.16 comparing CT/TT to CC in this lowest exposure tertile.
[43] POR = 0.85 comparing CT/TT to CC in this lowest exposure tertile.
[44] POR = 1.30 comparing CT/CC to TT in this lowest exposure tertile.
[45] POR = 1.07 comparing GA/AA to GG in this lowest exposure tertile.
[46] POR = 1.08 comparing 4-7 to 0-3 in this lowest exposure tertile.

TABLE 11. Prevalence Odds Ratios [47] and 95% CIs for Skin Lesion Prevalence in Relation to Antioxidant Enzyme SNPs by Smoking Status, Araihazar, Bangladesh, 2000-2002

SNP	Smoking Status		RERI
	Never	Ever	
SOD2			-0.41
CC	1.00	1.56 (0.97, 2.52)	(-1.4, 0.20)
CT/TT	1.00[48]	1.37 (0.89, 2.11)	
GPX1			-0.28
CC	1.00	1.39 (0.96, 2.01)	(-0.97, 0.30)
CT/TT	1.00[49]	1.26 (0.83, 1.90)	
CAT			0.19
TT	1.00	1.07 (0.31, 3.66)	(-1.44, 1.14)
CT/CC	1.00[50]	1.25 (0.47, 3.32)	
MPO			0.24
GG	1.00	1.19 (0.84, 1.69)	(-0.40, 0.78)
GA/AA	1.00[51]	1.26 (0.84, 1.91)	
Risk allele			-0.65
0–3	1.00	1.64 (1.11, 2.42)	(-1.65, 0.09)
4–7	1.00[52]	1.53 (1.01, 2.33)	

[47] Adjusted for sex, age, and well water arsenic concentration.
[48] POR = 1.21 comparing CT/TT to CC in this lowest exposure category.
[49] POR = 1.15 comparing CT/TT to CC in this lowest exposure category.
[50] POR = 0.99 comparing CT/CC to TT in this lowest exposure category.
[51] POR = 0.83 comparing GA/AA to GG in this lowest exposure category.
[52] POR = 1.54 comparing 4-7 to 0-3 in this lowest exposure category.

TABLE 12. Prevalence Odds Ratios [53] and 95% CIs for Skin Lesion Prevalence in Relation to Antioxidant Enzyme SNPs by Fruit and Vegetable Intake, Araihazar, Bangladesh, 2000-2002

SNP	Fruit and vegetable intake			RERI
	74.0-415.5	415.6-593.0	≥593.1	
SOD2				0.18
CC	1.00	0.66 (0.38, 1.16)	0.64 (0.37, 1.11)	(-0.15, 0.45)
CT/TT	1.00[54]	0.62 (0.39, 1.00)	0.52 (0.32, 0.84)	
GPX1				0.18
CC	1.00	0.67 (0.46, 0.99)	0.55 (0.37, 0.82)	(-0.17, 0.50)
CT/TT	1.00[55]	0.56 (0.35, 0.90)	0.55 (0.34, 0.87)	
CAT				-0.19
TT	1.00	0.39 (0.08, 1.89)	0.40 (0.09, 1.79)	(-2.07, 0.69)
CT/CC	1.00[56]	0.47 (0.15, 1.44)	0.42 (0.14, 1.30)	
MPO				-0.13
GG	1.00	0.56 (0.39, 0.81)	0.49 (0.33, 0.71)	(-0.54, 0.21)
GA/AA	1.00[57]	0.54 (0.33, 0.88)	0.52 (0.31, 0.88)	
Risk allele				0.18
0–3	1.00	0.64 (0.37, 1.12)	0.66 (0.38, 1.13)	(-0.01, 0.34)
4–7	1.00[58]	0.70 (0.45, 1.17)	0.61 (0.38, 0.97)	

[53] Adjusted for sex, age, well water arsenic concentration, total energy intake, and BMI.
[54] POR = 1.21 comparing CT/TT to CC in this lowest exposure tertile.
[55] POR = 1.36 comparing CT/TT to CC in this lowest exposure tertile.
[56] POR = 0.85 comparing CT/CC to TT in this lowest exposure tertile.
[57] POR = 0.82 comparing GA/AA to GG in this lowest exposure tertile.
[58] POR = 1.43 comparing 4-7 to 0-3 in this lowest exposure tertile.

Chapter 3

Arsenic Exposure, DNA Repair Genes, and Skin Lesion Prevalence in an Adult Bangladeshi Population

ABSTRACT

One of the proposed pathways that may be related to arsenic toxicity is deficient DNA repair and therefore, may play a role in the association between arsenic and skin lesion risk. The objective of this analysis was to assess the relationship between SNPs in DNA repair genes (*OGG1,* rs1052133; *XRCC1,* rs25487 and rs1799782; *XRCC3,* rs861539; *ERCC2,* rs1052559; *ERCC5,* rs17655; and *LIG4,* rs1805388) and skin lesion prevalence among 610 population-based prevalent skin lesion cases and 1,079 controls enrolled in the HEALS cohort. In logistic regression models controlling for sex, age, and well water arsenic concentration, no associations were observed between measured SNPs and skin lesion prevalence. The results did not vary by arsenic exposure levels (as measured by well water arsenic concentration, urinary total arsenic concentration, primary methylation index, and secondary methylation index), body mass index, or smoking status. However, we did observe a significant inverse association of total fruit and vegetable consumption with skin lesion prevalence, and its additive interaction with the polymorphism in *ERCC5*, which warrants investigation in future studies.

INTRODUCTION

DNA repair is a proposed pathway that may modify the association between arsenic and skin lesion risk (1). In a cross-sectional study, arsenic concentrations in drinking water and toenail clippings were positively associated with ERCC1 mRNA expression levels suggesting arsenic exposure may induce a DNA repair response (2); although, an inverse association was found in another study (3). In vitro experiments have shown that arsenic induces DNA strand breaks in a concentration dependent manner (4). Arsenic has been shown to inhibit nucleotide excision repair (5, 6). DNA repair deficiency has been associated with increased micronuclei frequency (7). Reduced DNA repair capacity has been associated with increased arsenical skin lesion risk (8).

DNA was genotyped for seven SNPs in the following DNA repair protein genes: 8-oxoguanine DNA glycosylase (*OGG1;* rs1052133*)*; X-ray repair complementing defective repair in Chinese hamster cells 1 (*XRCC1;* rs25487 and rs1799782); X-ray repair complementing defective repair in Chinese hamster cells 3 (*XRCC3;* rs861539); excision repair cross-complementing rodent repair deficiency, complementation group 2 (*ERCC2;* rs1052559); excision repair cross-complementing rodent repair deficiency, complementation group 5 (*ERCC5;* rs17655); and ligase IV, DNA, ATP-dependent (*LIG4;* rs1805388). OGG1 is a DNA glycosylase involved in base excision repair (BER) of 8-OHdG, an adduct formed from oxidative stress (9). XRCC1 is a DNA repair enzyme that interacts with polynucleotide kinase (PNK), DNA polymerase-beta (POLB) and DNA ligase III (LIG3) as part of a complex to repair single-strand breaks and functions in BER to repair damage caused by agents such as ROS (10). *XRCC3* encodes a protein

that forms a complex with RAD51 and RAD51C to repair double-strand DNA breaks through the homologous recombination pathway (11-14). ERCC2 is a DNA helicase involved in transcription and nucleotide excision repair (NER) (15). ERCC5 is a DNA repair protein involved in the NER pathway of UV-induced damage (16). LIG4 is a DNA double-strand break repair enzyme whose two main roles are the ligation step of the non-homologous end-joining pathway and V(D)J recombination (17).

In this study, we evaluate whether SNPs in genes encoding DNA repair enzymes (*OGG1, XRCC1, XRCC3, ERCC2, ERCC5, LIG4*) were associated with prevalent arsenical skin lesion status in 610 cases and 1,079 controls from the Health Effects of Arsenic Longitudinal Study (HEALS) cohort.

MATERIALS AND METHODS

Study population

HEALS is an ongoing, population-based cohort study examining both the short- and long-term health effects of arsenic exposure. The study was launched in Araihazar, Bangladesh, in 2000. The selection of cohort participants, study design and methods have been described in detail elsewhere (18).

The HEALS cohort served as the source of prevalent cases and controls utilized in these analyses. Between October 2000 and May 2002, married individuals were sampled—an eligibility criteria to minimize loss to follow-up—who were aged 18–75 years and residents of the study area for at least 5 years. From the 12,050 residents who met the eligibility criteria from an enumerated total 65,876 persons in the study area,

11,746 (97.5% response rate) men and women (4,801 married couples and 2,144 married individuals whose spouses did not participate) were enrolled into the HEALS cohort. Among the 11,746 enrolled baseline cohort participants, 11,224 participants (95.6%) provided urine samples. At baseline, 610 cases of skin lesions were identified among the 11,224 cohort members who provided urine samples and were included as prevalent cases in this study. Of the remaining participants with an available urine sample and known to be free of skin lesions at baseline (n=10,614), a ~10% random sample (n=1,079) was selected and included as controls in this study.

Structured baseline questionnaires were administered to participants in Bengali by trained interviewers, who were blind to the water arsenic concentration of the participants' wells. The questionnaires assessed sociodemographic characteristics, current and past tubewell use, typical water consumption patterns, food frequency of thirty-nine items common to the population (19), occupational exposures, and smoking habits. The study physicians also conducted a clinical examination, which included a skin evaluation for the detection of arsenical skin lesions based on a structured protocol (18, 20). Venous blood and spot urine samples were also collected from each participant on the day of the interview.

The study protocol was approved by the Institutional Review Boards of Columbia University, The University of Chicago, and the Bangladesh Medical Research Council. Informed consent was obtained from all participants prior to baseline interview.

Arsenic exposure assessment

Three primary measures of arsenic exposure were estimated for each case-control participant: well water arsenic concentration, creatinine-adjusted urinary total arsenic concentration, and urinary arsenic metabolites.

At the baseline survey, participants were asked to identify the well they primarily used as their source of drinking water, from which we were able to assign the appropriate well water arsenic concentration exposure. Well water arsenic concentrations of all 5,966 tubewells in the study area were measured by graphite furnace atomic absorption spectrometry, with a detection limit of 5 μg/L. Samples below the limit of detection were subsequently reanalyzed by inductively coupled plasma-mass spectrometry (ICP-MS), with a detection limit of 0.1 μg/L (21).

Urinary total arsenic concentration was measured in a spot urine sample collected at baseline by graphite furnace atomic absorption spectrometry, with a detection limit of 2 μg/L (22). Urinary creatinine was measured by a colorimetric Sigma Diagnostics Kit (Sigma, St. Louis, MO), and urinary total arsenic was subsequently divided by creatinine to obtain a creatinine-adjusted urinary total arsenic concentration, expressed as μg/g creatinine (23).

Urinary arsenic metabolites were measured based on the method by Reuter et al. (24) using high-performance liquid chromatography separation of arsenobetaine, arsenocholine, arsenate (As^{V}), arsenite (As^{III}), total monomethyl arsenic (MMA^{III} + MMA^{V}), and total dimethyl arsenic (DMA^{III} + DMA^{V}) followed by detection by ICP-

MS-dynamic reaction cell. Total inorganic arsenic (As^{III} + As^{V}) is utilized in these analyses since As^{III} can oxidize to As^{V} during sample transport, storage, and preparation.

For purposes of analysis, well water arsenic was categorized into quartiles based on the distribution in the controls; however, since the first and second quartiles roughly corresponded to the World Health Organization's guideline for arsenic in drinking water (10 μg/L) and the national standard for arsenic in drinking water in Bangladesh (50 μg/L), respectively, we adjusted the cut points slightly to correspond to these regulatory levels. Urinary total arsenic concentration and arsenic metabolite metrics were quartiled based on the distribution in the controls. The percentages of MMA, DMA and inorganic arsenic were calculated after exclusion of arsenobetaine and arsenocholine from the total. In addition, two methylation indices were constructed: primary methylation index (PMI)—the ratio of MMA to inorganic arsenic—and secondary methylation index (SMI)—the ratio of DMA to MMA.

Individuals who were missing data on a particular arsenic exposure measure were excluded from the analysis of that specific exposure. Well water arsenic and urinary total arsenic concentrations were available for all cases and controls. Arsenic metabolite data was missing for 18 case subjects and 37 control subjects.

Single nucleotide polymorphism assessment

High-throughput DNA extraction was completed in 96-well format using the QIAmp DNA 96 DNA Blood kit (Qiagen, Valencia, CA). Replica plates were made with 12.5 ng DNA in 2.5 μL per well. In the first step, the genomic DNA was amplified by

PCR using appropriate primers. After PCR amplification, the primers and deoxynucleotide triphosphates in 10 μL PCR product were digested with the 10 μL shrimp alkaline phosphatase cocktail containing 1.0 μL (1 unit/μL) of shrimp alkaline phosphatase, 0.1 μL of Escherichia coli exonuclease I (10 units/μL; U.S. Biochemical, Cleveland, OH), 1.0 μL of 10x shrimp alkaline phosphatase buffer, and 7.9 μL of DNase and RNase-free water for 45 min at 37°C followed by heating at 95°C for 15 min for enzyme deactivation. Then, single nucleotide extension was carried out in the presence of the appropriate allele-specific dideoxynucleotide triphosphates fluorescence labeled with either R110 or TAMRA (PerkinElmer, Waltham, MA). For single nucleotide extension reactions, both the forward and reverse probes were initially tested to select the better probe based on clear signal clustering. Reaction mixture (13 μL/well) containing 0.025 μL AcycloPrime enzyme, 0.5 μL terminator dye, 1 μL reaction buffer, 0.25 μL extension probe (10 pmol/μL), and 11.225 μL water was added to 7 μL of digested PCR product to make 20 μL reaction volume. Thermocycling was done at 95°C for 3 min followed by optimum number of cycles of 95°C for 15 s and 55°C for 30 s. Finally, the fluorescence was measured with Wallac 1420 Multilabel Counter Victor 3 (PerkinElmer, Waltham, MA). In addition to our assay-specific quality control samples, 10% of the samples were run in duplicate after relabeling to keep laboratory researchers blinded to its identity. Concordance based on the duplicates was >0.92. Call rates for the SNPs of interest ranged from 95.0% to 97.9%, as shown in **Table 1**.

Skin lesion assessment

Arsenical skin lesion assessment was conducted through skin examination at the baseline interview following a structured protocol by a trained study physician (18, 20). Arsenical skin lesions were categorized as the presence of melanosis on the body surface, leucomelanosis on the body surface, and keratosis on the hands or feet. For the purposes of this study, cases were selected as having the presence of any type of arsenical skin lesion.

The distribution of skin lesion severity among the 610 cases was as follows: 359 cases had melanosis only, 20 had leucomelanosis only, 10 had melanosis and leucomelanosis, 170 had melanosis and keratosis, 40 had leucomelanosis and keratosis, and 11 had melanosis, leucomelanosis, and keratosis.

Individuals without keratotic arsenical lesions are generally thought to have a milder form of skin lesions, manifest as changes in the dermal pigmentation of the skin. This suggests that there is some heterogeneity of the skin lesion definition; therefore, subset analyses were conducted to evaluate the association of SNPs with skin lesion severity. Skin lesion severity was stratified into absence of keratotic skin lesions (n=389) and presence of keratotic skin lesions (n=221).

Covariates

All covariate data was derived from the baseline interview. Demographic factors included sex and reported age at baseline. Self-reported smoking history was ascertained and categorized as ever versus never smoker for the purposes of these analyses. At the

time of the baseline interview, height and weight of each participant was measured by the study physician; body mass index was constructed as weight in kilograms divided by height in meters squared and categorized by the World Health Organization cutoff points for underweight, normal weight, and overweight/obese.

Fruit and vegetable consumption as well as total caloric intake was constructed based on the reported intake from the food frequency questionnaire (FFQ) (19). Participants were exclude from this analysis if they did not complete or reported an unlikely total caloric intake value on the FFQ (<500 or >4000 kcal/ day; n=47 missing and n=29 extreme values (total n=76 excluded)). A single variable for total fruit and vegetable intake was constructed by summing the average daily amount of fruit and vegetables consumed that was reported for 26 food items (**Table B.1**). The total fruit and vegetable intake was then tertiled based on the distribution in controls for the purposes of these analyses.

Statistical analysis

Hardy Weinberg Equilibrium (HWE) was calculated for the controls based on Pearson (χ^2) tests and α. Violation of HWE is calculated based on statistical departure from expected HWE frequencies based on the exact test. α is used to denote the magnitude of departure from HWE.

$$\alpha = \frac{1}{2}\log\left(\frac{4P_{gg}P_{GG}}{P_{Gg}^2}\right)$$

where P_{gg}, P_{Gg}, and P_{GG} are the genotype proportions (25).

The association between each SNP and arsenical skin lesion status was estimated by prevalence odds ratios (PORs) and their 95% confidence intervals (CIs) from logistic regression models employed by the LOGISTIC procedure in SAS. The PORs were adjusted for sex, age, and well water arsenic concentration, primarily because of the strong association of these covariates with the outcome. The genotype-specific PORs were estimated for each SNP using the homozygous major genotype as the reference category. Additionally, the per allele POR was estimated assuming a log-additive model. In these models, a single ordinal variable was included indicating the number of minor alleles (0, 1, 2) and the associated Wald statistic was interpreted as the P for trend.

Haplotypes were constructed for the 2 SNPs in *XRCC1* (rs25487 and rs1799782) using a maximum likelihood approach implemented in Hapstat, version 3.0 (26). Individuals who had genotype data for both SNPs were included (n=546 cases; n=926 controls). Logistic regression was used to assess the association of skin lesion prevalence with the specific haplotypes. The haplotype with major alleles was used as the reference category. PORs and their 95% confidence intervals were adjusted for sex, age, and well water arsenic concentration.

A summary risk allele count was created for the measured biallelic SNPs by summing the number of risk alleles carried by an individual, based on a priori knowledge of the SNP alleles on protein structure and function (OGG1=G allele, XRCC1 (rs25487)=A allele, XRCC1 (rs1799782)=T allele, XRCC3=T allele, ERCC2=C allele, ERCC5=C allele). For purposes of analysis, the risk allele count was dichotomized based on the distribution in the controls. Due to the fact that there was not a substantial amount

of variability in the *LIG4* SNP genotype in this population and the α statistic suggested a deviation from HWE in the control subjects, this SNP was excluded from the risk allele score and from evaluation of interaction.

We also evaluated possible additive interaction between arsenic exposure and each SNP and the risk allele count in separate analyses in relation to skin lesion prevalence. Multivariate adjusted estimates were used to estimate the relative excess risk for interaction (RERI), calculated as

$$\text{RERI} = \exp(\beta1+\beta2+\beta3) - \exp(\beta1) - \exp(\beta2) +1.$$

Here β1 is the coefficient of the ordinal arsenic exposure measure, β2 is the coefficient of the dichotomous SNP effect modifier measure, and β3 is the coefficient of the cross-product of the ordinal arsenic exposure and dichotomous SNP effect modifier (27, 28). Bias corrected and accelerated (BCa) 95% CIs of the RERI were estimated via 1000 bootstrap samples. CIs of the RERI were also calculated using the delta method described by Hosmer and Lemeshow with similar results (not shown) (29). We also evaluated effect modification of the associations between lifestyle factors (BMI, smoking status, and fruit and vegetable intake) and skin lesions by each SNP and risk allele count in separate analyses on the additive scale, adjusting for well water arsenic exposure and other covariates. In these analyses, β1 is the coefficient of the ordinal lifestyle measure, β2 is the coefficient of the dichotomous SNP effect modifier measure, and β3 is the coefficient of the cross-product of the ordinal lifestyle measure and dichotomous SNP effect modifier—with the exception of smoking status, which was modeled as a dichotomous lifestyle measure.

In exploratory subset analyses, skin lesion severity was evaluated by stratifying arsenical skin lesion status into absence of keratotic skin lesions and presence of keratotic skin lesions. Prevalent skin lesion cases with melanosis and/or leucomelanosis but not keratosis (n=389) were classified as non-keratotic skin lesions and, individuals with keratosis (n=221) were classified as keratotic skin lesions. Ordered polytomous regression models were employed by the LOGISTIC procedure in SAS, comparing non-keratotic skin lesions (coded=1) and keratotic skin lesions (coded=2) to the controls (coded=0, reference category).

RESULTS

Sample characteristics

Information on the genotyped DNA repair SNPs is shown in **Table 1**. Based on the α statistic, there was evidence that the *LIG4* SNP (rs1805388) deviated from Hardy-Weinberg equilibrium indicating a potential source of genotyping error for this SNP; therefore, it was excluded from interaction analyses and the risk allele score. The distribution of selected characteristics in the total HEALS cohort and by the 610 prevalent skin lesion cases and 1,079 controls are shown in **Table 2.** Since a random sample of the baseline cohort without skin lesions was selected as control participants for genotyping, the distribution of characteristics in the controls is not appreciably different from the total baseline cohort except for the distribution of age, where we see that control subjects were underrepresented in the 31–40 age range and overrepresented in the 41–50 age range relative to the total cohort. Similar to the previous HEALS cohort analysis for

skin lesion prevalence (20), we observed based on the selected cases and controls for this analysis that males, older age, low BMI (<18.5 kg/m^2), current or past cigarette smoking, and total fruit and vegetable intake were associated with increased skin lesion prevalence. Additionally, clear dose-response associations were observed with arsenic exposure as measured by well water arsenic and urinary total arsenic concentrations (**Table 2**), as well as with the distribution of urinary arsenic metabolites (**Table 3**).

Evaluation of associations with single and cumulative SNP effects

No significant associations were observed between each of the DNA repair SNPs and skin lesion prevalence, based on the genotype and per allele POR estimates (**Table 4**). Haplotypes were constructed for the 2 SNPs in *XRCC1* and no individuals were found to carry two copies of the minor alleles in both *XRCC1* polymorphisms. There were no significant associations of haplotypes with skin lesion prevalence (**Table 5**). A risk allele score for the genotyped SNPs was constructed based on *a priori* evidence of the risk allele for each SNP and was not associated with skin lesion prevalence (**Table 6**). These results did not vary by skin lesion severity, classified as non-keratotic and keratotic skin lesions based on polytomous regression analysis (**Table C.1**). Additionally, in exploratory analyses we evaluated these associations in male participants only since their baseline skin lesion prevalence was appreciably higher than females and, we did not observe notable differences in the distribution of SNPs or the associations of the SNPs with skin lesion prevalence in males as compared to the total study sample, precluding evidence of interaction by sex (**Table C.2**).

Additive interaction evaluation

We evaluated whether the associations between DNA repair SNPs and skin lesion prevalence were modified by various arsenic exposure measures and lifestyle factors on the additive scale. Departure from additivity was evaluated by well water arsenic concentration (**Table 7**), creatinine-adjusted urinary total arsenic concentration (**Table 8**), primary methylation index (**Table 9**), secondary methylation index (**Table 10**), BMI (**Table 11**), cigarette smoking status (**Table 12**), and total fruit and vegetable intake (**Table 13**). A significant departure from additivity was observed for the SNP in *ERCC5* and total fruit and vegetable intake on skin lesion prevalence (RERI=-0.78, 95% CI=1.43, -0.29; **Table 13**). On the additive scale, we observed that skin lesion prevalence was greater in individuals with the GC/CC genotype of the *ERCC5* SNP with each 1 tertile increase in total fruit and vegetable intake than would be expected based on the additive independent effects of the ERCC5 variant and total fruit and vegetable intake alone. There were also significant main effects for tertiles of total fruit and vegetable intake, adjusted for sex, age, well water arsenic concentration, BMI, and total caloric intake (tertile 2 versus 1 POR=0.68, 95% CI=0.50, 0.91; tertile 3 versus 1 POR=0.57, 95% CI=0.42, 0.78). We did not observed any other evidence of interaction on the additive scale by measures of arsenic exposure, arsenic metabolism, or lifestyle factors as characterized by RERI estimates that were not appreciably different from zero.

DISCUSSION

In this population-based case-control study of arsenical skin lesion prevalence in a Bangladeshi population, we systematically evaluated seven candidate SNPs in known DNA repair genes. There were no statistically significant associations between these SNPs (*OGG1,* rs1052133; *XRCC1,* rs25487 and rs1799782; *XRCC3,* rs861539; *ERCC2,* rs1052559; *ERCC5,* rs17655; and *LIG4,* rs1805388) and skin lesion prevalence. However, we observed a significant departure from additivity for the *ERCC5* variant and total fruit and vegetable intake on skin lesion prevalence, adjusted for arsenic exposure and other covariates.

The findings from this analysis further support previous research from this cohort. It was previously shown in a cross-sectional analysis of the prevalent skin lesions cases utilized in this analysis, that riboflavin, pyridozine, folic acid, vitamin A, vitamin C, and vitamin E intakes (as ascertained from the FFQ) were all inversely associated with skin lesion prevalence (30). In a recent analysis by Pierce et al (31) utilizing data from the HEALS cohort, it was shown that a diet comprised of vegetable intake was associated with a reduced risk of incident skin lesions. Additionally, it was also shown in a nested case-control study in the HEALS study population that folate deficiency was associated with skin lesion incidence (32).

ERCC5 is known to be involved in nucleotide excision repair (NER), particularly of UV-induced DNA damage (16). One mechanism by which arsenic is hypothesized to cause carcinogenesis is through an indirect mechanism, as a co-carcinogen, by inhibiting DNA repair pathways necessary to remove lesions generated by other environmental

carcinogens (33, 34). Numerous studies have shown that arsenic inhibits NER (5, 6, 35) and the repair of bulky DNA adducts *in vitro* and in laboratory animals (36-38). In recent epidemiologic studies by Andrew et al, arsenic exposure through drinking water was associated in a dose-dependent manner to decreased expression of NER genes and diminished repair of lesions in human lymphocytes (3, 39). Moreover, in a recent epidemiologic investigation of lung cancer in a Danish population, it was shown that polymorphisms in other NER-related genes interacted with intake of fruit on lung cancer risk (40). However, others have not show an effect of fruit and vegetable intake on DNA repair activity in leucocytes (41). To our knowledge, this is the first study to examine *ERCC5* in relation to arsenic toxicity; therefore, replication of a potential involvement of *ERCC5* in arsenic-related disease in future studies is needed.

In pilot work conducted by Ahsan et al (42) in the same cohort population as this analysis but using a non-overlapping set of prevalent skin lesion cases with keratosis and control subjects, evidence of an increased trend for the association between tertiles of creatinine-adjusted urinary total arsenic concentration and skin lesion prevalence were observed among individuals with the AA genotype of *ERCC2* (rs1052559) compared to the AC/CC genotype trends; however, a formal statistical test of interaction was not presented. In this current study, we did not observe any evidence of interaction between the same genotyped polymorphism in *ERCC2* and skin lesion prevalence. Based on the polytomous regression analysis conducted in this current study, stratifying cases by presence of keratosis did not reveal any significant difference in estimates between keratotic versus non-keratotic cases for *ERCC2*. Additionally, we evaluated the joint

effect of creatinine-adjusted urinary total arsenic concentration and the *ERCC2* SNP among the keratotic cases only compared to the control subjects and did not observe any evidence of a difference in prevalence trends (**Table C.3**). Since both the pilot work and the stratified analysis in this study were based on a small number of cases (n=29 and n=208, respectively), further investigation of this potential finding is warranted in a larger sample of well-defined keratotic skin lesion cases to draw more definitive conclusions.

The major strengths of this study were the use of a validated FFQ, measurement of genetic variants, large size of the study sample, the wide range of arsenic exposure, and the multiple constructs of arsenic exposure. First, we used a validated FFQ. Our study instrument contains the food items most commonly consumed by our study population based on comparison with food diaries in this population (19) and captures the major variability in diet. While the actual food intakes may not be accurately estimated by the FFQ within our study population, it is likely that it does rank participants reasonably well into tertiles of food intakes. Second, the advantage of measuring genetic variants is that they can be measured with little measurement error and bias. Additionally, SNPs are time-independent measures (43), which strengthens causal inference of the associations evaluated in this analysis between the measured SNPs and prevalent skin lesions. Conversely, the amount of DNA damage could have been measured but there would have been major limitations in the temporality of these measures with the use of prevalent cases. Additionally, in very early or mild cases of skin lesions (some forms of melanosis) may not be readily visible with clinical examination; therefore, biomarkers of DNA damage could also suffer from reverse causation if it was not ensured that the

measures were taken at an appropriate period before disease onset. Third, because we do not expect nutritional or genetics effects to be particularly large, the large sample size enhances our ability to detect associations. Finally, arsenic exposure was ascertained based on well water arsenic concentration, as well as urinary arsenic measures including total arsenic concentration and arsenic metabolites, which allowed us to evaluate various constructs of exposure including total exposure burden as well as exposure methylation ability.

While there are many strengths of this study, there are several limitations that are acknowledged. A limitation of this study is that the FFQ measures average diet; the food intake of the individual at the time of interview may have varied from the diet that the individual had at the time of skin lesion development for the prevalent cases or due to seasonal variability or fluctuations in household income for the cases and controls from the actual representative diet of the individual. Second, we conducted a series of analyses to evaluate main effects and interactions in this study that were pre-specified in our hypotheses. While there is evidence for the biologically plausibly of an interaction between intake of fruit and vegetable and the polymorphism in *ERCC5* on skin lesion prevalence, we urge caution in the interpretation of this finding, without replication in other studies. It is possible that measurement error in the fruit and vegetable intake variable or SNP assessment could have produced the heterogeneity that appears between this SNP and fruit and vegetable intake since it is the only significant result we find among several comparisons. Third, the assessment of the effect of SNPs on prevalent skin lesions is also limited for the evaluation of mechanistic hypotheses and phenotype

considerations since the association may be related to disease prevalence or progression and not disease incidence. Therefore, findings from these analyses should be replicated with incident cases. However, we deem the effect of disease prevalence to be minimal since arsenical skin lesions are typically not fatal. Additionally, in the subset analysis we conducted to evaluate the associations of SNPs with arsenical skin lesion severity, it did not appear that these SNPs were related to disease progression. Finally, the SNPs and genes for which there is data available were selected based on a candidate-gene approach. Therefore, only a single SNP in the majority of genes of interest were examined. Future studies should take a more comprehensive genomewide approach for evaluating genetic variants in the DNA repair pathway in relation to skin lesion status. Findings from this type of "discovery" approach could then be evaluated in candidate-gene studies employing SNPs which comprehensively tag the genes of interest.

The findings of this study have potentially important scientific and public health implications for arsenic in drinking water. Prior epidemiologic research has suggested DNA repair as a potential mechanism of arsenic toxicity, which observations from this study suggest is a viable mechanism for arsenic toxicity and warrants further investigation. Additionally, individuals deficient in fruit and vegetable intake and carried a risk allele in *ERCC5* had increased prevalence of skin lesions, which highlights potential pathways for intervention in individuals with manifest skin lesions.

In conclusion, our results suggest that the genetic polymorphisms in DNA repair genes measured in this study are not related to skin lesion prevalence in this Bangladeshi population. The significant inverse association of total fruit and vegetable consumption

with skin lesion prevalence, and its additive interaction with a polymorphism in *ERCC5* warrant investigation in future studies.

REFERENCES

1. Kitchin KT. Recent advances in arsenic carcinogenesis: modes of action, animal model systems, and methylated arsenic metabolites. *Toxicol Appl Pharmacol.* 2001;172(3):249-261.

2. Mo J, Xia Y, Ning Z, et al. Elevated ERCC1 gene expression in blood cells associated with exposure to arsenic from drinking water in Inner Mongolia. *Anticancer research.* 2009;29(8):3253-3259.

3. Andrew AS, Karagas MR, Hamilton JW. Decreased DNA repair gene expression among individuals exposed to arsenic in United States drinking water. *Int J Cancer.* 2003;104(3):263-268.

4. Dong JT, Luo XM. Arsenic-induced DNA-strand breaks associated with DNA-protein crosslinks in human fetal lung fibroblasts. *Mutat Res.* 1993;302(2):97-102.

5. Bau DT, Wang TS, Chung CH, et al. Oxidative DNA adducts and DNA-protein cross-links are the major DNA lesions induced by arsenite. *Environ Health Perspect.* 2002;110 Suppl 5753-756.

6. Schwerdtle T, Walter I, Mackiw I, et al. Induction of oxidative DNA damage by arsenite and its trivalent and pentavalent methylated metabolites in cultured human cells and isolated DNA. *Carcinogenesis.* 2003;24(5):967-974.

7. Mateuca RA, Roelants M, Iarmarcovai G, et al. hOGG1(326), XRCC1(399) and XRCC3(241) polymorphisms influence micronucleus frequencies in human lymphocytes in vivo. *Mutagenesis.* 2008;23(1):35-41.

8. Banerjee M, Sarma N, Biswas R, et al. DNA repair deficiency leads to susceptibility to develop arsenic-induced premalignant skin lesions. *Int J Cancer.* 2008;123(2):283-287.

9. Lu AL, Li X, Gu Y, et al. Repair of oxidative DNA damage: mechanisms and functions. *Cell Biochem Biophys.* 2001;35(2):141-170.

10. Whitehouse CJ, Taylor RM, Thistlethwaite A, et al. XRCC1 stimulates human polynucleotide kinase activity at damaged DNA termini and accelerates DNA single-strand break repair. *Cell.* 2001;104(1):107-117.

11. 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(6):783-793.

12. Liu N, Schild D, Thelen MP, et al. Involvement of Rad51C in two distinct protein complexes of Rad51 paralogs in human cells. *Nucleic Acids Res.* 2002;30(4):1009-1015.

13. Pierce AJ, Johnson RD, Thompson LH, et al. XRCC3 promotes homology-directed repair of DNA damage in mammalian cells. *Genes Dev.* 1999;13(20):2633-2638.

14. Brenneman MA, Weiss AE, Nickoloff JA, et al. XRCC3 is required for efficient repair of chromosome breaks by homologous recombination. *Mutat Res.* 2000;459(2):89-97.

15. Chen J, Suter B. Xpd, a structural bridge and a functional link. *Cell cycle.* 2003;2(6):503-506.

16. Thompson LH, Carrano AV, Sato K, et al. Identification of nucleotide-excision-repair genes on human chromosomes 2 and 13 by functional complementation in hamster-human hybrids. *Somatic cell and molecular genetics.* 1987;13(5):539-551.

17. Tomkinson AE, Mackey ZB. Structure and function of mammalian DNA ligases. *Mutat Res.* 1998;407(1):1-9.

18. Ahsan H, Chen Y, Parvez F, et al. Health Effects of Arsenic Longitudinal Study (HEALS): description of a multidisciplinary epidemiologic investigation. *J Expo Sci Environ Epidemiol.* 2006;16(2):191-205.

19. Chen Y, Ahsan H, Parvez F, et al. Validity of a food-frequency questionnaire for a large prospective cohort study in Bangladesh. *Br J Nutr.* 2004;92(5):851-859.

20. Ahsan H, Chen Y, Parvez F, et al. Arsenic exposure from drinking water and risk of premalignant skin lesions in Bangladesh: baseline results from the Health Effects of Arsenic Longitudinal Study. *Am J Epidemiol.* 2006;163(12):1138-1148.

21. Cheng Z, Zheng Y, Mortlock R, et al. Rapid multi-element analysis of groundwater by high-resolution inductively coupled plasma mass spectrometry. *Anal Bioanal Chem.* 2004;379(3):512-518.

22. Nixon DE, Mussmann GV, Eckdahl SJ, et al. Total arsenic in urine: palladium-persulfate vs nickel as a matrix modifier for graphite furnace atomic absorption spectrophotometry. *Clin Chem.* 1991;37(9):1575-1579.

23. Nermell B, Lindberg AL, Rahman M, et al. Urinary arsenic concentration adjustment factors and malnutrition. *Environmental research.* 2008;106(2):212-218.

24. Reuter W, Davidowski L, Neubauer K. Speciation of five arsenic compounds in urine by HPLC/ICP-MS. Perkin Elmer Application Notes: Norwalk, CT: Perkin-Elmer, 2003.

25. Minelli C, Thompson JR, Abrams KR, et al. How should we use information about HWE in the meta-analyses of genetic association studies? *Int J Epidemiol.* 2008;37(1):136-146.

26. Lin DY, Zeng D, Millikan R. Maximum likelihood estimation of haplotype effects and haplotype-environment interactions in association studies. *Genetic epidemiology.* 2005;29(4):299-312.

27. Knol MJ, van der Tweel I, Grobbee DE, et al. Estimating interaction on an additive scale between continuous determinants in a logistic regression model. *Int J Epidemiol.* 2007;36(5):1111-1118.

28. Li R, Chambless L. Test for additive interaction in proportional hazards models. *Annals of epidemiology.* 2007;17(3):227-236.

29. Hosmer DW, Lemeshow S. Confidence interval estimation of interaction. *Epidemiology.* 1992;3(5):452-456.

30. Zablotska LB, Chen Y, Graziano JH, et al. Protective effects of B vitamins and antioxidants on the risk of arsenic-related skin lesions in Bangladesh. *Environ Health Perspect.* 2008;116(8):1056-1062.

31. Pierce BL, Argos M, Chen Y, et al. Arsenic exposure, dietary patterns, and skin lesion risk in Bangladesh: a prospective study. *Am J Epidemiol.* 2011;173(3):345-354.

32. Pilsner JR, Liu X, Ahsan H, et al. Folate deficiency, hyperhomocysteinemia, low urinary creatinine, and hypomethylation of leukocyte DNA are risk factors for arsenic-induced skin lesions. *Environ Health Perspect.* 2009;117(2):254-260.

33. Rossman TG. Enhancement of UV-mutagenesis by low concentrations of arsenite in E. coli. *Mutat Res.* 1981;91(3):207-211.

34. Rossman TG, Uddin AN, Burns FJ. Evidence that arsenite acts as a cocarcinogen in skin cancer. *Toxicol Appl Pharmacol.* 2004;198(3):394-404.

35. Nollen M, Ebert F, Moser J, et al. Impact of arsenic on nucleotide excision repair: XPC function, protein level, and gene expression. *Molecular nutrition & food research.* 2009;53(5):572-582.

36. Tran HP, Prakash AS, Barnard R, et al. Arsenic inhibits the repair of DNA damage induced by benzo(a)pyrene. *Toxicol Lett.* 2002;133(1):59-67.

37. Danaee H, Nelson HH, Liber H, et al. Low dose exposure to sodium arsenite synergistically interacts with UV radiation to induce mutations and alter DNA repair in human cells. *Mutagenesis.* 2004;19(2):143-148.

38. Schwerdtle T, Walter I, Hartwig A. Arsenite and its biomethylated metabolites interfere with the formation and repair of stable BPDE-induced DNA adducts in human cells and impair XPAzf and Fpg. *DNA repair.* 2003;2(12):1449-1463.

39. Andrew AS, Burgess JL, Meza MM, et al. Arsenic exposure is associated with decreased DNA repair in vitro and in individuals exposed to drinking water arsenic. *Environ Health Perspect.* 2006;114(8):1193-1198.

40. Raaschou-Nielsen O, Sorensen M, Overvad K, et al. Polymorphisms in nucleotide excision repair genes, smoking and intake of fruit and vegetables in relation to lung cancer. *Lung cancer.* 2008;59(2):171-179.

41. Dragsted LO, Krath B, Ravn-Haren G, et al. Biological effects of fruit and vegetables. *The Proceedings of the Nutrition Society.* 2006;65(1):61-67.

42. Ahsan H, Chen Y, Wang Q, et al. DNA repair gene XPD and susceptibility to arsenic-induced hyperkeratosis. *Toxicol Lett.* 2003;143(2):123-131.

43. Le Marchand L, Wilkens LR. Design considerations for genomic association studies: importance of gene-environment interactions. *Cancer Epidemiol Biomarkers Prev.* 2008;17(2):263-267.

44. Witte JS. Genetic analysis with hierarchical models. *Genetic epidemiology.* 1997;14(6):1137-1142.

TABLE 1. Single Nucleotide Polymorphisms

Gene	SNP rs number	Chromosome	SNP	Amino acid substitution	Genotyping success rate (%)	α
OGG1	1052133	3	Ex6-315C>G	S326C	97.7	-0.01
ERCC2	1052559 (13181)	19	Ex23+61A>C	K751Q	97.9	0.01
ERCC5	17655	13	Ex15-344G>C	D1104H	96.8	-0.05
LIG4	1805388	13	Ex3+54C>T	T9I	96.6	-0.18
XRCC1	1799782	19	Ex6-22C>T	R194W	95.2	0.07
XRCC1	25487	19	Ex10-4A>G	Q399R	96.3	-0.01
XRCC3	861539	14	Ex8-53C>T	T241M	95.0	0.05

TABLE 2. Selected Characteristics for the Study Sample, Araihazar, Bangladesh, 2000-2002

Characteristic	HEALS Cohort (n=11,746)		Cases (n=610)		Controls (n=1,079)		POR[59] (95% CI)
	N	%	N	%	N	%	
Sex							
Male	5,042	42.9	507	82.6	440	40.8	1.0
Female	6,704	57.1	106	17.4	639	59.2	0.1 (0.1, 0.1)
Age, years							
18-30	3,653	31.1	161	26.4	386	35.8	1.0
31-40	4,186	35.7	224	36.7	237	22.0	2.5 (1.8, 3.4)
41-50	2,730	23.2	60	9.8	355	32.9	5.6 (4.0, 7.8)
51-75	1,176	10.0	165	27.0	100	9.3	9.8 (6.7, 14.1)
BMI[60], kg/m^2							
<18.50	4,555	39.7	305	50.4	402	37.5	1.0
18-50-24.99	6,107	53.3	280	46.3	589	54.9	0.6 (0.5, 0.8)
≥25.00	805	7.0	20	3.3	81	7.6	0.3 (0.2, 0.5)
Cigarette smoking							
Never	7,568	64.5	177	29.0	698	64.7	1.0
Ever	4,173	35.5	433	71.0	381	35.3	4.5 (3.6, 5.6)
Fruit/vegetable intake, g/day							
74.0-415.5	3,868	33.4	232	39.9	344	33.3	1.0
415.6-593.0	3,752	32.4	168	28.9	344	33.3	0.7 (0.6, 0.9)
≥593.1	3,950	34.2	181	31.2	344	33.3	0.8 (0.6, 1.0)
Well water arsenic, μg/L[61]							
0.1-10	2,743	23.4	72	11.8	265	24.6	1.0
10.1-50	2,511	21.4	97	15.9	248	23.0	1.4 (1.0, 2.0)
50.1-150	3,600	30.7	202	33.1	330	30.6	2.2 (1.6, 3.1)
150.1-864	2,889	24.6	239	39.2	236	21.9	3.7 (2.7, 5.1)
Urinary total arsenic, μg/g[62]							
16-103	2,725	24.2	64	10.5	269	24.9	1.0
104-192	2,713	24.2	121	19.8	271	25.1	1.9 (1.3, 2.6)
193-339	2,822	25.1	164	26.9	269	24.9	2.6 (1.8, 3.6)
340-8556	2,964	26.4	261	42.8	270	25.1	4.1 (2.9, 5.6)

[59] POR, Prevalence odds ratio; unadjusted.
[60] BMI cut-point defined by WHO BMI classification for underweight, normal, and overweight/obese.
[61] Water arsenic concentration cut-points roughly correspond to quartiles of the exposure distribution for the controls but have been slightly adjust to reflect policy relevant parameters.
[62] Creatinine-adjusted urinary total arsenic concentration.

TABLE 3. Urinary Arsenic Metabolite Distribution for the Study Sample, Araihazar, Bangladesh, 2000-2002

Characteristic	Cases (n=610)[63]		Controls (n=1,079)[64]		Crude Effect Estimate	
	N	%	N	%	POR	95% CI
% Total MMA						
<8.650	59	10.0	260	25.0	1.0	
8.650-11.765	130	22.0	261	25.0	2.2	1.5, 3.1
11.766-14.903	146	24.7	269	25.0	2.5	1.7, 3.5
14.904-33.415	257	43.4	260	25.0	4.4	3.1, 6.0
% Total DMA						
<63.500	192	32.4	260	25.0	1.0	
63.500-69.215	152	25.7	260	25.0	0.8	0.6, 1.0
69.216-74.480	124	21.0	262	25.1	0.6	0.5, 0.8
74.481-90.828	124	21.0	260	24.9	0.6	0.5, 0.9
% Inorganic arsenic[65]						
<10.81	152	25.7	260	25.0	1.0	
10.81-14.20	145	24.5	261	25.0	0.9	0.7, 1.3
14.21-18.31	141	23.8	262	25.1	0.9	0.7, 1.2
18.32-69.30	154	26.0	259	24.9	1.0	0.7, 1.3
Primary methylation index[66,67]						
<0.580	70	11.8	260	25.0	1.0	
0.580-0.835	132	22.3	259	25.0	1.9	1.3, 2.6
0.836-1.154	171	28.9	260	25.0	2.4	1.8, 3.4
1.155-19.570	218	36.9	260	25.0	3.1	2.3, 4.3
Secondary methylation index[68,69]						
<4.316	236	39.9	259	25.0	1.0	
4.316-5.855	153	25.9	259	25.0	0.6	0.5, 0.8
5.856-8.211	128	21.7	259	25.0	0.5	0.4, 0.7
8.212-32.300	74	12.5	259	25.0	0.3	0.2, 0.4

[63] All arsenic metabolites data missing for 18 case subjects.
[64] All arsenic metabolite data missing for 37 control subjects.
[65] Inorganic arsenic is sum of arsenate and arsenite.
[66] The primary methylation index was calculated as % MMA / % Inorganic arsenic.
[67] Primary methylation index missing for 1 additional case and 3 additional controls subjects due to 0 values.
[68] The secondary methylation index was calculated as % DMA / % MMA.
[69] Secondary methylation index missing for 1 additional case and 6 additional control subjects due to 0 values.

TABLE 4. Prevalence Odds Ratios and 95% CIs for Skin Lesion Prevalence in Relation to DNA Repair SNPs, Araihazar, Bangladesh, 2000-2002

Characteristic	MAF[70]	Cases (n=610)		Controls (n=1,079)		Crude Estimate POR (95% CI)	*P*[71]	Multivariate Estimate[72] POR (95% CI)	*P*[71]
		N	%	N	%				
OGG1	G=0.31								
CC		264	46.0	460	47.2	1.0		1.0	
CG		259	45.1	423	43.4	1.1 (0.9, 1.3)		0.9 (0.7, 1.2)	
GG		51	8.9	92	9.4	1.0 (0.7, 1.4)		0.8 (0.5, 1.3)	
Per allele						1.0 (0.9, 1.2)	0.85	0.9 (0.8, 1.1)	0.42
ERCC2	C=0.33								
AA		242	43.4	421	44.5	1.0		1.0	
AC		262	46.9	419	44.3	1.1 (0.9, 1.4)		1.1 (0.8, 1.4)	
CC		54	9.7	105	11.1	0.9 (0.6, 1.3)		0.9 (0.6, 1.3)	
Per allele						1.0 (0.9, 1.2)	0.94	1.0 (0.8, 1.2)	0.91
ERCC5	C=0.27								
GG		298	52.8	503	52.0	1.0		1.0	
GC		223	39.5	400	41.4	0.9 (0.8, 1.2)		0.9 (0.7, 1.2)	
CC		43	7.6	64	6.6	1.1 (0.8, 1.7)		1.1 (0.7, 1.7)	
Per allele						1.0 (0.8, 1.2)	0.95	1.0 (0.8, 1.2)	0.98
LIG4	T=0.09								
CC		473	82.8	788	82.3	1.0		1.0	
CT		92	16.1	165	17.2	0.9 (0.7, 1.2)		1.0 (0.7, 1.3)	
TT		6	1.0	4	0.4	2.5 (0.7, 8.9)		2.6 (0.6, 11.4)	
Per allele						1.0 (0.8, 1.3)	0.95	1.1 (0.8, 1.4)	0.70
XRCC1 (Ex6)	T=0.10								
CC		445	79.5	763	80.6	1.0		1.0	
CT		104	18.6	170	18.0	1.0 (0.8, 1.4)		1.2 (0.8, 1.6)	
TT		11	2.0	13	1.4	1.4 (0.6, 3.3)		1.5 (0.5, 4.1)	
Per allele						1.1 (0.9, 1.4)	0.45	1.2 (0.9, 1.6)	0.23
XRCC1 (Ex10)	A=0.33								
GG		240	42.5	423	44.1	1.0		1.0	
GA		251	44.4	429	44.8	1.0 (0.8, 1.3)		1.0 (0.8, 1.3)	
AA		74	13.1	106	11.1	1.2 (0.9, 1.7)		1.3 (0.9, 1.9)	
Per allele						1.1 (0.9, 1.3)	0.30	1.1 (0.9, 1.3)	0.29
XRCC3	T=0.20								
CC		352	63.4	619	65.3	1.0		1.0	
CT		180	32.4	287	30.3	1.1 (0.9, 1.4)		1.1 (0.8, 1.4)	
TT		23	4.1	42	4.4	1.0 (0.6, 1.6)		0.7 (0.4, 1.3)	
Per allele						1.0 (0.9, 1.3)	0.60	1.0 (0.8, 1.2)	0.95

[70] Minor allele frequency in the controls.
[71] *P* for trend.
[72] Adjusted for sex, age, and well water arsenic concentration.

TABLE 5. Prevalence Odds Ratios and 95% CIs for Skin Lesion Prevalence in Relation to *XRCC1* Haplotypes, Araihazar, Bangladesh, 2000-2002

rs1799782/rs25487	Case Frequency (N=546)	Control Frequency (N=926)	Crude POR (95% CI)	POR (95% CI)[73]
CG	0.537	0.559	1.0	1.0
CA	0.352	0.336	1.1 (0.9, 1.3)	1.1 (0.9, 1.3)
TG	0.112	0.105	1.1 (0.9, 1.4)	1.1 (0.9, 1.4)

[73] Adjusted for sex, age, and well water arsenic concentration.

TABLE 6. Prevalence Odds Ratios and 95% CIs for Skin Lesion Prevalence in Relation to DNA Repair Risk Allele Count, Araihazar, Bangladesh, 2000-2002

Risk Allele Count[74]	Cases (n=481)		Controls (n=824)		Crude Estimate		Multivariate Estimate[75]	
	N	%	N	%	POR	95% CI	POR	95% CI
0–3	287	59.7	516	62.6	1.0		1.0	
4–8	194	40.3	308	37.4	1.1	0.9, 1.4	1.1	0.8, 1.4

[74] *LIG4* excluded from the risk allele count.
[75] Adjusted for sex, age, and well water arsenic concentration.

TABLE 7. Prevalence Odds Ratios[76] and 95% CIs for Skin Lesion Prevalence in Relation to DNA Repair SNPs by Well Water Arsenic Concentration, Araihazar, Bangladesh, 2000-2002

SNP	Well Water Arsenic Concentration (µg/L)				RERI
	0.1-50	50.1-100	100.1-150	150.1-864	
OGG1					-0.03
CC	1.00	1.90 (1.06, 3.39)	2.39 (1.41, 4.06)	4.40 (2.54, 7.63)	(-0.25, 0.16)
CG/GG	1.00[77]	1.40 (0.78, 2.49)	2.50 (1.47, 4.22)	4.15 (2.46, 7.01)	
ERCC2					-0.01
AA	1.00	1.56 (0.83, 2.93)	3.31 (1.87, 5.86)	4.91 (2.74, 8.77)	(-0.25, 0.25)
AC/CC	1.00[78]	2.03 (1.12, 3.71)	2.83 (1.62, 4.93)	5.36 (3.05, 9.41)	
ERCC5					-0.05
GG	1.00	1.79 (1.01, 3.16)	2.95 (1.78, 4.91)	4.28 (2.58, 7.10)	(-0.25, 0.16)
GC/CC	1.00[79]	1.63 (0.93, 2.86)	2.16 (1.29, 3.61)	4.82 (2.84, 8.18)	
XRCC1 (Ex6)					0.03
CC	1.00	1.32 (0.84, 2.09)	2.26 (1.49, 3.40)	4.01 (2.64, 6.08)	(-0.39, 0.36)
CT/TT	1.00[80]	1.97 (0.97, 4.01)	2.60 (1.47, 4.59)	4.15 (2.25, 7.64)	
XRCC1 (Ex10)					0.04
GG	1.00	1.13 (0.61, 2.10)	2.06 (1.19, 3.55)	3.19 (1.84, 5.54)	(-0.15, 0.23)
GA/AA	1.00[81]	1.42 (0.80, 2.50)	1.91 (1.13, 3.22)	3.65 (2.15, 6.22)	
XRCC3					0.02
CC	1.00	1.24 (0.75, 2.06)	2.15 (1.36, 3.39)	3.88 (2.45, 6.16)	(-0.22, 0.25)
CT/TT	1.00[82]	1.76 (0.99, 3.13)	2.47 (1.49, 4.09)	3.54 (2.10, 5.97)	
Risk allele					0.08
0-3	1.00	1.35 (0.76, 2.39)	2.16 (1.31, 3.56)	4.47 (2.68, 7.46)	(-0.16, 0.36)
4-8	1.00[83]	1.57 (0.86, 2.87)	2.59 (1.50, 4.48)	4.77 (2.69, 8.45)	

[76] Adjusted for sex and age.
[77] POR = 1.19 comparing CG/GG to CC in this lowest exposure quartile.
[78] POR = 1.55 comparing AC/CC to AA in this lowest exposure quartile.
[79] POR = 1.35 comparing GC/CC to GG in this lowest exposure quartile.
[80] POR = 1.40 comparing CT/TT to CC in this lowest exposure quartile.
[81] POR = 0.87 comparing GA/AA to GG in this lowest exposure quartile.
[82] POR = 0.85 comparing CT/TT to CC in this lowest exposure quartile.
[83] POR = 1.12 comparing 4-8 to 0-3 in this lowest exposure quartile.

TABLE 8. Prevalence Odds Ratios[84] and 95% CIs for Skin Lesion Prevalence in Relation to DNA Repair SNPs by Creatinine-adjusted Urinary Total Arsenic Concentration, Araihazar, Bangladesh, 2000-2002

SNP	Creatinine-adjusted urinary total arsenic concentration (μg/g)				RERI
	16-103	104-192	193-339	340-8556	
OGG1					-0.01
CC	1.00	2.98 (1.65, 5.36)	3.26 (1.86, 5.70)	6.09 (3.50, 10.60)	(-0.21, 0.19)
CG/GG	1.00[85]	1.72 (0.98, 3.03)	3.12 (1.77, 5.49)	6.76 (3.96, 11.54)	
ERCC2					0.03
AA	1.00	2.34 (1.24, 4.39)	3.66 (1.98, 6.78)	6.79 (3.71, 12.43)	(-0.22, 0.29)
AC/CC	1.00[86]	2.45 (1.33, 4.53)	3.48 (1.90, 6.37)	7.18 (4.00, 12.87)	
ERCC5					-0.03
GG	1.00	1.50 (0.86, 2.64)	2.59 (1.52, 4.42)	5.11 (3.07, 8.51)	(-0.27, 0.17)
GC/CC	1.00[87]	1.73 (1.00, 3.00)	2.36 (1.37, 4.08)	4.85 (2.86, 8.23)	
XRCC1 (Ex6)					0.08
CC	1.00	1.82 (1.17, 2.84)	2.88 (1.86, 4.46)	5.15 (3.38, 7.85)	(-0.22, 0.45)
CT/TT	1.00[88]	2.20 (1.13, 4.28)	2.26 (1.21, 4.23)	6.93 (3.83, 12.53)	
XRCC1 (Ex10)					0.06
GG	1.00	1.55 (0.85, 2.80)	2.49 (1.40, 4.41)	3.98 (2.30, 6.90)	(-0.13, 0.23)
GA/AA	1.00[89]	1.56 (0.90, 2.71)	2.18 (1.27, 3.75)	5.05 (2.97, 8.60)	
XRCC3					0.11
CC	1.00	1.72 (1.05, 2.81)	2.08 (1.28, 3.39)	4.34 (2.73, 6.90)	(-0.07, 0.33)
CT/TT	1.00[90]	1.33 (0.75, 2.36)	3.05 (1.77, 5.27)	5.23 (3.10, 8.82)	
Risk allele					0.16
0-3	1.00	1.49 (0.85, 2.59)	2.65 (1.56, 4.53)	4.60 (2.75, 7.69)	(-0.08, 0.45)
4-8	1.00[91]	1.69 (0.93, 3.04)	2.93 (1.61, 5.33)	6.32 (3.57, 11.21)	

[84] Adjusted for sex and age.
[85] POR = 1.54 comparing CG/GG to CC in this lowest exposure quartile.
[86] POR = 1.64 comparing AC/CC to AA in this lowest exposure quartile.
[87] POR = 0.95 comparing GC/CC to GG in this lowest exposure quartile.
[88] POR = 1.67 comparing CT/TT to CC in this lowest exposure quartile.
[89] POR = 0.87 comparing GA/AA to GG in this lowest exposure quartile.
[90] POR = 0.76 comparing CT/TT to CC in this lowest exposure quartile.
[91] POR = 0.90 comparing 4-8 to 0-3 in this lowest exposure quartile.

TABLE 9. Prevalence Odds Ratios[92] and 95% CIs for Skin Lesion Prevalence in Relation to DNA Repair SNPs by Primary Methylation Index, Araihazar, Bangladesh, 2000-2002

SNP	Primary Methylation Index				RERI
	<0.580	0.580-0.835	0.836-1.154	1.155-19.570	
OGG1					-0.11
CC	1.00	1.17 (0.66, 2.10)	1.49 (0.86, 2.59)	1.52 (0.88, 2.62)	(-0.42, 0.10)
CG/GG	1.00[93]	1.74 (1.01, 3.01)	1.36 (0.79, 2.34)	1.35 (0.79, 2.31)	
ERCC2					0.11
AA	1.00	1.20 (0.66, 2.20)	1.33 (0.74, 2.41)	1.05 (0.59, 1.87)	(-0.12, 0.26)
AC/CC	1.00[94]	1.38 (0.77, 2.47)	1.18 (0.67, 2.09)	1.34 (0.76, 2.36)	
ERCC5					0.13
GG	1.00	1.51 (0.89, 2.58)	1.67 (0.99, 2.82)	1.11 (0.65, 1.88)	(-0.03, 0.27)
GC/CC	1.00[95]	1.25 (0.71, 2.19)	1.03 (0.60, 1.79)	1.64 (0.97, 2.77)	
XRCC1 (Ex6)					0.18
CC	1.00	1.42 (0.91, 2.22)	1.47 (0.95, 2.27)	1.30 (0.84, 2.00)	(-0.12, 0.36)
CT/TT	1.00[96]	1.63 (0.85, 3.13)	1.51 (0.82, 2.79)	2.16 (1.17, 3.97)	
XRCC1 (Ex10)					0.06
GG	1.00	1.67 (0.91, 3.04)	1.55 (0.87, 2.78)	1.36 (0.77, 2.39)	(-0.17, 0.23)
GA/AA	1.00[97]	1.49 (0.85, 2.60)	1.55 (0.89, 2.68)	1.66 (0.96, 2.88)	
XRCC3					-0.05
CC	1.00	1.46 (0.89, 2.41)	1.43 (0.88, 2.32)	1.55 (0.97, 2.50)	(-0.38, 0.16)
CT/TT	1.00[98]	1.56 (0.88, 2.77)	1.62 (0.95, 2.79)	1.50 (0.87, 2.58)	
Risk allele					0.10
0-3	1.00	1.38 (0.80, 2.39)	1.39 (0.82, 2.38)	1.23 (0.73, 2.07)	(-0.18, 0.28)
4-8	1.00[99]	1.48 (0.81, 2.69)	1.31 (0.74, 2.30)	1.74 (0.96, 3.13)	

[92] Adjusted for sex and age.
[93] POR = 0.99 comparing CG/GG to CC in this lowest exposure quartile.
[94] POR = 0.80 comparing AC/CC to AA in this lowest exposure quartile.
[95] POR = 0.83 comparing GC/CC to GG in this lowest exposure quartile.
[96] POR = 0.87 comparing CT/TT to CC in this lowest exposure quartile.
[97] POR = 1.11 comparing GA/AA to GG in this lowest exposure quartile.
[98] POR = 1.20 comparing CT/TT to CC in this lowest exposure quartile.
[99] POR = 1.02 comparing 4-8 to 0-3 in this lowest exposure quartile.

TABLE 10. Prevalence Odds Ratios[100] and 95% CIs for Skin Lesion Prevalence in Relation to DNA Repair SNPs by Secondary Methylation Index, Araihazar, Bangladesh, 2000-2002

SNP	Secondary Methylation Index				RERI
	<4.316	4.316-5.855	5.856-8.211	8.212-32.300	
OGG1					0.01
CC	1.00	0.84 (0.54, 1.33)	0.59 (0.37, 0.94)	0.71 (0.42, 1.21)	(-0.21, 0.18)
CG/GG	1.00[101]	0.76 (0.49, 1.17)	0.94 (0.59, 1.48)	0.47 (0.27, 0.80)	
ERCC2					-0.03
AA	1.00	0.82 (0.51, 1.29)	0.77 (0.47, 1.27)	0.54 (0.30, 0.97)	(-0.29, 0.17)
AC/CC	1.00[102]	0.79 (0.50, 1.23)	0.82 (0.52, 1.30)	0.65 (0.39, 1.09)	
ERCC5					0.09
GG	1.00	0.93 (0.61, 1.42)	0.86 (0.55, 1.34)	0.72 (0.43, 1.18)	(-0.10, 0.23)
GC/CC	1.00[103]	0.78 (0.50, 1.21)	0.77 (0.49, 1.21)	0.47 (0.27, 0.82)	
XRCC1 (Ex6)					-0.06
CC	1.00	0.75 (0.53, 1.07)	0.71 (0.50, 1.03)	0.56 (0.37, 0.85)	(-0.50, 0.20)
CT/TT	1.00[104]	1.13 (0.64, 2.02)	1.09 (0.60, 1.98)	0.57 (0.27, 1.21)	
XRCC1 (Ex10)					-0.06
GG	1.00	0.81 (0.50, 1.30)	0.60 (0.36, 1.00)	0.55 (0.32, 0.94)	(-0.31, 0.14)
GA/AA	1.00[105]	0.75 (0.49, 1.15)	0.83 (0.53, 1.29)	0.58 (0.34, 0.98)	
XRCC3					-0.22
CC	1.00	0.78 (0.53, 1.14)	0.70 (0.47, 1.06)	0.41 (0.26, 0.67)	(-0.67, 0.05)
CT/TT	1.00[106]	0.83 (0.52, 1.31)	0.71 (0.44, 1.14)	0.81 (0.46, 1.41)	
Risk allele					0.10
0-3	1.00	0.78 (0.50, 1.20)	0.85 (0.55, 1.32)	0.63 (0.38, 1.05)	(-0.16, 0.27)
4-8	1.00[107]	1.00 (0.62, 1.61)	0.78 (0.46, 1.32)	0.57 (0.30, 1.07)	

[100] Adjusted for sex and age.
[101] POR = 0.99 comparing CG/GG to CC in this lowest exposure quartile.
[102] POR = 1.03 comparing AC/CC to AA in this lowest exposure quartile.
[103] POR = 1.14 comparing GC/CC to GG in this lowest exposure quartile.
[104] POR = 0.90 comparing CT/TT to CC in this lowest exposure quartile.
[105] POR = 0.98 comparing GA/AA to GG in this lowest exposure quartile.
[106] POR = 0.86 comparing CT/TT to CC in this lowest exposure quartile.
[107] POR = 1.18 comparing 4-8 to 0-3 in this lowest exposure quartile.

TABLE 11. Prevalence Odds Ratios[108] and 95% CIs for Skin Lesion Prevalence in Relation to DNA Repair SNPs by Body Mass Index, Araihazar, Bangladesh, 2000-2002

SNP	Body Mass Index			RERI
	<18.5	18.5-24.9	>25.0	
OGG1				-0.16
CC	1.00	0.75 (0.52, 1.08)	0.37 (0.15, 0.92)	(-0.71, 0.17)
CG/GG	1.00[109]	0.72 (0.51, 1.03)	0.46 (0.21, 1.02)	
ERCC2				-0.05
AA	1.00	0.72 (0.50, 1.05)	0.40 (0.14, 1.12)	(-0.67, 0.30)
AC/CC	1.00[110]	0.79 (0.56, 1.13)	0.52 (0.24, 1.09)	
ERCC5				-0.24
GG	1.00	0.75 (0.53, 1.05)	0.25 (0.10, 0.65)	(-0.94, 0.16)
GC/CC	1.00[111]	0.73 (0.52, 1.02)	0.61 (0.27, 1.36)	
XRCC1 (Ex6)				0.24
CC	1.00	0.88 (0.66, 1.16)	0.47 (0.24, 0.92)	(-0.41, 0.53)
CT/TT	1.00[112]	0.80 (0.51, 1.26)	1.06 (0.27, 4.12)	
XRCC1 (Ex10)				-0.06
GG	1.00	0.79 (0.54, 1.16)	0.36 (0.14, 0.91)	(-0.65, 0.28)
GA/AA	1.00[113]	0.83 (0.58, 1.18)	0.59 (0.26, 1.34)	
XRCC3				0.10
CC	1.00	0.89 (0.65, 1.22)	0.37 (0.17, 0.84)	(-0.39, 0.38)
CT/TT	1.00[114]	0.80 (0.56, 1.15)	0.71 (0.29, 1.74)	
Risk allele				0.07
0-3	1.00	0.72 (0.51, 1.02)	0.47 (0.20, 1.11)	(-0.64, 0.39)
4-8	1.00[115]	0.81 (0.55, 1.18)	0.53 (0.20, 1.36)	

[108] Adjusted for sex, age, and well water arsenic concentration.
[109] POR = 0.85 comparing CG/GG to CC in this lowest exposure tertile.
[110] POR = 1.02 comparing AC/CC to AA in this lowest exposure tertile.
[111] POR = 0.89 comparing GC/CC to GG in this lowest exposure tertile.
[112] POR = 1.48 comparing CT/TT to CC in this lowest exposure tertile.
[113] POR = 1.05 comparing GA/AA to GG in this lowest exposure tertile.
[114] POR = 1.21 comparing CT/TT to CC in this lowest exposure tertile.
[115] POR = 1.14 comparing 4-8 to 0-3 in this lowest exposure tertile.

TABLE 12. Prevalence Odds Ratios[116] and 95% CIs for Skin Lesion Prevalence in Relation to DNA Repair SNPs by Smoking Status, Araihazar, Bangladesh, 2000-2002

SNP	Smoking Status		RERI
	Never	Ever	
OGG1			-0.23
CC	1.00	1.33 (0.88, 2.00)	(-0.86, 0.26)
CG/GG	1.00[117]	1.13 (0.76, 1.68)	
ERCC2			0.34
AA	1.00	1.01 (0.65, 1.55)	(-0.22, 0.77)
AC/CC	1.00[118]	1.21 (0.80, 1.84)	
ERCC5			0.36
GG	1.00	1.06 (0.71, 1.57)	(-0.14, 0.79)
GC/CC	1.00[119]	1.20 (0.80, 1.80)	
XRCC1 (Ex6)			-0.55
CC	1.00	1.41 (1.00, 1.98)	(-1.59, 0.31)
CT/TT	1.00[120]	1.39 (0.87, 2.22)	
XRCC1 (Ex10)			-0.35
GG	1.00	1.40 (0.92, 2.16)	(-1.24, 0.25)
GA/AA	1.00[121]	1.32 (0.87, 2.00)	
XRCC3			0.35
CC	1.00	1.16 (0.80, 1.69)	(-0.28, 0.90)
CT/TT	1.00[122]	1.38 (0.92, 2.07)	
Risk allele			-0.12
0-3	1.00	1.35 (0.90, 2.05)	(-0.96, 0.49)
4-8	1.00[123]	1.44 (0.92, 2.24)	

[116] Adjusted for sex, age, and well water arsenic concentration.
[117] POR = 1.04 comparing CG/GG to CC in this lowest exposure category.
[118] POR = 0.86 comparing AC/CC to AA in this lowest exposure category.
[119] POR = 0.78 comparing GC/CC to GG in this lowest exposure category.
[120] POR = 1.54 comparing CT/TT to CC in this lowest exposure category.
[121] POR = 1.27 comparing GA/AA to GG in this lowest exposure category.
[122] POR = 0.87 comparing CT/TT to CC in this lowest exposure category.
[123] POR = 1.20 comparing 4-8 to 0-3 in this lowest exposure category.

TABLE 13. Prevalence Odds Ratios[124] and 95% CIs for Skin Lesion Prevalence in Relation to DNA Repair SNPs by Fruit and Vegetable Intake, Araihazar, Bangladesh, 2000-2002

SNP	Fruit and vegetable intake, g/day			RERI
	74.0-415.5	415.6-593.0	≥593.1	
OGG1				-0.19
CC	1.00	0.65 (0.41, 1.01)	0.48 (0.30, 0.76)	(-0.63, 0.12)
CG/GG	1.00[125]	0.56 (0.36, 0.87)	0.53 (0.34, 0.82)	
ERCC2				-0.16
AA	1.00	0.54 (0.33, 0.86)	0.46 (0.29, 0.74)	(-0.61, 0.17)
AC/CC	1.00[126]	0.66 (0.43, 1.02)	0.49 (0.31, 0.77)	
ERCC5				-0.78
GG	1.00	0.48 (0.31, 0.73)	0.32 (0.20, 0.50)	(-1.43, -0.29)
GC/CC	1.00[127]	0.48 (0.30, 0.75)	0.51 (0.33, 0.79)	
XRCC1 (Ex6)				0.04
CC	1.00	0.57 (0.40, 0.81)	0.54 (0.38, 0.78)	(-0.56, 0.49)
CT/TT	1.00[128]	1.01 (0.58, 1.74)	0.52 (0.29, 0.94)	
XRCC1 (Ex10)				-0.04
GG	1.00	0.67 (0.42, 1.05)	0.47 (0.29, 0.77)	(-0.47, 0.27)
GA/AA	1.00[129]	0.66 (0.43, 1.02)	0.61 (0.40, 0.93)	
XRCC3				0.18
CC	1.00	0.69 (0.47, 1.03)	0.65 (0.44, 0.96)	(-0.18, 0.47)
CT/TT	1.00[130]	0.80 (0.51, 1.28)	0.53 (0.33, 0.85)	
Risk allele				-0.34
0-3	1.00	0.51 (0.33, 0.79)	0.44 (0.28, 0.68)	(-0.99, 0.07)
4-8	1.00[131]	0.73 (0.45, 1.19)	0.33 (0.32, 0.87)	

[124] Adjusted for sex, age, well water arsenic concentration, total caloric intake, and BMI.
[125] POR = 0.83 comparing CG/GG to CC in this lowest exposure tertile.
[126] POR = 0.85 comparing AC/CC to AA in this lowest exposure tertile.
[127] POR = 0.52 comparing GC/CC to GG in this lowest exposure tertile.
[128] POR = 0.99 comparing CT/TT to CC in this lowest exposure tertile.
[129] POR = 1.05 comparing GA/AA to GG in this lowest exposure tertile.
[130] POR = 1.33 comparing CT/TT to CC in this lowest exposure tertile.
[131] POR = 0.70 comparing 4-8 to 0-3 in this lowest exposure tertile.

Chapter 4

Gene-Gene, Gene-Environment, and Higher Order Interactions in Relation to Arsenic-related Skin Lesions in an Adult Bangladeshi Population

ABSTRACT

The oxidative stress and DNA repair pathways have been implicated in arsenic toxicity and have been hypothesized to underlie arsenic carcinogenesis. The object of this analysis was to explore gene-gene, gene-environment, and higher-order interactions among 10 single nucleotide polymorphisms (SNPs) related to the oxidative stress and DNA repair pathways by multifactor dimensionality reduction (MDR), classification and regression trees (CART) and logistic regression models. Data from 610 prevalent skin lesion cases and 1,079 controls from the Health Effects of Arsenic Longitudinal Study (HEALS) were utilized in these analyses. Based on single SNP evaluation in logistic regression models, none of the SNPs were associated with skin lesion prevalence, however, were evaluated for potential epistatic effects. Our analytic method utilized MDR and CART modeling approaches for the selection of potential gene-gene and gene-environment interactions. Considerable overlap of the interactions detected by both these methods was observed, which were further evaluated by logistic regression. Results from logistic regression modeling indicated evidence for some of these statistical interactions; however, further investigation and replication of the statistical interactions detected in this study is needed.

INTRODUCTION

Globally, more than 100 million people, including approximately 28–57 million in Bangladesh, are chronically exposed to arsenic through naturally contaminated drinking water (1). The International Agency for Research on Cancer has categorized arsenic as a class I human carcinogen (2). Arsenic in drinking water has been associated with increased risk of a wide range of health outcomes including cancers of the skin, lung, bladder, liver, and kidney (3-7), neurological disease (8), cardiovascular disease (9), as well as other non-malignant diseases (10, 11). While most arsenic-related cancers have long average latency periods, skin lesions appear within a relatively shorter period of time following exposure to arsenic (12, 13). Additionally, skin lesions are considered precursors to a majority of the arsenic-induced basal and squamous cell skin cancers (14).

Due to the absence of a suitable animal model for the evaluation of arsenic toxicity, a mode of action has yet to be determined for arsenic carcinogenesis (15, 16). Although, the oxidative stress and DNA repair pathways have been implicated in arsenic toxicity and have been hypothesized to underlie arsenic carcinogenesis (17). To date, few epidemiologic studies have evaluated genetic susceptibility to arsenical skin lesions based on single nucleotide polymorphisms (SNPs) in antioxidant enzyme or DNA repair genes. Only a single prior study by Ahsan et al. has evaluated SNPs in antioxidant enzyme genes (*MPO* and *CAT*) in relation to skin lesion prevalence and no significant main effects of the SNPs were observed (18). There have been 5 epidemiologic studies which have evaluated polymorphisms in DNA repair genes (*ERCC2*, *APEX1*, *XRCC1*, *XRCC3*, and *OGG1*) in relation to skin lesion prevalence and observations have been conflicting (19-23). Banerjee et al. observed a significant increased risk of hyperkeratosis in relation to

the C allele of a polymorphism in *ERCC2* (rs13181) (20), while others have not (19, 23). Breton et al. observed a significant increased risk of skin lesions in relation to the GG genotype compared to the TT genotype of a polymorphism in APEX1 (rs3136820) (21); however, this SNP has not been replicated in other studies with respect to skin lesion risk. Significant marginal effects of *XRCC1*, *XRCC3*, and *OGG1* in relation to skin lesions have not been observed (21, 22).

It has been previously described that genes may exhibit epistatic effects particularly for complex diseases, when no marginal effect of the gene is detected (24). There are several statistical approaches for evaluating gene-gene as well as gene-environment interactions. Historically, epidemiologists have favored logistic regression for evaluating interaction within a case-control setting. With technological advances in genotyping, the number of SNPs that can be typed within a study has increasingly grown, which in turn has made the number of possible interactions to explore very large. For example, all possible pair-wise interactions between SNPs would equal N!/[2! (N−2)!], where N is the number of SNPs. For the 11 SNPs examined in this study, there would be 55 SNP-pair interactions to evaluate. Therefore, for hypotheses where an *a priori* interaction is not specified, logistic regression is not an optimal tool for exploratory analyses of SNP-SNP interactions. Recently, dimensionality reduction approaches have been applied to these research questions, including multifactor dimensionality reduction (MDR) (25-29), classification and regression trees (CART) (30), and random forests (RF) (31). García-Magariños et al. (32) conducted a simulation study to evaluate the ability of these three methods as well as logistic regression to detect SNP-SNP interactions under different modeling scenarios (e.g., sample size, missing data, minor allele frequencies,

and penetrance); they concluded that CART performed well under all scenarios and was less computationally intensive than the other approaches. Additionally, Schwender et al. (33) conducted several classification methods in an analysis of 25 SNPs in a case-control study of breast cancer and also showed that the misclassification rate for the various methods applied was similar.

In this study, we utilized data from 610 prevalent skin lesion cases and 1,079 controls from the Health Effects of Arsenic Longitudinal Study (HEALS) cohort to evaluate gene-gene and gene-environment interactions in relation to skin lesion prevalence based on MDR, CART, and logistic regression analyses for SNPs in the oxidative stress (*SOD2,* rs4880; *CAT,* rs1001179; *GPX1,* rs1050450; and *MPO*, rs2333227) and DNA repair pathways (*OGG1,* rs1052133; *XRCC1,* rs25487 and rs1799782; *XRCC3,* rs861539; *ERCC2,* rs1052559; *ERCC5,* rs17655; and *LIG4,* rs1805388), as well as various measures of arsenic exposure in relation to skin lesion prevalence.

MATERIALS AND METHODS

Study Population

HEALS is an ongoing, population-based cohort study examining both the short- and long-term health effects of arsenic exposure. The study was launched in Araihazar, Bangladesh, in 2000. The selection of cohort participants, study design and methods have been described in detail elsewhere (34).

The prevalent cases and controls utilized in these analyses were selected from the baseline survey of the HEALS cohort. Between October 2000 and May 2002, married

individuals were sampled—an eligibility criteria to minimize loss to follow-up—who were aged between 18 and 75 years and were residents of the study area for at least 5 years. From the 12,050 residents who met the eligibility criteria from an enumerated total 65,876 persons in the study area, 11,746 (97.5% response rate) men and women (4,801 married couples and 2,144 married individuals whose spouses did not participate) were enrolled into the HEALS cohort. Among the 11,746 enrolled baseline cohort participants, 11,224 participants (95.6%) provided urine samples. At baseline, 610 cases of arsenical skin lesions were identified among the 11,224 cohort members who provided urine samples and were included as prevalent cases in this study. Of the remaining participants with an available urine sample and known to be free of skin lesions at baseline (n=10,614), a ~10% random sample (n=1,079) was selected and included as controls in this study.

The distribution of skin lesion severity among the 610 cases was as follows: 359 cases had melanosis only, 20 had leucomelanosis only, 10 had melanosis and leucomelanosis, 170 had melanosis and keratosis, 40 had leucomelanosis and keratosis, and 11 had melanosis, leucomelanosis, and keratosis.

The study protocol was approved by the Institutional Review Boards of Columbia University, The University of Chicago, and the Bangladesh Medical Research Council. Informed consent was obtained from all participants prior to baseline interview.

Single Nucleotide Polymorphism Assessment

High-throughput DNA extraction was completed in 96-well format using the QIAmp DNA 96 DNA Blood kit (Qiagen, Valencia, CA). Replica plates were made with

12.5 ng DNA in 2.5 μL per well. In the first step, the genomic DNA was amplified by PCR using appropriate primers. After PCR amplification, the primers and deoxynucleotide triphosphates in 10 μL PCR product were digested with the 10 μL shrimp alkaline phosphatase cocktail containing 1.0 μL (1 unit/μL) of shrimp alkaline phosphatase, 0.1 μL of Escherichia coli exonuclease I (10 units/μL; U.S. Biochemical, Cleveland, OH), 1.0 μL of 10x shrimp alkaline phosphatase buffer, and 7.9 μL of DNase and RNase-free water for 45 min at 37°C followed by heating at 95°C for 15 min for enzyme deactivation. Then, single nucleotide extension was carried out in the presence of the appropriate allele-specific dideoxynucleotide triphosphates fluorescence labeled with either R110 or TAMRA (PerkinElmer, Waltham, MA). For single nucleotide extension reactions, both the forward and reverse probes were initially tested to select the better probe based on clear signal clustering. Reaction mixture (13 μL/well) containing 0.025 μL AcycloPrime enzyme, 0.5 μL terminator dye, 1 μL reaction buffer, 0.25 μL extension probe (10 pmol/μL), and 11.225 μL water was added to 7 μL of digested PCR product to make 20 μL reaction volume. Thermocycling was done at 95°C for 3 min followed by optimum number of cycles of 95°C for 15 s and 55°C for 30 s. Finally, the fluorescence was measured with Wallac 1420 Multilabel Counter Victor 3 (PerkinElmer, Waltham, MA). In addition to our assay-specific quality control samples, 10% of the samples were run in duplicate after relabeling to keep laboratory researchers blinded to its identity. Concordance based on the duplicates was >0.92. Call rates for the SNPs of interest ranged from 95.0% to 97.9%, as shown in **Table 1**.

Skin Lesion Assessment

Arsenical skin lesion assessment was conducted through skin examination at the baseline interview following a structured protocol by a trained study physician (34, 35). Arsenical skin lesions were categorized as the presence of melanosis on the body surface, leucomelanosis on the body surface, and keratosis on the hands or feet. For the purposes of this study, cases were selected as having the presence of any type of arsenical skin lesion.

Arsenic exposure assessment

Three primary measures of arsenic exposure were estimated for each case-control participant: well water arsenic concentration, creatinine-adjusted urinary total arsenic concentration, and urinary arsenic metabolites.

At the baseline survey, participants were asked to identify the well they primarily used as their source of drinking water, from which we were able to assign the appropriate well water arsenic concentration exposure. Well water arsenic concentrations of all 5,966 tubewells in the study area were measured by graphite furnace atomic absorption spectrometry, with a detection limit of 5 μg/L. Samples below the limit of detection were subsequently reanalyzed by inductively coupled plasma-mass spectrometry (ICP-MS), with a detection limit of 0.1 μg/L (36).

Urinary total arsenic concentration was measured in a spot urine sample collected at baseline by graphite furnace atomic absorption spectrometry, with a detection limit of 2 μg/L (37). Urinary creatinine was measured by a colorimetric Sigma Diagnostics Kit (Sigma, St. Louis, MO), and urinary total arsenic was subsequently divided by creatinine

to obtain a creatinine-adjusted urinary total arsenic concentration, expressed as μg/g creatinine (38).

Urinary arsenic metabolites were measured based on the method by Reuter et al. (39) using high-performance liquid chromatography separation of arsenobetaine, arsenocholine, arsenate (As^{V}), arsenite (As^{III}), total monomethyl arsenic (MMA^{III} + MMA^{V}), and total dimethyl arsenic (DMA^{III} + DMA^{V}) followed by detection by ICP-MS-dynamic reaction cell. Total inorganic arsenic (As^{III} + As^{V}) is utilized in these analyses since As^{III} can oxidize to As^{V} during sample transport, storage, and preparation.

For purposes of analysis, well water arsenic was categorized into quartiles based on the distribution in the controls; however, since the first and second quartiles roughly corresponded to the World Health Organization's guideline for arsenic in drinking water (10 μg/L) and the national standard for arsenic in drinking water in Bangladesh (50 μg/L), respectively, we adjusted the cutoff points slightly to correspond to these regulatory levels. Urinary total arsenic concentration and arsenic metabolite metrics were quartiled based on the distribution in the controls. The percentages of MMA, DMA and inorganic arsenic were calculated after exclusion of arsenobetaine and arsenocholine from the total. In addition, two methylation indices were constructed: primary methylation index (PMI)—the ratio of MMA to inorganic arsenic—and secondary methylation index (SMI)—the ratio of DMA to MMA.

Statistical Analysis

Hardy Weinberg Equilibrium (HWE) was calculated for the controls based on Pearson (χ^2) tests and α. Violation of HWE is calculated based on statistical departure

from expected HWE frequencies based on the exact test. α is used to denote the magnitude of departure from HWE.

$$\alpha = \frac{1}{2}\log\left(\frac{4P_{gg}P_{GG}}{P_{Gg}^2}\right)$$

where P_{gg}, P_{Gg}, and P_{GG} are the genotype proportions (40).

The association between each SNP and arsenical skin lesion status was estimated by prevalence odds ratios (PORs) and their 95% confidence intervals (CIs) from logistic regression models employed by the LOGISTIC procedure in SAS. The PORs were adjusted for sex, age, and well water arsenic concentration, primarily because of the strong association of these covariates with the outcome. The genotype-specific PORs were estimated for each SNP using the homozygous major genotype as the reference category.

Two approaches were selected to evaluate SNP-SNP and higher order interactions (i.e., MDR and CART). Due to the fact that there was not a substantial amount of variability in the *LIG4* SNP genotype in this population and the α statistic suggested a deviation from HWE in the control subjects, this SNP was excluded from evaluation of interaction.

SNP-SNP interactions were first evaluated using MDR among the 10 genotyped SNPs, with no additional predictors specified. SNP-SNP interactions were then evaluated with the addition of various measures of arsenic exposure to the interaction models (i.e., well water arsenic concentration, creatinine-adjusted urinary total arsenic concentration, primary methylation index, and secondary methylation index). Each construct of arsenic exposure was modeled separately with the set of 10 SNPs. MDR is a nonparametric method that reduces data into a single dimensional variable (classifying individuals as

high and low risk groups), determined by the ratio of diseased to non-diseased individuals (28). Each model is constructed using ten-fold cross-validation, where 9/10 of the data are used to construct the models and 1/10 of the data are used to estimate the testing accuracy (i.e., the proportion of individuals the model correctly classifies). All SNPs were modeled as ordinal variables, coded as 0, 1, or 2, which represented the number of minor alleles carried. All measures of arsenic exposure were modeled as ordinal variables, coded as 0, 1, 2, or 3, which represented increasing quartile of exposure level. Individuals with data missing for any of the variables were excluded from the analysis. MDR analyses were conducted using MDR software (version 2.0 beta 8.3) (41).

CART was utilized to assess SNP-SNP interactions based on the same strategy; first, interactions were evaluated among the 10 SNPs and then various constructs of arsenic exposure were integrated into the interaction models. CART analysis is a nonparametric recursive partitioning approach that constructs a decision tree using ten-fold cross-validation (42). It is characterized by two central features in tree construction—1) recursive partitioning which splits the root node into offspring nodes and continues into further generation of nodes and 2) pruning which removes from the bottom up splits that are based on unreliably small sample sizes (42). All SNPs were modeled as ordinal variables, coded as 0, 1, or 2, which represented the number of minor alleles carried. All measures of arsenic exposure were modeled as ordinal variables, coded as 0, 1, 2, or 3, which represented increasing quartiles of exposure level. Participants with data missing for any of the variables were excluded from the analysis. Individuals were classified in the tree based on the Gini splitting rule and, the tree was

pruned to minimum cross-validated error in the terminal nodes (30). CART analyses were conducted using DTREG software (43).

Logistic regression was finally employed to assess interactions that were predicted by both MDR and CART to evaluate how well the results converged. Logistic regression was conducted using the LOGISTIC procedure in SAS. The likelihood ratio test was used to evaluate interaction on the multiplicative scale by comparing the full model containing interactions terms to the reduced model without the interaction terms.

RESULTS

Deviation from the expected genotype frequency distribution was observed for *LIG4* so we elected to eliminate this SNP from evaluation of interaction analyses (**Table 1**). Characteristics of the study populations are shown in **Table 2**. There was a higher prevalence of skin lesions among males and older individuals as well as evidence of a dose-response trend with increasing quartiles of well water arsenic concentration, creatinine-adjusted urinary total arsenic concentration, and the primary methylation index. An inverse trend was observed with increasing quartiles of the secondary methylation index.

Genotype frequencies of the SNPs as well as associations with skin lesion prevalence are shown in **Table 3**. There were no significant marginal effects observed; adjustment for sex, age, and well water arsenic concentration did not appreciably change the effect estimates (data not shown). The conditional effects of the SNPs (e.g., with the inclusion of all the SNPs in a single logistic regression model) were also evaluated and, we did not observe any significant associations with skin lesions (**Table D.2**).

Furthermore, utilizing Hierarchical Bayes modeling (44) to integrate information on the shared pathways (i.e., oxidative stress and DNA repair, **Table D.1**) of the SNPs did not yield considerably different results from the conditional model (**Table D.2**).

The results of the MDR interaction models are shown in **Table 4**, with the results for the best 1, 2, and 3 factor models indicated. Five different interaction modeling scenarios were evaluated, which included SNP-SNP and environment-SNP interactions. A different pattern of interactions was generally observed for each of the exposure modeling scenarios; however, in each case, the measure of arsenic exposure was the single most important predictor of skin lesions. Overall, the interaction model between well water arsenic concentration and the polymorphism in *MPO* provided the highest testing accuracy (0.6152) and maximum cross-validation consistency (10/10) for skin lesion prevalence. Based on these 2 factors, individuals who were classified as high risk were 2.6 (95% CI=2.0, 3.3) times more likely to have skin lesions than individuals who were classified as low risk (shown in **Figure D.1**).

The results of the CART analyses are shown in **Figures D.4-D.8** for all prediction models. Again, a different pattern of interactions was generally observed for each of the exposure modeling scenarios; however, in each model that included arsenic exposure, arsenic was the single most important predictor of skin lesions indicated by the first split in each of the graphs by the arsenic variable.

A comparison of the MDR and CART interaction models revealed consistency for the following interactions in the SNPs only model: 1) *ERCC5*XRCC1* (Ex10) and 2) *OGG1*ERCC5*XRCC1* (Ex10). In the model that included well water arsenic concentration, only the interaction between arsenic and *MPO* was observed in both

methods. The model that included primary methylation index showed consistency by both methods for: 1) PMI*_OGG1_ and 2) PMI*_SOD2*OGG1_. These interactions were subsequently evaluated by logistic regression.

Both MDR and CART predicted an interaction between the polymorphisms in *ERCC5* and *XRCC1* (Ex10), which was confirmed by logistic regression (**Table 5**). Overall, the inclusion of interaction terms between *ERCC5* and *XRCC1* (Ex10) in the model were significant (adjusted model: χ^2=13.8, 4 d.f., *P* for interaction=0.008). There was a significant antagonist interaction between the GC genotype of *ERCC5* and the GA genotype of *XRCC1* (Ex10). The data were too sparse to evaluate the three-way interaction between the polymorphisms in *OGG1*, *ERCC5*, and *XRCC1* (Ex10), as shown in **Figure D.2b**. Evaluation of the interaction between well water arsenic concentration (dichotomized into lowest two quartiles versus highest two quartiles) and the polymorphism in *MPO* did not reveal any significant interaction overall by logistic regression modeling (**Table 6**); although, there appeared to be a slight reduction in risk among individuals with the GA genotype of *MPO*. When well water arsenic concentration was included in the model by 3 indicator variables to specify quartiles of exposure, the interaction term for the third quartile of exposure and the GA genotype of *MPO* showed statistical evidence of antagonism (β_{GxE}=-1.1708, *P* value=0.0028). This antagonistic interaction was predicted by MDR, as shown in **Figure D.1**, with the individuals in the corresponding cell labeled low risk. Evaluation of the interaction between the primary methylation index (dichotomized into lowest two quartiles versus highest two quartiles) and the polymorphism in *OGG1* did not reveal any significant interaction overall by logistic regression modeling (**Table 7**); nor was there any evidence

of interaction when the primary methylation index was modeled by 3 indicator variables (data not shown). Based on the MDR model (**Figure D.3a**), it appears there was positive interaction between low PMI and the CG genotype of *OGG1*, with those individuals classified as high risk, and an antagonistic interaction predicted between high PMI and the GG genotype of *OGG1*, with those individuals classified as low risk, that was not statistically observable in the logistic regression modeling. Finally, the three-way interaction between PMI, *OGG1* and *SOD2* was evaluated by logistic regression (**Table 8**). There was evidence of a significant multiplicative SNP-SNP interaction between *OGG1* and *SOD2* (χ^2=12.4, 4 d.f., *P* for interaction=0.01), with evidence for a significant antagonistic interaction between the CG genotype of *OGG1* and the CT genotype of *SOD2* as well as a significant antagonistic interaction between the GG genotype of *OGG1* and the CT genotype of *SOD2*, also depicted in the MDR model (**Figure D.3b**). In an effort to evaluate the three-way interaction with PMI in the logistic regression models, the models were stratified by low and high PMI. While the multiplicative interaction was not statistically significant, the SNP-SNP interaction marginally appeared in the high PMI exposure group and was not apparent in the low PMI exposure group (*P* for interaction=0.06 and 0.29, respectively).

DISCUSSION

In this study, we evaluated potential gene-gene interactions of 10 SNPs genotyped in candidate genes in the oxidative stress and DNA repair pathways. We additionally evaluated gene-environment interactions through the inclusion of various constructs of arsenic exposure (well water arsenic concentration, creatinine-adjusted urinary total

arsenic concentration, primary methylation index, and secondary methylation index) with the SNPs in the predictor variable sets for MDR and CART. The approach taken was to use the MDR and CART methods to select concordant interactions for evaluation by logistic regression. Several gene-gene and gene-environment interactions were consistent between MDR and CART including: *ERCC5*XRCC1* (Ex10); *OGG1*ERCC5*XRCC1* (Ex10); well water arsenic**MPO*; PMI**OGG1*; and, PMI**SOD2*OGG1*. These selected interactions were evaluated for statistical evidence of multiplicative interaction by logistic regression, which revealed significant antagonistic interactions for *ERCC5*XRCC1* (Ex10) and PMI**SOD2*OGG1*.

The *ERCC5* gene codes for a DNA repair protein involved in the NER pathway of UV-induced damage (45). This gene is located in chromosome region 13q33. A nonsynonymous SNP (G>C, Asp1104His, rs17655) in *ERCC5* was examined in this study. While there is no current literature on the association of this gene with arsenic or skin lesions, there is evidence of a potential interaction between arsenic and UV (46, 47) in relation to skin lesions. XRCC1 is a DNA repair enzyme that interacts with polynucleotide kinase (PNK), DNA polymerase-beta (POLB) and DNA ligase III (LIG3) as part of a complex to repair single-strand breaks and functions in BER to repair damage caused by agents such as reactive oxygen species (48). This gene is located in chromosome region 19q.13.2. A nonsynonymous SNP (C>T, Arg194Trp, rs1799782) in this gene was genotyped in this study, of which the T allele has been associated with deficient DNA repair (49). In this study, there was statistically significant antagonistic interaction between the heterozygous genotypes of both *ERCC5* and *XRCC1* (Ex10). Both these genes have been previously examined in epidemiologic investigations related

to carcinogenesis (50-52); although, an interaction between these genes has not been previously reported.

In this study we also observed significant antagonistic interaction between *OGG1* and *SOD2*, particularly among individuals with higher levels of the primary methylation index. OGG1 is a DNA glycosylase involved in base excision repair (BER) of 8-hydroxydeoxyguanosine (8-OHdG), an adduct formed from oxidative stress (53). During BER, OGG1 removes damaged bases by cleaving N-glycosylic bonds (53). Prior *in vitro* (54) and animal (55, 56) studies have shown OGG1 to play a role in mitigating arsenic-induced oxidative damage. This gene is located in chromosome region 3p26.2. A nonsynonymous SNP (C>G, Ser326Cys, rs1052133) in *OGG1* was genotyped in this study. Proteins encoded with the G allele exhibit reduced 8-OHdG repair activity (57). Additionally, SOD2 protein plays a major role in maintaining oxidative balance by converting superoxide (O_2^-)—a precursor molecule for all other reactive oxygen species—into hydrogen peroxide and oxygen (58). Additionally, prior *in vitro* experiments have shown SOD2 to play a role in mitigating arsenic-induced DNA damage from oxidative stress (59, 60). This gene is located in chromosome region 6q25.3. A nonsynonymous SNP (C>T, Ala16Val, rs4880) in *SOD2* was genotyped in this study. The C allele retains the alpha helical structure of the protein for normal activity of the enzyme (61). Finally, inefficient arsenic methylation (i.e., represented by elevated concentrations of MMA to inorganic arsenic in urine) has been consistently observed as a risk factor for skin lesion prevalence in previous epidemiologic studies (62-65). A recent cross-sectional study among arsenic-exposed individuals showed that lower primary methylation index was associated with lower plasma antioxidant capacity (66). Moreover,

in vitro studies have shown that MMA induces reactive oxygen species, particularly the trivalent form of this organic arsenic species (67, 68). In a study by Bailey et al. (68), *in vitro* experiments using normal human epidermal keratinocytes showed that MMA^{III} exhibited the greatest potential for skin carcinogenesis through the induction of oxidative stress, compared to the trivalent forms of inorganic arsenic and DMA. In a comparative functional genomics analysis in yeast, Jo et al. (69) showed that glutathione (related to cellular antioxidant status) was more important in MMA^{III} toxicity than trivalent inorganic arsenic toxicity. We observed a significant antagonistic interaction between the CG genotype of *OGG1* and the CT genotype of *SOD2* as well as a significant antagonistic interaction between the GG genotype of *OGG1* and the CT genotype of *SOD2*. Both these genes have been associated with cancer outcomes (70-72), although an interaction between them has not been previously reported.

There are several strengths of this study. First, the main strength of this study was our multi-analytic approach to the assessment of gene-gene, gene-environment, and higher-order interactions. We utilized two nonparametric approaches (i.e., MDR and CART) to select promising interactions for evaluation by logistic regression. MDR and CART proved to be powerful analytic methods for the detection of statistical gene-gene and gene-environment interactions, including higher-order interactions, with a limited sample size. MDR has been previously shown to undercover gene-gene and gene-environment interactions in the absence of any significant marginal effects of the factors (73, 74). Even with the potential noise present in the predictor variable sets—with the inclusion of several SNPs unrelated to the outcome—both the MDR and CART methods consistently selected arsenic exposure as the most important predictor of skin lesions

when included in the predictor variable sets. Second, the measurement of various constructs of arsenic exposure was evaluated. Different components of arsenic have been assessed in this study population including well water arsenic concentration (a measure of external exposure), urinary total arsenic concentration (a measure of internal dose), as well as arsenic metabolites (a measure of biological response). It is possible that genes interact differently with each of the components of arsenic exposure, which we were able to systematically evaluate in this study.

There are several limitations of this study that we also consider. First, significant interactions observed in these analyses may not be causal, but simply due to chance. We did not observe any significant marginal effects of the SNPs, and while it is possible that epistatic effects could still be present, we may have simply added noise to the predictor variable sets by including these SNPs. Second, the interactions observed were all antagonistic interactions between the heterozygous genotypes. The biological basis of this type of relationship is difficult to conceive. Third, the evaluation of three-way interactions using logistic regression in this study was hampered by the small sample size to asses these higher order interactions.

In summary, our analytic method utilized MDR and CART modeling approaches for the selection of potential gene-gene and gene-environment interactions. We observed considerable overlap of the interactions detected by both these methods, which were further evaluated by logistic regression. Results from logistic regression modeling, confirmed some of these statistical interactions; however, further investigation and replication of the statistical interactions detected in this study is needed.

REFERENCES

1. Smith AH, Lingas EO, Rahman M. Contamination of drinking-water by arsenic in Bangladesh: a public health emergency. *Bulletin of the World Health Organization.* 2000;78(9):1093-1103.

2. IARC Working Group on the Evaluation of Carcinogenic Risks to Humans. Some Drinking-water Disinfectants and Contaminants, including Arsenic. Volume 84. Lyon, France, 2004.

3. Smith AH, Goycolea M, Haque R, et al. Marked increase in bladder and lung cancer mortality in a region of Northern Chile due to arsenic in drinking water. *Am J Epidemiol.* 1998;147(7):660-669.

4. Chen CJ, Chen CW, Wu MM, et al. Cancer potential in liver, lung, bladder and kidney due to ingested inorganic arsenic in drinking water. *Br J Cancer.* 1992;66(5):888-892.

5. Hopenhayn-Rich C, Biggs ML, Fuchs A, et al. Bladder cancer mortality associated with arsenic in drinking water in Argentina. *Epidemiology.* 1996;7(2):117-124.

6. Hopenhayn-Rich C, Biggs ML, Smith AH. Lung and kidney cancer mortality associated with arsenic in drinking water in Cordoba, Argentina. *Int J Epidemiol.* 1998;27(4):561-569.

7. Tseng WP. Effects and dose--response relationships of skin cancer and blackfoot disease with arsenic. *Environ Health Perspect.* 1977;19109-119.

8. Brouwer OF, Onkenhout W, Edelbroek PM, et al. Increased neurotoxicity of arsenic in methylenetetrahydrofolate reductase deficiency. *Clin Neurol Neurosurg.* 1992;94(4):307-310.

9. Wu MM, Kuo TL, Hwang YH, et al. Dose-response relation between arsenic concentration in well water and mortality from cancers and vascular diseases. *Am J Epidemiol.* 1989;130(6):1123-1132.

10. Rahman M, Tondel M, Ahmad SA, et al. Hypertension and arsenic exposure in Bangladesh. *Hypertension.* 1999;33(1):74-78.

11. Heck JE, Chen Y, Grann VR, et al. Arsenic exposure and anemia in Bangladesh: a population-based study. *Journal of occupational and environmental medicine.* 2008;50(1):80-87.

12. Wong SS, Tan KC, Goh CL. Cutaneous manifestations of chronic arsenicism: review of seventeen cases. *J Am Acad Dermatol.* 1998;38(2 Pt 1):179-185.

13. Haque R, Mazumder DN, Samanta S, et al. Arsenic in drinking water and skin lesions: dose-response data from West Bengal, India. *Epidemiology.* 2003;14(2):174-182.

14. Council National Research. Subcommittee on Asenic in Drinking Water. *Arsenic in Drinking Water*. Washington, Dc: National Academy Press, 2001.

15. Rossman TG. Mechanism of arsenic carcinogenesis: an integrated approach. *Mutat Res.* 2003;533(1-2):37-65.

16. Tokar EJ, Benbrahim-Tallaa L, Ward JM, et al. Cancer in experimental animals exposed to arsenic and arsenic compounds. *Critical reviews in toxicology.*40(10):912-927.

17. Kitchin KT. Recent advances in arsenic carcinogenesis: modes of action, animal model systems, and methylated arsenic metabolites. *Toxicol Appl Pharmacol.* 2001;172(3):249-261.

18. Ahsan H, Chen Y, Kibriya MG, et al. Susceptibility to arsenic-induced hyperkeratosis and oxidative stress genes myeloperoxidase and catalase. *Cancer Lett.* 2003;201(1):57-65.

19. Ahsan H, Chen Y, Wang Q, et al. DNA repair gene XPD and susceptibility to arsenic-induced hyperkeratosis. *Toxicol Lett.* 2003;143(2):123-131.

20. Banerjee M, Sarkar J, Das JK, et al. Polymorphism in the ERCC2 codon 751 is associated with arsenic-induced premalignant hyperkeratosis and significant chromosome aberrations. *Carcinogenesis.* 2007;28(3):672-676.

21. Breton CV, Zhou W, Kile ML, et al. Susceptibility to arsenic-induced skin lesions from polymorphisms in base excision repair genes. *Carcinogenesis.* 2007;28(7):1520-1525.

22. Kundu M, Ghosh P, Mitra S, et al. Precancerous and non-cancer disease endpoints of chronic arsenic exposure: the level of chromosomal damage and XRCC3 T241M polymorphism. *Mutat Res.* 2011;706(1-2):7-12.

23. McCarty KM, Smith TJ, Zhou W, et al. Polymorphisms in XPD (Asp312Asn and Lys751Gln) genes, sunburn and arsenic-related skin lesions. *Carcinogenesis.* 2007;28(8):1697-1702.

24. Moore JH, Williams SM. New strategies for identifying gene-gene interactions in hypertension. *Annals of medicine.* 2002;34(2):88-95.

25. Ritchie MD, Hahn LW, Moore JH. Power of multifactor dimensionality reduction for detecting gene-gene interactions in the presence of genotyping error, missing data, phenocopy, and genetic heterogeneity. *Genetic epidemiology.* 2003;24(2):150-157.

26. Ritchie MD, Motsinger AA. Multifactor dimensionality reduction for detecting gene-gene and gene-environment interactions in pharmacogenomics studies. *Pharmacogenomics.* 2005;6(8):823-834.

27. Motsinger AA, Ritchie MD. Multifactor dimensionality reduction: an analysis strategy for modelling and detecting gene-gene interactions in human genetics and pharmacogenomics studies. *Human genomics.* 2006;2(5):318-328.

28. Moore JH. Computational analysis of gene-gene interactions using multifactor dimensionality reduction. *Expert review of molecular diagnostics.* 2004;4(6):795-803.

29. Ritchie MD. Bioinformatics approaches for detecting gene-gene and gene-environment interactions in studies of human disease. *Neurosurgical focus.* 2005;19(4):E2.

30. Breiman L, Friedman JH, Olshen RA, et al. Classification and Regression Trees. Belmont: Wadsworth, 1984.

31. Breiman L. Random Forests. *Machine Learning.* 2001;45(1):5-32.

32. Garcia-Magarinos M, Lopez-de-Ullibarri I, Cao R, et al. Evaluating the ability of tree-based methods and logistic regression for the detection of SNP-SNP interaction. *Annals of human genetics.* 2009;73(Pt 3):360-369.

33. Schwender H, Zucknick M, Ickstadt K, et al. A pilot study on the application of statistical classification procedures to molecular epidemiological data. *Toxicol Lett.* 2004;151(1):291-299.

34. Ahsan H, Chen Y, Parvez F, et al. Health Effects of Arsenic Longitudinal Study (HEALS): description of a multidisciplinary epidemiologic investigation. *J Expo Sci Environ Epidemiol.* 2006;16(2):191-205.

35. Ahsan H, Chen Y, Parvez F, et al. Arsenic exposure from drinking water and risk of premalignant skin lesions in Bangladesh: baseline results from the Health Effects of Arsenic Longitudinal Study. *Am J Epidemiol.* 2006;163(12):1138-1148.

36. Cheng Z, Zheng Y, Mortlock R, et al. Rapid multi-element analysis of groundwater by high-resolution inductively coupled plasma mass spectrometry. *Anal Bioanal Chem.* 2004;379(3):512-518.

37. Nixon DE, Mussmann GV, Eckdahl SJ, et al. Total arsenic in urine: palladium-persulfate vs nickel as a matrix modifier for graphite furnace atomic absorption spectrophotometry. *Clin Chem.* 1991;37(9):1575-1579.

38. Nermell B, Lindberg AL, Rahman M, et al. Urinary arsenic concentration adjustment factors and malnutrition. *Environmental research.* 2008;106(2):212-218.

39. Reuter W, Davidowski L, Neubauer K. Speciation of five arsenic compounds in urine by HPLC/ICP-MS. Perkin Elmer Application Notes: Norwalk, CT: Perkin-Elmer, 2003.

40. Minelli C, Thompson JR, Abrams KR, et al. How should we use information about HWE in the meta-analyses of genetic association studies? *Int J Epidemiol.* 2008;37(1):136-146.

41. Moore J, Andrews P. MDR, 2011. (www.multifactordimensionalityreduction.org).

42. Zhang H, Bonney G. Use of classification trees for association studies. *Genetic epidemiology.* 2000;19(4):323-332.

43. Sherrod P. DTREG, 2011. (www.dtreg.com).

44. Witte JS. Genetic analysis with hierarchical models. *Genetic epidemiology.* 1997;14(6):1137-1142.

45. Thompson LH, Carrano AV, Sato K, et al. Identification of nucleotide-excision-repair genes on human chromosomes 2 and 13 by functional complementation in hamster-human hybrids. *Somatic cell and molecular genetics.* 1987;13(5):539-551.

46. Rossman TG. Enhancement of UV-mutagenesis by low concentrations of arsenite in E. coli. *Mutat Res.* 1981;91(3):207-211.

47. Rossman TG, Uddin AN, Burns FJ. Evidence that arsenite acts as a cocarcinogen in skin cancer. *Toxicol Appl Pharmacol.* 2004;198(3):394-404.

48. Whitehouse CJ, Taylor RM, Thistlethwaite A, et al. XRCC1 stimulates human polynucleotide kinase activity at damaged DNA termini and accelerates DNA single-strand break repair. *Cell.* 2001;104(1):107-117.

49. 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(4):604-608.

50. Moreno V, Gemignani F, Landi S, et al. Polymorphisms in genes of nucleotide and base excision repair: risk and prognosis of colorectal cancer. *Clin Cancer Res.* 2006;12(7 Pt 1):2101-2108.

51. Ming-Shiean H, Yu JC, Wang HW, et al. Synergistic effects of polymorphisms in DNA repair genes and endogenous estrogen exposure on female breast cancer risk. *Annals of surgical oncology.* 2011;17(3):760-771.

52. Rajaraman P, Hutchinson A, Wichner S, et al. DNA repair gene polymorphisms and risk of adult meningioma, glioma, and acoustic neuroma. *Neuro Oncol.* 2010;12(1):37-48.

53. Lu AL, Li X, Gu Y, et al. Repair of oxidative DNA damage: mechanisms and functions. *Cell Biochem Biophys.* 2001;35(2):141-170.

54. Pu YS, Jan KY, Wang TC, et al. 8-Oxoguanine DNA glycosylase and MutY homolog are involved in the incision of arsenite-induced DNA adducts. *Toxicol Sci.* 2007;95(2):376-382.

55. Wanibuchi H, Salim EI, Kinoshita A, et al. Understanding arsenic carcinogenicity by the use of animal models. *Toxicol Appl Pharmacol.* 2004;198(3):366-376.

56. Kinoshita A, Wanibuchi H, Morimura K, et al. Carcinogenicity of dimethylarsinic acid in Ogg1-deficient mice. *Cancer science.* 2007;98(6):803-814.

57. 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(25):3219-3225.

58. Skrzycki M, Majewska M, Podsiad M, et al. Expression and activity of superoxide dismutase isoenzymes in colorectal cancer. *Acta biochimica Polonica.* 2009;56(4):663-670.

59. Nordenson I, Beckman L. Is the genotoxic effect of arsenic mediated by oxygen free radicals? *Hum Hered.* 1991;41(1):71-73.

60. Wang TS, Hsu TY, Chung CH, et al. Arsenite induces oxidative DNA adducts and DNA-protein cross-links in mammalian cells. *Free Radic Biol Med.* 2001;31(3):321-330.

61. Shimoda-Matsubayashi S, Matsumine H, Kobayashi T, et al. Structural dimorphism in the mitochondrial targeting sequence in the human manganese superoxide dismutase gene. A predictive evidence for conformational change to influence mitochondrial transport and a study of allelic association in Parkinson's disease. *Biochem Biophys Res Commun.* 1996;226(2):561-565.

62. Del Razo LM, Garcia-Vargas GG, Vargas H, et al. Altered profile of urinary arsenic metabolites in adults with chronic arsenicism. A pilot study. *Archives of toxicology.* 1997;71(4):211-217.

63. Valenzuela OL, Borja-Aburto VH, Garcia-Vargas GG, et al. Urinary trivalent methylated arsenic species in a population chronically exposed to inorganic arsenic. *Environ Health Perspect.* 2005;113(3):250-254.

64. Yu RC, Hsu KH, Chen CJ, et al. Arsenic methylation capacity and skin cancer. *Cancer Epidemiol Biomarkers Prev.* 2000;9(11):1259-1262.

65. Ahsan H, Chen Y, Kibriya MG, et al. Arsenic metabolism, genetic susceptibility, and risk of premalignant skin lesions in Bangladesh. *Cancer Epidemiol Biomarkers Prev.* 2007;16(6):1270-1278.

66. Wu MM, Chiou HY, Wang TW, et al. Association of blood arsenic levels with increased reactive oxidants and decreased antioxidant capacity in a human population of northeastern Taiwan. *Environ Health Perspect.* 2001;109(10):1011-1017.

67. Nesnow S, Roop BC, Lambert G, et al. DNA damage induced by methylated trivalent arsenicals is mediated by reactive oxygen species. *Chem Res Toxicol.* 2002;15(12):1627-1634.

68. Bailey KA, Hester SD, Knapp GW, et al. Gene expression of normal human epidermal keratinocytes modulated by trivalent arsenicals. *Molecular carcinogenesis.* 2010;49(12):981-998.

69. Jo WJ, Loguinov A, Wintz H, et al. Comparative functional genomic analysis identifies distinct and overlapping sets of genes required for resistance to monomethylarsonous acid (MMAIII) and arsenite (AsIII) in yeast. *Toxicol Sci.* 2009;111(2):424-436.

70. Wang SS, Davis S, Cerhan JR, et al. Polymorphisms in oxidative stress genes and risk for non-Hodgkin lymphoma. *Carcinogenesis.* 2006;27(9):1828-1834.

71. Lan Q, Mumford JL, Shen M, et al. Oxidative damage-related genes AKR1C3 and OGG1 modulate risks for lung cancer due to exposure to PAH-rich coal combustion emissions. *Carcinogenesis.* 2004;25(11):2177-2181.

72. Yi JF, Li YM, Liu T, et al. Mn-SOD and CuZn-SOD polymorphisms and interactions with risk factors in gastric cancer. *World J Gastroenterol.*16(37):4738-4746.

73. Chen M, Kamat AM, Huang M, et al. High-order interactions among genetic polymorphisms in nucleotide excision repair pathway genes and smoking in modulating bladder cancer risk. *Carcinogenesis.* 2007;28(10):2160-2165.

74. Duell EJ, Bracci PM, Moore JH, et al. Detecting pathway-based gene-gene and gene-environment interactions in pancreatic cancer. *Cancer Epidemiol Biomarkers Prev.* 2008;17(6):1470-1479.

TABLE 1. Single Nucleotide Polymorphisms

Gene	SNP rs number	Chromosome	SNP	Amino acid substitution	Genotyping success rate (%)	α
Antioxidant enzymes genes						
SOD2	4880	6	Ex2+24C>T	A16V	96.5	0.03
CAT	1001179	11	-329T>C		96.3	0.01
GPX1	1050450	3	Ex1-226C>T	P200L	96.4	-0.03
MPO	2333227	17	-642G>A		96.8	-0.03
DNA repair genes						
OGG1	1052133	3	Ex6-315C>G	S326C	97.7	-0.01
ERCC2	1052559	19	Ex23+61A>C	K751Q	97.9	0.01
ERCC5	17655	13	Ex15-344G>C	D1104H	96.8	-0.05
LIG4	1805388	13	Ex3+54C>T	T9I	96.6	-0.18
XRCC1	1799782	19	Ex6-22C>T	R194W	95.2	0.07
XRCC1	25487	19	Ex10-4A>G	Q399R	96.3	-0.01
XRCC3	861539	14	Ex8-53C>T	T241M	95.0	0.05

TABLE 2. Selected Characteristics of the Study Sample, Araihazar, Bangladesh, 2000-2002

Characteristic	Cases (n=610)		Controls (n=1,079)		POR[132] (95% CI)
	N	%	N	%	
Sex					
Male	507	82.6	440	40.8	1.0
Female	106	17.4	639	59.2	0.1 (0.1, 0.1)
Age, years					
18-30	161	26.4	386	35.8	1.0
31-40	224	36.7	237	22.0	2.5 (1.8, 3.4)
41-50	60	9.8	355	32.9	5.6 (4.0, 7.8)
51-75	165	27.0	100	9.3	9.8 (6.7, 14.1)
Well water arsenic, μg/L[133]					
0.1-10	72	11.8	265	24.6	1.0
10.1-50	97	15.9	248	23.0	1.4 (1.0, 2.0)
50.1-150	202	33.1	330	30.6	2.2 (1.6, 3.1)
150.1-864	239	39.2	236	21.9	3.7 (2.7, 5.1)
Urinary total arsenic, μg/g[134]					
16-103	64	10.5	269	24.9	1.0
104-192	121	19.8	271	25.1	1.9 (1.3, 2.6)
193-339	164	26.9	269	24.9	2.6 (1.8, 3.6)
340-8556	261	42.8	270	25.1	4.1 (2.9, 5.6)
Primary methylation index[135]					
<0.580	70	11.8	260	25.0	1.0
0.580-0.835	132	22.3	259	25.0	1.9 (1.3, 2.6)
0.836-1.154	171	28.9	260	25.0	2.4 (1.8, 3.4)
1.155-19.570	218	36.9	260	25.0	3.1 (2.3, 4.3)
Secondary methylation index[136]					
<4.316	236	39.9	259	25.0	1.0
4.316-5.855	153	25.9	259	25.0	0.6 (0.5, 0.8)
5.856-8.211	128	21.7	259	25.0	0.5 (0.4, 0.7)
8.212-32.300	74	12.5	259	25.0	0.3 (0.2, 0.4)

[132] POR, Prevalence odds ratio; unadjusted.
[133] Water arsenic concentration cut-points roughly correspond to quartiles of the exposure distribution for the controls but have been slightly adjust to reflect policy relevant parameters.
[134] Creatinine-adjusted urinary total arsenic concentration cut-points are quartiles of the exposure distribution in control subjects.
[135] The primary methylation index was calculated as % MMA / % Inorganic arsenic.
[136] The secondary methylation index was calculated as % DMA / % MMA.

TABLE 3. Prevalence Odds Ratios and 95% CIs for Skin Lesion Prevalence in Relation to SNPs

SNP	MAF[137]	Cases		Controls		POR[138]
		N	%	N	%	(95% CI)
Antioxidant enzymes genes						
SOD2	T=0.45					
CC		173	30.6	302	31.5	1.0
CT		268	47.3	456	47.5	1.0 (0.8, 1.3)
TT		125	22.1	202	21.0	1.1 (0.8, 1.4)
GPX1	T=0.19					
CC		372	65.4	623	65.2	1.0
CT		174	30.6	297	31.1	1.0 (0.8, 1.2)
TT		23	4.0	36	3.8	1.1 (0.6, 1.8)
CAT	T=0.23					
CC		358	63.5	571	59.5	1.0
CT		185	32.8	343	35.8	0.9 (0.7, 1.1)
TT		21	3.7	45	4.7	0.7 (0.4, 1.3)
MPO	A=0.14					
GG		423	74.3	709	73.6	1.0
GA		134	23.6	237	24.6	0.9 (0.7, 1.2)
AA		12	2.1	17	1.8	1.2 (0.6, 2.5)
DNA repair genes						
OGG1	G=0.31					
CC		264	46.0	460	47.2	1.0
CG		259	45.1	423	43.4	1.1 (0.9, 1.3)
GG		51	8.9	92	9.4	1.0 (0.7, 1.4)
ERCC2	C=0.33					
AA		242	43.4	421	44.5	1.0
AC		262	46.9	419	44.3	1.1 (0.9, 1.4)
CC		54	9.7	105	11.1	0.9 (0.6, 1.3)
ERCC5	C=0.27					
GG		298	52.8	503	52.0	1.0
GC		223	39.5	400	41.4	0.9 (0.8, 1.2)
CC		43	7.6	64	6.6	1.1 (0.8, 1.7)
LIG4	T=0.09					
CC		473	82.8	788	82.3	1.0
CT		92	16.1	165	17.2	0.9 (0.7, 8.9)
TT		6	1.0	4	0.4	2.5 (0.7, 8.9)
XRCC1 (Ex6)	T=0.10					
CC		445	79.5	763	80.6	1.0
CT		104	18.6	170	18.0	1.0 (0.8, 1.4)
TT		11	2.0	13	1.4	1.4 (0.6, 3.3)
XRCC1 (Ex10)	A=0.33					
GG		240	42.5	423	44.1	1.0
GA		251	44.4	429	44.8	1.0 (0.8, 1.3)
AA		74	13.1	106	11.1	1.2 (0.9, 1.7)
XRCC3	T=0.20					
CC		352	63.4	619	65.3	1.0
CT		180	32.4	287	30.3	1.1 (0.9, 1.4)
TT		23	4.1	42	4.4	1.0 (0.6, 1.6)

[137] Minor allele frequency in the controls.
[138] Unadjusted.

TABLE 4. MDR Interaction Models for Single Nucleotide Polymorphisms and Arsenic Exposure in Relation to Skin Lesion Prevalence

No. of factors	Model	Testing accuracy	Cross-validation consistency	POR (95% CI)[139]
SNPs only				
1	CAT	0.4916	4/10	1.2 (0.9, 1.5)
2	ERCC5, XRCC1 (Ex10)	0.5172	8/10	1.5 (1.2, 1.9)
3	OGG1, ERCC5, XRCC1 (Ex10)	0.4921	6/10	1.8 (1.4, 2.3)
Well water arsenic and SNPs				
1	Water arsenic	0.5957	10/10	2.3 (1.8, 2.9)
2	Water arsenic, MPO	0.6152	10/10	2.6 (2.0, 3.3)
3	Water arsenic, MPO, ERCC2	0.5704	3/10	2.8 (2.2, 3.5)
Creatinine-adjusted urinary total arsenic and SNPs				
1	Urinary arsenic	0.6011	10/10	2.3 (1.8, 3.0)
2	Urinary arsenic, XRCC1 (Ex6)	0.6077	10/10	2.5 (1.9, 3.2)
3	Urinary arsenic, ERCC5, XRCC1 (Ex10)	0.5616	3/10	2.7 (2.1, 3.5)
Primary methylation index and SNPs				
1	PMI	0.5650	10/10	1.8 (1.4, 2.3)
2	PMI, OGG1	0.5606	6/10	2.3 (1.8, 3.0)
3	PMI, SOD2, OGG1	0.5461	5/10	2.6 (2.0, 3.3)
Secondary methylation index and SNPs				
1	SMI	0.5630	10/10	2.0 (1.5, 2.5)
2	SMI, MPO	0.5758	5/10	2.3 (1.8, 2.9)
3	SMI, CAT, ERCC2	0.5813	8/10	2.8 (2.2, 3.6)

[139] High risk versus low risk groups.

TABLE 5. Logistic Regression Evaluation of *ERCC5*XRCC1* (Ex10) Interaction Detected by Both MDR and CART for Skin Lesion Prevalence

SNPs		**Cases N (%)**	**Controls N (%)**	**Crude POR (95% CI)**	**Adjusted POR[140] (95% CI)**
ERCC5	***XRCC1* (Ex10)**				
GG	GG	139 (25.4)	197 (21.1)	1.0	1.0
GG	GA	115 (21.0)	231 (24.8)	0.7 (0.5, 0.9)	0.7 (0.5, 0.9)
GG	AA	38 (6.9)	54 (5.8)	1.0 (0.6, 1.6)	1.1 (0.6, 1.9)
GC	GG	77 (14.0)	186 (19.9)	0.6 (0.4, 0.8)	0.6 (0.4, 0.8)
GC	GA	110 (20.1)	157 (16.8)	1.0 (0.7, 1.4)	1.1 (0.7, 1.5)
GC	AA	28 (5.1)	46 (4.9)	0.9 (0.5, 1.5)	0.9 (0.5, 1.6)
CC	GG	16 (2.9)	28 (3.0)	0.8 (0.4, 1.5)	0.9 (0.4, 1.8)
CC	GA	21 (3.8)	31 (3.3)	1.0 (0.5, 1.7)	0.8 (0.4, 1.5)
CC	AA	4 (0.7)	3 (0.3)	1.9 (0.4, 8.6)	2.1 (0.4, 12.2)
	P for interaction			0.008[141]	0.008

[140] Adjusted for well water arsenic concentration, sex and age.
[141] χ^2 test for likelihood ratio test for interaction terms, P value.

TABLE 6. Logistic Regression Evaluation of Well Water Arsenic**MPO* Interaction Detected by Both MDR and CART for Skin Lesion Prevalence

Factors		**Cases N (%)**	**Controls N (%)**	**Crude POR (95% CI)**	**Adjusted POR[142] (95% CI)**
Well water arsenic	***MPO***				
0.1-50	GG	112 (19.7)	335 (34.8)	1.0	1.0
0.1-50	GA	43 (7.6)	110 (11.4)	1.2 (0.8, 1.8)	1.1 (0.7, 1.8)
0.1-50	AA	6 (1.1)	10 (1)	1.8 (0.6, 5.0)	1.9 (0.6, 6.3)
50.1-864	GG	311 (54.7)	374 (38.8)	2.5 (1.9, 3.2)	2.7 (2.0, 3.6)
50.1-864	GA	91 (16)	127 (13.2)	2.1 (1.5, 3.0)	2.3 (1.6, 3.4)
50.1-864	AA	6 (1.1)	7 (0.7)	2.6 (0.8, 7.8)	3.0 (0.8, 10.6)
	P for interaction			0.19[143]	0.29

[142] Adjusted for sex and age.
[143] χ^2 test for likelihood ratio test for interaction terms, P value.

TABLE 7. Logistic Regression Evaluation of Primary Methylation Index*_OGG1_ Interaction Detected by Both MDR and CART for Skin Lesion Prevalence

Factors		**Cases N (%)**	**Controls N (%)**	**Crude POR (95% CI)**	**Adjusted POR[144] (95% CI)**
Primary methylation index	***OGG1***				
<0.580-0.835	CC	78 (14)	216 (23.1)	1.0	1.0
<0.580-0.835	CG	95 (17.1)	197 (21)	1.3 (0.9, 1.9)	1.3 (0.9, 2.0)
<0.580-0.835	GG	19 (3.4)	46 (4.9)	1.1 (0.6, 2.1)	1.1 (0.6, 2.1)
0.836-19.570	CC	176 (31.7)	228 (24.3)	2.1 (1.5, 3.0)	1.4 (0.9, 2.0)
0.836-19.570	CG	157 (28.2)	210 (22.4)	2.1 (1.5, 2.9)	1.2 (0.9, 1.8)
0.836-19.570	GG	31 (5.6)	40(4.3)	2.1 (1.3, 3.7)	1.1 (0.6, 2.0)
P for interaction				0.38[145]	0.36

[144] Adjusted for sex and age.
[145] χ^2 test for likelihood ratio test for interaction terms, P value.

TABLE 8. Logistic Regression Evaluation of Primary Methylation Index*_OGG1*SOD2_ Interaction Detected by Both MDR and CART

		Cases N (%)	Controls N (%)	Overall POR (95% CI)	Low PMI	High PMI
OGG1	**SOD2**					
CC	CC	65 (11.7)	150 (16.0)	1.0	1.0	1.0
CG	CC	88 (15.9)	120 (12.8)	1.7 (1.1, 2.5)	2.1 (1.1, 4.1)	1.6 (0.9, 2.7)
GG	CC	18 (3.2)	27 (2.9)	1.5 (0.8, 3.0)	1.5 (0.5, 4.2)	2.2 (0.9, 5.8)
CC	CT	129 (23.2)	188 (20.0)	1.6 (1.1, 2.3)	1.5 (0.8, 2.8)	1.7 (1.1, 2.8)
CG	CT	115 (20.6)	206 (21.9)	1.3 (0.9, 1.9)	1.6 (0.9, 2.9)	1.2 (0.7, 1.9)
GG	CT	18 (3.2)	48 (5.1)	0.9 (0.5, 1.6)	0.8 (0.3, 2.4)	1.0 (0.4, 2.2)
CC	TT	61 (10.9)	103 (11.0)	1.4 (0.9, 2.1)	1.1 (0.5, 2.3)	1.6 (0.9, 2.9)
CG	TT	50 (9.0)	84 (8.9)	1.4 (0.9, 2.2)	1.4 (0.6, 3.0)	1.5 (0.8, 2.7)
GG	TT	13 (2.3)	14 (1.5)	2.1 (0.9, 4.8)	2.7 (0.7, 11.2)	2.1 (0.7, 6.2)
	P for interaction			0.01[146]	0.29	0.06

[146] χ^2 test for Likelihood ratio test for interaction terms, P value.

Chapter 5

Conclusions

OBJECTIVES

The aims of this dissertation were to 1) evaluate the effect of single nucleotide polymorphisms (SNPs) in antioxidant enzymes genes in relation to skin lesion prevalence, 2) to evaluate the effect of SNPs in DNA repair genes in relation to skin lesion prevalence, 3) to evaluate SNP-environment interactions between SNPs on these pathways and environmental factors (i.e., arsenic exposure measures and lifestyle factors) using logistic regression to assess interaction on the additive scale, and 4) to evaluate gene-gene, gene-environment, and higher-order interactions of these genetic and environmental factors using a multi-analytic approach including nonparametric (i.e, MDR and CART) as well as parametric (i.e, logistic regression) methods.

SUMMARY OF RESULTS

In Chapter 2, we assessed the relationship between SNPs in antioxidant enzyme genes and skin lesion prevalence, as well as possible interactions of these associations on the additive scale by various environmental factors. There were no statistically significant associations between these SNPs (*SOD2,* rs4880; *CAT,* rs1001179; *GPX1,* rs1050450; and *MPO*, rs2333227) and skin lesion prevalence. The results did not vary by arsenic exposure levels (as measured by well water arsenic concentration, urinary total arsenic concentration, primary methylation index, and secondary methylation index), body mass index, smoking status, or fruit and vegetable intake. However, there was marginal evidence that skin lesion prevalence was increased among individuals who carried 4 or more risk alleles compared to individuals carrying 0-3 risk alleles in these SNPs.

Additionally, we observed a significant departure from additivity for the risk allele score and primary methylation index on skin lesion prevalence. There was no evidence that these SNPs were associated with skin lesion severity, such as pigmentation changes and keratosis.

In Chapter 3, we assessed the relationship between SNPs in DNA repair genes (*OGG1,* rs1052133; *XRCC1,* rs25487 and rs1799782; *XRCC3,* rs861539; *ERCC2,* rs1052559; *ERCC5,* rs17655; and *LIG4,* rs1805388) and skin lesion prevalence, as well as possible interactions of these associations on the additive scale by various environmental factors. In logistic regression models controlling for sex, age, and well water arsenic concentration, no associations were observed between measured SNPs and skin lesion prevalence. The results did not vary by arsenic exposure levels (as measured by well water arsenic concentration, urinary total arsenic concentration, primary methylation index, and secondary methylation index), body mass index, or smoking status. However, we did observe a significant inverse association of total fruit and vegetable consumption with skin lesion prevalence, and its additive interaction with the polymorphism in *ERCC5*.

In Chapter 4, we utilized a multi-analytic approach to explore gene-gene, gene-environment, and higher-order interactions among 10 single nucleotide polymorphisms (SNPs) related to the oxidative stress and DNA repair pathways by MDR, CART, and logistic regression models. As shown in Chapters 2 and 3, none of these SNPs were associated with skin lesion prevalence, however, were evaluated for potential epistatic effects. Our analytic method utilized MDR and CART modeling approaches for the

selection of potential gene-gene and gene-environment interactions. Considerable overlap of the interactions detected by both these methods was observed, which were further evaluated by logistic regression. Results from logistic regression modeling, confirmed some of these statistical interactions; however, their biological interpretation is limited. Therefore, further investigation and replication of the statistical interactions detected in this study is needed.

STRENGTHS AND LIMITATIONS

To our knowledge, this is the first study to assess polymorphisms in a number of oxidative stress and DNA repair genes in relation to skin lesion prevalence, which have not been previously examined but have been implicated in arsenic toxicity and serve as good candidate genes for evaluation of genetic susceptibility. There are several strengths and weakness of the study that we consider.

Arsenic Exposure Constructs

A strength of the analyses conducted in this dissertation was the availability of various constructs of arsenic exposure assessment. Arsenic exposure was ascertained based on well water arsenic concentration, as well as urinary arsenic measures including total arsenic concentration and arsenic metabolites. Different components of arsenic were captured by these various measurements including a measure of external exposure by well water arsenic concentration, a measure of internal dose by urinary total arsenic concentration, as well as measures of biological response by the urinary arsenic

metabolites. While the arsenic exposure assessments were reasonably correlated within this study, there was still some variability across measures in this population and, it is possible that genes may interact differently with each of the components of arsenic exposure, which we were able to systematically evaluate this in this study.

With respect to the measurement of arsenic metabolites, due to laboratory limitations, total MMA and DMA species (i.e., trivalent and pentavalent species) were assessed jointly. However, there is evidence in the literature to show that the trivalent organic arsenic species may be more toxic (1, 2) and the estimation of the distribution of the trivalent metabolites would be useful in future studies.

Additionally, in the context of oxidative stress, arsenic exposure served as a proxy measure to the actual measure of interest, which is arsenic-induced oxidative stress. For the purposes of this dissertation, the assumption was made that individuals with a higher concentration of arsenic exposure experienced higher levels of oxidative stress. Alternatively, biomarkers of oxidative stress could have been measured.

Skin Lesion Assessment

For the purposes of this study, prevalent skin lesion cases were selected as having the presence of any type of arsenical skin lesion (i.e., melanosis, leucomelanosis, or keratosis). In exploratory subset analyses, skin lesion severity was evaluated by stratifying arsenical skin lesion status into absence of keratotic skin lesions and presence of keratotic skin lesions. While we did not see any significantly different effects by skin lesion severity, the sample size to detect any small effects that would be expected for

single SNPs was actually rather quite small. Future studies should be conducted with precise case definitions and be well-powered to examine such differences.

The assessment of the effect of SNPs on prevalent skin lesions is also a limitation for the evaluation of mechanistic hypotheses and phenotype considerations since the association may be related to disease prevalence or progression and not disease incidence. Therefore, findings from these analyses should be replicated with incident cases. However, we deem the effect of disease prevalence to be minimal since arsenical skin lesions are typically not fatal.

Genetic Susceptibility Assessment

The advantage of using genetic variants is a strength of this type of research because they can be measured with little measurement error and bias. Additionally, SNPs are time-independent measures, which strengthen causal inference of the associations evaluated in this dissertation between the measured SNPs and skin lesion prevalence. However, a limitation in causal inference is that assumptions regarding the association of the gene with skin lesion prevalence cannot be derived from the null findings of the single SNPs in this study, since they may not be correlated with a causal variant in the gene and do not fully tag the variability in the gene. Therefore, our inferences are limited to the measured SNPs themselves.

Conversely, the amount of oxidative stress or DNA breaks could have been measured but there would have been major limitations in the temporality of these measures with the use of prevalent cases. Additionally, in very early or mild cases of skin

lesions (some forms of melanosis) may not be readily visible with clinical examination; therefore, oxidative stress or DNA repair biomarkers could also suffer from reverse causation if it was not ensured that the measures were taken at an appropriate period before disease onset. The antioxidant defense system is comprised of both enzymatic and non-enzymatic antioxidants. The focus of this dissertation was to examine the influence of polymorphisms of selected enzymatic antioxidants. The role of non-enzymatic antioxidants and activity of enzymatic antioxidants were unmeasured, therefore the total antioxidant potential of the cell remains unknown.

The effect of any single SNP on skin lesion prevalence is likely to be small; therefore, consideration of the cumulative effect of SNPs through the construction of the risk allele score was a strength in the analytic method of this dissertation.

Finally, the SNPs and genes for which there is data available were selected based on a candidate-gene approach. Therefore, a limited number of SNPs and genes were examined in this dissertation research. Future studies could take a comprehensive genomewide approach for evaluating associations between tens of thousands of SNPs simultaneously (using a microarray platform) and skin lesion status. Findings from this type of "discovery" approach could then be evaluated in candidate-gene studies employing SNPs which comprehensively tag the genes of interest.

Analytic Methods for Gene-Gene and Gene-Environment Interactions

The evaluation of SNP-SNP interactions in this study had several complications. We evaluated 10 SNPs in 9 genes that are known key players in the oxidative stress and

DNA repair pathways. However, the genes themselves have yet to be implicated in skin lesion prevalence although the pathways have been hypothesized to have a role in arsenic toxicity with supporting evidence from *in vitro* and human studies. Therefore, it is not clear whether these SNPs could possibly have epistatic effects in relation to skin lesion prevalence or are noise SNPs. Based on the evidence from this study, we conclude that the genotyped SNPs had no independent marginal or epistatic effects with respect to skin lesion prevalence. As discussed previously, we cannot derive conclusions based on the SNPs as to whether gene-gene interactions are present between genes included in this study since the SNPs measured do not sufficiently represent the variability in the genes that may be related to susceptibility, either through linkage with a single causal SNP or tagging the gene in manner to represent the full variability of the coding variation.

A statistical issue with the evaluation of gene-gene and gene-environment interactions is that while the nonparametric methods (i.e., MDR and CART) were quite powerful and efficient in evaluating two- and three-way interactions, the potential interactions these methods yielded were statistical interactions and their interpretation had limited biological relevance, based on the fact that significant interaction effects were primarily observed between heterozygous genotypes. Moore and Williams have discussed the challenge in interpreting statistical epistasis as biological epistasis (3), which has also been a theme in the epidemiologic literature with respect to interactions (4).

However, a major strength of this work was the use of the multi-analytic approach to assess possible interactions, with final evaluation in logistic regression modeling. The

evaluation of interactions in logistic regression, however, was on the multiplicative scale. The additive scale may also be of interest for the evaluation of gene-gene interactions; although, assessment of interactions on this scale currently has limited implementation due to lack of flexibility in the modeling of parameters for additive interaction estimation and is limited to evaluation of two-way interactions. Additionally, logistic regression is suited to evaluate only lower-order interactions. With higher-order interactions the models become too complex and the sample size becomes an issue for their evaluation (5, 6).

Statistical Conclusion Validity

Multiple comparisons were a possible threat to statistical validity in this study. In Chapters 2 and 3, we conducted several tests of SNP-environment interactions. Since we had specific pre-specified hypotheses regarding these interactions, we did not take measures to adjust the *P* values for multiple comparisons. However, we acknowledge that significant findings in this study could have arisen by chance and require replication by other studies.

Study Sample

Individuals in this study were selected from the Health Effects of Arsenic Longitudinal Study (HEALS). This sampling scheme was efficient because it provided comprehensive exposure, covariate, and outcome data as well as biological samples for all study participants. However, participants in HEALS were recruited from several

villages in Araihazar, a rural 25 km^2 area in Bangladesh. The customary cultural practice among individuals in the region is subsequent to marriage for the wife to move to the spouse's village; therefore, men remain in their home village and there is a patrilineal family structure within the villages. Therefore, it is possible that participants selected for this study may have had some degree of relatedness which would have introduced confounding into this study similar to population stratification (7, 8). Since the pedigree information for these families was not collected, the number and extent to which participants were related is unknown. However, evaluation of the number of individuals who shared a tubewell (a proxy for family unit) among the selected cases and controls shows that tubewells were utilized by a range of 1 to 6 individuals, with 1.3 users on average per tubewell. Thus, we do not deem relatedness to be a major issue in this study.

PUBLIC HEALTH RELEVANCE

While there is substantial evidence of dose-response associations between arsenic in drinking water and various health outcomes, studies have shown that remediation of arsenic exposure alone does not reduce arsenic-related health risks in the population (9-11). In a recent publication using prospective data from the HEALS cohort, we showed utilizing repeated measures of urinary total arsenic exposure over time that once chronically exposed decreasing exposure for a short amount of time did not reduce one's risk of mortality (9). Additionally, other studies have shown that mortality attributed to cancers and heart disease did not begin to decline until approximately 2 decades after preventing exposure to high-arsenic well water (10, 11). Therefore, evidence from these

prior studies and our data suggest that it may be important to consider other health prevention and promotion strategies in conjunction with remediation for arsenic-exposed populations. Evaluation of genetic determinants of skin lesion prevalence could contribute to a greater understanding of the genetic and molecular pathways that underlie arsenic toxicity and may inform future interventions of arsenic-exposed populations. In addition to elucidating biological mechanisms of action, investigating genetic susceptibility may help identify individuals with higher risk to arsenic-related toxicity, aiding in prevention and intervention of arsenic-exposed populations.

In summary, there was marginal evidence that skin lesion prevalence was increased among individuals who carried 4 or more risk alleles in genotyped SNPs related to the oxidative stress pathway compared to individuals carrying 0-2 risk alleles in these SNPs and, we observed a significant departure from additivity for the risk allele score and primary methylation index on skin lesion prevalence. Additionally, we observed a significant inverse association of total fruit and vegetable consumption with skin lesion prevalence, and an additive interaction with the polymorphism in *ERCC5* in the DNA repair pathway and fruit and vegetable intake in relation to skin lesion prevalence. However, these finding require replication in other studies.

REFERENCES

1. Nesnow S, Roop BC, Lambert G, et al. DNA damage induced by methylated trivalent arsenicals is mediated by reactive oxygen species. *Chem Res Toxicol.* 2002;15(12):1627-1634.

2. Bailey KA, Hester SD, Knapp GW, et al. Gene expression of normal human epidermal keratinocytes modulated by trivalent arsenicals. *Molecular carcinogenesis.* 2010;49(12):981-998.

3. Moore JH, Williams SM. Traversing the conceptual divide between biological and statistical epistasis: systems biology and a more modern synthesis. *Bioessays.* 2005;27(6):637-646.

4. Clayton D, McKeigue PM. Epidemiological methods for studying genes and environmental factors in complex diseases. *Lancet.* 2001;358(9290):1356-1360.

5. Peduzzi P, Concato J, Kemper E, et al. A simulation study of the number of events per variable in logistic regression analysis. *Journal of clinical epidemiology.* 1996;49(12):1373-1379.

6. Concato J, Feinstein AR, Holford TR. The risk of determining risk with multivariable models. *Annals of internal medicine.* 1993;118(3):201-210.

7. Sillanpaa MJ. Overview of techniques to account for confounding due to population stratification and cryptic relatedness in genomic data association analyses. *Heredity.* 2010.

8. Zhang F, Deng HW. Correcting for cryptic relatedness in population-based association studies of continuous traits. *Hum Hered.* 2010;69(1):28-33.

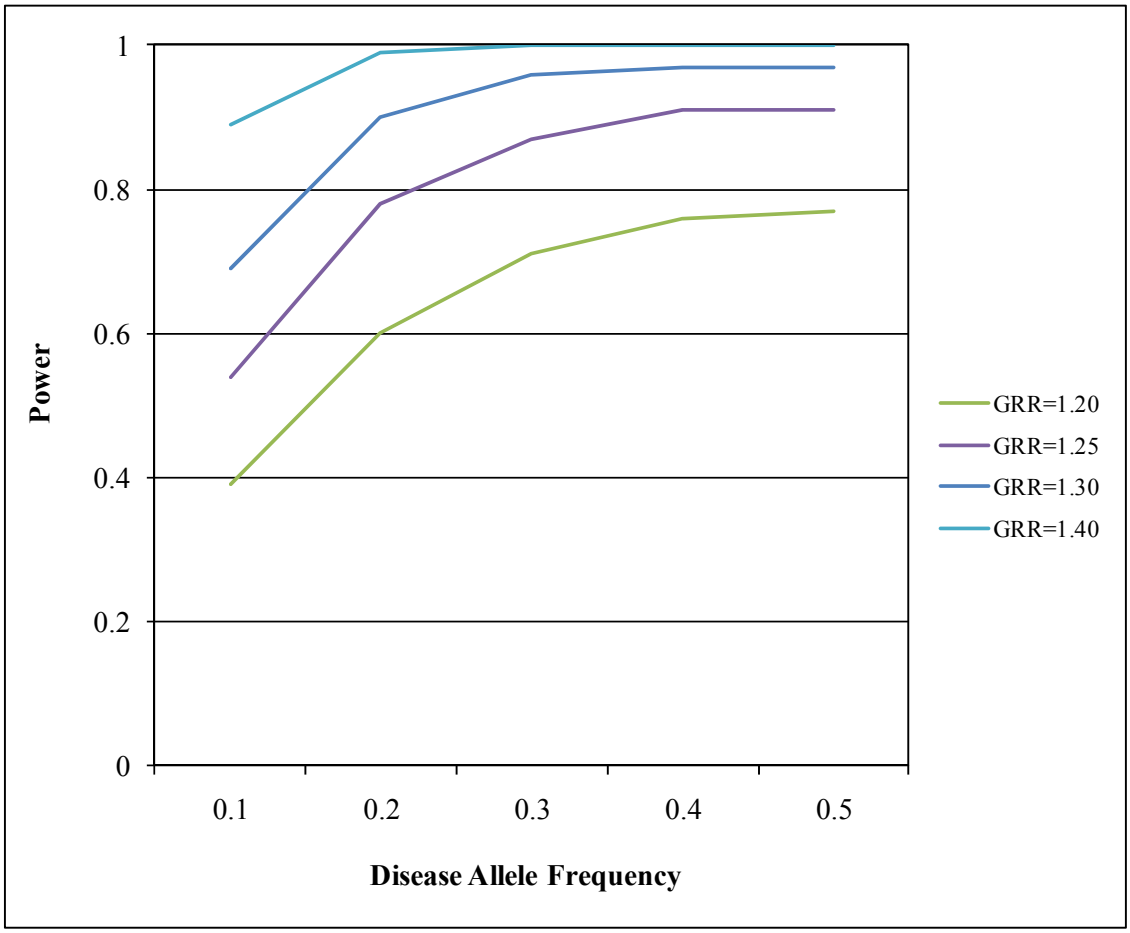
9. Argos M, Kalra T, Rathouz PJ, et al. Arsenic exposure from drinking water, and all-cause and chronic-disease mortalities in Bangladesh (HEALS): a prospective cohort study. *Lancet.* 2010;376(9737):252-258.

10. Chang CC, Ho SC, Tsai SS, et al. Ischemic heart disease mortality reduction in an arseniasis-endemic area in southwestern Taiwan after a switch in the tap-water supply system. *J Toxicol Environ Health A.* 2004;67(17):1353-1361.

11. Yang CY, Chiu HF, Chang CC, et al. Bladder cancer mortality reduction after installation of a tap-water supply system in an arsenious-endemic area in southwestern Taiwan. *Environmental research.* 2005;98(1):127-132.

APPENDIX A

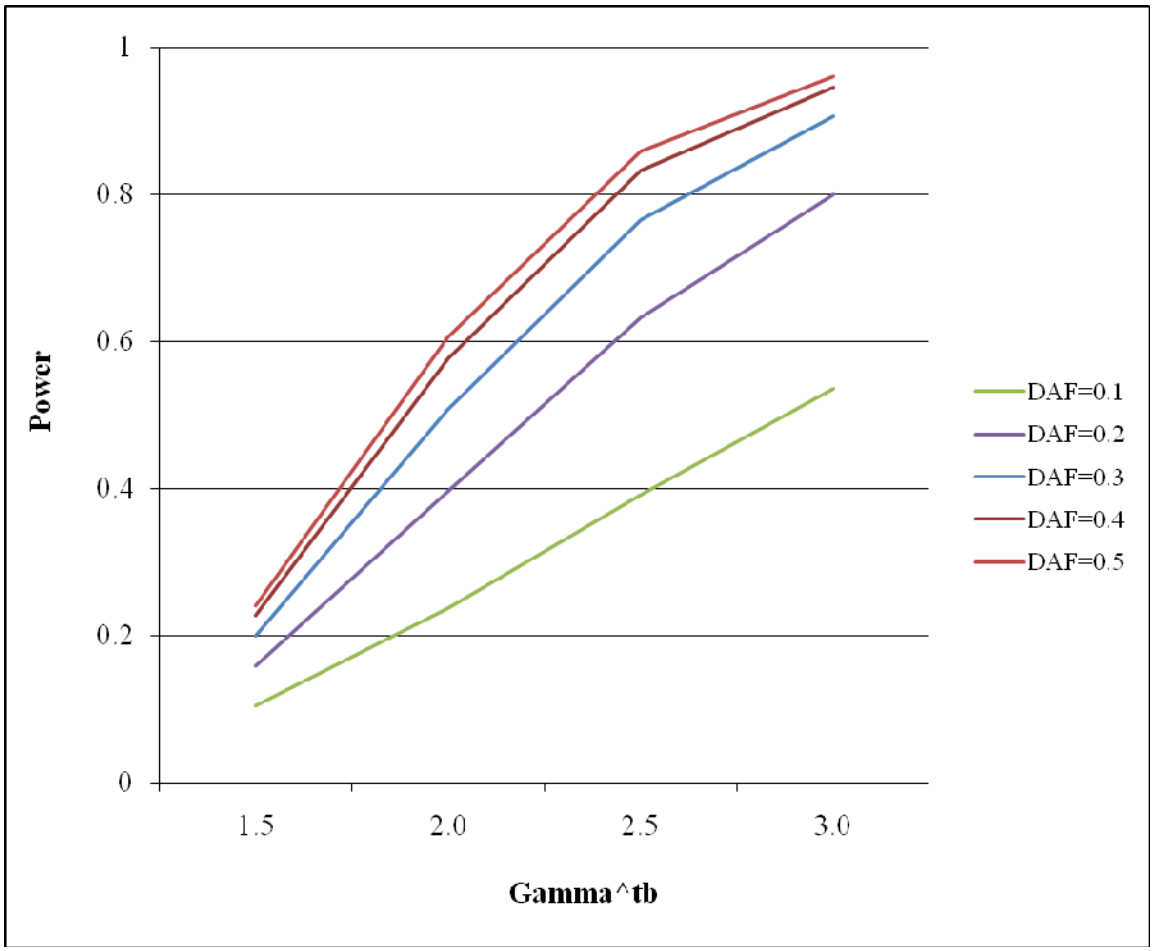
FIGURE A.1. Power calculations for case-control association of SNPs[147]



GRR=Genotype Relative Risk

[147] Power estimated based on 610 cases, 1,079 controls, α=0.05, and prevalence of disease=0.0624.

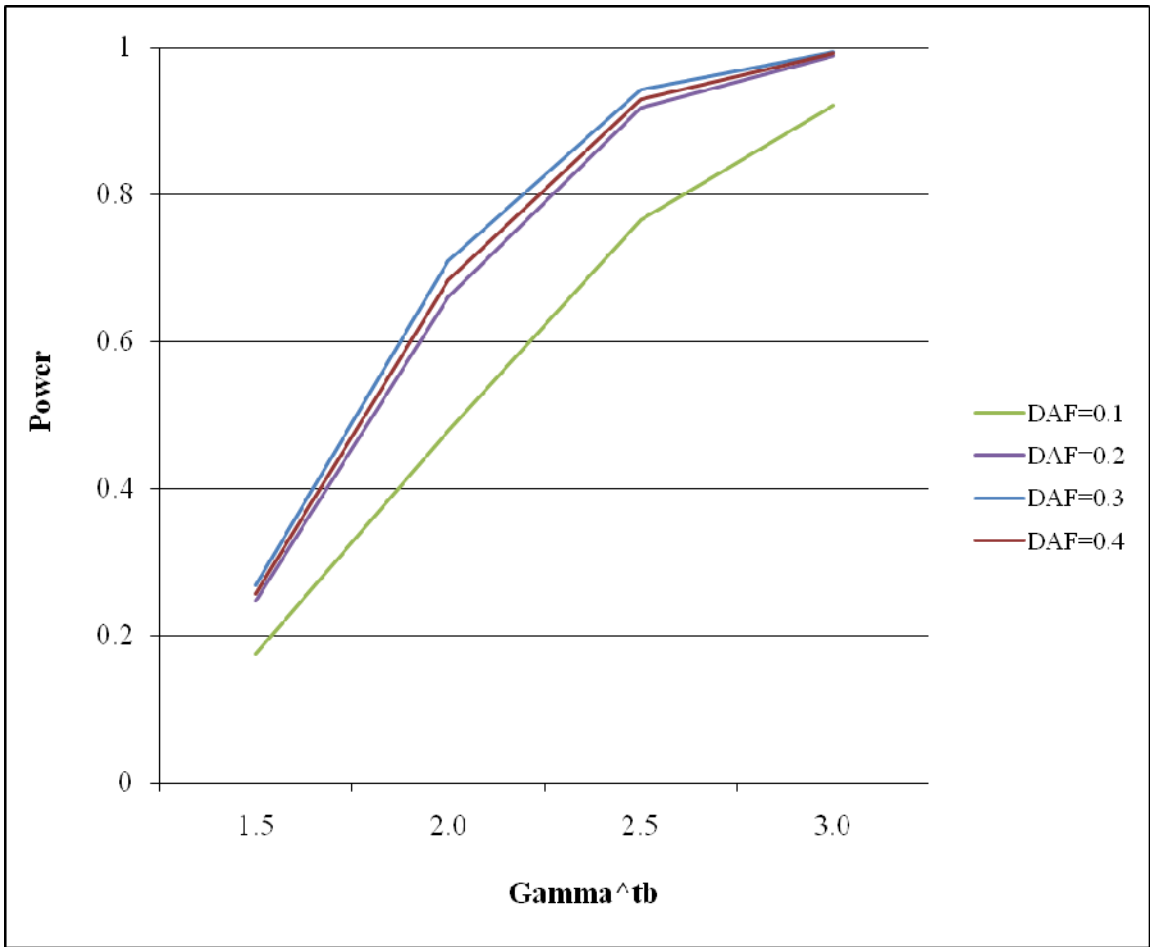
FIGURE A.2. Power calculations for additive interaction of ordinal SNPs and effect modifiers[148]



DAF=Disease Allele Frequency

[148] Power estimated based on 610 cases, 1,079 controls, α=0.05, prevalence of disease=0.0624, OR_G=1.25, OR_E=3.50.

FIGURE A.3. Power calculations for additive interaction of dichotomous SNPs and effect modifiers[149]



DAF=Disease Allele Frequency

[149] Power estimated based on 610 cases, 1,079 controls, α=0.05, prevalence of disease=0.0624, OR_G=1.25, OR_E=1.60.

APPENDIX B

TABLE B.1. Prevalence Odds Ratios and 95% CIs for Skin Lesion Prevalence in Relation to Tertile of Food Item Intake, Araihazar, Bangladesh, 2000-2002

Food Item	Food Item Tertile of Intake					*P* for trend
	Q1	Q2		Q3		
	POR	POR[150]	95% CI	POR	95% CI	
Lentils	1.0	0.7	0.5, 0.9	0.3	0.3, 0.6	0.0001
Watermelon	1.0	0.2	0.07, 0.9	0.4	0.3, 0.6	0.0001
Mango	1.0	0.5	0.4, 0.7	0.8	0.6, 1.1	0.1572
Banana	1.0	0.5	0.4, 0.7	0.6	0.5, 0.9	0.0030
Jackfruit	1.0	0.8	0.6, 1.1	1.1	0.9, 1.5	0.3272
Guava	1.0	0.3	0.2, 0.5	0.4	0.3, 0.6	0.0001
Potato	1.0	0.6	0.4, 0.8	1.0	0.7, 1.3	0.8624
Spinach	1.0	0.7	0.5, 0.9	0.6	0.5, 0.9	0.0025
Bottle gourd	1.0	0.7	0.5, 0.9	0.8	0.6, 1.1	0.0868
Pumpkin	1.0	0.5	0.4, 0.7	--	--	0.0001
Beans	1.0	0.6	0.4, 0.8	1.0	0.8, 1.4	0.7550
Eggplant	1.0	1.0	0.7, 1.4	0.5	0.4, 0.6	0.0001
Okra	1.0	0.7	0.5, 0.9	0.6	0.5, 0.8	0.0013
Spinach stalks	1.0	0.6	0.5, 0.8	0.4	0.3, 0.5	0.0001
Bitter gourd	1.0	1.1	0.8, 1.4	1.0	0.8, 1.3	0.9300
Green papaya	1.0	0.4	0.3, 0.6	--	--	0.0001
Ridge gourd	1.0	1.2	0.8, 1.8	0.8	0.6, 1.1	0.1613
Snake gourd	1.0	0.8	0.6, 1.0	0.6	0.5, 0.8	0.0008
Yam	1.0	1.0	0.7, 1.3	0.6	0.4, 0.8	0.0005
Cauliflower	1.0	0.6	0.4, 0.8	0.5	0.4, 0.7	0.0001
Cabbage	1.0	0.4	0.3, 0.5	0.3	0.2, 0.4	0.0001
Tomato	1.0	0.9	0.7, 1.3	1.3	1.0, 1.8	0.0505
Parwar (squash)	1.0	0.7	0.5, 0.9	0.4	0.3, 0.6	0.0001
Ghosala (squash)	1.0	0.6	0.4, 0.9	0.7	0.6, 0.9	0.0157
Radish	1.0	0.8	0.6, 1.1	0.6	0.4, 0.8	0.0009
Sweet potato	1.0	0.6	0.5, 0.8	0.7	0.6, 0.9	0.0202

[150] Adjusted for sex, age, well water arsenic concentration, total caloric intake, and BMI.

TABLE B.2. Prevalence Odds Ratios and 95% CIs for Skin Lesion Prevalence in Relation to Antioxidant Enzyme SNPs by Skin Lesion Severity, Araihazar, Bangladesh, 2000-2002

SNP	Controls (n=1,079)		Non-keratotic lesions (n=389)				Keratotic lesions (n=221)			
	N	%	N	%	POR[151]	95% CI	N	%	POR[23]	95% CI
SOD2										
CC	302	31.5	111	31.0	1.0		62	29.8	1.0	
CT	456	47.5	162	45.3	0.9	0.7, 1.3	106	51.0	1.1	0.8, 1.6
TT	202	21.0	85	23.7	1.1	0.8, 1.6	40	19.2	0.9	0.6, 1.5
Per allele					1.0	0.9, 1.3			1.0	0.8, 1.2
GPX1										
CC	623	65.2	239	66.6	1.0		133	63.3	1.0	
CT	297	31.1	106	29.5	0.9	0.7, 1.2	68	32.4	1.1	0.7, 1.5
TT	36	3.7	14	3.9	1.1	0.5, 2.3	9	4.3	1.4	0.6, 3.2
Per allele					1.0	0.8, 1.2			1.1	0.8, 1.5
CAT										
CC	571	59.5	242	67.4	1.0		116	56.6	1.0	
CT	343	35.8	107	29.8	0.8	0.6, 1.1	78	38.0	1.3	0.9, 1.9
TT	45	4.7	10	2.8	0.6	0.3, 1.3	11	5.4	1.6	0.7, 3.4
Per allele					0.8	0.6, 1.1			1.3	1.0, 1.7
MPO										
GG	709	73.6	280	77.1	1.0		143	69.4	1.0	
GA	237	24.6	76	20.9	0.8	0.6, 1.2	58	28.2	1.2	0.8, 1.8
AA	17	1.8	7	2.0	1.1	0.4, 3.0	5	2.4	1.7	0.5, 5.1
Per allele					0.9	0.7, 1.2			1.2	0.9, 1.7
Risk allele count										
0–3	559	64.4	194	59.9	1.0		115	59.6	1.0	
4–7	309	35.6	130	40.1	1.1	0.9, 1.5	78	40.4	1.2	0.8, 1.6

[151] Adjusted for sex, age, and well water arsenic concentration.

TABLE B.3. Prevalence Odds Ratios and 95% CIs for Skin Lesion Prevalence in Relation to Antioxidant Enzyme SNPs Among Male Study Participants, Araihazar, Bangladesh, 2000-2002

SNP	Cases (n=507)		Controls (n=440)		Crude Estimate		*P* for trend	Multivariate Estimate[152]		*P* for trend
	N	%	N	%	POR	95% CI		POR	95% CI	
SOD2										
CC	141	30.2	125	31.3	1.0			1.0		
CT	225	48.2	180	45.0	1.1	0.8, 1.5		1.1	0.8, 1.5	
TT	101	21.6	95	23.7	0.9	0.6, 1.4		1.0	0.7, 1.4	
Per allele					1.0	0.8, 1.2	0.83	1.0	0.8, 1.2	0.95
GPX1										
CC	307	65.5	255	65.5	1.0			1.0		
CT	146	31.1	122	31.4	1.0	0.7, 1.3		1.0	0.7, 1.4	
TT	16	3.4	12	3.1	1.1	0.5, 2.4		1.0	0.4, 2.3	
Per allele					1.0	0.8, 1.3	0.91	1.0	0.8, 1.3	0.97
CAT										
CC	298	63.7	240	61.2	1.0			1.0		
CT	154	32.9	133	33.9	0.9	0.7, 1.2		1.0	0.7, 1.4	
TT	16	3.4	19	4.9	0.7	0.3, 1.3		0.7	0.4, 1.6	
Per allele					0.9	0.7, 1.1	0.32	0.9	0.7, 1.2	0.70
MPO										
GG	349	74.0	293	73.6	1.0			1.0		
GA+AA	123	26.0	105	26.4	1.0	0.7, 1.3	0.91	1.0	0.7, 1.4	0.94
Risk allele count										
0–3	262	60.8	223	62.3	1.0			1.0		
4–7	169	39.2	135	37.7	1.1	0.8, 1.4	0.67	1.0	0.8, 1.4	0.74

[152] Adjusted for age and well water arsenic concentration.

APPENDIX C

TABLE C.1. Prevalence Odds Ratios and 95% CIs for Skin Lesion Prevalence in Relation to DNA Repair SNPs by Skin Lesion Severity, Araihazar, Bangladesh, 2000-2002

SNP	Controls (n=1,079)		Non-keratotic lesions (n=389)				Keratotic lesions (n=221)			
	N	%	N	%	POR [153]	95% CI	N	%	POR [25]	95% CI
OGG1										
CC	460	47.2	164	44.8	1.0		100	48.1	1.0	
CG	423	43.4	163	44.5	1.0	0.7, 1.3	96	46.1	0.9	0.6, 1.3
GG	92	9.4	39	10.7	1.0	0.6, 1.6	12	5.8	0.5	0.2, 0.9
Per allele					1.0	0.8, 1.2			0.8	0.6, 1.0
ERCC2										
AA	421	44.5	156	44.6	1.0		86	41.4	1.0	
AC	419	44.3	160	45.7	1.1	0.8, 1.4	102	49.0	1.2	0.8, 1.7
CC	105	11.1	34	9.7	0.8	0.5, 1.4	20	9.6	0.9	0.5, 1.6
Per allele					1.0	0.8, 1.2			1.0	0.8, 1.3
ERCC5										
GG	503	52.0	194	54.2	1.0		104	50.5	1.0	
GC	400	41.4	144	40.2	0.9	0.7, 1.2	79	38.3	1.0	0.7, 1.4
CC	64	6.6	20	5.6	0.8	0.4, 1.4	23	11.2	1.7	0.9, 3.1
Per allele					0.9	0.7, 1.1			1.2	0.9, 1.5
LIG4										
CC	788	82.3	307	84.3	1.0		166	80.2	1.0	
CT/TT	169	17.7	57	15.7	0.3	0.1, 1.4	41	19.8	0.8	0.1, 8.8
XRCC1 (Ex6)										
CC	763	80.6	278	78.3	1.0		167	81.5	1.0	
CT	170	18.0	69	19.4	1.2	0.9, 1.7	35	17.1	1.0	0.7, 1.6
TT	13	1.4	8	2.3	1.7	0.6, 5.0	3	1.4	1.1	0.3, 4.5
Per allele					1.3	0.9, 1.7			1.0	0.7, 1.5
XRCC1 (Ex10)										
GG	423	44.1	154	43.0	1.0		86	41.6	1.0	
GA	429	44.8	152	42.5	0.9	0.7, 1.3	99	47.8	1.1	0.8, 1.6
AA	106	11.1	52	14.5	1.4	0.9, 2.1	22	10.6	1.1	0.6, 1.9
Per allele					1.1	0.9, 1.4			1.1	0.8, 1.4
XRCC3										
CC	619	65.3	222	62.9	1.0		130	64.4	1.0	
CT	287	30.3	114	32.3	1.1	0.8, 1.5	66	32.7	1.1	0.8, 1.6
TT	42	4.4	17	4.8	0.9	0.4, 1.6	6	2.9	0.5	0.2, 1.3
Per allele					1.0	0.8, 1.3			0.9	0.7, 1.2
Risk allele count										
0–3	516	62.6	180	60.2	1.0		107	58.8	1.0	
4–8	308	37.4	119	39.8	1.1	0.8, 1.5	75	41.2	1.2	0.8, 1.7

[153] Adjusted for sex, age, and well water arsenic concentration.

TABLE C.2. Prevalence Odds Ratios and 95% CIs for Skin Lesion Prevalence in Relation to DNA Repair SNPs Among Male Study Participants, Araihazar, Bangladesh, 2000-2002

SNP	Cases (n=504)		Controls (n=440)		Crude Estimate		*P* for trend	Multivariate Estimate[154]		*P* for trend
	N	%	N	%	POR	95% CI		POR	95% CI	
OGG1										
CC	215	45.4	181	44.8	1.0			1.0		
CG	215	45.4	180	45.6	1.0	0.8, 1.3		0.9	0.7, 1.2	
GG	44	9.2	43	10.6	0.9	0.5, 1.4		0.8	0.5, 1.3	
Per allele					1.0	0.8, 1.2	0.66	0.9	0.7, 1.1	0.30
ERCC2										
AA	206	44.3	175	45.4	1.0			1.0		
AC	213	45.8	164	42.6	1.1	0.8, 1.5		1.1	0.8, 1.4	
CC	46	9.9	46	12.0	0.8	0.5, 1.3		0.9	0.5, 1.4	
Per allele					0.9	0.8, 1.2	0.84	1.0	0.8, 1.2	0.77
ERCC5										
GG	237	51.3	217	54.4	1.0			1.0		
GC	188	40.7	153	38.3	1.1	0.8, 1.5		1.1	0.8, 1.5	
CC	37	8.0	29	7.3	1.2	0.7, 2.0		1.2	0.7, 2.0	
Per allele					1.1	0.9, 1.4	0.38	1.1	0.9, 1.4	0.70
LIG4										
CC	384	81.7	343	86.4	1.0			1.0		
CT	81	17.2	51	12.8	1.4	1.0, 2.1		1.4	0.9, 2.1	
TT	5	1.1	3	0.8	1.5	0.3, 6.3		1.7	0.4, 7.7	
Per allele					1.4	1.0, 1.9	0.07	1.4	1.0, 2.0	0.07
XRCC1 (Ex6)										
CC	374	80.6	312	81.5	1.0			1.0		
CT	80	17.2	68	17.7	1.0	0.7, 1.4		1.0	0.7, 1.4	
TT	10	2.2	3	0.8	2.8	0.8, 10.2		2.6	0.6, 10.9	
Per allele					1.1	0.8, 1.5	0.46	1.1	0.8, 1.5	0.58
XRCC1 (Ex10)										
GG	197	42.2	165	42.0	1.0			1.0		
GA	212	45.4	183	46.6	1.0	0.7, 1.3		1.0	0.7, 1.3	
AA	58	12.4	45	11.4	1.1	0.7, 1.7		1.1	0.7, 1.8	
Per allele					1.0	0.8, 1.2	0.87	1.0	0.8, 1.3	0.70
XRCC3										
CC	292	63.5	254	65.0	1.0			1.0		
CT	150	32.6	113	28.9	1.1	0.9, 1.5		1.1	0.8, 1.5	
TT	18	3.9	24	6.1	0.6	0.3, 1.2		0.6	0.3, 1.1	
Per allele					1.0	0.8, 1.2	0.85	0.9	0.7, 1.2	0.62
Risk allele count										
0–3	241	60.2	207	61.1	1.0			1.0		
4–8	159	39.8	132	38.9	1.0	0.8, 1.4		1.0	0.7, 1.4	

[154] Adjusted for age and well water arsenic concentration.

TABLE C.3. Prevalence Odds Ratios[155] and 95% CIs for Keratosis in Relation to *ERCC2* by Creatinine-adjusted Urinary Total Arsenic Concentration, Araihazar, Bangladesh, 2000-2002

SNP	Creatinine-adjusted urinary total arsenic concentration (µg/g)			
	16-103	**104-192**	**193-339**	**340-8556**
ERCC2				
AA	1.00	1.97 (0.78, 4.97)	3.57 (1.46, 8.73)	5.41 (2.26, 12.96)
AC/CC	1.00	2.93 (1.22, 7.04)	2.58 (1.03, 6.33)	6.31 (2.73, 14.62)

[155] Adjusted for sex and age.

APPENDIX D

Hierarchial Bayes modeling. In supplementary analyses, we attempted to exploit the information of the pathway information of the SNPs utilizing Hierarchial Bayes modeling. The marginal effects of the SNPs are summarized in **Table D.2** based on dichotomous SNP classifications, as indicated in the table with the heterozygous and homozygous genotypes containing the risk allele combined. A logistic regression model was run for each SNP individually to yield the marginal effects. For simplicity, unadjusted estimates are presented for all analyses, since adjustment did not appreciably change the PORs. We then summarized the conditional effect of the SNPs by running a single logistic regression model with all the SNPs simultaneously (**Table D.2**). Finally, we specified a two-stage model to conduct Hierarchical Bayes modeling (44). In the first-stage we ran the conditional logistic regression model to generate the log odds estimates and the covariance matrix between the SNPs. Based on existing knowledge of the gene functions, we specified the design matrix (**Table D.1**) for the pathways for use in the second-stage regression model. A value of 1 was assigned to SNPs that were involved in a particularly pathway, otherwise 0 was assigned. In the second-stage model, we regressed the prior information of the pathways on the log odds of each SNP, specifying a residual variance (τ^2) of 0.25. In general, we found Hierchical modeling yielded similar results to the conditional approach (**Table D.2**) and did not produce any significant associations between SNPs and skin lesion prevalence.

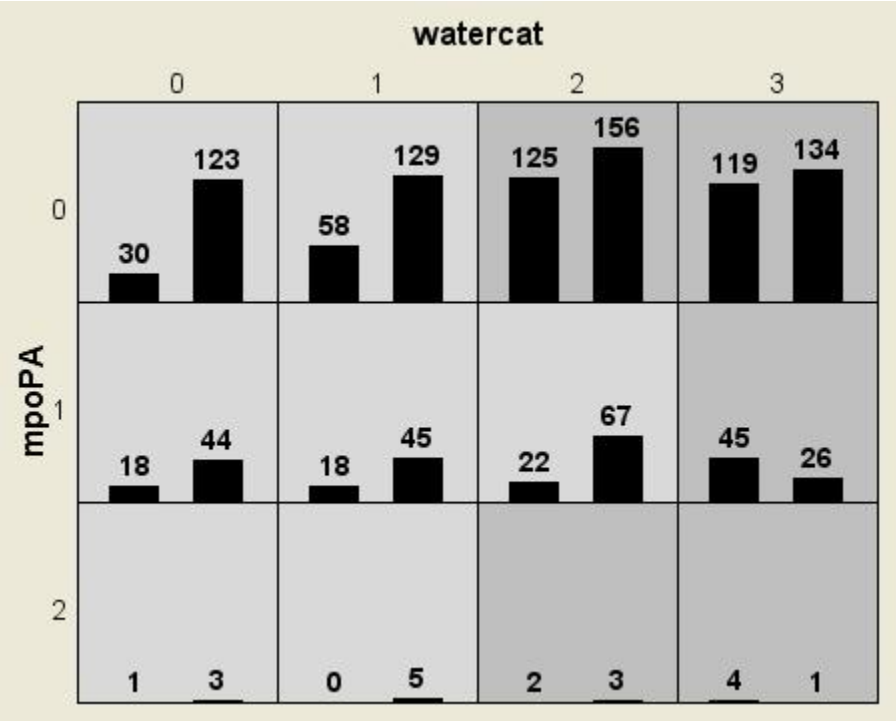
TABLE D.1. Second-stage Design Matrix for Multilevel Model

SNP	Pathway	
	Antioxidant	**DNA Repair**
SOD2	1	0
GPX1	1	0
CAT	1	0
MPO	1	0
OGG1	1	1
ERCC2	0	1
ERCC5	0	1
LIG4	0	1
XRCC1 (Ex6)	0	1
XRCC1 (Ex10)	0	1
XRCC3	0	1

TABLE D.2. Hierarchical Bayes Regression Estimates and 95% CIs for the Study Sample, Araihazar, Bangladesh, 2000-2002

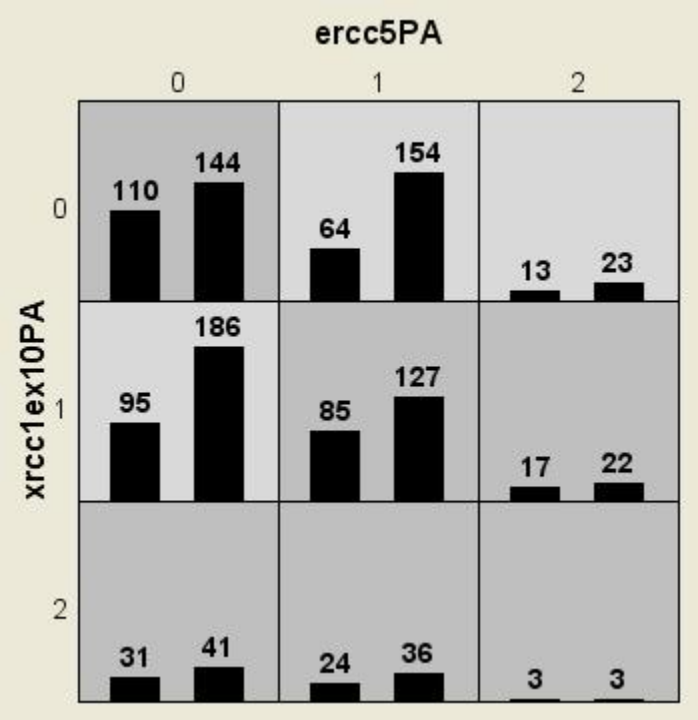
SNP	Marginal Model		Conditional Model		Hierarchical Model	
	POR	95% CI	POR	95% CI	POR	95% CI
SOD2 - CT/TT	1.04	0.83, 1.30	1.06	0.82, 1.38	1.06	0.83, 1.37
GPX1 - CT/TT	0.99	0.80, 1.23	1.07	0.83, 1.37	1.06	0.83, 1.36
CAT - TC/CC	1.27	0.75, 2.16	1.44	0.77, 2.67	1.32	0.77, 2.27
MPO - GA/AA	0.96	0.76, 1.22	0.91	0.69, 1.20	0.92	0.70, 1.20
OGG1 - CG/GG	1.05	0.85, 1.29	1.06	0.83, 1.34	1.06	0.84, 1.34
ERCC2 - AC/CC	1.05	0.85, 1.29	1.02	0.80, 1.30	1.02	0.81, 1.29
ERCC5 - GC/CC	0.97	0.79, 1.19	0.91	0.72, 1.16	0.92	0.72, 1.16
LIG4 - TC/CC	0.40	0.11, 1.41	1.26	0.23, 6.92	1.06	0.43, 2.61
XRCC1 (Ex6) - CT/TT	1.08	0.83, 1.40	1.02	0.75, 1.38	1.02	0.76, 1.36
XRCC1 (Ex10) - GA/AA	1.07	0.87, 1.32	1.07	0.84, 1.37	1.07	0.84, 1.36
XRCC3 - CT/TT	1.09	0.87, 1.35	0.94	0.73, 1.20	0.94	0.74, 1.20

FIGURE D.1. Graphical Presentation of MDR Interaction Model for Well Water Arsenic Concentration and *MPO* for Skin Lesion Prevalence



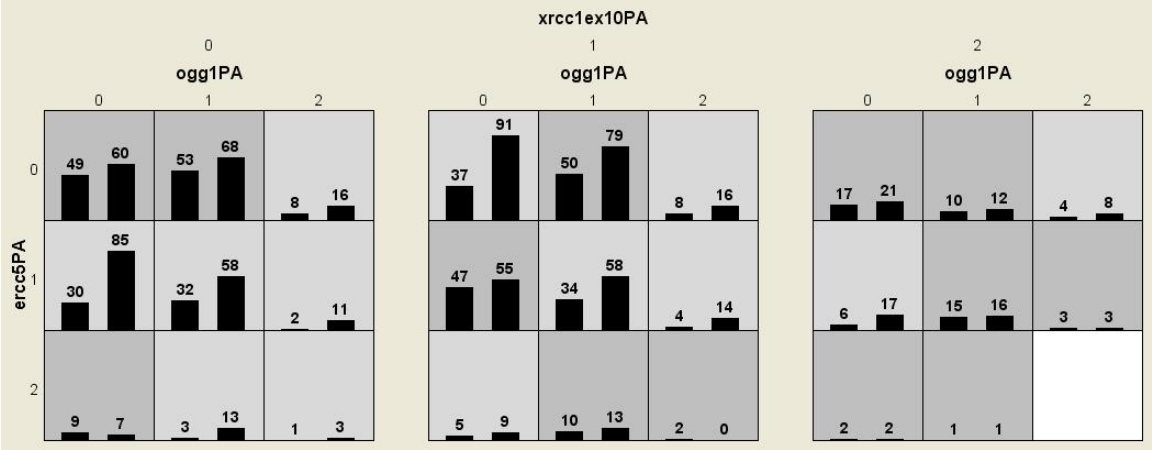
Overall best model was the two factor model, which classified individuals based on well water arsenic concentration and *MPO*. The low risk group is indicated in light grey and the high risk group is indicated in dark grey.

FIGURE D.2a. Graphical Presentation of MDR Interaction Model for SNPs



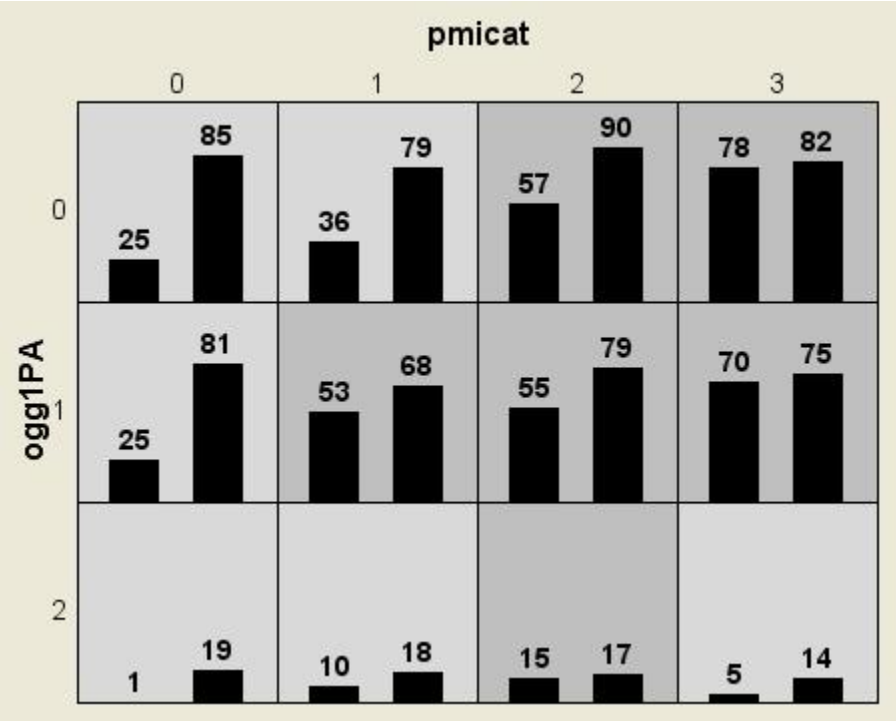
a. Two factor model yielded *XRCC1* (Ex10) (rs25487) and *ERCC5* (rs17655) as the best model. The low risk group is indicated in light grey and the high risk group is indicated in dark grey.

FIGURE D.2b. Graphical Presentation of MDR Interaction Model for SNPs



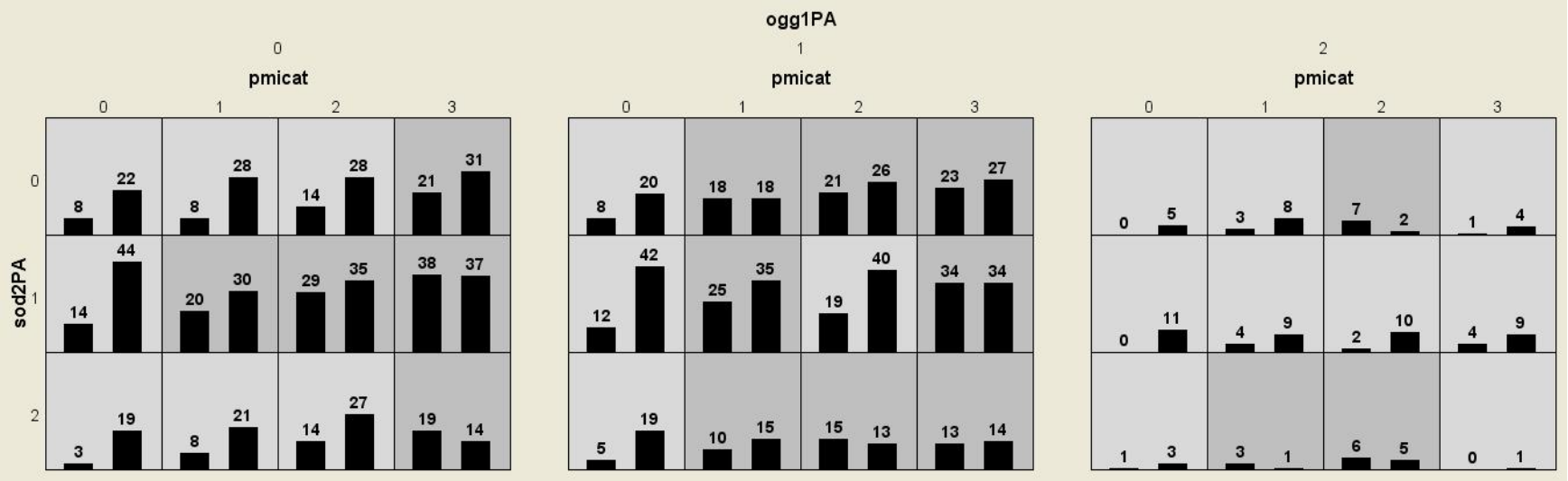
b. Three factor model yielded *OGG1* (rs1052133), *XRCC1* (Ex10) (rs25487), and *ERCC5* (rs17655) as the best model. The low risk group is indicated in light grey, the high risk group in dark grey, and missing cells in white.

FIGURE D.3a. Graphical Presentation of MDR Interaction Models for Primary Methylation Index and SNPs



a. Two factor model yielded primary methylation index and *OGG1* (rs1052133) as the best model. The low risk group is indicated in light grey and the high risk group is indicated in dark grey.

FIGURE D.3b. Graphical Presentation of MDR Interaction Models for Primary Methylation Index and SNPs



b. Three factor model yielded primary methylation index, *OGG1* (rs1052133), and *SOD2* (rs4880) as the best model. The low risk group is indicated in light grey and the high risk group is indicated in dark grey.

FIGURE D.4. Graphical Presentation of CART Model for Single Nucleotide Polymorphisms

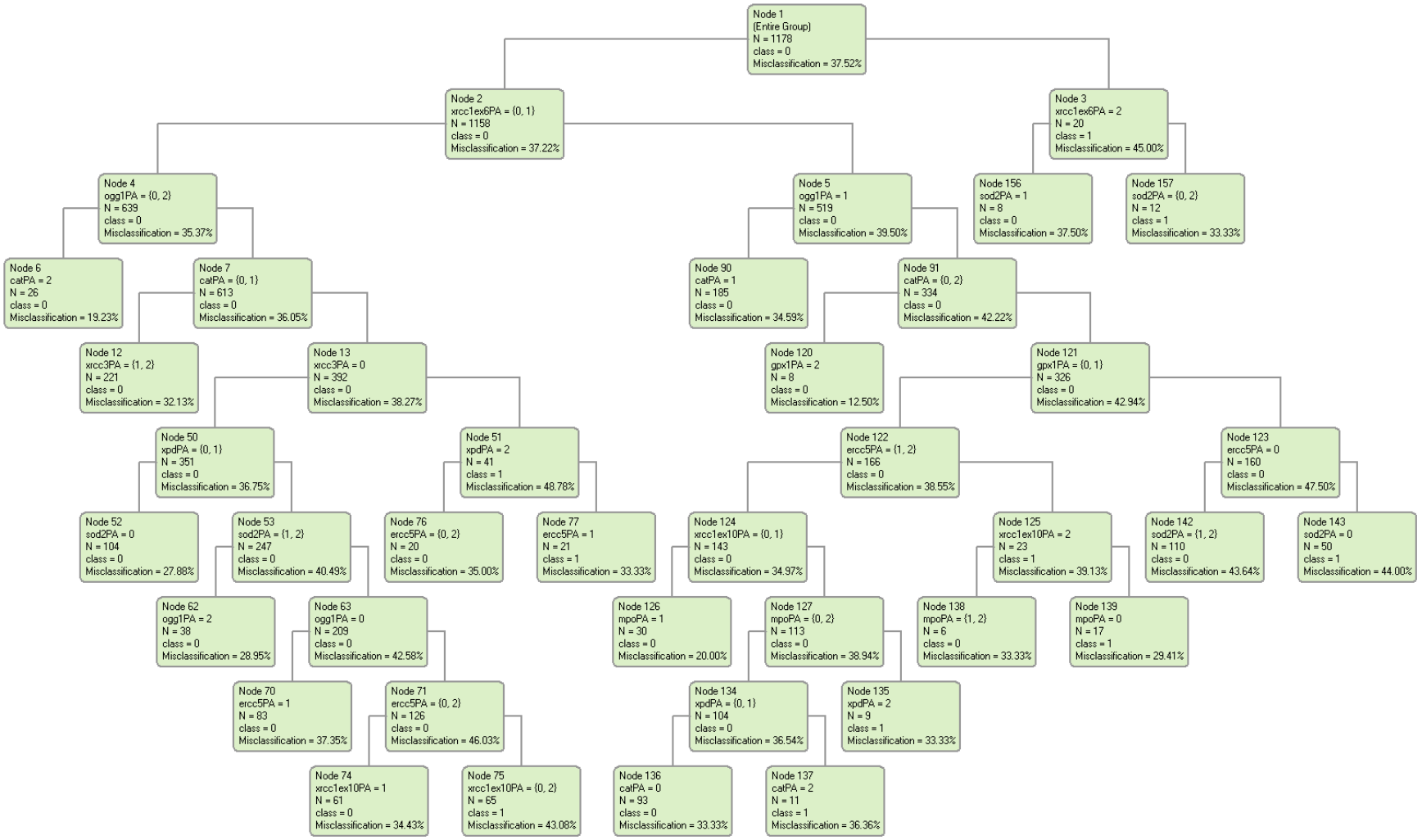


FIGURE D.5. Graphical Presentation of CART Model for Well Water Arsenic and Single Nucleotide Polymorphisms

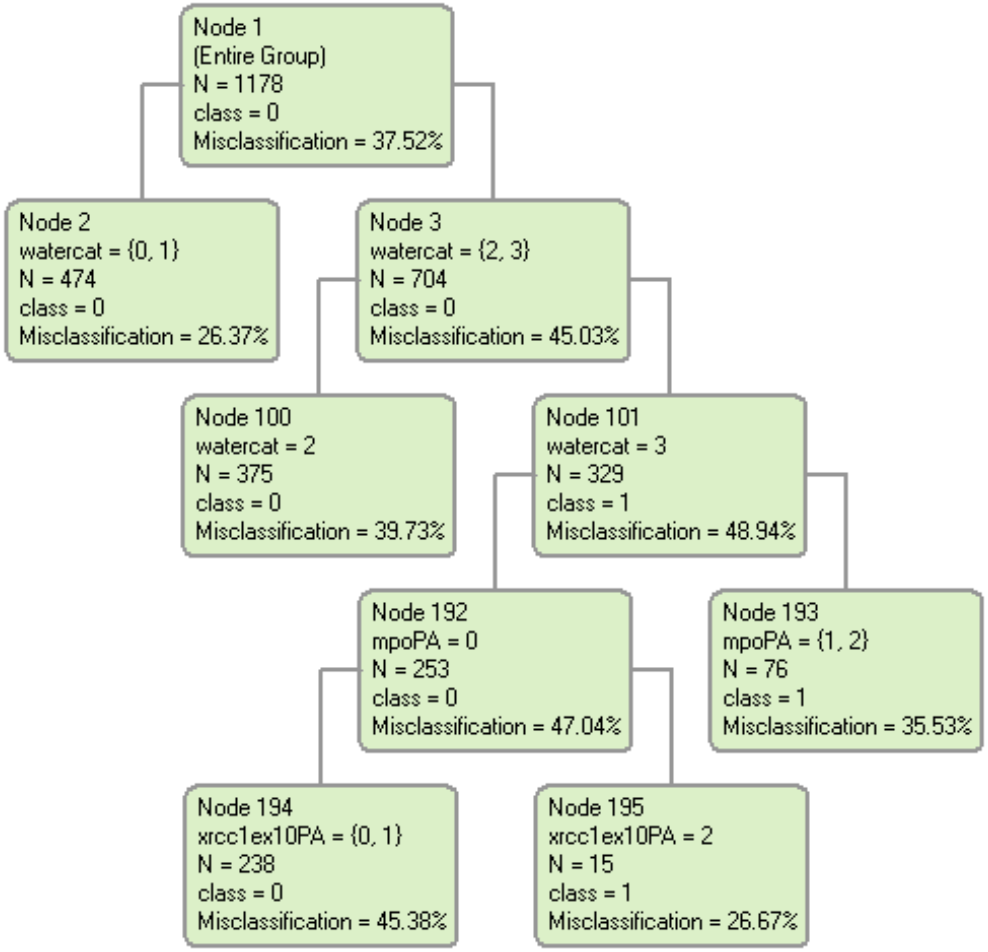


FIGURE D.6. Graphical Presentation of CART Model for Urinary Total Arsenic and Single Nucleotide Polymorphisms

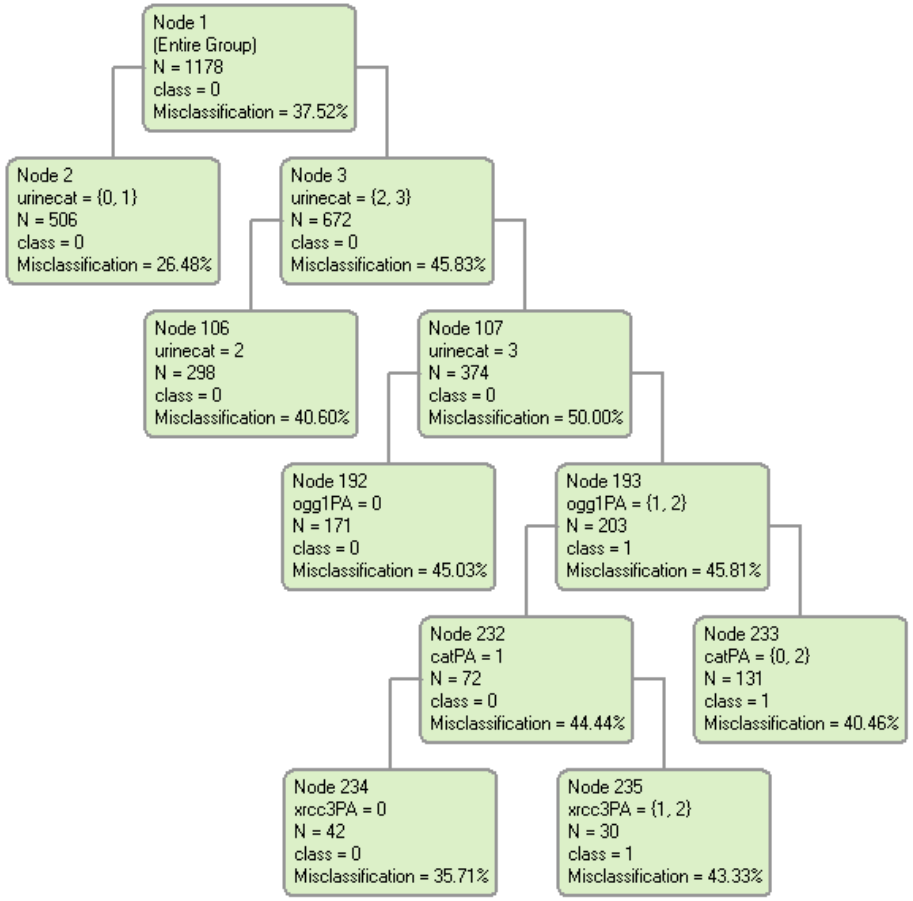


FIGURE D.7. Graphical Presentation of CART Model for Primary Methylation Index and Single Nucleotide Polymorphisms

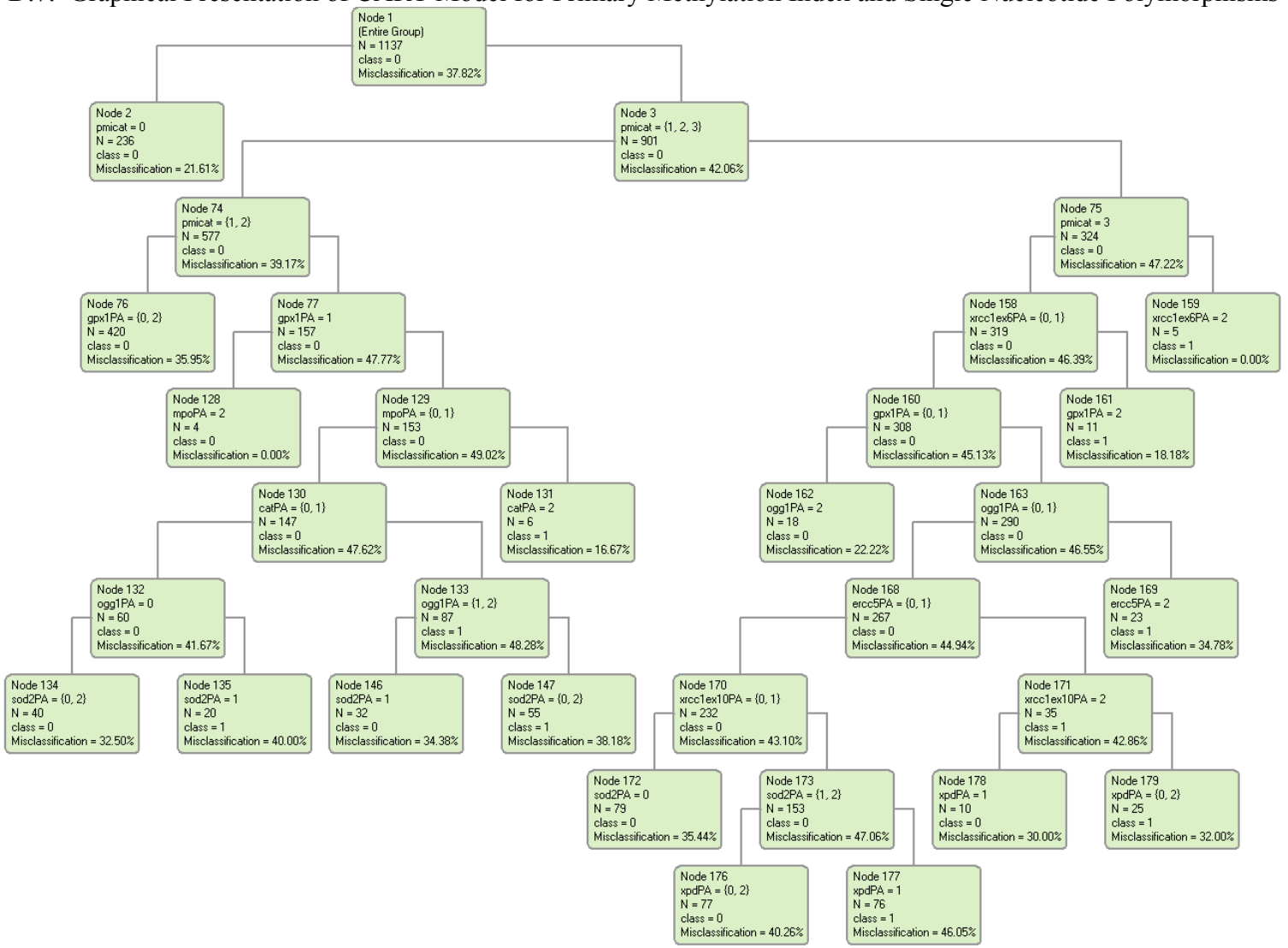


FIGURE D.8. Graphical Presentation of CART Model for Secondary Methylation Index and Single Nucleotide Polymorphisms