MODIFYING EFFECTS OF OXIDATIVE STRESS AND DNA REPAIR VARIANTS ON PHYSICAL ACTIVITY AND BREAST CANCER RISK

Lauren E. McCullough

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

Chapel Hill
2012

Approved by:

Marilie Gammon

Regina Santella

Andrew Olshan

Kari North

Patrick Bradshaw

Abstract

LAUREN E. MCCULLOUGH: Modifying Effects of Oxidative Stress and DNA repair Variants on Physical Activity and Breast Cancer Risk
(Under the direction of Marilie Gammon)

Purpose. The mechanisms driving the inverse association between recreational physical activity (RPA) and breast cancer risk are unclear. Exercise both increases reactive oxygen species production, which may transform normal epithelium to a malignant phenotype, and enhances antioxidant capacity, which could protect against subsequent oxidative insult. Physical activity may also improve damage repair systems, particularly those that operate on oxidative damage. Given the paradoxical and complex effects of physical activity both oxidative stress and DNA repair pathways are of interest. Polymorphisms in these pathways may modify the association between RPA and breast cancer incidence. **Methods**. We estimated interactions between RPA and several polymorphisms in oxidative stress-related genes (*CAT, COMT, GPX, GSTP1, GSTA1, GSTM1, GSTP1, MPO,* and *MnSOD*) as well as DNA repair genes (*ERCC1, MGMT, MLH1, MSH2, MSH3, OGG1, XPA, XPC, XPD, XPF, XPG,* and *XRCC1*). Data were from the Long Island Breast Cancer Study Project, a population-based, case-control study with interview and biomarker data available on 1053 cases and 1102 controls. **Results**. Six variants in antioxidant and DNA repair pathway genes (*CAT* rs1001179, *GSTP1*-lle105Val, *XPC*-Ala499Val, *XPF*-Arg415Gln, *XPG*-Asp1104His and *MLH1*-lle219Val) interacted with postmenopausal RPA (p=0.043, 0.006, 0.048, 0.022, 0.012, and 0.010, respectively). Highly active women with genotypes related to reduced antioxidant capacity were at increased risk of breast cancer (*CAT* OR=1.61; 95% CI, 1.06-2.45) while risk reductions were observed among moderately active women with genotypes related to enhanced antioxidant capacity

(*GSTP1* OR=0.56; 95% CI, 0.38-0.84). With respect to DNA repair we found risk reductions for highly active women with common genotypes for *XPC* (OR=0.57; 95% CI, 0.38-0.84) and *XPF* (OR=0.64; 95% CI, 0.46-0.89) compared to non-active women homozygous for the major alleles. Non-significant risk reductions were observed among active women with at least one variant allele for *XPG* and *MLH1*, respectively. **Conclusions**. Genes involved in antioxidant and DNA repair pathways may modify the RPA-breast cancer risk association. While the functional significance of many polymorphisms with respect to breast cancer remains largely unknown, the observed associations are biologically plausible and consistent across multiple indicators of physical activity reducing the likelihood that these findings are attributable to chance. Our results merit further investigation.

To my mother, who believed in me when I couldn't believe in myself.

To my father, who lost his battle with cancer but whose spirit continues to work through me.

To my husband, who has accompanied me on this long journey and who has always been the wind beneath my wings.

Acknowledgements

This dissertation is something I am deeply proud of and is a product of the numerous people who have supported me over the past several years.

I would like to first give praise and honor to my Lord and Savior Jesus Christ. For without him none of this would be possible.

I acknowledge the members of my dissertation committee who challenged me when appropriate, provided thoughtful feedback at every step, and supported me always. I am particularly grateful to my advisor Marilie Gammon whose dedication to my education and development as an independent scientist far surpassed any expectation. She has enriched my educational experience at UNC beyond measure and I am forever indebted to her.

The support and encouragement of my parents and extended family throughout the years has played an incredible role in this achievement. When faced with adversity and failure they picked me up and cheered me on. I am forever grateful. I am thankful to my friends Sirin Yaemsiri, Christina Ludema, and Sarah Reah who know the difficulties and challenges of this process first hand. Your friendship has made this process bearable. I appreciate all the criticism, guidance and distractions you three have provided throughout the years.

Finally, I must acknowledge my husband and best friend Christopher Young who has shared my trials, joys, failures and comebacks for over a decade. Words cannot begin to describe how much your support has meant to me. This accomplishment is as much yours as it is mine. I am also thankful to my son, Jayden Christopher, who keeps me focused on what is important.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

8-oxo-dG	8-oxodeoxyguanosine
95% CI	95% confidence interval
ADP	Adenosine diphosphate
AP	Apurinic/apyrimidinic
APE1	AP endonuclease
ATM	Ataxia telangiectasia mutated
ATP	Adenosine triphosphate
BER	Base excision repair
BMI	Body mass index (BMI)
BRCA1	Breast cancer 1
BRCA2	Breast cancer 2
BRFSS	Behavioral Risk Factor Surveillance System
CAT	Catalase
CDC	Centers for Disease Control and Prevention
CHEK2	CHK2 checkpoint homolog
CRP	C-reactive protein
DAG	Directed acyclic graphs
DSB	Double strand break
EPT	Estrogen-progestin therapy
ER	Estrogen receptor
ERCC	DNA excision repair protein
GPx	Glutathione peroxidase
GST	Glutathione S-transferases
GWAS	Genome-wide association studies
HCFA	Health Care Finance Administration

HER2	Human epidermal growth factor receptor 2
HNPCC	Hereditary non-polyposis colorectal cancers
HR	Hormone receptor
HRT	Hormone replacement therapy
HWE	Hardy-Weinberg equilibrium
ICR	Interaction contrast ratio
IGF	Insulin-like growth factor
IGF-BP	Insulin-like growth factor binding protein
IL-1β	Interleukin-1-beta
IL-6	Interleukin-6
LD	Linkage disequilibrium
LIBCSP	Long Island Breast Cancer Study Project
LRT	Likelihood ratio test
MAF	Minor allele frequencies
MET	Metabolic equivalents of energy expenditure
MLH1	MutL homologue 1
MLH3	MutL homologue 3
MMR	Mismatch repair
MnSOD	Manganese superoxide dismutase
MPO	Myeloperoxidase
MSH2	MutS homologue 2
MSH3	MutS homologue 3
MSH6	MutS homologue 6
MSI	Microsatellite instability
MUTYH	MutY homologue
NER	Nucleotide excision repair

NHEJ	Non-homologous end joining
OC	Oral contraceptives
OGG1	8-oxoguanine glycosylase
OR	Odds ratio
P53	Tumor protein 53
PAH	Polycyclic aromatic hydrocarbon
PCNA	Proliferating cell nuclear antigen
PCR	Polymerase chain reaction
PMS2	Postmeiotic segregation increased 2
PR	Progesterone receptor
PTEN	Phosphatase and tensin homolog
RDD	Random digit dialing
RNS	Reactive nitrogen species
ROS	Reactive oxygen species
RPA	Recreational physical activity
RR	Relative Risk
SHBG	Sex-hormone-binding globulin
SNP	Single nucleotide polymorphisms
STK11	Serine/threonine kinase 11
TNF	Tumor necrosis factor
US	United States
XRCC	X-ray repair cross-complementing protein
XP	Xeroderma pigmentosum

Chapter I: Background

This dissertation aimed to examine effect modification of the physical activity-breast cancer association by single nucleotide polymorphisms (SNP) in two different pathways: (1) oxidative stress and (2) DNA repair. This investigation began by estimating the associations between genetic variants in catalase (*CAT*); mismatch repair (MMR) and breast cancer risk, using resources from the Long Island Breast Cancer Study Project (LIBCSP), a population-based, case-control study with detailed data on recreational physical activity (RPA) and genetic markers. I subsequently examined how functional variants in candidate genes from the aforementioned pathways modified the effect of physical activity on breast cancer incidence. Secondary aims involved examining the effects of multiple genetic polymorphisms on breast cancer risk using SNP*SNP analyses, gene*gene analyses and a pathway based approach, as well as evaluating associations between genetic polymorphisms and breast cancer risk in hormonally distinct case groups (e.g. hormone receptor positive and hormone receptor negative).

Chapter one provides a detailed review of the relevant literature, to inform discussion of the study rationale and approach. Specifically, this chapter summarizes the breast cancer literature (1.1), details the physical activity-breast cancer mechanisms described to date (1.2), and provides biologic rationale for two alternative pathways (e.g. oxidative stress [1.3] and DNA repair [1.4]) that may, in part, drive the observed inverse association between exercise and breast cancer risk.

1.1 Breast Cancer Biology and Epidemiology

Although much of what is understood about the etiology of breast cancer has evolved from risk factor epidemiology, many of the underlying mechanisms of the disease remain unknown. Many established breast cancer risk factors, such as reproductive characteristics, are not easily modifiable but known to play an important role in breast cancer development. A large proportion of breast cancer risk may, however, be attributed to factors which are amendable to intervention. The case for these modifiable or lifestyle risk factors are most clearly revealed by the large differences in breast cancer incidence between countries [1] and the monotonic increase in risk among immigrants across generations. Physical activity and obesity have emerged as two important, potentially modifiable, targets for breast cancer risk reduction and may account for a large proportion of breast cancer cases. Identifying pathways through which these factors operate could play an important role in advancing the knowledge of breast cancer etiology, and improving risk reduction strategies for breast cancer. The following section outlines the trends in breast cancer incidence, potential mechanisms, and risk factors for the disease.

1.1.1 Trends in Breast Cancer Incidence

Breast cancer is the leading cause of global cancer incidence and mortality among women. Among women in the United States (US), breast cancer is the primary cause of, non-melanoma, cancer related illness with an estimated 226,870 new cases and 39,510 deaths attributable to breast cancer in 2012 [2]. Breast cancer differentially affects women by age and race (**Figure A.1**). From 2000-2004 94% of incident cases in the US occurred in women over the age of 40 and most women were of European decent [3]. While Caucasian women show a higher rate of breast cancer incidence after age 40, African American women experience the highest rates of premenopausal breast cancer. From 2000-2004 the annual incidence rate of breast cancer for all ages was 132.5 and 118.3 per 100,000 person-years in Caucasian and African American women, respectively [3].

2

1.1.2 Mechanisms of Breast Cancer Risk

Prolonged exposure to circulating estrogens and progestins has long been suspected as a primary mechanism in breast cancer carcinogenesis. Increased endogenous hormone levels are proposed to elevate risk through enhanced cell proliferation which occurs at multiple points along the cancer continuum from initiation to tumor metastasis [4]. Increased cellular proliferation is likely to result in a greater number of deleterious mutations that, if un-repaired, could result in breast malignancy [5]. Estrogen may directly influence cell proliferation via induction of proteins involved in nucleic acid synthesis or through the activation of oncogenes. It can, similarly, indirectly impact proliferation by stimulating the secretion of prolactin or enhancing growth factor production [6]. Closely linked to estrogen are insulin and insulin-like growth factor (IGF). They have been proposed to work with estrogen to influence breast cancer risk by both increasing cell proliferation and preventing apoptosis [7]. Insulin is additionally known to reduce synthesis of sex-hormone-binding globulin (SHBG), a transporter of testosterone and estradiol, in the liver [7]. Reduced levels of SHBG result in increased availability of bioavailable estradiol.

Another commonly proposed mechanism is reduced genomic integrity. There are a number of highly penetrant breast cancer susceptibility genes that have been implicated in familial breast cancer. Well established are breast cancer 1 (*BRCA1*) and breast cancer 2 (*BRCA2*) which account for 15-20% of ancestral breast cancer clustering [8], and are commonly associated with early-onset breast cancer [9]. A 2003 pooled analysis reported an average 65% cumulative risk of breast cancer among *BRCA1* carriers by age 70. The reported cumulative risk among *BRCA2* carriers was 45% [10]. There are a number of other mutations considered to have middle penetrance. Ataxia telangiectasia mutated (*ATM),* CHK2 checkpoint homolog *(CHEK2),* tumor protein 53 *(P53),* phosphatase and tensin homolog *(PTEN),* and serine/threonine kinase 11 (*STK11)* have been classified as breast cancer associated genetic defects responsible for various aspects of genomic integrity [11].

While mid to high penetrant genetic polymorphisms have been implicated in the etiology of heritable breast cancer, these variants are relatively uncommon in the population. Focus has shifted to finding common, low penetrance, polymorphisms which may contribute only a slight increase in risk. Alterations in these genes can serve as triggers for genomic instability, increased mutation, and could mediate, in part, the effects of physical activity and other breast cancer risk factors.

1.1.3 Breast Cancer Risk Factor Epidemiology

Reproductive Factors. Unlike many cancers breast cancer has a number of well-established risk factors (**Table A.1**). As previously described, cumulative exposure to estrogens and progestins appears to play a large role in breast cancer carcinogenesis. Many of the established risk factors for breast cancer are therefore related to, or serve as proxies for, endogenous hormone levels [12]. Early age at menarche [12, 13], age at first birth [14], parity [14, 15], lactation history [16], and late age at menopause [12] are known to influence breast cancer incidence. These reproductive risk factors are hypothesized to primarily impact breast cancer risk by influencing the cumulative lifetime exposure of breast tissue to circulating estrogens and progestins [5]. For example, the inverse association between parity and breast cancer risk may be due to a reduction in the number of ovulatory cycles and thus decreased estradiol exposure, although other mechanisms (mammary cell differentiation and estrogen responsiveness) are likely to account for some portion of this association [17]. Findings in the LIBCSP [18] indicate that reproductive practices have an important role in breast cancer etiology. Investigators reported adjusted odds ratio (OR) and 95% confidence intervals (95% CI) for several risk factors including: parity (OR=0.63 for 4+ children vs. none 95% CI: 0.48, 0.82), breastfeeding (OR=0.70 for 14 months vs. none, 95% CI: 0.53, 0.89), and age at first birth (OR=1.36 for 28+years vs. <22 years, 95% CI: 1.10, 1.69). Age at menarche was not found to influence risk in this study population.

Exogenous Hormones. Oral contraceptives (OC) are the most commonly used contraceptive method for US women. Oral contraceptives generally contain 20 to 35 µg of ethinyl estradiol and are thus a frequent source of exposure to exogenous hormones in premenopausal women [19]. The most comprehensive assessment of the association between OC use and breast cancer risk is the 1996 Oxford pooled analysis of 54 epidemiologic studies with 53,297 breast cancer cases and 100,239 controls. The analysis showed a slightly increased risk of breast cancer among current users (RR=1.24; 95% CI: 1.15-1.33) compared to never users [20]. A risk reduction was observed after stopping OC use (Relative Risk [RR]=1.16 for 1-4 years after stopping and 1.07 for 5-9 years after stopping). No increased risk was found 10 or more years after discontinuation of OCs (RR=1.01; 95% CI: 0.96-1.05) [20]. A more recent 2002 study [21] reported no excess breast cancer risk among current or former OC users (OR=1.0; 95% CI: 0.8-1.3 and OR=0.9; 95% CI: 0.8-1.0, respectively). Differences in study results may be due to changes in OC formulations with newer contraceptives having lower-dose estrogen.

Hormone replacement therapy (HRT) may be another source of exogenous hormones. It is frequently used among postmenopausal women to delay bone density loss and alleviate menopausal symptoms. A 2009 review of postmenopausal hormone therapy and breast cancer risk reported that while combined estrogen-progestin therapy (EPT) moderately increased the risk of breast cancer (20-40%), the evidence for unopposed estrogen showed no increase in risk [22]. This study did not account for relative duration of use. The Nurses' Health Study found monotonic increases in breast cancer risk with current use of unopposed estrogen (*p* for trend <0.001) and a 42% increased risk among women with 20+ years of use compared to never users [23]. Exogenous hormone use was associated with breast cancer risk among Long Island study participants. The odds of breast cancer were elevated for OC use, HRT use, and long term HRT use [24]. The authors note

that the timing of exogenous hormone use is important in understanding risk, specifically among postmenopausal women.

Environmental Factors. While a considerable number of risk factors have been linked to estrogen the etiology of breast cancer is multi-factorial and increased breast cancer risk could occur through many other pathways. It is well established that exposure to ionizing radiation leads to increased cancer risk. Women exposed to the atomic bomb in Nagasaki and Hiroshima Japan experienced as much as a 9-fold increase in breast cancer risk [25].

Polycyclic aromatic hydrocarbon (PAH)-DNA adducts are among the most consistently reported environmental factors associated with breast cancer. Human PAH exposure primarily comes from byproducts of fuel burning, cigarette smoke, and the consumption of grilled and smoked foods [26]. LIBCSP investigators reported a 50% increased risk (95% CI: 1.04, 2.20) among women with highest PAH quantile compared to women in the lowest quantile [26].

Cigarette Smoking. The association between cigarette smoking and breast cancer risk has been inconsistent, likely due to competing biologic mechanisms. Smoking is known to increase exposure to carcinogens. However, it has been suggested that smoking is related to appetite suppression, early initiation of menopause, and altered hormone metabolism which would therefore decrease overall exposure to estrogens [25]. The exact association between smoking and breast cancer remains to be elucidated, as active smoking has not consistently been found to be associated with breast cancer in the epidemiologic literature. Neither self-reported former or current smoking was associated with breast cancer risk among Long Island women [18]. However, LIBCSP data did indicate a positive association among women who resided with a smoking spouse for greater than 27 years (OR=2.10; 95% CI: 1.47-3.02) [27].

6

Alcohol Consumption. The consumption of alcohol has been associated with a modest increased risk of breast cancer [28]. It is suggested that the alcohol-breast cancer association may be directly related to alcohol metabolism and its effects on the levels of estrogen and estrogen receptors in breast cells [29]. Other mechanisms of the alcohol-breast cancer relationship have been proposed. Mechanisms include an increase in reactive oxygen species (ROS), hydroxyl radicals, and DNA modification [28]. Reports from a 2006 meta-analyses of high quality studies showed a 22% (95% CI: 1.09, 1.37) increased risk of breast cancer among women classified as drinkers, compared to abstainers [30]. Contrary to these results Long Island investigators found no association between ever alcohol use and never users [18].

Diet. A number of dietary factors including: fat, fiber, and fruit and vegetable consumption have been evaluated in association with breast cancer risk. However, the role of diet in breast cancer etiology remains controversial [31] as the magnitude and direction of effect are known vary by study design [32]. Moreover, studies which observe an association between diet and breast cancer incidence show that dietary factors may only slightly modify risk. There are several rationales for null or weak associations in the diet-breast cancer literature. Lof and Widerpass (2009) suggest that: (1) measurement error may disguise existing associations, (2) dietary exposures may not be ascertained during the etiologically relevant time period, and (3) there may be differences in risk according to tumor characteristics or genetics [33]. Alternatively, there may also be no causal association between diet and breast cancer risk.

Obesity. The relationship between obesity and breast cancer varies by menopausal status. Studies indicate that obesity decreases risk among premenopausal women while increasing risk among postmenopausal women [34]. Pichard and colleagues report a positive relationship between body mass index (BMI, the ratio of weight in kg squared to height in meters squared) and breast cancer with RRs ranging from 1.26 to 2.52 among

7

postmenopausal women [35]. The positive association between obesity and postmenopausal breast cancer is thought to occur through the aromatization of androgens in adipose tissue. These androgens are subsequently converted to estradiol, the most metabolically active form of estrogen, thereby increasing breast cancer risk [7, 36]. Compared to ovarian estrogen production among premenopausal women, adipose mediated estrogen production is highly unregulated [36]. In combination with reductions of SHBG (a frequently observed phenomena of obesity-related hyperinsulemia), unregulated estrogen production results a greater than 2-fold increase in free estradiol among postmenopausal women [36].

Physical Activity. There has been an overwhelming amount of epidemiological literature describing the beneficial effects of exercise in breast cancer risk reduction [37-45]. Studies suggest that risk reduction is approximately 25% when the most physically active woman is compared to the least physically active woman, even among high risk populations [44, 46]. The most obvious mechanism through which physical activity may influence cancer risk is by reducing adipose tissue, and consequently the hormonal milieu that occurs with postmenopausal obesity (discussed above) [41]. However, physical activity has been shown to reduce the risk of premenopausal breast cancer and the inverse association with postmenopausal breast cancer risk persists even upon controlling for body weight [47, 48]. These observations indicate that in addition to an obesity-mediated pathway, physical activity is likely to influence breast cancer carcinogenesis through independent mechanisms.

1.1.4 Conclusions

Breast cancer remains an important public health concern in both the US and abroad. There has been considerable research in identifying the epidemiologic risk factors associated with breast cancer, but the underlying mechanisms of the disease remain unknown. While many of the established risk factors have been linked to hormone pathways, they are primarily reproductive and unlikely to be intervened upon. There are a

8

number of non-reproductive risk factors known to play a role in breast cancer development. While they may not be as predictive as those which are reproductive, they can ultimately be controlled, aiding in the reduction of breast malignancy development. Physical activity appears to play an important role in the reduction of both pre- and postmenopausal breast cancer risk. Given the widespread accessibility of physical activity, identifying mechanisms through which it acts, independent of obesity, has become increasingly important.

This dissertation focused on pathways relevant to breast cancer and physical activity, specifically oxidative stress and DNA repair. In the following sections I describe the current state of knowledge of the inverse association observed between physical activity and breast cancer risk, summarize the mechanistic pathways which have been evaluated to date and present a conceptual model of the hypothesized mechanisms through which physical activity is proposed to act. Finally, I review two pathways (i.e. oxidative stress and DNA repair) that play a potential role in the etiology of breast cancer, as well as provide a rationale for examining these two pathways in the molecular epidemiology of physical activity.

1.2 Recreational Physical Activity

Physical activity has been associated with reduced incidence of a number of chronic diseases including heart disease [49, 50], diabetes [50, 51], stroke [50], osteoporosis [52] and cancer [50]. Increased activity has also been proposed to counter disability, and improve cognitive functioning. Interest in physical activity for the primary prevention of breast cancer has increased, as there are convincing epidemiologic data that show a beneficial effect of exercise on breast cancer risk reduction [53]. While most of the established risk factors for breast cancer such as family history and reproductive characteristics are not easily amenable to intervention, physical activity may be one of few risk factors for cancer that can be modified through lifestyle and behavior change. It is unclear, however, whether activity alone provides a protective effect or if it serves as a proxy for overall health status.

Elucidation of the underlying mechanism linking physical activity inversely to breast cancer risk would strengthen the biologically plausibility of the association. Mechanistic insight could additionally aid in identifying targets for intervention, inform the recommendations for lowering breast cancer risk, and provide new clues to cancer biology.

1.2.1 Definitions and Measures of Physical Activity

Physical activity is defined as bodily movement produced by skeletal muscles resulting in a quantifiable form of energy expenditure [54]. Physical activity can be broadly classified as either cardiorespiratory (aerobic activity) or resistance (anabolic activity), each distinct in their physiological effects. Aerobic activities profoundly impact the cardiovascular and respiratory systems while anabolic activities influence the neural and muscular systems [55]. All individuals are exposed to physical activity in several domains across the life span. The Centers for Disease Control and Prevention (CDC) defines regular physical activity among adults as at least 150 minutes a week of moderate-intensity aerobic activity or 75 minutes a week of vigorous-intensity aerobic activity. These recommendations are based on the Healthy People 2010 physical activity objectives and the 2008 Physical Activity Guidelines for Americans. The CDC's analysis from the Behavioral Risk Factor Surveillance System (BRFSS) indicated that 76% of the US population participated in some type of physical activity during month prior to interview [56]. With a nationwide prevalence of approximately 75% physical activity may conceivably be one of the most pervasive modifiable exposures associated with breast cancer risk.

There are three dimensions to physical activity, each of which may be varied in their effects on carcinogenesis. Several investigators advocate that a complete assessment of an individual's energy expenditure from physical activity would include information on all of these important dimensions [37, 40]. The first component is frequency, which reflects the number of times the activity is performed (e.g., times per month/week/year). The second dimension is duration, broadly defined as the length of each activity session (e.g., minutes

or hours per episode). The final element is the intensity or rate of energy expenditure required to execute the activity. Metabolic equivalents of energy expenditure (MET) are commonly used to assess intensity. One MET unit is defined as the energy expended sitting quietly. This is equal to 3.5 millimeters of oxygen per kilogram of body weight per minute [57]. According to the CDC and the American College of Sports Medicine, light, moderate, and vigorous activities are classified as <3 METs, 3-6 METs, and >6 METs, respectively [58]. A summary measure of the three components is MET-hours, obtained by multiplying all three dimensions of activity. MET-hours may reflect the activity dose of one session, day, week, or month and is useful when grouping participants or comparing activity levels across populations. The type of activity may also be important in quantifying physical activity as recreational activities tend to be higher intensity and shorter duration compared to activities related to occupation or daily living which are traditionally low intensity and longer in duration.

Due to the heterogeneity of physical activity and little standardization in assessment methods, it is often difficult to obtain valid estimates of energy expenditure [55]. While physiological measures of physical activity (e.g. resting heart rate and aerobic capacity) perform particularly well and are commonly regarded as a gold standard, they are not frequently employed in epidemiologic studies because they fail to capture the etiologically relevant time period and are often too expensive for large population-based designs [59]. Questionnaire based assessments are common practice in observational epidemiology and may query participants all three dimensions of physical activity. While there are qualitative differences in physical activity assessment across studies, the physical activity-breast cancer literature overwhelmingly supports an inverse association.

1.2.2 Epidemiology of Physical Activity and Breast Cancer

The association between physical activity and breast cancer has been studied at length. The overall findings are detailed in several reviews [37-45]. Risk reductions reported

11

in these studies range from 20-40% among active women [40, 41, 43, 44]. Generally slightly stronger associations have been found in case-control compared with cohort studies [37, 42], although the Nurse's Health Study reported a slightly stronger association for lifetime activity in the prospective design (prospective OR=0.58, 95% CI: 0.37, 0.93 versus retrospective OR=0.80; 95% CI: 0.50, 1.29) [60]. While early reports show risk reductions for both occupational and recreational activity, a comprehensive 2008 review of physical activity parameters and breast cancer risk [42] showed the greatest risk reductions for RPA (20% risk reduction). Activity related to occupation, transportation, and daily living each resulted in approximately 14% reduced risk. The same review reported similar risk reductions from vigorous activities (average 26%) and moderate intensity activities (average 22%).

Timing. Determining the time period during which physical activity most influences breast cancer risk is of paramount importance in making recommendations for lowering incidence. The strength of the physical activity-breast cancer association may vary across the life course, as is observed for other established risk factors. In a 2001 review Latikka and colleagues [45] assessed the effect of physical activity at various phases of life on breast cancer outcomes. The literature primarily focused on current activity, of which the vast majority of studies (~82%) reported inverse associations. Two of three studies showed adolescent activity or activity during college may protect against breast cancer while inconsistencies were observed for studies that examined both historical and current activity [45]. Dorn et al. examined the physical activity-breast cancer association at two, ten, and twenty years prior to interview as well as across the lifetime [61]. Although most of the estimates were imprecise, they reported risk reductions for all physical activity categories above the referent in each time period with strongest effects observed for women active at least 20 yr prior to interview and among postmenopausal women who were consistently active throughout their lifetime [61]. A recent prospective study found moderate-to-vigorous activity during the past 10 years to be associated with decreased postmenopausal breast

12

cancer risk (RR=0.84; 95%CI=0.76, 0.93). The associations for activities performed during other periods of life were near null (RR=1.03, 1.01, and 0.97 for 15-18, 19-29, and 35-39 years of age, respectively), and lifetime activity was not assessed [62]. Although the optimal timing of physical activity for breast cancer protection remains to be resolved, it appears that activity performed through adulthood and the postmenopausal years provide the greatest risk reductions. This was particularly true in our own study population [63]. In the only other study known to use the same comprehensive physical activity assessment as the LIBCSP only lifetime activity was reported [64] and exercise was shown to be inversely associated with breast cancer risk.

Dose Response. There is some evidence for a dose response relationship between increasing activity levels and decreasing breast cancer risk. Thune and colleagues [43] observed graded dose-response relationships in 57% of studies evaluated (N=28). The proportion is as high as 87% in some reviews [40]. Evidence of dose-response relationships are more frequently observed in case-control compared to cohort studies [42, 43]. One review reported that 47% of case-control studies and 39% of cohort studies found linear tends for decreasing risk with increasing activity [42]. Linear trend analysis performed by Monninkhof and colleagues indicated a 6% (95% CI: 3%, 8%) decrease in breast cancer risk for each additional hour of physical activity per week [38]. In the LIBCSP we found a non-linear dose response association between exercise and breast cancer with the greatest risk reductions observed among women in the third quartile of RPA [63].

Effect Measure Modification. Many of the observed differences in the effect of physical activity can be, in part, ascribed to methodological differences in evaluating activity across studies. It is also likely that these differences can be attributed to the heterogeneity of effects among subgroups of women. It is important to consider the association between physical activity and breast cancer risk within strata of menopausal status, body mass index, family history, and other potential effect measure modifiers. These analyses not only help to

identify at-risk subgroups, but they may aid in further understanding the biology driving the physical activity-breast cancer association.

The body of literature to date indicates that there are stronger and more consistent effects of physical activity among postmenopausal women compared to premenopausal women [61, 65-67], although CIs overlap in many studies. A 2004 review study [41] showed that among 26 studies examining the association between physical activity and premenopausal breast cancer risk 50% found no association. Thirteen studies reported risk reductions, with seven being statistically significant. Among the 27 studies conducted in postmenopausal women, 22 (81%) found risk reductions when comparing the most active women to women who were least active. Sixteen of these studies reported statistically significant risk reductions. More recent reviews observed similar trends among pre and postmenopausal women [38]. While it is clear that an effect of physical activity persists in both pre and postmenopausal strata, risk reductions are greater in magnitude for postmenopausal women (40% average risk reduction) compared to premenopausal women (33% average risk reduction) [42].

Both independent and review studies demonstrate a protective effect of physical activity in low and high BMI categories. However, the magnitude of effect within strata of BMI has been shown to vary. One study reports significant decreases in invasive breast cancer risk with increasing levels of long-term strenuous recreational physical activity among women with a BMI < 25 (P trend=.03) but not among women with BMI ≥ 25 [68]. Investigators of the E3N cohort reported no effect modification by BMI on the physical activity-breast cancer association [46] as did investigators of the MARIE study, even upon stratifying by menopausal status [69]. Some review studies have drawn similar conclusions [38, 59]. Friedenreich and colleagues observed a trend of decreasing breast cancer risk with decreasing BMI and increasing physical activity. Risk reductions were approximately 25% among women with normal BMI (22–25 kg/m^2) and 20% among women with high BMI (≥25

14

kg/m^2). There were near null effects of physical activity on breast cancer risk among women classified as obese (≥30 kg/m^2), although few studies reported effects in this strata [42]. Data from the LIBCSP showed that among postmenopausal women who engaged in more than 9 hrs/wk of RPA (on average) risk reductions were 27% for normal BMI, 15% for overweight women, and near null (OR=0.99) for women with BMI ≥30 kg/m^2 [63]. Few studies have assessed modification by family history, but the greatest risk reductions have been reported for women without a family history of breast cancer. Dallal et al. [68] reported significant decreases in breast cancer risk with increasing levels of strenuous recreational physical activity among women with no first-degree family history of breast cancer (P trend=.01). This trend was not observed among women with a family history of breast cancer (P trend=.72). Similar associations for family history have been observed in other studies [64, 70] and Friedenreich [42] reported an average risk reduction of 6% among women without a family history, compared to an average increased risk of 20% for women with a family history of breast malignancy. Contrary to these findings, Pijpe and colleagues [71] reported a reduced risk of breast cancer among *BRCA1/2* mutation carriers engaging in sports participation (HR=0.59, 95%CI=0.36-0.95 among women with a medium versus low level of sport intensity and duration). Additionally, one study reported delayed onset of breast cancer among *BRCA1* mutation carriers who were physically active [72].

Energy intake and energy expenditure together determine an individual's overall energy balance which greatly affects adiposity and breast cancer risk. It is therefore important to determine if energy intake modifies the physical activity-breast cancer relationship. Studies which have been able to control for energy report no difference in the effects of physical activity for high vs. low consumers of total energy [42, 59]. An average risk reduction of 21% has been reported for both groups [42].

Etiologic Heterogeneity. While hormone receptor status is not an effect modifier, but a marker of potential etiologic heterogeneity, studies have assessed the physical

activity-breast cancer association within strata of these breast cancer outcomes. Inconsistencies have been observed in the literature. Dallal and colleagues [68] report significant decreases in breast cancer risk with increasing levels of both moderate (P trend=0.003) and strenuous (P trend=0.003) RPA among women with estrogen receptor (ER) negative tumors. No associations were observed for ER+, ER+/progesterone receptor (PR)+, or ER+/PR- cancers. These results are contrary to similar analyses which report no difference in the physical activity-breast cancer association by receptor status [64], and still others showing stronger associations for ER+ tumors [73]. Friedenreich and colleges reported risk reductions of 29% and 14% for hormone receptor negative and positive tumors, respectively [42].

1.2.3 Physical Activity Mechanisms

The biologic pathways influencing the physical activity-breast cancer association are less understood than its epidemiology. While it is consistently observed that physical activity aids in the reduction of breast cancer risk, there has been little molecular evidence demonstrating the physical activity cancer prevention paradigm [44, 74, 75]. The lack of information on the molecular epidemiology of physical activity and the strong potential for confounding effects emphasize the need for mechanistic data to facilitate causal inference and identify new targets for intervention. There are several mechanisms that may mediate the association between physical activity and breast cancer risk. While reduction in hormonal pathways and obesity related mechanisms offer the most convincing explanation to date, other physiological effects of exercise may prove beneficial as well. Mechanisms commonly cited in the literature include: changes in body weight, sex steroid hormones, insulin, growth factors, inflammation, and immune functioning (**Figure A.2**) [74-78].

Body Weight. The most commonly proposed mechanism for the physical activity breast cancer association is through a reduction in body weight. Increased body weight is an independent risk factor for postmenopausal breast cancer and its effects on endogenous

16

estrogens are well documented. Research has shown that visceral fat (fat accumulated over the central abdomen) is metabolically active and has numerous physiological corollaries thought to influence carcinogenesis [79]. Physical activity is known to preferentially reduce central obesity [59] and is likely one mechanism through which exercise may prevent cancer development. It is, however, difficult to disentangle physical activity dependent mechanisms from those associated with obesity, as the two are closely linked.

Sex Steroid Hormones. Sex steroid hormones (particularly estrogens) are known to play an important role in pathogenesis of breast cancer. Physical activity has been shown to influence the amount circulating reproductive hormones in both pre- and postmenopausal women.

The total number of menstrual cycles is a well-established risk factor for breast cancer. Among premenopausal women some epidemiologic studies have shown that excessive amounts of energy expenditure can cause temporary suppression of gonadal hormones, delayed menses, menstrual cycle irregularities, anovulation, and amenorrhea [80-83], primarily as a result of luteal phase inadequacy [39]. While these data suggest that exercise reduces cumulative exposure to sex steroid hormones thereby decreasing the risk of breast cancer [45, 80], more recent research indicates that beyond its effect on energy availability exercise has little disruptive effect on the hormonal milieu of premenopausal women [84]. Moreover, there is little evidence of a direct association between disturbances in menstrual characteristics and breast cancer incidence.

Among premenopausal women the ovaries are the principal source of estrogens (primarily estradiol). Following menopause the ovaries produce very little estrogen. However, through the aromatization of androgens in fat tissue postmenopausal women may still be exposed to high endogenous estrogen levels [85]. The P450 enzyme aromatase is known to be involved in the biosynthesis of estrogens from androgenic precursors [77] and has been correlated with breast cancer [86]. In both pre and postmenopausal women

17

estradiol forms a reversible redox reaction with estrone which is subsequently metabolized along one of two pathways: 2-hydroxyestrone or 16α-hydroxyestrone [87]. Adipose mediated estrogen production is preferential to the bioactive 16α-hydroxyestrone pathway and is associated with decreased levels of the less bioactive 2-hydroxyestrone [88]. Among obese postmenopausal women the increased propensity for the 16α-hydroxyestrone pathway results in an even greater amount of circulating estrogen. In contrast, premenopausal women experience consistently high levels of estrogen from menarche, regardless of weight, so any additional exposure from the aromatization of androgens or 16α-hydroxyestrone metabolizing pathway would not greatly impact overall estrogen levels.

Bioavailability of sex steroids also increases after menopause as the levels of SHBG, the predominant carrier of estradiol, decreases [89]. This is of particular importance among women who are obese because circulating triglycerides from fat stores are able to dislodge estradiol from SHBG thereby increasing bioavailability of sex steroids [90]. Physical activity has been shown to reduce serum concentrations of both estrogens and androgens among postmenopausal women [83, 91-93], although there may be the possibility of a U-shaped relation between activity and hormone levels [92]. These changes are likely to occur through indirect effects of physical activity on obesity.

Insulin. Insulin, a peptide hormone secreted by the beta cells of the pancreas, is primarily responsible for the regulation of blood glucose. Insulin is positively associated with breast cancer risk [94] and influenced by both central adiposity and physical activity [95]. Increases in visceral fat are associated with elevated levels of serum free fatty acids in the blood. These fatty acids are thought to cause a reduction in glucose uptake and an obligatory rise in insulin secretion in an effort to maintain glucose homeostasis [76, 77, 96, 97]. Increased levels of insulin cause a cascade of deleterious events that include: (1) decreased production of insulin-like growth factor binding proteins (IGF-BPs), (2) amplified levels of IGF-1, and (3) reduced availability of SHBG [77, 98]. Increases in both insulin and

18

IGF-1 are proposed to inhibit apoptosis and stimulate the progression of neoplastic mammary cells from the G1 to S phase of the cell cycle [99]. Reduced SHBG increases the fraction of bioavailable estradiol and testosterone. These mechanisms, collectively, provide an opportunity for tumor development and progression.

Insulin resistance occurs when insulin becomes less effective at lowering blood sugars. Physical activity may influence insulin resistance in multiple ways. Acute bouts of exercise are reported to increase glucose uptake by skeletal muscle resulting in improved insulin sensitivity [100, 101]. However, intervention studies show comparable improvements in insulin resistance for diet-induced or exercised-induced weight loss [101] suggesting that an overall reduction in adiposity, irrespective of the mechanism, increases insulin resistance. In addition to a reduction of insulin levels from physical activity, regular exercise has been proposed to influence cancer risk through its effects on IGF-1 [102]. The relation between physical activity and IGF-1 is not clear however; experimental studies have not shown an effect of exercise on circulating IGF-1 levels [103]. Similarly, animal models indicate that exercise training does not affect basal levels of IGF-1 [78]. The evidence is more consistent that physical activity increases levels of IGF-BP resulting in a decrease in overall IGF-1 bioavailability and activity [99, 102]. Greater levels of physical activity could result in lower levels of endogenous sex hormones via any number of metabolic events including: reduced insulin resistance, reduced IGF-1, and increased production of SHBG. The overall effect is a lower risk of hormone-related cancers. These changes are also observed for reduced adiposity and may not be independently related to physical activity.

Inflammation. Adipokines (a variety of cytokines) are biologically active polypeptides secreted from white adipose tissue. A number of adipokines are considered indicators of inflammation including: leptin, adiopenctin, tumor necrosis factor alpha (TNF-α), and interleukin-6 (IL-6) [104]. Inflammation is a complex biological response of vascular tissues to harmful stimuli (i.e. invading pathogens) and a necessary step in the initiation of wound

healing. Pro-inflammatory cytokines TNF-α and interleukin-1-beta (IL-1β) are produced at the infection site in response to leptin. These markers stimulate the release of IL-6 and collectively activate C-reactive protein (CRP) during acute phase response. Following clearance signal transduction of IL-1β and TNF- α are blocked by the release of IL-1 receptor antagonist and soluble TNF- α receptors [99]. Chronic inflammation is a condition that results in an increase in the circulating levels of pro-inflammatory markers TNF-α, IL-6, and leptin (often highly expressed in adipose tissue) and a decrease in anti-inflammatory markers, such as adiponectin, which are inversely correlated with adiposity [105]. Chronic inflammation is not only a risk factor for obesity, but has been associated with several other conditions including metabolic syndrome [106], type 2 diabetes [107], and some cancers [108, 109]. Increases in cancer risk due to chronic inflammation may occur because of changes in the microenvironment, increased proliferative activity, and oxidative stress [110]. These factors are likely to work together to deregulate normal cell development thus increasing the propensity for malignancy.

While obesity has clear mechanistic associations with inflammation it has been suggested that regular exercise may reduce inflammation independent of weight loss. Studies support an inverse association between chronic physical activity and inflammatory markers CRP, TNFα, and IL-6 [44, 111, 112]. One study reported that while there are small increases of pro-inflammatory markers, a surge of cytokine inhibitors are initiated following exercise [113]. In contrast to acute infection the cytokine response to exercise does not involve amplification of pro-inflammatory cytokines. Rather, it is initiated by IL-6 and followed by an increase in anti-inflammatory markers thereby decreasing the likelihood of low grade chronic inflammation. Although these observations are indicative of an independent effect of exercise on inflammation, not all intervention studies of exercise have shown reductions in these inflammatory markers [44].

Immune Function. It is hypothesized that the immune system aids in cancer risk reduction by recognizing and eliminating abnormal cells [114]. It is also probable that the immune system modulates susceptibility to tumor formation by hindering cell growth or by counteracting the effects of tumor growth promoters [87]. Regular moderate physical activity may therefore reduce risk of cancer by stimulating components of the innate immune response. Studies indicate that in response to moderate physical activity a number of innate immune parameters (i.e. macrophages, monocytes, lymphocytes, neutrophils, eosinophils, killer cells, and acute-phase proteins) increase in function and/or quantity [78, 115, 116]. These effects are transient, with levels of immune parameters dropping below pre-exercise levels following activity [116]. The improved immune function that is hypothesized to occur with physical activity may vary by exercise type, intensity, or duration. An inverted 'J'-shaped dose-response relationship between intensity of physical activity and immune function is frequently reported with the greatest benefit occurring among individuals who undertake regular moderate exercise [44, 87]. Depression of immune function may be induced by excessive exercise at high intensity levels [44]. While there appears to be evidence for the immune system contributing to the association between physical activity and breast cancer susceptibility, this relationship is largely untested and requires further investigation.

Other Mechanisms. The physical activity-breast cancer mechanisms described above primarily involve targets of adiposity and weight loss. It is likely that many of these mechanisms work synergistically acting through a common obesity related pathway, but even early epidemiologic studies show that upon control for body weight an effect for physical activity persists [47, 48]. Similarly, animal models have shown that the negative energy balance induced by exercise does not, alone, explain the cancer preventive effects of physical activity [78]. These results suggest that while physical activity and weight reduction are strongly linked, each is likely to confer an independent benefit to reduce breast cancer risk. In addition to being targets of obesity, the proposed mechanisms are primarily

related to the promotion and progression of postmenopausal breast cancer. Physical activity may, however, influence risk at multiple points along the cancer continuum (**Figure A.3**). Exercise may influence the development of cancer by decreasing the rates of genetic and epigenetic alterations or by shifting the equilibrium of growth and death in cancer cells.

There have been few attempts to disentangle the effect of exercise from those of energy balance. While obesity related pathways are biologically plausible and likely to influence the physical activity-breast cancer association, other pathways important in carcinogenesis should be considered. Some investigators propose that physical activity may work through mechanisms further upstream including pathways related to early stages of malignant transformation [74, 75]. Markers associated with oxidative stress and DNA repair have been shown to be pertinent to breast cancer, but they may also be modifiable by exercise [42, 74, 83]. To date there are no epidemiologic studies which have evaluated the biologic plausibility of these pathways.

1.2.4 Conclusions

It is well established that physical activity reduces the risk of breast cancer. While the epidemiologic literature shows decreases in risk from all types of physical activity, moderate recreational activity appears to have the strongest association with risk reduction. Similarly, activities done throughout the lifetime have been consistently associated with breast cancer risk reduction compared to activities performed around the time of diagnosis [42]. Several mechanisms have been proposed to mediate the association between physical activity and breast cancer risk. They include sex steroid hormones, insulin resistance, growth factors, inflammation, and immune function. While these mechanisms are biologically plausible and likely to contribute to the inverse association between physical activity and breast cancer risk, they are greatly influenced by body weight – a consistent risk factor for postmenopausal breast cancer. Other pathways, independent of obesity, should be considered in the physical activity-breast cancer paradigm. The remaining sections will

explore the role of oxidative stress and DNA repair in the physical activity-breast cancer association

1.3 Oxidative Stress

Reactive oxygen species may be generated through any number of endogenous or exogenous processes. The term ROS is commonly used to describe certain reactive oxygen metabolites containing unpaired electrons in their outer orbit [117, 118]. ROS includes free radical derivatives of molecular oxygen (e.g. superoxide radical $O_2^{\bullet-}$, hydrogen peroxide H_2O_2, hydroxyl radical $^{\bullet}OH$) which are continuously generated within a cell as a result of oxidative metabolism [119, 120]. ROS may be produced by other endogenous processes such as estrogen metabolism, peroxisomes activity, or inflammatory cell activation. ROS and ROS-generating compounds are ubiquitous in the environment commonly found in inhaled smoke, alcohol, and ingested goods [121, 122]. While modest levels of ROS are useful for cell signaling processes [123] excess ROS may result in DNA damage, lipid peroxidation, and protein modification [121, 124-126]. These changes are known as oxidative stress, a term used to refer to the global burden of harmful reactive biochemical species present in tissue as a consequence of the regular cellular oxidative metabolism of endogenous and exogenous compounds [127].

1.3.1 Endogenous Responses to Oxidative Stress

When endogenous or exogenous ROS production occurs in an environment with sufficient *in vivo* defense mechanisms to scavenge the ROS, there are seemingly few harmful effects. When there is excess ROS production or insufficient defense mechanisms, oxidative stress may ensue. There are several antioxidant defenses that can protect against increases in lipid peroxidation counteracting oxidative damage. Enzymes responsible for neutralizing ROS endogenously include catalase (CAT), manganese superoxide dismutase (MnSOD), glutathione peroxidase (GPx), as well as glutathione *S*-transferases (GST) [127].

These enzymes form the first line of defense against superoxide and hydrogen peroxide. Secondary defenses include reduced *GPx* and *GST* [128, 129], which play a central role in the defense against free radicals, peroxides, as well as a wide array of xenobiotics and carcinogens [130]. If peroxide or other free radical derivatives of molecular oxygen are not neutralized by the above mechanisms, they may contribute to additional ROS generation by myeloperoxidase (*MPO*) [131]. *MPO* generates ROS endogenously performing as an antimicrobial enzyme, catalyzing a reaction between H_2O_2 and chloride to generate hypochlorous acid [132]. Hypochlorous acid further reacts with other biological molecules to generate secondary radicals [133]. Endothelial nitric oxide synthase is also responsible for the generation of reactive nitrogen species (RNS), catalyzing the production of the NO radical [134]. The levels of potentially cytotoxic reactive species within the body may therefore depend on the balance between endogenous pro- and anti-oxidants.

1.3.2 Biologic Plausibility of the Oxidative Stress-Breast Cancer Association

Oxidative stress may play an important role in the risk of many chronic diseases including cardiovascular disease, aging, and human cancer [135-138]. The function of oxidative stress in carcinogenesis has been widely demonstrated in small human and animal studies, providing increasing evidence that it may be involved in the pathophysiology of breast cancer [127, 139-141]. Oxidative damage has been suggested to contribute to the formation of DNA adducts [142] and has been implicated in neoplastic transformation [143]. Oxidative damage has frequently been reported to be higher among women with breast cancer, compared to controls [127]. Similarly, studies have observed enhanced lipid peroxidation in breast tumor tissue compared to uninvolved adjacent tissue. Elevations in both enzymatic (SOD, CAT, GPx, GST) and nonenzymatic (GST) antioxidants in tumor cell lines have been observed in some [126, 144-146] but not all [126, 147] studies.

Although the exact mechanisms remain to be elucidated it is likely that oxidative damage to DNA results in genetic mutations or alterations in gene expression [148]. Genetic

24

variants that influence pro or anti-oxidant mechanisms may therefore play an important role in breast cancer carcinogenesis [127]. CAT is a heme enzyme that has a primary role in neutralizing ROS by converting H_2O_2 into H_2O and O_2 [127]. Mouse models have shown that acatalasemic mice (C3H strain), with approximately one tenth the CAT blood and tissue levels of normal mice, were more susceptible to spontaneous mammary carcinoma after feeding on regular laboratory chow for 15 months [149]. While these studies are limited to the animal literature such associations could be assessed by evaluating the impact of catalase activity (using genotype as a proxy) on breast cancer risk.

1.3.3 Epidemiology of Catalase and Breast Cancer Risk

Activity levels of the CAT enzyme is likely affected by functional polymorphisms in the gene. A common 262 C/T polymorphism (rs1001179) has been identified in the promoter region of the human *CAT* gene [150]. Endogenous variability associated with this SNP may play an important role in host response to oxidative stress. Studies have shown that the variant *CAT* allele (T) is associated with hypertension [151] and vitiligo [152], both conditions related to oxidative stress. The 262 C/T polymorphism has been shown to affect the transcriptional activity of the promoter [150] and is thought to result in reduced enzyme activity [153-155]. Using a sample of 420 controls Ahn and colleagues examined the functional effects of this variant on catalase enzyme activity. They found a dose-response reduction in activity by *CAT* genotypes, with geometric means of 115.4, 82.1, and 73.5 units/mg hemoglobin for CC, CT, and TT genotypes, respectively. The reported % activity difference for CT genotype versus CC genotype was 28.8% (P = 0.002) and was 36.3% (P = 0.02) for TT genotype versus CC genotypes [153]. These associations were also observed in a smaller, yet independent, study population [156].

There have been few reports of the association between the catalase-262 C/T polymorphism and breast cancer risk in the epidemiologic literature. Investigators from the LIBCSP found that women with the common CC genotype had a 17% reduction in risk of

25

breast cancer (95% CI: 0.69, 1.00) compared to women with at least one T allele (CT and TT genotypes) [156]. Follow-up studies reported approximately null associations between *CAT* genotypes and breast cancer risk [157, 158], although there were less participants in both studies (N=569 and N=505 for Quick et al., 2008 and Li et al., 2009, respectively). While no other *CAT* variants have been reported in the epidemiologic literature, other polymorphisms may be independently associated or act, in combination, with this functional variant to influence breast cancer incidence.

1.3.4 Oxidative Stress and Physical Activity

Several risk factors associated with breast cancer (i.e. alcohol consumption, cigarette smoking, exposure to environmental tobacco smoke, and reduced estrogen metabolism) may exert their harmful effects via generation of ROS [144, 159-161] while others (i.e. consumption of fruits vegetables and antioxidants) are suggested to oppose ROS formation [127] therefore reducing the risk of breast malignancy. Several studies have examined the association between ROS related risk factors, genetic variants and breast cancer risk – hypothesizing that ROS generating risk factors act, in combination, with reduced antioxidant expression to increase the risk of breast cancer. Ahn and colleagues reported that current smokers with at risk *GSTA1* polymorphisms (B/B genotypes) had a 1.89-fold increase in risk (OR=1.89; 95% CI=1.09–3.25), compared with never smokers with the common A/A genotype [162]. Among ever users of HRT the increased risk of breast cancer was more pronounced among women with variant CT or TT *CAT* genotypes (OR=1.88; 95% CI=1.29-2.75) than among women with the CC genotype (OR=1.15; 95% CI=0.86-1.54) although *CAT* genotype alone was not associated with breast cancer risk [157]. Exposures known to oppose ROS generation have been shown to obliterate the increased risk associated with genotype. Among women with the at risk B/B *GSTA1* genotype, Ahn et al. observed an inverse trend between cruciferous vegetable consumption and breast cancer risk (P for trend = 0.05) [162]. Similarly, high fruit consumers with the

common *CAT* genotype (CC) have been shown to have the lowest risk of breast cancer (OR=0.59; 95%CI=0.38-0.89). Women who were low consumer/common genotype or high consumer/variant genotype had OR's (95%CI's) = 0.94 (0.65-1.37) and 1.06 (0.66-1.73), respectively [156]. These data indicate that genotypes related to reduced antioxidant expression may be associated with increased breast cancer through risk factors that increase ROS generation.

Studies have shown that physical activity is a strong inducer of lipid peroxidation and free radicals [119, 163, 164]. While the dimensions of exercise-induced ROS production are unknown, there is some evidence that strenuous activity [165-167], endurance and resistance training [168] increase lipid peroxidation. Three mechanisms of exercise induced ROS production have been proposed [169]. One mechanism is via an electron 'leak' of the mitochondrial electron transport chain. During activity whole body oxygen consumption increases 10 to 20-fold and local muscle consumption 100 to 200-fold. While the majority of oxygen binds to hydrogen transforming it to water through the electron transport chain, an electron leak at the ubiquinone–cytochrome b level may result in the formation of superoxide radicals. Thus, as oxygen consumption increases during exercise there is a parallel increase in free radical production and lipid peroxidation. Another possible mechanism is ischaemia-reperfusion. During exercise, blood flow is restricted in many organs and tissues (e.g. kidneys, splanchnic region) to increase blood supply to the working muscles. As a result, the regions with restricted blood flow may experience a hypoxic state which heightens as the intensity of exercise increases. At the cessation of exercise these regions undergo re-oxygenation, which may lead to a burst of ROS production typical of ischemia-reperfusion. A third mechanism is auto-oxidation of catecholamines, whose levels are amplified many-fold during exercise leading to increased oxidative stress.

Regular physical activity is also known to stimulate endogenous antioxidants as a physiological response to the oxidative stress [170, 171]. Up-regulation of antioxidant

27

enzymes may render cells more resistant to subsequent oxidative insult [172] thereby

neutralizing the potentially mutagenic effects of lipid peroxidation [173]. Regular exercise

has been shown to enhance antioxidant status at multiple levels in both animal models and

clinic studies. Rodent studies showed increased levels of hepatic CAT [174-176] with

exercise training. Amplified levels of hepatic SOD [175, 176] and cytosolic SOD activity [174]

have also been documented. Clinical studies showed that among runners, there is a positive

correlation between exercise training and GST [171], GPX, as well as CAT activity [177]. In

addition to changes in antioxidant enzymes, exercise induced oxidative damage may also

lead to increased cellular apoptosis providing yet another potential mechanism driving the

inverse association between physical activity and breast cancer [178].

1.3.5 Conclusions

ROS induced oxidative damage generates products that have the potential to react

with DNA, which may lead to mutations in proto-oncogenes and tumor-suppressor genes.

These changes (if unrepaired) could result in the transformation of normal epithelium to a

malignant phenotype [139, 140] making it a potentially important contributor to the etiology

of breast cancer. Levels of oxidative stress in the body are ultimately determined by

variability in exposure to endogenous or exogenous factors that could increase ROS, as well

as cellular response to ROS [127]. Catalase and other antioxidant enzymes are responsible

for neutralizing free oxygen derivatives. Enzymatic levels have been shown to vary by

malignancy status in breast tissue which lends support to the oxidative stress-breast cancer

hypothesis. This hypothesis may be further tested by examining the association between

inherited genetic variants in *CAT* and breast cancer risk, as studies have shown a dose-

response reduction in activity with increasing number of variant alleles.

While the immediate systemic response to physical activity is an increase in ROS

production, the lasting effect of regular endurance training is adaptation of antioxidant

capacity. Increased antioxidant capacity may protect against the unfavorable effects of

oxygen free radicals and prevent oxidative damage. These changes seem to occur after moderate to exhaustive exercise, which parallels current knowledge of the inverse association between physical activity and breast cancer. It is therefore plausible that lipid peroxidation and subsequent increases in antioxidant capacity are important mechanisms for physical activity. This could be assessed by examining the extent to which antioxidant genotypes in multiple oxidative stress related genes modify the effect of physical activity on breast cancer risk.

1.4 DNA Repair

Models of causation are important to distinguish epidemiological risk factors and associations with disease. Cancer research dating back to the 1920s has shown that tumor initiation begins with DNA alterations resulting from inherent, spontaneous, or carcinogen induced genetic changes. These changes may lead to DNA damage and mutations that, in the absence of apoptosis, may propagate through the genome leading to unregulated cell growth. One mechanism of DNA damage is through the generation of ROS [179] which may damage DNA both directly and indirectly by forming lipid peroxidation products. Several types of oxidative DNA damage have been identified [125], but the most common oxidative stress-induced DNA lesion is 8-oxodeoxyguanosine (8-oxo-dG) [122, 180]. These lesions are excised exclusively by 8-oxoguanine DNA glycosylase, the enzyme encoded by DNA repair gene 8-Oxoguanine glycosylase (*OGG1*) [180].

DNA integrity may be compromised by both endogenous and exogenous processes, but damage may be prevented or repaired through any number of innate defenses including: neutralization by endogenous antioxidants, detoxification of reactive metabolites, apoptosis, cell cycle arrest, and DNA repair [181]. If these mechanisms are insufficient however, uncorrected oxidative damage may ensue, contributing to the formation of somatic mutations in tumor suppressor genes or proto-oncogenes [182]. Persistent mutations in critical genes may result in genomic instability and ultimately, carcinogenesis [183].

Polymorphic sites in DNA repair pathway genes are therefore strong candidates for cancer susceptibility genes [184].

DNA is primarily maintained by repair mechanisms, which recognize, excise, and replace damaged nucleotides. The mechanism of repair is contingent on both the structure of the damage and its location within the genome. There are at least four repair pathways (**Table A.2**) that operate on damaged DNA: double strand break (DSB) repair, base excision repair (BER), nucleotide excision repair (NER) and mismatch repair (MMR) [181, 185]. Oxidative DNA damage is repaired mainly through the BER pathway [181]. However, research suggests that oxidative stress may lead to inactivation of both NER [186] and MMR activity [187]. The following text will give an overview of the DNA repair pathways, summarize associations between DNA repair variants and breast cancer, provide biologic rationale for exploring the role of MMR in breast carcinogenesis, as well as review the evidence for the association between DNA repair and physical activity.

1.4.1 DNA Repair Pathways

Double Strand Break Repair. Double-strand breaks can be produced by exogenous agents, including ionizing radiation, chemotherapeutic agents, endogenously generated ROS and replication errors [188]. Double strand breaks are among the most dangerous type of DNA damage, as the unavailability of a damage-free template makes repair more difficult [181]. There are two main DSB repair pathways: homologous recombination (HR) and non-homologous end joining (NHEJ) [189]. The HR pathway involves greater than 15 molecules including the breast and ovarian cancer genes *BRCA1* and *BRCA2*, as well as x-ray repair cross-complementing protein (*XRCC)1*, *XRCC2*, and *ATM*. DNA-dependent protein kinase A is an essential component of the NHEJ pathway, although other molecules (e.g. Ligase IV) are important [188]. In HR the damaged ends are removed and strands extended using a homologous sequence to guide repair. NHEJ is not contingent upon homologies between the two recombining ends and repairs damage by directly ligating the broken ends –

30

potentially resulting in more error [190]. NHEJ additionally has two known sub-pathways: classic NHEJ (described above) and alternative NHEJ [191]. Little is known about the mechanisms and factors involved in alternative NHEJ.

Base Excision Repair. BER is essential to repairing oxidative DNA damage and other small lesions such as those produced by methylating agents and nonbulky adducts [181]. Repair is initiated by DNA glycosylases, which recognize and remove damaged nucleotides by cleavage of the N-glycosylic bond between the base and its associated deoxyribose group, generating an apurinic/apyrimidinic (AP) site [192]. The AP site is further processed by AP endonuclease which converts the base lesion into a single strand break. The break is repaired via DNA synthesis and ligation which occurs along one of two pathways: short-patch BER (replacement of a single nucleotide) or long-patch BER (synthesis of multiple nucleotides) [193, 194]. Each pathway necessitates its own set of enzymes. The short-patch BER pathway requires four proteins: AP endonuclease (APE1), DNA polymerase β, and DNA ligase III/XRCC1 heterodimer. The long-patch BER pathway involves six: APE1, replication factor C, proliferating cell nuclear antigen, flap endonuclease 1, DNA polymerases δ/ε, and the DNA ligase I [192, 194, 195].

Nucelotide Excision Repair. The NER pathway repairs bulky lesions such as pyrimidine dimmers and other UV-induced photoproducts, oxidative damage, cross-links, and chemical adducts [181, 189, 196, 197]. This pathway is essential in removing UV-induced DNA damage as evidenced by *Xeroderma pigmentosum* (XP), an autosomal recessive genetic disorder characterized by extreme sensitivity to sunlight that results from germline mutations of NER proteins [198, 199]. The NER pathway consists of, at minimum, four major steps: damage recognition, unwinding of DNA, removal of the damaged fragment, and synthesis of DNA (reviewed in [189]). Damage is recognized by a protein called XPC, which is bound to the protein HHRAD23B (R23). Together they form the XPC–HHRAD23 heterodimeric subcomplex. Several other proteins bind to the complex, assisting

31

with base damage recognition (XPA and replication protein A), unwinding of DNA (transcription factor IIH (TFIIH)), and excision (XPG). Six subunits containing two DNA helicase activities (XPB and XPD) make up TFIIH and are responsible for unwinding DNA near the site of damage. Binding of the ERCC1–XPF heterodimeric subcomplex produces the final NER multiprotein. Excision occurs at junctions both 3' and 5' to the site of base damage by XPG and the ERCC1–XPF complex, respectively. The resulting oligonucleotide fragment is excised from the genome. DNA repair synthesis involved polymerases δ or ε, a number of replication proteins, and DNA ligase [189, 200].

Mismatch Repair. The MMR genes repair incorrect pairings of nucleotide bases (base-base or insertion-deletion mismatches) which may occur as a result of genetic recombination, replicative errors in DNA polymerase, deamination of 5-methylcytosine to thymine, or environmental mutagens [201, 202]. MMR is essential for maintaining genomic stability as its proteins have been shown to encourage cytotoxicity [203], p53 phosphorylation [204], cell-cycle arrest [205], and cell death [204, 206, 207] in DNA damaged cells. Loss of MMR function thwarts the correction of replicative errors leading to genomic instability. These changes can be detected by polymorphisms in microsatellites which are characterized by repeated regions of one to six nucleotide units scattered throughout the genome. Microsatellite instability (MSI) is a hallmark of MMR dysfunction in colorectal cancer and other malignancies (discussed below) and occurs during replication when the two strands of DNA become misaligned, resulting in small loops of unpaired DNA [208, 209]. Loss of MMR function may occur because of mutations in one of six genes associated with MMR: *MLH1, MLH3, MSH2, MSH3, MSH6* and postmeiotic segregation increased 2 (*PMS2*). These genes make up the MutS and MutL homologue proteins involved in MMR (**Figure A.4**). MutS homologues include MSH2-MSH6 (MutSα) and MSH2-MSH3 (MutSβ). It is suggested that MutSα is responsible for the repair of base:base mispairs and MutSβ is responsible for insertion/deletion mispairs [210]. The MutL

homologues include MLH1-PMS2 (MutLα) and MLH1-MLH3 (MutLβ). Once bound to the

mismatch, MutS associates with its complementing MutL heterodimeric complex. The

associated complex exchanges bound adenosine diphosphate (ADP) from adenosine

triphosphate (ATP) resulting in a conformational change and the formation of a clamp that

translocates along DNA. Proliferating cell nuclear antigen (PCNA), interacting with MSH3

and/or MSH6, corrects the strand that retains the primer. The mismatch strand is

subsequently degraded by a 3->5 exonuclease and re-synthesized by DNA polymerase γ

and PCNA [201].

1.4.2 Epidemiology of DNA Repair and Breast Cancer

Approximately 130 human genes are involved in the four DNA repair pathways previously

described, each playing an important role in the maintenance of genomic integrity [211].

Knowing that reduced DNA repair may lead to genetic instability and carcinogenesis, genes

involved in these pathways may potentially serve as candidate cancer-susceptibility markers

[181, 183]. In a 2000 review by Berwick and Vineis, investigators cite consistent

associations (OR range 2.0-10.0) between DNA repair capacity and cancer occurrence

[183]. Reductions in repair capacity are likely to be associated with functional polymorphic

sites in DNA repair genes.

A comprehensive review of all studies examining associations between DNA repair

polymorphisms and risk of several types of cancers was conducted by Goode and

colleagues [181]. By April 2002 investigators had identified 30 published studies of adult

glioma, bladder cancer, breast cancer, esophageal cancer, lung cancer, prostate cancer,

skin cancer (melanoma and nonmelanoma), squamous cell carcinoma of the head and

neck, skin cancer, and stomach cancer and the following DNA repair variants: *OGG1* and

XRCC1 (BER genes); DNA excision repair protein (*ERCC*)*1,* xeroderma pigmentosum

(*XP*)*C, XPD,* and *XPF* (NER genes); and *BRCA2* and *XRCC3* (DSB repair genes) [181]. At

the time of the Goode 2002 review only *XRCC1* and *BRCA2* variants had been evaluated in

association with breast cancer incidence. A single study of the Arg194Trp polymorphism in *XRCC3* showed a reduced risk of breast cancer among study participants with Trp (OR=0.7, 95%CI: 0.4, 1.3 for at least one Trp allele vs. Arg/Arg) [212]. The Asn372His polymorphism of *BRCA2* was associated with breast cancer risk in several studies. The combined odds of breast cancer among women with the His/His genotype was 1.3 (95% CI: 1.1, 1.6) times the odds of breast cancer among women with Asn/Asn genotype [181].

Since the Goode 2002 review, a number of studies have examined the effects of DNA repair polymorphisms in breast cancer incidence. There are mixed results for the function of DSB repair genes in the etiology of breast cancer [213], but the vast majority of the evidence is supportive of an association [214-216]. Among all DSB polymorphisms only one in *XRCC3* is consistently associated with breast cancer susceptibility, although its effect is likely small [215]. Epidemiologic studies conducted among Caucasian women have not supported a role for genetic variations in the BER pathway and breast carcinogenesis [217, 218]. However, current data suggest that associations may be positive and more stable when considering specific ethnic groups including Asian [217, 219] and Indian [220, 221] populations. Variants in NER pathway genes have been most readily studied in both individual epidemiologic investigations [216, 222-229] and meta analyses [230-232]. The literature, overall, is inconsistent which may reflect differences in study populations, low statistical power or other methodological issues.

There remains a dearth of literature examining polymorphisms in MMR and breast cancer outcomes with only five published studies to date [225, 228, 233-235]. Poplawski and colleagues examined two common polymorphisms of *MSH2*: an A → G transition at 127 position producing an Asn → Ser substitution at codon 127 and a G → A transition at 1032 position resulting in a Gly → Asp change at codon 322. Both polymorphisms are capable of changing the biological properties, structure, and function of the MSH2 protein [233]. Study results showed a strong, although imprecise, association between the G/G genotype of the

Gly322Asp polymorphism and breast cancer occurrence (OR=8.39; 95% CI: 1.44, 48.8) as well as an inverse association with the G/A genotype (OR=0.13; 95% CI: 0.02, 0.83) [233]. The associations with Asn127Ser were imprecise and not statistically significant. In a hospital-based sample of Korean women Lee and colleagues reported that the risk of breast cancer increased in a dose response manner with the number of G alleles in the *MLH1* –93 G>A polymorphism (OR=1.33; 95% CI=0.81-2.19; OR=2.24; 95% CI=1.21-4.17, respectively; p for trend = 0.01). This effect was observed only among postmenopausal women, and upon correcting for multiple comparisons the associations were no longer significant [225]. A 2008 case control study reported a decrease in breast cancer risk in carriers of the variant alleles in the Ile219Val polymorphism of *MLH1* (OR=0.49 95% CI=0.29-0.85) [228]. While several other polymorphisms were examined (*MSH3*: Arg940Gln, Thr1036Ala and *MSH6*: Gly39Glu) no other statistically significant associations emerged among Caucasian women. The most recent study [234], conducted in a hospital setting among Portuguese women, estimated the effects of polymorphism in several MMR genes (*MSH3, MSH4, MSH6, MLH1, MLH3, PMS1* and mutY homologue [*MUTYH*]). The Leu844Pro, polymorphism of *MLH3* was found to be associated with breast cancer incidence. The OR's (95% CI's) for Leu/Leu and Pro/Pro genotypes were 0.65 (0.45-0.95) and 0.62 (0.41-0.94), respectively. The most recent study [235] found an inverse association between *MSH3* rs6151838 and breast cancer (OR=0.73; adjusted p=0.048) as well as an inverse association between *PMS1* rs5743030 and PR+ breast cancer (OR=0.61; adjusted p=0.047). Collectively, these data show that putative at risk alleles in MMR may be associated with breast cancer outcomes. However, these studies are plagued with small samples and less than optimal design schemes making additional studies warranted.

1.4.3 Biologic Plausibility of the Mismatch Repair-Breast Cancer Association

Available evidence indicates that that a considerable fraction of breast tumors show instability in sequence motifs of mono- and di-nucleotide repeats [236-240]. This

phenonema is known as MSI and has been shown to be closely associated with MMR deficiency [181, 241]. These deficiencies occur frequently in hereditary non-polyposis colorectal cancers (HNPCC) [242-245] as well as other MSI related cancers including some sporadic colorectal tumors, hematological malignancies, endometrial, prostatic, and gastric cancers [202, 246-248]. In comparison to HNPCC syndrome the role of MSI and MMR gene dysfunction in breast cancer development is less well established.

MMR gene dysfunction is proposed to occur through two mechanisms: epigenetic gene silencing through hypermethylation and genetic mutations in MMR genes [249, 250]. These changes may lead to increased mutations of oncogenes, tumor suppressor genes, and loss of DNA damage-induced apoptosis, therefore facilitating carcinogenesis [251]. While several MMR genes are associated with cancer predisposition, *MSH2* and *MLH1* genes are central to all mismatch recognition and alterations in them have been shown to be the most common mechanism inducing cancer-related MSI [252-254]. In a study of 32 sporadic breast tumors Murata and colleagues identified MSI in approximately 47% of samples. Greater than 90% of MSI tumors showed reduced protein expression of MSH2 (N=2) or MLH1 (N=5) and approximately 50% had genetic alterations is both genes [255]. While these findings indicate high correlations between MMR dysfunction and MSI in breast cancer, not all studies observed this relationship. The rates of MSI in sporadic breast cancer have been shown to vary greatly between studies [240, 255, 256]. Chintamani and colleagues report a range from 5-30% [202]. This variation is most likely due to differences in microsatellie markers used for analysis. It may alternatively indicate that progression differs, mechanistically, in primary breast cancer compared to HNPCC [202]. For example, the lack of a correlation between MMR loss and MSI in breast cancer may be due to involvement of MMR genes that do not induce MSI or potential interactions within population subgroups.

MMR gene expression also appears to be associated with clinicopathological parameters of breast cancer. There is accumulating evidence that reduced expression of *MSH2* and *MLH1* are related to tumor progression and invasion [202, 255, 257] although some studies report null effects [257, 258]. Although modest, the evidence to date indicates that breast cancer is associated with MSI, MSI is primarily caused by variations in MMR genes, and polymorphisms in MMR genes are potentially important modulators of breast cancer progression.

1.4.4 DNA Repair and Physical Activity

While few meaningful effects of DNA repair variants on breast cancer outcomes have been reported in epidemiologic studies it is plausible that these polymorphisms may more profoundly influence carcinogenesis by modifying the effect of environmental exposures on cancer risk [183, 184]. The effects of smoking [228, 229, 259-262]; PAH-DNA adducts [222]; radiation exposure [263, 264]; and dietary factors [259, 260, 265, 266] have each been shown to be modified by SNPs in DNA repair genes. I identified no epidemiologic studies which examined associations between DNA repair variants and breast cancer with respect to physical activity.

Physical activity has been shown, repeatedly, to increase the formation of reactive oxygen and nitrogen species [119, 163, 164], which may influence carcinogenesis. Regular exercise training may also result in improved damage repair systems [170, 171, 267, 268]. As early as 1999, Radak and colleagues showed that exercise decreased the degree of oxidative damage in lipids, proteins, and DNA (namely 8-oxo-dG) in the skeletal muscle of trained rats [269]. The investigators hypothesized that the observed reduction in oxidative damage could be attributed to the increased regulation of repair systems [267]. A 2002 follow up investigation reported similar findings [270]. In addition to observing a reduction in the number of skeletal 8-oxo-dG lesions in exercised rodents, investigators confirmed increases in 8-oxo-dG repair which was measured by the nicking of a ^{32}P-labeled damaged

oligonucleotide. This study also found increased chymotrypsin-like activity of the proteasome complex, a repair enzyme important in the degradation of proteins modified by oxidative stress [270].

Much like the animal literature, clinical studies show significantly reduced levels of basal 8-oxo-dG among physically active study participants compared to individuals classified as sedentary [268](Sato 2003). Sato and colleagues [268] reported significant reductions in the 8-oxo-dG levels of sedentary participants following 30 minutes of mild exercise. There was no change among active men. These data suggest that physical activity may reduce interim levels of 8-oxo-dG and that over time, sustained involvement in physical activity could result in systemic reductions of 8-oxo-dG. In addition to reductions of 8-oxo-dG, data from clinical studies of trained cyclists [271] and marathon runners [272] showed that DNA excision repair enzymes NESP and RAD23A [271], as well as OGG1 [272] are up-regulated with exercise training.

The extent of physical activity-induced DNA repair may vary based on exposure frequency, endogenous activation, detoxification, antioxidant capacity or other defense mechanisms. Although reductions in oxidative induced DNA damage could be an artifact of any of the above pathways, the animal and clinical evidence to date suggests that exercise could result in up-regulation of DNA repair enzymes and may be an important part of the exercise induced adaptation process.

1.4.5 Conclusions

While it has been well established that select DNA repair mechanisms are relevant to breast cancer incidence, few studies have examined the association with MMR variants. Previously conducted breast cancer studies have reported significant associations with minor alleles in MMR SNPs, which may vary by menopausal status. But results are inconsistent, are conducted among different international populations with varying genetic profiles, and are based on small, select samples. Additional investigations are of interest as

MMR greatly contributes to the overall fidelity of replication and genomic integrity. Moreover, there is an increasing amount of evidence suggesting that both MMR dysfunction and MSI are correlated with clinical markers of breast cancer.

Gene-environment interactions have been frequently examined in the breast cancer literature. Investigators have published several positive results for an interaction between DNA repair variants and cigarette smoking, radiation exposure, PAH-DNA adducts, dietary antioxidants, and fruit and vegetable consumption. While it has been recognized that physical activity may exert its effects via DNA repair, no study had considered a possible interaction with physical activity levels. Animal and clinical studies showed that DNA repair enzymes are up regulated with physical activity which is likely to result in diminished DNA damage. An epidemiologic model was tested by assessing the joint effects of low physical activity and reduced DNA repair capacity on breast cancer incidence.

1.5 Summary and Specific Aims

Breast cancer is the leading cause of global cancer incidence and mortality among women. Among US women it is the primary cause of, non-melanoma, cancer-related illness and is second to lung cancer in mortality. With almost half of women engaging in some type of physical activity, it is conceivably one of the most prevalent environmental exposures associated with breast cancer risk. Physical activity has been suggested to reduce the risk of breast cancer by at least 25%; however the inverse association seen with physical activity could be due to a healthy person effect and serve as a proxy for other healthy behaviors that are associated with breast cancer risk (e.g. low BMI and healthy diet). Identifying women who are particularly susceptible to the beneficial effects of physical activity based on genetic characteristics could aid in validating the biologic plausibility of this association. Moreover, it could facilitate the ability to elucidate the mechanism through which physical activity exerts its effects, and ultimately allow us to better tailor our public health messages.

Physical activity may intervene along multiple paths in the stages of carcinogenesis. It is likely to operate in obesity-related pathways (e.g. insulin resistance and hormonal pathways) that are related to both promotion and progression and as well as through pathways like oxidative stress and DNA repair that are more closely linked to initiation. There is a dearth of literature on how physical activity is modified by genetic variability in these latter pathways. Physical activity, a known inducer of ROS, may influence breast cancer risk by up-regulating antioxidant enzymes and overall antioxidant capacity. It could also impact breast carcinogenesis by improving damage-repair systems, particularly those that operate on ROS induced single-strand breaks including (e.g. BER, NER and MMR). While these mechanisms are not mutually exclusive, they may work to diminish DNA damage lowering the likelihood of cancer initiating events. Polymorphic sites in oxidative stress and DNA repair pathways were therefore strong candidates for cancer susceptibility genes and may modify the effects of exercise on breast cancer.

This dissertation research evaluated associations between novel polymorphisms in oxidative stress (*CAT*), and DNA repair (*MLH1, MSH2,* and *MSH3*) and breast cancer. Importantly, this research was the first to assess interactions between RPA, oxidative stress and DNA repair polymorphisms with respect to breast cancer risk. In addition to the SNPs evaluated for main effects, selected oxidative stress and DNA repair polymorphisms of interest for gene-physical activity interactions were those previously genotyped in the LIBCSP: *CAT* (-262C>T)*, COMT* (Val158Met and rs737865)*, GPX* (Pro198Leu)*, GSTA1* (rs3957396)*, GSTM1* (gene deletion)*, GSTP1* (Ile105Val)*, GSTT1* (gene deletion)*, MnSOD* (Val16Ala)*, MPO* (463G>A)*, ERCC1* (C8092A)*, MGMT* (Leu84Phe, Ile143Val, Lys178Arg)*, OGG1* (Ser326Cys)*, XPA* (4G>A)*, XPC* (Ala499Val, Lys939Gln)*, XPD* (Asp312Asn, Lys751Gln)*, XPF* (Arg415Gln)*, XPG* (Asp1104His)*,* and *XRCC1* (Arg194Trp, Gln399Arg). The study aims were defined by the biologic pathway under investigation:

AIM 1. Oxidative Stress

AIM 1A: to determine the main effect of select SNPs in *CAT* on breast cancer risk.

AIM 1B: to explore the potential interactions between SNPs in *CAT* and breast cancer risk (SNP-SNP interactions).

AIM 1C: to evaluate interactions between polymorphisms in oxidative stress genes (*CAT, COMT, GPX, GSTA1, GSTM1, GSTP1, GSTT1, MnSOD and MPO*) and self-reported recreational physical activity on breast cancer risk (gene-environment interactions)

AIM 2. DNA Repair

AIM 2A: to determine the main effect of select SNPs in three candidate genes (*MLH1, MSH2, and MSH3*) of the mismatch repair (MMR) pathway on breast cancer risk.

AIM 2B: to explore potential interactions between genes in the MMR pathway (*MLH1, MSH2, and MSH3*) and breast cancer risk (gene-gene interactions).

AIM 2C: to evaluate interactions between polymorphisms in DNA repair genes (*ERCC1, MGMT, MLH1, MSH2, MSH3, OGG1, XPA, XPC, XPD, XPF, XPG and XRCC1*) and self-reported recreational physical activity on breast cancer risk (gene-environment interactions)

These aims were accomplished through the analysis of extant data from the LIBCSP; a population-based case-control study developed to investigate the association between environmental factors and breast cancer risk in Long Island, NY. This dissertation employed pre-existing physical activity and biomarker data, as well as newly genotyped SNP data to examine the aforementioned aims.

Chapter II: Materials and Methods

2.1 Study Rationale, Research Aims and Hypotheses

Breast cancer is the most common malignancy and the second-leading cause of cancer-related death among women in the United States. Physical activity is known to reduce the risk of breast cancer, particularly among post-menopausal women, and may be a modifiable protective factor important for intervention. The mechanisms driving the observed inverse association remain unresolved. While much attention has been given to obesity-related pathways in the attempt to understand the molecular effects of physical activity on breast cancer risk, strong evidence is lacking. It is well documented that exercise may work through obesity driven pathways but independent and review studies alike suggest that these mechanisms may not fully account for observed association with breast cancer risk. The numerous physiologic consequences of physical activity necessitate a more complete understanding of the mechanisms driving the inverse association. Given exercise is a known inducer of ROS and lipid peroxidation, polymorphisms in oxidative stress or DNA repair may modify the exercise-breast association contributing to disease risk.

To my knowledge this dissertation was the first investigation to explore the potential modifying effects of DNA repair and oxidative stress variants on physical activity. While these pathways have been well examined in their association with breast cancer risk, and modifying effects assessed among a number of other environmental modulators, their impact on physical activity had not previously been evaluated in the epidemiologic literature. The aims of this ancillary study were therefore two-fold. I first estimated effects of several novel SNPs in oxidative stress and DNA repair pathways using the resources of a large population-based case-control study: the Long Island Breast Cancer Study Project. This

research could increase understanding of breast cancer etiology with respect to oxidative stress and DNA repair pathways. I subsequently examined whether the effect of physical activity on breast cancer risk was modified by individual variability in the genetic variants of these pathways. Understanding the association between physical activity and breast cancer has important public health implications due to the high incidence of breast cancer and the relative ease of access to physical activity in the US general population. Interactions between physical activity, oxidative stress and DNA repair polymorphisms may help elucidate underlying mechanisms linking exercise to breast cancer. Further, it could aid in identifying subgroups of women who are particularly susceptible to the beneficial effects of physical activity, based on genetic characteristics, both strengthening the biological plausibility of this relationship and informing public health recommendations for lowering breast cancer risk.

The specific aims of this project were outlined in section 1.5 and are as follows:

AIM 1A: to determine the main effect of select SNPs in *CAT* on breast cancer risk

Hypothesis: genotypes related to reduce antioxidant capacity will be positively associated with breast cancer

AIM 1B: to explore the potential interactions between SNPs in *CAT* and breast cancer risk

Hypothesis: joint effects of risk genotypes will be greater than multiplicative.

AIM 1C: to evaluate interactions between polymorphisms in oxidative stress genes and self-reported recreational physical activity on breast cancer

Hypothesis: genotypes related to reduce antioxidant capacity will antagonistically reverse the beneficial effects of high physical activity.

AIM 2A: to determine the main effect of select SNPs in three candidate genes (*MLH1, MSH2, and MSH3*) of the mismatch repair (MMR) pathway on breast cancer risk

Hypothesis: genotypes related to reduce DNA repair capacity will be positively associated with breast cancer

AIM 2B: to explore potential interactions between genes in the MMR pathway (*MLH1, MSH2, and MSH3*) and breast cancer risk

Hypothesis: joint effects among risk genotypes will be greater than multiplicative

AIM 2C: to evaluate interactions between polymorphisms in DNA repair genes and self-reported recreational physical activity on breast cancer risk

Hypothesis: genotypes related to related to reduced DNA repair capacity will act synergistically with low levels of physical activity to increase the risk of breast cancer greater than would be expected by their individual effects

These aims were accomplished through the analysis of extant data from the LIBCSP; a large a population-based case-control study rich in measures of lifetime physical activity and biomarker data [18]. This dissertation employed the pre-existing physical activity measures well as newly genotyped MMR and *CAT* data obtained from the banked DNA of approximately 1,053 breast cancer cases and 1,102 controls. The following sections detail the LIBCSP study population and parent study design, provides a description of the laboratory assays, covariates, and statistical analyses for this analysis, and reviews the perceived advantages and limitations of this research.

2.2 Long Island Breast Cancer Study Project

2.2.1 Eligibility

The LIBCSP is a federally-funded population based study conducted among adult English-speaking female residents of Nassau and Suffolk counties, Long Island, New York. The LIBCSP case-control study was federally mandated and generally supported by Long Island activists, as well as New York State government. Eligible cases for the study were

English-speaking women of all ages and races newly diagnosed with first primary in situ or invasive breast cancer between August 1, 1996, and July 31, 1997, and were residents of either Nassau or Suffolk counties at the time of diagnosis. Eligible controls for the study were English-speaking female residents of Nassau and Suffolk counties at the time of identification, without a personal history of breast cancer. Controls were frequency matched to the expected age distribution of case subjects by 5-year age group.

2.2.2 Case Identification

As a part of the data collection procedures for the parent study, newly diagnosed cases were identified through a 'super-rapid' identification network with a goal to ascertain potentially eligible cases prior to commencing chemotherapy. This network consisted of 28 hospitals on Long Island, as well as three large tertiary care hospitals in New York City. Pathology departments of most hospitals were contacted on a weekly basis (two to three times per week), although institutions with a large proportion of diagnosed cases were contacted daily. Study personnel contacted physicians of potentially eligible case women to confirm diagnosis, date of diagnosis, and to seek permission to contact the patient for potential participation in the study. Prior to case identification investigators contacted over 400 primary care physicians, internists, surgeons, and oncologists who could potentially be involved in the diagnosis or treatment of Long Island breast cancer cases. Physicians were mailed information regarding the study and asked for a documented approval and cooperation. No physician indicated refusal to participate. A total of 2,271 women were initially identified as potential eligible cases. Approximately 73% (2,030) were determined to be eligible according to the study's criteria and physician consent was obtained for 90.5% (1,837). Physician refusal was most often due to poor health status because of age-related co-morbidity. For cases, the average length of time between date of diagnosis and interview was 96 days.

2.2.3 Control Identification

Potentially eligible control women under the age of 65 were identified by Waksberg's method of random digit dialing (RDD) [273]. RDD selection began July 1, 1996, and continued in eight waves over the subsequent twelve months. For women 65 and older investigators used the Health Care Finance Administration (HCFA) rosters for control recruitment. HCFA selection occurred twice during the 12-month ascertainment period and coincided with the 12 months of case identification. The response rate to the RDD telephone screener in Long Island was 77.9%. However, when applied to participants under age 65 the response rate is approximately 57.9% of all control respondents. The average length of time between control identification and interview date was 167 days.

2.2.4 Subject Recruitment and Participation

Eligible case and control women were first contacted by mail which included a letter detailing the study as well as a descriptive brochure. The letter was followed up with contact from a trained recruiter who telephoned the subject to answer questions and arrange for a study interview. The main questionnaire was completed by 1,508 (82.1%) of eligible case women (n=235 with in situ breast cancer) and 1,556 (62.7%) of eligible control women (**Figure A.5**). Motives for non-response to the interview among cases and controls included subject refusal (n = 218 (12.4%) and 573 (21.6%), respectively); too ill, cognitively impaired, or deceased (n = 76 (4.1%) and n = 193 (7.8%)), and non-locatable, moved out of area, or other (n = 26 (1.4%) and n = 195 (7.9%)). Study participants ranged from age 24 to 98 years. Response to interview varied by age with 88.9% of cases and 76.1% of controls under age 65 years participating versus 71.6% of cases and 43.3% and controls over 65 years of age.

2.2.5 Study Interview

Prior to conducting any component of the interview, written signed informed consent was obtained from participants. The interview consisted of (1) the interviewer-administered main questionnaire (2) a self-administered food frequency questionnaire and (3) collection of a biologic sample (blood) and completion of a specimen check-list. All interviews were conducted by a certified phlebotomist or nurse who underwent a week long, standardized, training course in interview administration. Interviews took place in the participant's home.

Main Questionnaire. Data were collected through an interviewer-administered questionnaire which took an average of 101 minutes to complete. Respondents were asked about their demographic characteristics, residential history in Nassau and Suffolk counties, occupational history, medical history, family history of cancer, menstrual history, use of exogenous hormones, reproductive history, body size changes by decade of life, active and passive cigarette smoking, and use of alcohol by decade of life.

For quality control, a random 20% of all respondents were re-contacted via phone to insure that the interview occurred, verify the length of the interview, and to briefly re-interview the participants. Completed questionnaires were shipped to Westat, Inc., Bethesda, MD, for data verification, coding, and data entry, as well as initial range and logic checks.

Assessment of Recreational Physical Activity. As part of the main LIBCSP questionnaire administration, interviewers asked subjects about their participation in recreational physical activity. The recreational physical activity instrument used for the parent study was a modification of that developed by Bernstein and colleagues [274]. A recreational physical activity (RPA) screener was used to query participant's regular involvement in physical activity or exercise during any period throughout the life-course. Participants were asked: "Have you ever participated in any physical activities or exercises on a regular basis – that is, for at least 1 hour per week for 3 months or more in any year?"

These activities included participation as a member of a sports team; participation in individual sports, such as swimming, gymnastics, running, jogging, or walking for exercise; gym workouts and trainings; as well as participation in dance or exercise classes.

Biologic Sample Collection. After completing an additional informed consent form, participants were asked to donate blood sample. Additionally, women were asked to complete a self-administered specimen checklist which queried participants about foods, drugs, and behaviors they may have engaged in the few days prior to the sample donation. For 73.1% (n=1,102) of case and 73.3% (n=1,141) of control respondents who had completed the main interview, a nonfasting 40 mL blood sample was obtained (**Table A.3**). The blood samples were collected in 5 EDTA-treated lavender-top tubes and shipped at room temperature, overnight, to Dr. Regina Santella's laboratory at Columbia University for processing. For most subjects, processing and aliquoting of the biologic samples occurred within 24 h of collection. Aliquots of plasma, red blood cells, mononuclear cells, and granulocytes from 40ml of blood were stored at –80 degrees centigrade with bar-code labels, which were preprinted with the subjects' randomly selected study identification number. Based on previous analysis of DNA in LIBCSP, I anticipated approximately 1053 cases and 1102 controls with blood available for genotyping [224]. The final sample size was primarily dependent on sufficient DNA to complete the assays and the number of failed samples within each SNP. Donation of biologic samples varied with age, with a lower proportion of older control women donating blood. However, case-control status was not a predictor of blood donation among interview respondents.

2.2.6 Medical Record Retrieval and Abstraction

Cases were asked to sign medical record release forms to assess clinical characteristics of the primary breast cancer diagnosis (e.g., stage of disease [in situ vs. invasive], hormone receptor status). Signed medical record release forms were obtained for

1,473 case respondents. Records were successfully located and abstracted for 1,402 participants (**Table A.3**).

2.2.7 Population Characteristics

Age at reference was approximately normally distributed across the study population with the greatest percentage of women falling within the 45-54 age range for both cases and controls. The majority of the women were white: 93.8% and 91.8% of cases and controls, respectively. The sample population was well educated with roughly 87% of all cases and 90% of all controls completing high school. In both groups greater than 95% of the women were currently or previously married and approximately 65% of cases and 68% of controls had an income greater than $35,000 per year.

Many well established risk factors for breast cancer were confirmed to affect risk among women of all ages on Long Island [18]. These include age adjusted parity (OR=0.63 for 4+ children vs. none, 95%CI=0.48, 0.82), breastfeeding (OR=0.70 for 14 months vs. none, 95% CI=0.53, 0.89), age at first birth (OR=1.36 for 28+years vs. <22 years, 95% CI=1.10, 1.69), and family history of breast cancer in mother or sister (OR=1.66 vs. none, 95% CI=1.36, 2.02).

2.2.8 Conclusions

LIBCSP is a large population based case-control study. Unique to the LIBCSP data is the wide range of ages (20-98 years) for both breast cancer cases and controls. Overall, the study had good response rates for both groups and was able to obtain DNA for most of its participants facilitating laboratory assays. Other population-based case-control studies rich in biomarker data could be used for the analysis but none have reliable measures of lifetime physical activity in relation to breast cancer risk. Moreover, the investigators of the LIBCSP obtained a wealth of additional questionnaire data beyond the initial scope of the parent study, enhancing the ability to assess and control for potential modifiers and confounders.

2.3 Genotyping

Biomarker studies have often been proposed to assess the role of physical activity in the etiology of cancer [59, 74, 83]. To better understand the physical activity-breast cancer association one would ideally want to: (1) measure some biomarker of physical activity exposure in association with breast cancer outcomes; (2) use biomarkers of altered function to estimate the extent to which normal cellular processes have been impacted by physical activity; and (3) determine what biomarkers of susceptibility modify the causal pathway from exposure to disease [275]. A general lack of biomarkers of exposure impedes the ability to establish a true causal association between physical activity and chronic disease outcomes [74]. Studies using biomarkers of altered function, primarily conducted in experimental and clinical settings, have shown that correlates of DNA repair are up-regulated with exercise. Similarly, some antioxidant enzymes are increased with regular physical activity. These studies are however, infeasible in large population-based designs typical of many modern epidemiologic studies. At best biomarkers of susceptibility can be employed in molecular epidemiology studies of physical activity to better comprehend the "black box" from exposure to disease [275]. Correctly specifying what biomarkers influence or modify the effect of physical activity on breast cancer outcomes could aid in strengthening the argument for a causal association.

The following sections detail the genetic approach considered for analyses, including strengths and limitations of each approach, the methods used for selecting SNPs, as well as the genotyping procedures employed in the LIBCSP.

2.3.1 Genetic Approach

To study associations between inter-individual variation of MMR, oxidative stress variants and breast cancer, I examined polymorphisms in three MMR genes (*MLH1, MSH2, MSH3*) and one oxidative stress gene (*CAT*) using a candidate gene approach. There were

two approaches considered for these analyses: (1) the candidate gene approach which directly estimate the effects of genetic variants hypothesized to play a role in a disease and (2) genome-wide scanning which scans markers across the genomes of thousands of individuals to identify genetic variation associated with a particular disease. Each approach has specific advantages and disadvantages.

Genome Wide Association Studies. Genome-wide scanning has been successful, enabling the discovery of new associations that have been replicated in multiple studies. Interestingly, many of the loci identified to date were not listed as candidate genes for their respective disease [276]. This approach is therefore useful for identifying candidate genes that were previously unknown or that occur in unexplored regions of DNA. Genome-wide scanning methods are based primarily on statistical associations between a single SNP and a phenotypic group, ignoring prior knowledge about disease pathobiology [277, 278]. The disease-associated SNPs found in genome-wide association studies (GWAS) are therefore unlikely to be the functional variants, instead serving as proxies for the true causal variant contributing to disease. The modeling framework employed in most GWAS tests only one SNP at a time in association with the phenotype under investigation. It thus requires novel bostatistical methods to control for multiple comparisons and new analytic approaches that account for heterogeneity across populations, gene-gene, and gene-environment interactions [277].

Candidate Gene Approach. It is commonly hypothesized that a large component of the genetic variation observed in disease phenotype is due to functional mutations in putative genes. These putative genes are also known as candidate genes and may either directly or indirectly regulate the developmental processes of investigated traits [278]. The candidate gene approach has been commonly applied in gene-disease research primarily by evaluating the effects of the causative genetic variants, or those in linkage disequilibrium (LD) with the functional variants, using association studies. In candidate-gene studies genes

51

are selected a priori based on their etiological role in disease and are conducted in population-based samples [279]. This approach to uncovering the genetic architecture of complex traits capitalizes on both the biological understanding of the phenotype, as well as the increased statistical efficiency of association analysis [279].

In spite of this, candidate-gene studies have a number of limitations. While association studies are well powered to detect genes of small effect, they often result in spurious positive associations that fail to be replicated when followed up in subsequent analyses [278]. There are, however, many reasons for the lack of reproducibility seen with some candidate genes studies. Tabor and colleagues (2002) suggest that discrepant findings may be due to variations in study design, heterogeneity across populations, and phenotype definition. The candidate gene approach is also criticized because of its highly subjective method of choosing specific candidates from a number of potential causative genes. These selections are based heavily on existing knowledge of the known or presumed biology of the phenotype under investigation which is most often finite or unknown creating, in some applications, an information bottleneck [278].

Both GWAS and the candidate gene approach have strengths and limitations. Given the aim of this dissertation was to explore interactions between variants, physical activity and breast cancer risk, the optimal approach for this work was a candidate gene model. A candidate-gene study is better able to assess gene-environment interactions and is rooted in biology of both the disease phenotype and effect measure modifier. It was therefore a useful first step in exploring biological mechanisms between genetic determinants and complex disease, a primary goal of this dissertation.

2.3.2 SNP Selection and Tests of Hardy Weinberg Equilibrium

In an effort to maximize the likelihood of finding a biologically important association and reduce the chance of detecting false positives, I prioritized the selection of both genes and variants in oxidative stress and DNA repair pathways. The *MLH1, MSH2, MSH3* and

52

CAT genes were chosen for this dissertation because they play a critical role in the DNA mismatch repair process and against lipid peroxidation, respectively. While I did not attempt to capture all genetic variability within these genes, targeted SNP selection was informed by functional data, association studies in the breast cancer literature, and patterns of LD within each gene. The additional SNPs selected for the gene*physical activity interaction were based primarily on probable functional application as no prior studies had assessed interactions with physical activity. The genes and respective SNPs for gene-environment interactions are detailed in **Table A.4**.

SNP Function. Identifying functional SNPs is important because these SNPs are likely to play an essential role in gene expression and therefore cell phenotype [280]. Functional SNPs may additionally help to define a biological mechanism through which genotype is causally associated with disease. A single base pair change affecting polyphen prediction, transcription factor binding prediction, miRNA binding, 3D conformation, or splicing regulation were defined as potentially functional SNPs. Similarly, base pair changes that were nonsynonymous or resulted in a stop codon were also classified as potentially functional. These polymorphisms were identified through the breast cancer literature and the SNPinfo web server (SNP function prediction).

SNPs Identified in the Breast Cancer Literature. A total of 6 SNPs in the genes of interest had been evaluated in the breast cancer literature. Four of these SNPs (rs1001179, rs1799977, rs1800734, and rs4987188) were associated with breast cancer risk in at least one study. The *CAT* SNP (rs1001179) was previously genotyped as part of the LIBCSP. Both rs1799977 and rs1800734 are located in *MLH1*. The former SNP was selected for genotyping because: (1) it is known to have functional properties and (2) the latter has only been associated in Korean populations. While rs4987188 is known to have functional properties, its MAF is 3% and was therefore excluded from this analysis.

Tag SNPs. Tag SNPs are polymorphisms that are highly correlated with other SNPs in a gene [281] which, upon genotyping, can be used to infer characteristics of un-typed SNPs. This method is based on the degree of LD between the tag and un-typed SNPs which is specified a priori. A tagSNP approach maximizes the ability to capture genetic variation across a genomic region while reducing costs. Two programs were used to identify tagSNPs: the Tagger SNP selection program in Haploview version 4.2 [282, 283] and the SNPinfo web server from the National Institute of Environmental Health Service [284]. For both programs tagSNPs were selected using data from phase II of the International HapMap Project database [285]. Given the racial homogeneity of the LIBCSP population with DNA available for the proposed analyses (93.4% White and 6.6% Non-White [224]), the CEU population (30 Utah trios with ancestry from northern and western Europe) was used as the reference panel for SNP selection.

Both programs use pairwise tagging methods to select a maximally informative set of common SNPs incorporating LD information based on the r^2 statistic [286]. The r^2 statistic is used to assess the degree of correlation between SNPs. It is a measure of how well the identity of one allele at a polymorphic locus predicts the identity of the allele at another polymorphic locus. An $r^2=1.0$ indicates that the examined loci are in "perfect LD". Other measures of linkage disequilibrium (i.e. D') are often used, but fail to obligate identical minor allele frequencies (MAF) among SNPs when D'=1 [286]. The tagging algorithm begins by calculating the r^2 between all pairs of SNPs in the gene region (including 1000 base pairs up and down stream) above a pre-specified MAF threshold. The single SNP that is correlated with the greatest number of other SNPs at a specified r^2 is identified and grouped with its correlated SNPs into a bin. The best tag SNP in each bin is then selected based on all pairwise r^2. This process is repeated using the remaining un-binned SNPs until only SNPs not in high LD with other SNPs remain. These are placed into their own singleton bin. This combined group of tag SNPs represents the minimum set of informative SNPs for the gene.

54

For the present study an r^2 of 0.80 and MAF of 0.05 or greater were imposed on SNP selection procedures. The tagging procedures described above were used to select SNPs for *CAT, MLH1, MSH2, and MSH3.* Due to limited resources TagSNP selection was prioritized based on a combination of factors including: location within the gene, bin size, and MAF. A total of 9 SNPs were identified for genotyping in the four genes under study (**Table A.5**). **Figures A.6-A.9** show the LD plots for each gene in the CEU population.

   Haplotypes. While the tagSNP approach is useful for examining the independent effect of common genetic variants on breast cancer risk, it fails to account for contributions among rare variants as well as the physical location of a potential causal allele relative to another causal allele. A broad multilocus haplotype approach examines variation across the gene by identifying a set of closely linked genetic markers present on one chromosome which tend to be inherited together [287]. Haplotype analyses may have greater power to detect susceptibility alleles compared to multiple single SNP analyses when the true causal allele is unknown or when disease is influenced by multiple causal alleles occurring in cis [288-291]. Assessment of haplotypes may be obligatory when two or more SNPs are in high LD ($r^2 \geq 0.7$) with one another. The SNP selection approach used in this project was designed to select the minimally sufficient number of tagSNPs to characterize each gene in the CEU HapMap population using an a priori r^2 value of 0.8. Using plink v1.06, I calculated pairwise LD between all genotyped SNPs and found none to be in high LD (**Table A.6**). The most highly correlated SNPs (*MSH2*: rs3732182 and rs4583514) had an r^2 value of 0.634.

   Hardy-Weinberg Equilibrium. Tests of Hardy-Weinberg equilibrium (HWE) were conducted for candidate gene SNPs to ensure assumptions of independent inheritance were upheld in the LIBCSP. Departures from HWE were assessed among Caucasian controls, as this demographic group best represented the source population [292]. Deviations from HWE may indicate genotyping error, selection bias, population stratification, new mutations, or a violation of the HWE population assumptions in controls, while among cases it may denote

55

an association between the putative at risk allele and disease [293]. For a biallelic locus in a randomly mating population, where the frequency of alleles are represented by 'p' (major allele) and 'q' (minor allele), there is a mathematical relationship between the frequency of alleles at a genetic locus and the genotypes resulting from those alleles: $p2 + 2pq + q2 = 1$ [292]. This equation was used to determine the expected genotypic frequencies under the conditions of HWE and subsequently compared to the observed frequencies using a one degree of freedom Pearson's chi-square test. All SNPs for MMR related genes were in HWE (**Table A.7**). Two *CAT* SNPs (rs4756146 and rs2284365) were observed to deviate significantly from HWE (**Table A.8**), although the respective MAFs were comparable to the CEU HapMap population.

2.3.3 Genotyping Procedures and Quality Control

LIBCSP genotyping was conducted by Dr. Regina Santella's laboratory at Columbia University, New York, NY. All DNA samples were available on 96 well master plates. Controls for genotype at each locus and two no-DNA controls were included on each plate. Plates have a 10% duplication rate with laboratory personnel blinded to both duplication and case control status. Genotyping was performed using Taqman (Applied Biosystems, Foster City, CA) assays. Briefly, TaqMan probes are hydrolysis probes designed to anneal within a DNA region amplified by a specific set of primers. They consist of a fluorophore and a quencher covalently attached to the 5' and 3'-end of the oligonucleotide, respectively. As the Taq polymerase extends the primer and synthesizes the lagging strand, the 5' to 3' nuclease activity of Taq polymerase degrades the dual-labeled probe that has annealed to the template. This releases the fluorophore, thereby relieving the quenching effect and allowing for the transmission of fluorescence signals from the flurophore. The resulting fluorescence signal allows one to quantify the amount of polymerase chain reaction (PCR) product formed during PCR [294]. The fluorescence profiles are subsequently quantified using Sequence Detection Software (Applied Biosystems). The rs numbers for the SNPs of

interest were given to Applied Biosystems for the preparation of the specific kits. Taqman, samples were run on an ABI 7900 Real Time PCR system.

2.4 Variable Construction and Covariates

The following section describes the selection and coding of covariates including the primary exposure, effect modifiers, and all potential confounders. SNP selection and genotyping for *MLH1*, *MSH2, MSH3,* and *CAT* were detailed in section 2.3. Similarly, section 2.2 described the data collection process for the RPA measurement. Confounders were selected *a priori* using directed acyclic graphs (DAG) [295] on the basis of subject matter knowledge. Final selection of confounders was based on modeling strategies discussed in section 2.5.

2.4.1 Exposure Variable Construction

Genotype information comes from two sources: the parent study's previously assayed DNA samples and new laboratory analyses using isolated DNA to genotype *CAT, MLH1*, *MSH2, and MSH3* SNPs. The details of ascertainment and analyses for these exposures have been described previously (Section 2.2). There are several model forms that may potentially be used to estimate genotype effects. The allelic identity at a particular locus on both copies of a chromosome determines the genotype. Since the SNPs selected for this analysis are biallelic, there are only three possible genotypes for each SNP: (1) homozygous for the common allele; (2) heterozygous; or (3) homozygous for the minor allele. A general model assumes no relationship between the three genotypes. In the dominant genetic model a single variant allele is sufficient to affect disease risk; the heterozygotes and minor allele homozygotes are therefore collectively considered the 'exposed' group. In a recessive genetic model two variant alleles are needed to affect disease risk; the heterozygotes and major allele homozygotes are together considered the 'referent' group. Under the additive genetic model, the effects of genotype are linear; the change in disease risk is thus proportional to the number of variant alleles in the genotype

[296]. All SNP associations were be initially estimated using the general model. The general model is known to have less power than the true, correctly specified mode of inheritance because it requires two degrees of freedom compared to the one degree of freedom required for other genetic models [297]. When the true model form is unknown however, the general model is more powerful than choosing any incorrectly specified model form. For this dissertation the mode of inheritance for most SNPs was not known; the general model was therefore most appropriate to estimate genotype effects. However, for most loci, a more flexible modeling approach was necessary.

For each analysis the SNPs were modeled categorically using indicator variables. Two indicator variables were created, one for the heterozygote genotype and another for the minor allele homozygote genotype. Estimated ORs contrasted the effect of heterozygote vs. major allele homozygote (referent group) and minor allele homozygote vs. major allele homozygote. In most instances data were too sparse among the minor allele homozygote genotype to specify the mode of inheritance (i.e. dominant, recessive, or additive). I was therefore forced to use a dominant model, particularly for assessing interactions with physical activity.

2.4.2 Effect Modifier Variable Construction

Biologic mechanisms that lead to the development of breast cancer are likely to have multiple interacting component causes. It was therefore important that, in addition to estimating main effects of candidate genes, I evaluated statistical interaction with non-genetic breast cancer risk factors. In doing this I was able to estimate the extent to which genotype associations may influence these risk factors. One of the central aspects of this project was to uncover mechanisms through which RPA exerts its protective effects on breast cancer outcomes. Using a candidate gene approach based on biologic plausibility I selected one oxidative stress gene (*CAT*) and three MMR genes (*MLH1, MSH2* and *MSH3*) to investigate. These genes were evaluated in addition to functional oxidative stress and

58

DNA repair genes already ascertained in the LIBCSP. Section 2.2.5 provided an overview of the physical activity data collection. Respondents were asked about all recreational physical activities in which they had engaged for at least one hour per week and at least three months or more in any year over their entire lifetime. Those participants who replied never having participated in RPA were classified as having no RPA. For those women who answered yes, a detailed lifetime history of physical activity was created using recall cues such as life events calendars and residential history. For each participant the investigators obtained the name of the activity, the ages the activity was started and stopped (if applicable), the number of hours per week and months per year the activity was usually performed. The total years of participation in the activity was also recorded. When activities were terminated and begun again at a later time each episode was coded separately, allowing for an evaluation of activity patterns at various ages. For participants who listed an activity without providing the number of months per year of participation, 12 months per year was imputed if the activity was deemed non-seasonal. If the activity was characterized as being seasonal the reported menopause specific mean months per year was imputed. The complete assessment provided a detailed self-reported lifetime history of each participant's RPA. These data were ultimately converted into hours per week and weeks per year of participation and the values summed across all activities for each year of a woman's life, providing a lifetime duration-frequency variable for RPA from menarche (left truncated) to reference date.

Similarly, for women classified as ever having participated in recreational physical activity, metabolic equivalents of energy expenditure (MET) scores were assigned to each reported activity according to a published database [57]. In scenarios where the activity reported was not in the published database, efforts were made to find activities in the database similar to that which was reported and use the corresponding MET score. These scores were multiplied by the number of hours per week the participant reported engaging in

the activity and were summed across all activities to create a lifetime intensity-duration-frequency variable for RPA from menarche to reference date.

A total of 149 subjects (4.9%) were missing the lifetime duration-frequency RPA variable. This includes 70 (4.6%) cases and 79 (5.1% controls). One hundred ninety two participants were missing MET scores including 90 (6.0%) cases and 102 (6.6%) controls. A sensitivity analysis was conducted imputing the highest, lowest, and mean reported activity level for postmenopausal women missing RPA from first live birth to menopause [298]. Results from the logistic models of the three datasets generated by the different imputations were not materially different from the complete case analysis. These individuals were therefore excluded from analyses.

Variable construction for the RPA was assessed in detail for a previous analysis [63]. For this project I explored several constructions for the physical activity variable including categorization based on control quartiles, parametric specifications (e.g., logit, linear), and flexible modeling (e.g, quadratic, higher order polynomials, or splines) (**Figures A.10-A.15**). RPA distributions based on control quartiles appeared to best described the shape of the data using the fewest number of parameters. In order to maintain reasonable cell sizes for gene*environment interactions on an additive scale, however, the RPA variable was modeled categorically using indicators for high (greater than or equal to control median), low (less than control median), and no (based on the RPA screener) physical activity. In all analyses the lowest RPA group (no activity) served as the referent.

In addition to the physical activity data the baseline questionnaire queried women on a number of other exposures including reproductive, medical and environmental histories; self-reported weight and height by decade of life; cigarette and alcohol use; use of exogenous hormones; energy intake; demographic characteristics; and, among cases, tumor receptor status. While other exposures have been shown to modify the effect of oxidative stress or DNA repair variants within this study population [156, 162, 224, 259, 299]

these analyses were considered beyond the scope of the project, as the primary aim of this dissertation was to assess the interactions with RPA.

2.4.3 Confounders

I expected minimal confounding of the association between breast cancer and *CAT, MLH1, MSH2,* and *MSH3.* Based on the DAG in **Figure A.16**, a minimally sufficient adjustment set for both oxidative stress and DNA repair variants included race, family history of breast cancer, and religion. The LIBCSP population for the proposed study is primarily Caucasian (93.4%) making race unlikely to play an important role in these analyses. Both family history of breast cancer and religion (specifically Judaism) may represent proxies for increased frequency of high penetrant, low prevalence breast cancer susceptibility genes (primarily *BRCA1*). Adjustment for these variables may be important as 19.2% of cases and 13.0% of controls report a family history of breast cancer among a first degree relative, and 17.2% of case and 15.4% of controls self-identified as Jewish in this study population [26]. Given these data I expected minimal confounding of the main genotype effect. I also anticipated little confounding of the gene-environment interaction, as a potential confounder must affect both components of the interaction on the multiplicative scale. If there were evidence for interaction on the additive scale however, some covariates could confound the association between physical activity and breast cancer risk within strata of genotype. There were no such instances in these analyses as interaction was only observed in multiplicative models (*See Chapters 3 and 4).*

2.5 Data Analysis

The primary objective of this dissertation was to elucidate the mechanism through which physical activity may act on breast cancer risk. This was accomplished by assessing potential interactions within two closely linked pathways: oxidative stress (AIM 1) and DNA repair (AIM 2). There are three principal analyses for each pathway: (A) estimation of the association between selected SNPs and breast cancer risk (main effects), (B) evaluation of

2-way interactions between genes/SNPs within a pathway and breast cancer outcomes (gene-gene or SNP-SNP interactions), and (C) evaluation of interactions between functional variants and self-reported lifetime physical activity on breast cancer risk (gene-environment interactions). All statistical analyses were conducted using SAS version 9.1.2 (SAS Institute, Cary, NC).

2.5.1 Descriptive Statistics

The distribution of participant characteristics (SNPs, potential confounders and effect modifiers) were calculated as counts and frequencies for categorical variables and reported separately for cases and controls in the total study population (**Tables A.9-A.10**) as well as among white women only (**Tables A.11-A.12**).

2.5.2 Main Effect of SNPs

Unconditional logistic regression models were used to estimate ORs and 95% CIs for the main effect of each candidate SNP genotype on breast cancer risk, with adjustments made for the frequency matching factor – age at reference date [300] (**Table A.13-A.14**). These estimates were also calculated among white women only to verify that results did not change materially by including women of other racial demographics in the model (**Table A.15-A.16**).

The binary logistic model function used to estimate OR's was: Logit [D=1|X=x] = α + β1X1 + β2X2, where

α = model intercept

X1 = presence or absence of heterozygote genotype

X2 = presence or absence of minor allele homozygote genotype

β1 = regression coefficient corresponding to heterozygote genotype

β2 = regression coefficient corresponding to minor allele homozygote genotype

D = case (1) or control (0) status

Covariates, selected based on DAG analyses, were considered confounders if they were associated with exposure in the source population, associated with disease among women who were unexposed and resulted in at least a 10% change in estimate when added to the model, compared to a model without the covariate [301]. None of the identified covariates met these criteria (**Tables A.17-A.18**). In addition to estimating the main effects I assessed potential heterogeneity of effects across strata of menopausal status, family history, and religion (**Tables A.19-A.20**). Race was also explored as a potential effect modifier, but given the low proportion of African American and 'other' women in the LIBCSP our analysis was limited in assessing potential modification by these variables. This dissertation also explored the possibility that both MMR and CAT SNP effects varied by tumor receptor status (**Tables A.21-A.22**) using broad categories or hormone receptor (HR) positive (any ER or PR+) and HR negative (ER-/PR-) to serve as proxies for the less aggressive luminal subtypes and more aggressive triple negative/ human epidermal growth factor receptor 2+ (HER2+) subtypes [302].

2.5.3 Interactions

A statistical interaction occurs when there is departure from additivity of effects on one chosen outcome scale [303]. Evidence for a biological interaction between two exposures may be inferred from measures of statistical interaction calculated from regression models. While multiplicative interactions are likely more reflective of a multistage disease like breast cancer, additive interactions on the risk scale may better reveal biological interactions [303].

Gene-Gene/SNP-SNP Interactions. Interaction between MMR genes, *CAT* SNPs, and breast cancer development was evaluated using a likelihood ratio test (LRT): the difference of two -2LogL values of logistic models calculated with and without the interaction terms for SNP1 and SNP2 (**Tables A.23-A.24**). I subsequently explored pathway effects by collapsing the number of 'at risk' alleles in the *CAT* gene (**Table A.25**) and variant alleles

63

within the MMR pathway (**Table A.26**) to calculate adjusted ORs for breast cancer (**Tables A.27-A.26**). Low-, intermediate-, and high-risk categories for each pathway were defined based on similar point estimates of breast cancer risk for each number of putative high-risk or variant alleles (**Tables A.27-A.28**). Similar analyses had been conducted among variants in the NER pathway using both LIBCSP [222] and Carolina Breast Cancer Study [229] study data.

Gene-Environment Interaction. To assess potential interaction between functional polymorphisms in oxidative stress and DNA repair genes, RPA and breast cancer risk I examined both multiplicative and additive interactions. Multiplicative interactions were assessed by including a multiplicative interaction term in the regression model and calculating departures from the multiplicative null using the LRT (**Tables A.29-A.32**). The LRT compares the -2 log likelihood (-2LL) of two models, one of which is nested within the other, to determine if the addition of the interaction term improves model fit.

The basic logistic regression model allowing for interaction is: logit [D=1|X=x] = α + β1X1 + β2X2 + β3(X1)(X2)

where α = model intercept

X1 = exposure

X2 = effect modifier

β1 = regression coefficient corresponding to the exposure

β2 = regression coefficient corresponding to the outcome

β3 = regression coefficient corresponding to the excess effect of joint exposure

Departures from the additive null were evaluated by the interaction contrast ratio (ICR) [303] using indicator terms for participants with the genotype only, exposure only, and both the genotype and exposure of interest (**Tables A.33-A.36**). The magnitude of additive interaction effect between SNPs and physical activity was determined by estimating the adjusted interaction contrast ratio (ICR) using the formula: ICR OR $_{exposed, variant}$ − OR $_{exposed}$ −

OR $_{variant}$ + 1 and its respective confidence interval obtained by ICR ± 1.96 standard error(ICR) [304]. ICRs less than zero indicated less than additive effects, ICRs of zero suggested no interaction on the additive scale, and ICRs greater than zero implied superadditivity [303]. While it would have been interesting to explore gene-environment interactions using the total variant approach discussed above, limited power in the LIBCSP hindered assessment of interactions using this method.

2.5.4 Sample Size and Power

The total sample with data available for genotyping included 1053 cases and 1102 controls. For these analyses, the estimated study power varied based on the genetic model selected, frequency of the at-risk genotype (10% to 50%), the expected ORs for the association between the genotypes and risk of breast cancer, and if women are categorized by menopausal status. Power calculations were based on the generally accepted standard two sided α=0.05 and were calculated using POWER version 3.0 software available through the National Cancer Institute and described in Garcia-Closas [305].

The study had more than adequate power to detect even modest associations for main gene effects (using an additive, dominant, or recessive model). Based on previous analyses I expected the genetic effect to range between 1.2 and 2.0. Power was estimated as ≥ 80% for ORs ≥ 1.5 for all women combined and for postmenopausal women alone assuming a two-sided α=0.05, 1% disease prevalence, and at-risk genotype prevalence of at least 10%. **Figure A.17** shows the expected study power for varying frequencies of the at-risk genotype among pre and postmenopausal women combined. To examine OR effect modification [305, 306] of most gene*gene or gene*physical activity combinations 80% power was expected to detect substantial interactions of OR ≥ 4.5 and OR ≥ 5 multiplicative (**Figure A.18**) and additive models (**Figure A.19**), respectively. These calculations assumed a 2-level genetic model, 3-level physical activity categorization, and an at risk genotype prevalence of at least 25%. Power declined with decreasing at-risk genotype prevalence

and increasing numbers of physical activity categories.

2.6 Strengths and Limitations

2.6.1 Study Design

One major advantage of using the Long Island data set was its large population-based design. This provided sufficient statistical power to examine gene-gene and gene-environment interactions. Moreover, LIBCSP is unique among epidemiologic studies of breast cancer given that there was no upper or lower age limit for subject eligibility, making it one of few studies that can provide extensive data on the epidemiology of breast cancer among women 65 years of age and over. Similar to other population based case-control studies LIBCSP experienced lower participation rates among controls compared to cases (62.7% vs. 82.1%, respectively), which may indicate the presence of participation bias. These differences are primarily attributed to poor response among elderly control women ≥ 65 years where a 43.3% response rate was achieved in comparison to the 71.2% response rate among case women. These differences were less evident among women under 65 years with an 88.9% and 76.1% response rate among cases and controls, respectively [18].

Despite the population-based nature of the sample, a limitation of the study is its relatively homogenous population, specifically with regard to race. The racial homogeneity of the study population restricts the ability to evaluate modification by breast cancer subtypes (e.g. Luminal A, Luminal B, HER2, and Basal-like) which are known to vary by both menopausal status and race [307]. The overwhelming majority of the women included in the parent LIBCSP study were classified as estrogen receptor positive, progesterone receptor positive, and HER-2/neu negative (Dr. Marilie Gammon, personal communication, 2009) which are indicative of the luminal A subtype [307]. Given the narrow range of breast cancer subtypes I was limited in my ability to detect differences by tumor status. Similarly, because of the differences in the ethnic distribution of this study population compared to the US population as a whole, study results may not be readily applicable to all women of the

US. While this limits external validity, the internal validity of the study will be enhanced and will apply to women with the highest risk of developing breast cancer, namely white, postmenopausal women. Importantly, this study may provide some clues about the underlying biologic mechanisms of physical activity which are unlikely to vary by race despite potential racial variation in the frequency of specific alleles and prevalence of exposure.

2.6.2 Exposure Assessment

The proposed ancillary project is efficient because the exposure data and biologic specimens have been obtained through the parent study. The detailed exposure information from LIBCSP will enable us to explore several different measures of physical activity (e.g. hours/wk and MET hrs/wk) as well as several time periods throughout the reproductive lifespan and potential interaction with *MLH1*, *MSH2*, *MSH3*, and *CAT* to influence breast cancer risk. The physical activity assessment is unique in that it is one of few population based studies to inquire about recreational physical activity by decade from menses to reference date.

Physical Activity. A perceived weakness of large scale studies of physical activity is the inability to accurately assess activity in the distant past. By design, a case-control study such as this must often rely on self-reported data to ascertain relevant exposure and covariate information. Errors in reporting or differential reporting by cases and controls have the potential to bias the study results. Ideally, etiologic studies of physical activity would assess exercise-related biomarkers of biologically effective dose [74]. These markers would represent the amount of an external exposure that has both entered the body and interacted with molecular targets. The use of these biomarkers is thought to increase the validity of exposure assessment, however in the realm of physical activity there are currently no such markers.

Activity levels in the LIBCSP were assessed via an interviewer administered questionnaire, which could be hampered by measurement error. To reduce these errors, Long Island interviewers were educated and trained to collect data in a systematic manner [18]. Nevertheless, differential recall of physical activity among cases and controls could potentially bias results. In these analyses such errors are potentially superfluous because I was interested in estimating the effect of gene*environment interactions whereby there is no differential reporting of physical activity based on genotype [308]. It is often a concern that physical activity questionnaires lack content validity and reliability specifically when employed in case control studies. These errors could potentially reduce the ability of a study to identify important relationships. More objective measures such as accelerometers and pedometers could be used, but they are not feasible in large population-based studies and would fail to capture the etiologically relevant time period for breast cancer outcomes.

While the physical activity measurement in this study has not been validated, it is reassuring to note that the instrument has been useful in revealing important relationships between exercise and breast cancer risk in several epidemiologic studies [274, 309]. Similarly, the results obtained in the Long Island study data [63] for the main effect of physical activity among post-menopausal women (30% reduction in risk) is consistent with the 25% risk reduction reported in other independent and review studies [44, 46].

Genetic Variants. Common sources of error for all biomarker studies are issues related to specimen collection, processing, and storage. A number of steps were taken to minimize these errors in LIBCSP. Samples from matched sets were assayed together in the same batch to ensure that effect estimates did not vary because of inter-assay variability. For quality control purposes 10% duplicates of the samples were distributed throughout DNA samples and laboratory personnel were blinded to case control status. Additionally, computerized algorithms in SAS were used to cross-check genetic data for inconsistencies.

68

Small sample size is an inherent limitation of most molecular epidemiology studies chiefly because of the infrequency of some minor alleles. The LIBCSP was well powered to investigate 2-way interactions (gene*environment and gene*gene among genes or physical activity. Higher order interactions were not assessed in this study.

Finally, bias may arise if cases and controls differentially donated blood samples or if physical activity status was a predictor of blood donation. Among respondents who completed the interview there was no case-control differential between case and control participants who donated a blood sample.

2.6.3 Data Analysis

Multiple comparisons is a consideration in the proposed study, as there are no a priori evidence indicating an association exists between specific oxidative stress/DNA repair variants, physical activity, and breast cancer. There are a number of approaches to correct for these errors. The first approach would be to make no correction for multiple testing. This approach would obligate that the SNPs assessed are not highly correlated. While MMR and *CAT* SNPs selected for main effects analysis (N=9) are not highly correlated (highest pairwise R2 = 0.634), functional SNPs, chosen a priori for the gene-environment analyses (N=29) will likely be related. Further, the additional gene-environment component of the proposed study increases the number of comparisons to be performed (as some GxE interactions will be evaluated using multiple genetic models) making some correction for multiple comparisons warranted. Permutation methods can become computationally demanding with many SNPs and may not be applicable when covariates are included in the model. Similarly, more sophisticated approaches (e.g. Bayesian or Monte Carlo methods) involve many assumptions, are computationally intensive [287, 310] and may provide no added benefit over more direct approaches. To correct for multiple testing I proposed to employ the Bonferroni correction, or step down procedure [311], but given the biologically based hypothesis I made no correction for multiple comparisons.

69

2.6.4 Data Interpretation

Identifying susceptibility or resistance alleles in gene-environment studies may help to uncover biologic pathways that are most relevant for environmental exposures known to impact disease susceptibility. While the term "gene-environment" often implies a specific biologic relationship between the genes and the environment I recognize that in epidemiologic studies these interactions are statistical, and do not indicate a specific disease mechanism. However, in using interactions as a best effort to uncover underlying biology some specific methods can be employed: (1) it is important to assess interactions on both multiplicative and additive scales given multistage diseases like cancer which are likely to have factors acting in both the same and different stages along the cancer continuum (2) to reduce the likelihood of spurious findings it is suggested that gene products and exposures be in the same biologic pathway. Further, selecting variants that alter gene function are likely more useful than screening large numbers of SNPs [312] (3) assessment of gene*environment interactions required multiple level stratification, invariably creating small cells. Caution was therefore employed in data interpretation.

2.7 Summary

The LIBCSP is a large population-based case-control investigation with a comprehensive, life-course assessment of recreational physical activity and is rich in genetic data. Moreover, the LIBCSP has extensive data on a number of breast cancer risk factors, confounders, and effect modifiers including information regarding breast tumor characteristics. The proposed research aimed to (1) evaluate associations between variants in *CAT*, the MMR pathway and breast cancer using multivariable logistic regression, adjusting for potential confounders (2) examine interactions between SNPs and/or genes in *CAT* and MMR, respectively, and (3) assess interactions between recreational physical activity, oxidative stress and DNA repair polymorphisms in BER, NER and MMR pathways with respect to breast cancer risk using multiplicative and additive models. No previous

epidemiologic research had evaluated potential interactions between physical activity and genetic variants on risk of breast cancer.

This research may have widespread implications for breast cancer. Although there is evidence that some breast cancers exhibit MSI few studies have sought to explore the effects of MMR variants on breast cancer risk. These associations are important to understanding the etiology of breast cancer because MMR is involved in the overall maintenance of genomic stability. This dissertation would be the first to systematically evaluate these associations in a large population-base sample. Similarly, some studies show that the C/T polymorphism (rs1001179) in the promoter region of the *CAT* gene is associated with increased breast cancer risk, however few other polymorphisms have been evaluated. Both the animal and *in vitro* literature suggests that the generation of lipid peroxidation products play a large role in breast cancer susceptibility. Uncovering other risk variants in this gene may help to better understand the role of *CAT* in the maintenance of oxidative balance.

Most observational studies show an inverse association between physical activity and breast cancer risk. Although the exact frequency, duration, and intensity are not well established there is adequate evidence that lifetime physical activity is an important modifiable factor for breast cancer. The importance of uncovering the underlying mechanisms of physical activity is frequently cited in the physical activity-breast cancer literature. However, little research has been conducted outside the obesity related pathways despite the strong biologic plausibility of both antioxidant and DNA repair pathways. Identifying women who are particularly susceptible to the beneficial effects of physical activity based on genetic characteristics could aid in validating the biologic plausibility of this association, helping to better identify new targets for intervention and inform public health recommendations for lowering breast cancer risk.

Chapter III: Polymorphisms in Oxidative Stress Genes, Recreational Physical Activity and Breast Cancer Risk

3.1 Introduction

Oxidative stress is hypothesized to play an important role in breast carcinogenesis [127, 139-141] and is caused by the imbalance of reactive oxygen species (ROS) production and antioxidant defenses which neutralize these molecules [172]. ROS may be generated through any number of endogenous or exogenous mechanisms. While modest levels of ROS are useful for cell signaling processes [123] excess production may result in DNA damage, lipid peroxidation, and protein modification [121, 125, 126]. When endogenous or exogenous ROS production occurs in an environment with sufficient in vivo defense mechanisms to scavenge the ROS, there are seemingly few harmful effects. When there is excess ROS production and/or insufficient defense mechanisms, oxidative stress may ensue. There are several antioxidant defenses that can protect against oxidative damage including catalase (CAT), manganese superoxide dismutase (MnSOD), glutathione peroxidase (GPX), glutathione *S*-transferases (GSTs), myeloperoxidase (MPO) and catechol- O-methyltransferase (COMT) [127]. CAT plays an important role in neutralizing ROS by converting H_2O_2 into H_2O and O_2 [127]. Activity levels of the CAT enzyme are likely affected by a functional polymorphism (rs1001179) in the promoter region of the gene [156]. While this polymorphism has been associated with decreased enzyme activity [153-156], its association with breast cancer risk is unclear [156-158]. Other polymorphisms in CAT may be important in understanding the underlying association with breast cancer incidence and should be considered.

While physical inactivity is a well-established risk factor for breast cancer [42], the mechanisms driving the association are not well described [44, 74, 75]. Given the biological

adaptation of enhanced antioxidant enzymatic capacity that occurs with regular exercise and its contribution to ROS, the oxidative stress pathway may be of interest. Physical activity may therefore interact with antioxidant-related genetic polymorphisms to influence breast carcinogenesis. No previous epidemiologic investigations have explored this possibility. We hypothesized that genotypes related to reduced antioxidant expression may have an antagonistic effect on the benefits of physical activity. In this report, we aimed to: (1) examine the independent main effects of three variants in *CAT* (rs4756146, rs2284365, and rs480575) on breast cancer risk; (2) examine two-way interactions between SNPs in *CAT* and breast cancer risk; and (3) examine potential interaction between recreational physical activity (RPA) and several oxidative stress related genes (*CAT, COMT, GPX, GSTA1, GSTM1, GSTP1, GSTT1, MnSOD*, and *MPO*) with respect to breast cancer incidence. Secondary aims were to evaluate associations between *CAT* polymorphisms and breast cancer with cases categorized according to tumor hormone receptor status. These aims were accomplished through the use of existing biomarker and questionnaire data from the LIBCSP.

3.2 Materials and Methods

3.2.1 Study population

Study participants were from the LIBCSP, a population-based case-control study conducted among English-speaking female residents of Nassau and Suffolk counties, Long Island, New York. Details of the study methods have been described previously [18]. Briefly, LIBCSP cases were women aged 20-98 years diagnosed with a first primary in-situ or invasive breast cancer between August 1, 1996 and July 31, 1997. Case women were identified through daily or weekly contact with local hospital pathology departments. Population-based controls were women without a personal history of breast cancer randomly selected using random digit dialing for those under age 65 and the Health Care Finance Administration rosters for women ages 65 and older. Controls were frequency

matched to the expected age distribution of case women by 5-year age groups. All data were collected through a two-hour, interviewer-administered, structured questionnaire. Interview response among eligible cases and controls were 82.1% (n=1508) and 62.8% (n=1556), respectively. Respondents were more likely to be older (median age = 57 years in cases and 56 years in controls), postmenopausal (n = 1003 cases and 989 controls), and white (93.4% white, which reflects the underlying distribution of the source population).

Of those who completed an interview, 73.1% of cases and 73.3% of controls donated a blood sample. Among women who donated blood, genotyping was unavailable for 4.4% of cases and 3.4% of controls primarily due to insufficient DNA. Our final sample therefore includes 1053 cases and 1102 controls. Written informed consent was obtained from all participants. This study was approved by the Institutional Review Board of the collaborating institutions.

3.2.2 SNP Selection and Genotyping

We selected three SNPs in *CAT* for genotyping (rs4756146, rs2284365, and rs480575). A tagging strategy was employed to maximize our ability to capture genetic variation across the gene (gene and 1000bp upstream and downstream). Tag SNPs were selected using the SNPinfo web server from the National Institute of Environmental Health Science [284] based on data from phase I and II of the International HapMap Project database [285]. Given the racial homogeneity of the LIBCSP population with DNA available for the proposed analyses [224]; the CEU population (30 Utah trios with ancestry from northern and western Europe) was used as the reference panel for SNP selection. We imposed a minor allele frequency (MAF) cutoff value of 10% and r2 threshold minimum of 0.80 on SNP selection procedures. From the 11 tag SNPs identified to capture the *CAT* region, three were singleton or double bins, two had MAF < 10%, and one was previously genotyped in LIBCSP. Of the remaining 5 tag SNPs we selected three based on location, bin size, and linkage disequilibrium with functional variants.

In addition to the newly genotyped *CAT* variants, we selected 10 functional polymorphisms from 9 genes in the oxidative stress pathway to assess gene*environment (G*E) interactions with RPA: (rs1001179), *COMT* (rs4680 and rs737865), *GPX* (rs1050450), *GSTA1* (rs3957356), *GSTM1* (gene deletion), *GSTP1* (rs1695), *GSTT1* (gene deletion), *MnSOD* (rs4880), and *MPO* (rs2333227). A single base pair change affecting polyphen prediction (*GPX*), transcription factor binding prediction (*CAT*, *COMT* rs737865, *MNSOD*, and *GSTA1*), miRNA binding (*GPX*), 3D conformation (*COMT* rs4680), or splicing regulation (*GPX*, *COMT* rs4680, *MPO*, and *GSTP1*) were defined as potentially functional SNPs. Similarly, base pair changes that were nonsynonymous (*GPX*, *COMT* rs4680, *MPO*, and *GSTP1*) or resulted in a stop codon were also classified as potentially functional. These polymorphisms were identified through the breast cancer literature and the SNPinfo web server [284]. The main effects of these associations with breast cancer risk have been previously reported in the LIBCSP study population [156, 162, 299, 313-316]. However, interactions with physical activity have not been considered. Previously published SNP-specific effects among postmenopausal women for the genes of interest are provided in **Table A.37** to offer a full pathway context for our findings on the effect of RPA and ROS-related polymorphisms.

A non-fasting 40 mL blood sample was obtained from participants at time of interview and shipped at room temperature, overnight, for processing. Genomic DNA was extracted from mononuclear cells in whole blood separated by Ficoll (Sigma Chemical Co., St. Louis, Missouri). Pelleted cells were frozen at -80° C until DNA isolation by standard phenol, and chloroform isoamyl alcohol extraction and RNase treatment were performed [18]. Genotyping of newly selected CAT SNPs was accomplished using Taqman assays (Applied Biosystems, Foster City, CA) in 384-well plates. For the remaining SNPs genotyping was performed by BioServe Biotechnologies (Laurel, MD) using Sequenom's high-throughput matrix-assisted laser desorption/ionization time-of-flight mass spectrometry, as described

75

previously [299]. Controls for genotype and two non-template controls were included on each plate. Samples that were outside the variables defined by the controls were identified as non-informative and retested. For quality control, 10% of samples were distributed throughout the DNA samples as blinded duplicates. Laboratory personnel were blinded to case-control status, and all genotyping results were reviewed manually.

3.2.3 Recreational Physical Activity and Covariate Assessment

Exposure information was obtained from two sources: the interviewer-administered structured questionnaire and laboratory analyses using blood samples to obtain genotypes for CAT and oxidative stress genes. As part of the structured questionnaire participants were asked about their involvement in RPA using a modified instrument developed by Bernstein and colleagues [274]. Women were asked about all activities in which they had engaged for at least one hour per week and three months or more in any year over the life-course. Women who reported never having participated in activity were classified as having no RPA. Among ever RPA participators information on the activity name, the ages the activity was started and stopped, and the number of hours per week and months per year the activity was performed was obtained. Activity data for ever participators were summed across all activities for each year of a woman's life, providing a lifetime composite score of exercise duration from menarche (left truncated) to reference date. We previously reported the main effects of RPA during four etiologically relevant time periods (adolescent, reproductive, postmenopausal, and lifetime RPA) [63]. For the present analyses we assessed the interaction between polymorphisms in oxidative stress genes and two time periods for which the effects for breast cancer were strongest: postmenopausal and lifetime RPA. Given our previous analysis showed no substantial differences by intensity [63], we report RPA using average hours only.

During the interviewer-administered structured questionnaire, participants were additionally queried about their demographic characteristics; reproductive, medical and

environmental histories; cigarette smoking and alcohol use; use of exogenous hormones; energy intake; and select anthropometric measurements. Among eligible cases, clinical data on the characteristics of their breast cancer diagnosis, including hormone receptor (HR) status, were obtained from medical records.

3.2.4 Statistical Methods

We first conducted tests for Hardy-Weinberg equilibrium (HWE) using observed genotype frequencies among Caucasian controls and X2 test with 1 degree of freedom [292]. Unconditional logistic regression [300] was used to estimate odds ratios (ORs) and 95% confidence intervals (CIs) for the independent effects of CAT SNPs, their interactions, and the joint effect of oxidative stress variants and RPA. All CAT SNPs were initially evaluated using a general genetic model, but due to sparse data among women with the homozygous variant genotype, a dominant model (at least one variant allele vs. no variant alleles) was used for analyses of main effects and subsequent interactions.

We identified potential confounders based on the known epidemiology of breast cancer and analysis of causal diagrams [295]. For CAT variants, potential confounders were first degree family history of breast cancer (yes/no), race (categorical), and religion (categorical). As reported in our recently published manuscript [63] that examines effects for RPA on breast cancer incidence, we considered the following potential confounders: education (categorical), family history of breast cancer (yes/no), history of benign breast disease (yes/no), income (categorical), lactation history (ever/never), use of oral contraceptives (ever/never), parity (categorical), and smoking history (never, current, former). Confounders were included in the final model if their inclusion changed the exposure estimate by greater than 10% [301]. None of the above covariates met our criterion (which is consistent with the lack of confounding noted in our previous examination of the main effect of physical activity on breast cancer risk [63]. Additionally adjustment for

body mass did not alter our estimate by greater than 10%. Final models were therefore adjusted only for 5-year age group.

The main effect of CAT variants on breast cancer risk was assessed among all women and within strata of menopausal status when the Breslow-Day p for homogeneity was <0.10 [317]. The effect of each CAT variant was evaluated by HR status stratifying cases into two groups using information on estrogen receptor (ER) and progesterone receptor (PR) status [302]: women with tumors that showed any hormone responsiveness (ER+/PR+, ER+/PR-, and ER-/PR+) and women who showed none (ER-/PR-).

We evaluated potential G*E interactions (both additive and multiplicative) by using indicator terms for those with the genotype only, exposure only, and both the genotype and exposure of interest. For genotype we assessed interactions using a dominant genetic model and for RPA we classified participants into three categories with cut points based on the median value among controls: no RPA, low RPA (< control median), and high RPA (≥ control median). Departures from the multiplicative null were assessed using the likelihood ratio test, comparing a model with and without the interaction terms [317]. Departures from the additive null were estimated by the interaction contrast ratio (ICR). The magnitude of the additive interaction effect was estimated based on the following formula: ICR= OR_{11}–OR_{01}–OR_{10}+1 and its respective confidence interval obtained by ICR ±1.96 SE (ICR) [304]. If the relative risk, as approximated by the OR, for both genotype and exposure differed significantly from the relative risk of either factor alone added together minus 1, we concluded that there was evidence of additive interaction [303]. All analyses were conducted using SAS 9.1 (Cary, NC).

3.3 Results

3.3.1 Hardy Weinberg Equilibrium

Among Caucasian controls, the MAF for CAT SNPs rs4756146 C-, rs2284365 C-
and rs480575 G-alleles were 14%, 25% and 31%, respectively. Allele frequencies were
comparable to those of the CEU HapMap population (8%, 20%, and 30%), although control
genotype distributions for rs4756146 (p=0.02) and rs2284365 (p=0.01) deviated significantly
from HWE. Call rates were >95% for CAT SNPs and we observed good agreement in the
randomly selected duplicates included for quality control (N discordant for rs4756146 [N=3,
8.6%] and rs2284365 [N=0, 0%]) suggesting that deviation from HWE was not due to
genotyping error.

3.3.2 Main SNP Effects

The genotype frequencies and age-adjusted associations with breast cancer risk for
CAT polymorphisms are reported in **Table 3.1**. We observed no substantial associations
between the CAT SNPs rs4756146, rs2284365 or rs480575 and breast cancer risk when
genes were examined individually. However, the Breslow-day test for homogeneity revealed
modification by menopausal status for rs4756146 (p=0.0419): strongest effects were
observed among postmenopausal women; those with CT or CC genotypes had decreased
risk of breast cancer compared to women with TT genotypes (OR=0.77; 95% CI, 0.59-0.99)
(**Table 3.2**). We observed a non-significant increase in risk of breast cancer among
premenopausal women carrying at least one variant allele (OR=1.27; 95% CI, 0.88-1.85).
We also found some suggestion of difference in the effect of rs4756146 by HR status. There
was an 11% risk reduction among pre and postmenopausal HR positive cases combined
(OR=0.89; 95% CI, 0.69-1.15) and a 34% risk reduction among HR negative cases
(OR=0.66; 95% CI, 0.40-1.08) compared to all controls. There was no modification by family
history or religion for any SNP, and no heterogeneity by menopausal status or across tumor

types for CAT rs2284365 and CAT rs480575 (**Table A.20**). Our results did not vary upon restriction to Caucasian women.

3.3.3 SNP-SNP Interactions

We assessed all potential multiplicative interactions between the three newly genotyped CAT polymorphisms described above and a functional CAT polymorphism (rs1001179) previously reported by Ahn and colleagues [156]. Of the six possible 2-way combinations, we found only one potential interaction between rs480575 and rs2284365 (**Table A.38**), although this interaction did not reach statistical significance (a priori α=0.05). We observed a significantly decreased risk of breast cancer among women who carried at least one variant allele for CAT rs480575 and were homozygous for common alleles for CAT rs2284365 (OR=0.69; 95% CI, 0.49-0.96).

3.3.4 Gene-Environment (GxE) Interactions

The OR (95% CI) for breast cancer risk by genotype and RPA are shown in **Table 3.3** along with previously reported postmenopausal age-adjusted main effects for RPA and genetic variants. While we observed similar GxE results for lifetime RPA among all women, the effects were stronger once restricted to postmenopausal participants. This is likely due to the strength of the main effect as we previously found stronger inverse associations for postmenopausal RPA than lifetime RPA [63]. Models are therefore presented among postmenopausal women, using reduced variables for RPA (none, < control median, ≥ control median) and a dominant genetic model.

The association between postmenopausal breast cancer risk and carrying at least one variant CAT allele (rs1001179; CT and TT genotypes) was increased among women who had engaged in > 9.23 hrs/wk of RPA from menopause to reference date (OR=1.61; 95% CI, 1.06-2.45; p for multiplicative interaction = 0.043). There was a modest risk reduction (OR=0.89; 95% CI, 0.59-1.34) among women who were heterozygous or homozygous for the variant allele and moderately active (0.01-9.23 hrs/wk). Despite the

significant interaction the estimate in the no activity group was 1.36 (95% CI, 0.83-1.21). We observed a significant 44% risk reduction (OR=0.56; 95% CI, 0.38-0.84) among postmenopausal women who engaged in low to moderate RPA with at least one G allele (AG and GG combined) in GSTP1 Ile105Val, compared with those with the AA genotype (p for multiplicative interaction = 0.006). Among highly active women there was little effect of genotype on breast cancer risk (OR=1.08; 95% CI, 0.71-1.64). There was some suggestion of an inverse association between the TC and CC genotypes of CAT SNP rs4756146 and postmenopausal breast cancer risk among non-active women (OR=0.57; 95% CI, 0.33-0.98), however, we observed no significant interaction on the multiplicative scale for this or any of the remaining SNP-RPA combinations. Stratum specific effects of genotype were also assessed using splines for RPA. These analyses revealed similar results as our categorical classification of RPA (**Figures A.20-A.21**). Additionally, our models did not support the presence of an additive interaction between any of the 13 polymorphisms and RPA (**Table A.36**).

3.4 Discussion

In this population-based study, women with at least one variant allele in CAT rs4756146 had a 23% reduced risk of postmenopausal breast cancer compared to women with the common TT genotype. The association was not observed among premenopausal women, or when both pre- and postmenopausal women were considered together. Examination of potential interactions between CAT SNPs revealed a significantly decreased risk of breast cancer among women who carried at least one variant allele for rs480575 and were homozygous for common alleles for rs2284365, although a test of formal interaction was not significant. When we examined joint effects of polymorphisms in oxidative stress genes and RPA from menopause to reference date in relation to postmenopausal breast cancer risk, we found some evidence for modification of genotype effect by activity level. A non-statistically significant positive association was observed among women with more than

81

one variant CAT allele (rs1001179). This association was stronger and statistically significant among participants who were highly active. The inverse association between GSTP1 Ile105Val and breast cancer was more pronounced among women who were moderately active. These findings could indicate that lower neutralization of ROS may augment breast cancer risk among a background of high RPA, whereas higher enzymatic activity may result in enhanced risk reduction among women who are moderately physically active. However, given the lack of evidence across other oxidative stress markers these results require additional confirmation.

There are multiple reports of the association between the functional catalase-262 C/T polymorphism and breast cancer incidence in the epidemiologic literature [156-158], but no study to date has assessed the individual or combined effects of CAT tag SNPs (rs4756146, rs2284365, rs480575) and breast cancer risk. We found that the association among women carrying at least one variant allele in CAT SNP rs4756146 varied by menopausal status. While the exact mechanisms needs to be further investigated, it is possible that postmenopausal women (with a lower estrogen milieu) may more greatly benefit from ROS removal. Given the important role of CAT in neutralizing ROS [127], polymorphisms resulting in reduced enzyme activity may alter an individual's ability to counter lipid peroxidation and DNA oxidation thereby influencing cancer risk. However, in light of the marginally significant odds ratios and little evidence for association among the remaining polymorphisms, these results may be due to chance.

Many [156, 157, 162, 299], but not all [315], studies which examined the association between ROS related exposures, genotype and breast cancer risk have shown that both ROS-generating (e.g., cigarette smoking and exogenous hormones) and ROS-opposing factors (e.g., consumption of fruits and vegetables) may interact with endogenous sources of pro- and antioxidants to modify the effects of oxidative stress related genetic polymorphisms on breast cancer risk. Given the more complex physiological effects of

physical activity any interactions with SNPs in the oxidative stress pathway may be challenging to disentangle.

Physical activity is a known inducer of ROS [119, 163, 164] and has been associated with lipid peroxidation among trained athletes [165-167]. The seemingly paradoxical inverse association between physical activity and breast cancer risk may be explained, in part, by the long-term effects of regular exercise. Some studies suggest that while exercise-induced ROS production may be an immediate systemic response to physical activity, the lasting effect of regular exercise training is adaptation of antioxidant capacity [170, 171]. Regular activity has been shown to enhance antioxidant status at multiple levels in both animal models and clinic studies [173, 318-321] and may render cells more resistant to subsequent oxidative insult [172] thereby neutralizing the potentially mutagenic effects of lipid peroxidation [173]. Changes in antioxidant status are proposed to occur even with moderate activity, which parallels our knowledge of the association between physical activity and breast cancer.

We previously reported a non-linear dose response association between RPA and breast cancer risk among post-menopausal women in the LIBCSP [63]. While a significant 30% risk reduction was observed among women in the third quartile of activity (OR=0.70; 95% CI, 0.52-0.95) women in the highest quartile experienced a modest 16% risk reduction (OR=0.84, 95% CI, 0.63-1.13). Our finding is contrary to many previous epidemiologic studies, which report an inverse dose-response association between physical activity and breast cancer risk [42]; however, the high levels of activity reported by women in the LIBCSP permitted us to consider a wider range of effects than prior investigations. One possible explanation for inconsistent findings among highly active women may be the presence of modification by biologically relevant genotypes. While the effect estimates for RPA in quartiles 3 and 4 were not substantially different one could posit that ROS induction among women with very high activity levels could be amplified by reduced antioxidant

83

capacity in relevant genes such as CAT. Moreover, moderate levels of RPA may enhance risk reduction among women who are carriers of alleles known to be related to higher endogenous enzymatic activity such as the GSTP1 Ile105Val variant [322-324]. Although there is a strong biologic rational for the role of exercise in oxidative stress, the lack of modification across other genes may suggest that the observed results are due to chance and should be interpreted with caution.

The effects observed in this study may be due to biases arising from sample selection, errors in recall, or misclassification of genotype. In the LIBCSP, blood donation varied by both age and race [18]. While genotype is likely associated with race, given the small number of non-white women (6.6% non-white) included in our study, racial variations in blood donation is likely negligible. Moreover, analyses restricted to Caucasian women resulted in little change to observed estimates. Inaccurate recall of exposure variables can similarly lead to biased results and is common in case-control studies. However, it is unlikely that misclassification of RPA is differential with respect to genotype. We therefore expect that recall differences by disease status would not substantially bias our interaction parameter estimates. Distributions of two CAT variants deviated significantly from HWE, which may inflate Type I error [325]. We anticipate that factors other than genotyping error (e.g. natural selection or nonrandom mating) may be responsible for the departure from HWE given the comparable allele frequency to the CEU HapMap population, the use of the high-throughput genotyping methods, as well as the high call and concordance rates. This study benefits from the relatively large sample size, which increased power to detect modest associations, perform subgroup analyses, and evaluate joint effects of genotype and RPA. However, even very large studies assessing main effects of genetic variants can quickly become underpowered when examining gene-environment interactions. Although the racial homogeneity of the LIBCSP population enhances internal validity, it is likely to reduce the generalizability of our study results. Despite potential racial variation in genotype frequency

and exposure prevalence, we believe that this study may provide clues about the underlying biologic mechanisms of oxidative stress and RPA which are unlikely to vary by race.

In summary, variant alleles in rs4756146 appear to be associated with reduced breast cancer risk among postmenopausal women. The statistical interaction, on a multiplicative scale, between CAT, postmenopausal RPA and breast cancer may support our biologically plausible hypothesis that ROS generating risk factors act in combination with reduced antioxidant expression to increase the risk of breast cancer. Similarly, the observed interaction between GSTP1 and RPA could suggest that ROS are best neutralized in environments where there is amplified antioxidant capacity either via endogenous or exogenous mechanisms. This study is, to our knowledge, the first to assess the interaction between oxidative stress genotypes and exercise. Our findings may support the link between physical activity, genetic polymorphisms in genes related to antioxidant capacity and breast cancer risk, but given the probability of chance findings, these hypotheses should be explored in other studies with adequate power and equally detailed exposure assessment. Although genotype is non-modifiable, it is encouraging to note that women who were moderately physically active had enhanced risk reduction when they were carriers of alleles related to higher enzymatic activity.

TABLE 3.1 Age Adjusted Odds Ratios and 95% Confidence Intervals for the Association between Catalase Variants and Breast Cancer Risk. The Long Island Breast Cancer Study Project (1996-1997).

Gene (rs)	Genotype	Cases (N=1053) N	%	Controls (N=1102) N	%	OR (95% CI) [a]	
CAT (rs4756146)	TT	774	77.87%	809	75.82%	1.00	Reference
	TC	201	20.22%	229	21.46%	0.93	(0.75, 1.16)
	CC	19	1.91%	29	2.72%	0.68	(0.37, 1.22)
	TC and CC	220	22.13%	258	24.18%	0.90	(0.74, 1.11)
CAT (rs2284365)	TT	589	59.43%	610	57.22%	1.00	Reference
	TC	344	34.71%	371	34.80%	0.98	(0.81, 1.18)
	CC	58	5.85%	85	7.97%	0.72	(0.50, 1.03)
	TC and CC	402	40.56%	456	42.77%	0.93	(0.78, 1.11)
CAT (rs480575)	AA	517	52.54%	504	48.60%	1.00	Reference
	AG	378	38.41%	422	40.69%	0.89	(0.74, 1.08)
	GG	89	9.04%	111	10.70%	0.78	(0.58, 1.06)
	AG and GG	467	47.45%	533	51.39%	0.87	(0.73, 1.04)

[a] Odds ratio (OR) and 95% confidence interval (CI)

TABLE 3.2 Age Adjusted Odds Ratios and 95% Confidence Intervals for the Association between Catalase SNP rs4756146 and Breast Cancer Risk by Menopausal Status. The Long Island Breast Cancer Study Project (1996-1997).

Gene (rs)	Cases	Controls	OR (95% CI) [a]	
CAT (rs4756146) Pre-Menopausal				
TT	241	282	1.00	Reference
CT	70	68	1.32	(0.89, 1.94)
CC	6	6	0.88	(0.27, 2.85)
CT and CC	76	74	1.27	(0.88, 1.85)
CAT (rs4756146) Post-Menopausal				
TT	514	491	1.00	Reference
CT	127	152	0.81	(0.62, 1.06)
CC	12	23	0.51	(0.25, 1.04)
CT and CC	139	175	0.77	(0.59, 0.99)

[a] Odds ratio (OR) and 95% confidence interval (CI)

TABLE 3.3 Age Adjusted Odds Ratios and 95% Confidence Intervals for Multiplicative Effects of Oxidative Stress SNPs and Postmenopausal Recreational Physical Activity on Postmenopausal Breast Cancer Risk. The Long Island Breast Cancer Study Project (1996-1997).

Gene (SNP) major/minor allele Postmenopausal RPA (average hrs/wk) [a]	Homozygous for major allele		At least one copy of minor allele			p for interaction
	Ca/Co [b]	Ref	Ca/Co	OR (95% CI) [c]		
CAT (rs4756146) T/C						
<0.01	133/104	1.00	33/44	0.57	(0.33, 0.98)	0.126
0.01-9.23	169/140	1.00	48/51	0.78	(0.49, 1.24)	
>9.23	137/163	1.00	36/45	1.05	(0.63, 1.76)	
CAT (rs2284365) T/C						
<0.01	104/80	1.00	64/69	0.70	(0.44, 1.13)	0.331
0.01-9.23	126/109	1.00	92/81	0.97	(0.65, 1.44)	
>9.23	104/117	1.00	64/89	0.87	(0.57, 1.35)	
CAT (rs480575) A/G						
<0.01	90/71	1.00	77/76	0.76	(0.48, 1.22)	0.692
0.01-9.23	107/85	1.00	104/98	0.83	(0.55, 1.24)	
>9.23	93/99	1.00	79/100	0.91	(0.60, 1.39)	
CAT (rs1001179) C/T [d]						
<0.01	103/95	1.00	70/53	1.36	(0.83, 2.21)	0.043
0.01-9.23	149/126	1.00	75/71	0.89	(0.59, 1.34)	
>9.23	100/143	1.00	82/69	1.61	(1.06, 2.45)	
COMT (rs4680) G/A [e]						
<0.01	46/37	1.00	128/116	0.81	(0.47, 1.40)	0.446
0.01-9.23	64/54	1.00	162/142	0.99	(0.64, 1.53)	
>9.23	56/52	1.00	130/160	0.78	(0.49, 1.23)	
COMT (rs737865) T/C [f]						
<0.01	89/77	1.00	80/76	0.97	(0.60, 1.56)	0.439
0.01-9.23	109/87	1.00	118/107	0.88	(0.59, 1.30)	
>9.23	77/98	1.00	102/114	1.11	(0.74, 1.68)	
GPX (rs1050450) C/T [g]						

<0.01	97/82	1.00	93/70	1.46	(0.92, 2.34)	0.349
0.01-9.23	107/97	1.00	120/96	1.08	(0.73, 1.61)	
>9.23	79/103	1.00	103/109	1.18	(0.78, 1.79)	
GSTA1 (rs3957356) G/A [h]						
<0.01	49/48	1.00	124/104	1.11	(0.67, 1.84)	0.295
0.01-9.23	76/68	1.00	151/126	1.07	(0.71, 1.63)	
>9.23	59/61	1.00	124/152	0.84	(0.54, 1.30)	
GSTP1 (rs1695) A/G [i]						
<0.01	84/80	1.00	90/68	1.22	(0.77, 1.95)	0.006
0.01-9.23	123/80	1.00	97/109	0.56	(0.38, 0.84)	
>9.23	76/92	1.00	103/120	1.08	(0.71, 1.64)	
GSTM1 (Null vs. Present) [j]						
<0.01	72/73	1.00	86/65	1.40	(0.86, 2.28)	0.387
0.01-9.23	105/99	1.00	111/78	1.38	(0.92, 2.08)	
>9.23	93/108	1.00	82/88	1.12	(0.73, 1.71)	
GSTT1 (Null vs. Present) [k]						
<0.01	124/109	1.00	34/31	0.92	(0.51, 1.64)	0.526
0.01-9.23	175/136	1.00	42/42	0.78	(0.48, 1.28)	
>9.23	139/153	1.00	36/47	0.96	(0.58, 1.60)	
MnSOD (rs4880) T/C [l]						
<0.01	46/39	1.00	128/110	0.87	(0.51, 1.48)	0.195
0.01-9.23	59/57	1.00	164/138	1.17	(0.76, 1.81)	
>9.23	57/51	1.00	125/161	0.74	(0.47, 1.16)	
MPO (rs2333227) G/A [m]						
<0.01	105/95	1.00	69/57	1.16	(0.72, 1.87)	0.119
0.01-9.23	144/112	1.00	82/84	0.76	(0.51, 1.14)	
>9.23	109/136	1.00	74/76	1.16	(0.76, 1.76)	

[a] Postmenopausal RPA ≤ 9.23 hrs/wk (OR=0.99; 95% CI, 0.77-1.26) RPA > 9.23 hrs/wk (OR=0.77; 95% CI, 0.60-0.99)

[b] Cases (Ca) and controls (Co)

[c] Odds ratio (OR) and 95% confidence interval (CI)

[d] Ahn et al. Am J Epidemiol. (2005) Postmenopausal OR=1.15; 95% CI, 0.92-1.43
Note: Previous report used recessive model

[e] Gaudet et al. Breast Cancer Res Treat. (2006) Postmenopausal OR=0.90; 95% CI, 0.70-1.14

[f] Gaudet et al. Breast Cancer Res Treat. (2006) Postmenopausal OR=0.91; 95% CI, 0.73-1.13

[g] Ahn et al. Cancer Epidemiol Biomarkers Prev. (2005) Postmenopausal OR=1.13; 95% CI, 0.91-1.41

[h] Ahn et al. Carcinogenesis. (2006) Postmenopausal OR=1.04; 95% CI, 0.83-1.31

[i] Steck et al. J Nutr. (2007) Postmenopausal OR=0.93; 95% CI, 0.74-1.15

[j] Steck et al. J Nutr. (2007) Postmenopausal OR=1.21; 95% CI, 0.97-1.52

[k] Steck et al. J Nutr. (2007) Postmenopausal OR=0.92; 95% CI, 0.70-1.21

[l] Gaudet et al. Cancer Causes Control. (2005) Postmenopausal OR=0.99; 95% CI, 0.81-1.21

[m] Ahn et al. Cancer Res. (2004) Postmenopausal OR=0.91; 95% CI, 0.73-1.14

Note: Previous report adjusted for age, family history and parity

Chapter IV: Polymorphisms in DNA Repair Genes, Recreational physical Activity and Breast Cancer Risk

4.1 Introduction

Damage to DNA may occur through a variety of endogenous or exogenous processes including oxidative damage through the generation of reactive oxygen species (ROS) [179]. If these compounds are not neutralized by endogenous antioxidants they have the capacity to react with biomolecules causing damage. The integrity of DNA is primarily maintained by four repair pathways that operate on impaired DNA: base excision repair (BER), double strand break (DSB) repair, nucleotide excision repair (NER) and mismatch repair (MMR) [181]. Of these, MMR is the least studied in breast carcinogenesis. MMR improves replication fidelity by correcting DNA polymerase-mediated replication errors [181, 326]. MMR gene dysfunction is proposed to occur through one of two mechanisms: epigenetic gene silencing or genetic mutations in one of six genes associated with MMR [210, 249, 250]. These changes may lead to increased mutations of oncogenes, tumor suppressor genes, and loss of DNA damage-induced apoptosis, therefore facilitating carcinogenesis [251].

Microsatellite instability (MSI), characterized as simple sequence repeats in DNA, is closely associated with MMR deficiency [181, 241] and is reported to be present in some breast tumors [236-240, 327]. Moreover, there is accumulating evidence that reduced tumor expression of MSH2 and MLH1 are related to breast tumor progression and invasion [202, 255, 257]. While these data suggest a potential role for MMR in breast cancer susceptibility [239] there are few epidemiologic studies examining associations between germline polymorphisms in MMR and breast cancer incidence [225, 228, 233-235]. These data are

mixed and previous investigations have been limited by small samples or less than optimal designs making additional studies warranted.

Polymorphisms in MMR or other DNA repair genes may interact with environmental exposures to influence carcinogenesis [183], even if the main effect of repair variants on breast cancer risk is small or null. Physical activity (PA) is known to be associated with breast cancer risk, particularly among postmenopausal women, but the complex physiological effects of exercise make the mechanisms driving the inverse association difficult to disentangle. While PA is an important contributor to ROS production, regular exercise may improve damage repair systems [170, 171, 267, 268], particularly those that operate on single-strand breaks induced by oxidation (i.e. BER, NER and MMR). Animal and clinical studies have shown that several DNA repair enzymes are up-regulated with long term exercise [269-272, 328]. These changes likely result in increased resistance to oxidative damage, diminished DNA impairment and a lower probability of initiating events. This could be tested through epidemiologic analyses assessing the joint effects of low PA and reduced DNA repair capacity on breast cancer incidence, but to date no study has considered this approach despite the biologic plausibility.

We hypothesized that MMR variants related to reduced repair capacity may be associated with increased breast cancer risk. We also anticipate that genotypes associated with reduced DNA repair act synergistically with low levels of PA to increase the risk of breast cancer greater than would be expected by their individual effects. In this report, we aimed to: (1) examine the main effect of variants in MMR on breast cancer risk; (2) examine two-way interactions between SNPs in the MMR pathway and breast cancer incidence; and (3) examine the joint effects of recreational physical activity (RPA) and variants in DNA repair genes from the BER, NER and MMR pathways on breast cancer risk. Secondary aims were to evaluate associations between CAT polymorphisms and breast cancer with cases categorized according to tumor hormone receptor status, as well as to examine the

combined effect of MMR SNPs on breast cancer incidence using a pathway-based approach.

4.2 Materials and Methods

Resources for this project were drawn from the Long Island Breast Cancer Study Project (LIBCSP), a population-based investigation conducted on Long Island, New York. Details of the case-control study have been previously described [18]. This study was approved by the Institutional Review Board of the collaborating institutions.

4.2.1 Study population

LIBCSP case women were English speaking residents of Nassau and Suffolk counties in New York (NY) aged 20-98 years and newly diagnosed with a first primary in-situ or invasive breast cancer between August 1, 1996 and July 31, 1997. Population-based controls, without a personal history of breast cancer, were randomly selected from the same two counties using random digit dialing for women under age 65 and rosters from the Health Care Finance Administration for women ages 65 and older. Controls were frequency matched to the expected age distribution of case women by 5-year age groups. Distributions by race were similar for cases and controls (94% white, 4% black, and 2% other), and are consistent with the resident populations for these NY counties [18].

Exposure and covariate data were obtained through an interviewer-administered structured questionnaire. Interviews were completed for 82.1% (n=1508) of eligible cases and 62.8% (n=1556) of controls. Among the women who completed an interview, blood samples were donated by 73.1% and 73.3% of cases and controls, respectively. Women who donated a blood sample differed somewhat from non-blood donors on key characteristics including age and race; blood donors more likely to be younger in age and white [18]. Genotyping was unavailable for 4.4% of cases and 3.4% of controls primarily due to insufficient DNA quantity. Thus, the final sample includes 1053 case and 1102 control women.

4.2.2 Single Nucleotide Polymorphism (SNP) Selection and Genotyping

We selected five tag SNPs (MSH3: rs1650663; MLH1: rs2286940; MSH2: rs2303428, rs3732182 and rs4583514) and one functional variant (MLH1: rs1799977, a single non-synonymous base pair change affecting splicing regulation) from three genes in the MMR pathway for genotyping. A tagging strategy was employed to maximize our ability to capture genetic variation across the each gene. The National Institute of Environmental Health Service SNPinfo database [284] was used to select Tag SNPs based on data from phase I and II of the International HapMap Project [285]. The CEU population (30 Utah trios with ancestry from northern and western Europe) was selected as the reference panel given the racial homogeneity of the LIBCSP population with DNA available for analyses (93.4% White and 6.6% Non-White) [224]. SNP selection procedures were based on a minor allele frequency (MAF) cutoff value of 5% and r2 threshold minimum of 0.80 as well as a combination of factors including SNP location, MAF and bin size. In addition to the aforementioned MMR SNPs, we selected 14 presumed functional variants from 9 genes in BER and NER pathways to assess interactions with RPA: *ERCC1* (rs3212986), *MGMT* (rs12917, rs2308321 and rs2308327), *OGG1* (rs1052113), *XPA* (rs1800975), *XPC* (rs2228000 and rs2228001), *XPD* (rs1799793 and rs13181), *XPF* (rs1800067), *XPG* (rs17655), *XRCC1* (rs1799782 and rs25487).

A non-fasting 40 mL blood sample was obtained from participants at time of interview and shipped at room temperature, overnight, for processing. Genomic DNA was extracted from mononuclear cells in whole blood separated by Ficoll (Sigma Chemical Co., St. Louis, Missouri). Pelleted cells were frozen at -80° centigrade until DNA isolation by standard phenol, and chloroform/isoamyl alcohol extraction and RNase treatment [18]. Genotyping of newly selected CAT SNPs was accomplished using Taqman assays (Applied Biosystems, Foster City, CA) with 384-well plates. For the remaining SNPs genotyping was performed using several high-throughput genotyping methods, which have varied over the course of

the study. The fluorescence polarization (FP) method, as described by Chen and colleague's [329], was used to genotype ERCC1, MGMT, OGG1, XPC, XPD (rs13181) and XRCC1. XPA and XPD (rs1799793) were genotyped using Taqman assays described above. Genotyping for XPF and XPG was accomplished using Sequenom's high-throughput matrix-assisted laser desorption/ionization time-of-flight mass spectrometry, previously reported by Ahn and colleagues [299]. Controls for genotype and two non-template controls were included on each plate. Samples that were outside the variables defined by the controls were identified as non-informative and retested. Ten percent of samples were distributed throughout the DNA samples for quality control and laboratory personnel blinded to case-control status.

4.2.3 RPA and Covariate Assessment

Other data for this project comes from the interviewer-administered LIBCSP questionnaire and, for cases, medical record abstraction. The comprehensive questionnaire lasted approximately 101 minutes, and was completed on average within 3 months of diagnosis for cases, and 6 months of study identification for controls. Study participants were asked about their demographic characteristics; reproductive, medical and environmental histories; cigarette smoking and alcohol use; use of exogenous hormones; energy intake; participation in PA, and select anthropometric measurements. RPA was assessed using a modified instrument developed by Bernstein and colleagues [274]. Women were screened for RPA participation by replying to the question: "Have you ever participated in any activities or exercises on a regular basis – that is, for at least 1 hour per week for 3 months or more in any year?" Women answering negatively were classified as having no RPA while subjects answering positively were further queried about their involvement. For these women the activity name, the ages the activity was started and stopped, and the number of hours per week and months per year the activity was performed were obtained. Activity data was summed across all activities for each year of a woman's

95

life, providing a composite score of exercise duration from menarche (left truncated) to reference date. In this analysis we assessed the interaction between variants in DNA repair genes and two PA variables for which the effects for breast cancer were strongest in the LIBCSP population [63]: average hours per week of postmenopausal and lifetime RPA. Our findings showed little or no heterogeneity by RPA-intensity [63].

Among case women, clinical data (including hormone receptor [HR] status of the first primary breast cancer) were obtained from abstracting medical records.

4.2.4 Statistical Methods

All analyses were conducted using SAS 9.1 (Cary, NC). Evaluation of Hardy-Weinberg equilibrium (HWE) was conducted using observed genotype frequencies among White controls and X^2 test with 1 degree of freedom [292]. Odds ratios (ORs) and corresponding 95% CIs for the independent effects of MMR SNPs, their interactions, and the joint effect of DNA repair variants and RPA were estimated by unconditional logistic regression [300]. SNPs effects were assessed using a dominant genetic model because of the low prevalence of the homozygous recessive genotype among MMR SNPs.

Main effects of MMR variants on breast cancer risk were assessed among all women combined, and within strata of menopausal status (pre vs. post) when the Breslow-Day p for homogeneity was <0.10 [317]. Effects were also evaluated by breast cancer subtype according to HR status, by stratifying cases into two HR groups using information on estrogen receptor (ER) and progesterone receptor (PR) status. Women who were ER or PR positive were classified as HR+ (which is consistent with considering luminal A and B intrinsic subtypes as a single group), while women with tumors that were negative for both ER and PR receptors were classified as HR- [302].

Two-way interactions between MMR genes and breast cancer risk were assessed using a likelihood ratio test (LRT): the difference of the -2LogL values of logistic models calculated with and without the interaction terms for SNP1 and SNP2 [303]. For our MMR

pathway-based analysis we combined MMR genotypes, and calculated age-adjusted ORs for breast cancer stratifying on the number of 'variant' genotypes: 0-3, 4-7, ≥8 using methods previously described by Mohrenweiser [330].

Additive and multiplicative gene by environment interactions were assessed using indicator terms for women with the genotype only, exposure only, and both the genotype and exposure of interest. A dominant genetic model was used for all SNPs. Among physically active women, participants were classified into categories based on the median average hours per week among controls creating three levels of RPA: no, low (< control medium), and high (≥ control median) RPA. Departures from the multiplicative null were assessed using the LRT, comparing a model with and without the interaction terms [317]. Departures from the additive null were estimated by the interaction contrast ratio (ICR) based on the formula: ICR= OR11–OR01–OR10+1 [303].

Covariates considered as potential confounders for MMR variants were first degree family history of breast cancer (yes/no), race (categorical), and religion (categorical). For RPA, potential covariates included education (categorical), family history of breast cancer (yes/no), history of benign breast disease (yes/no), income (categorical), lactation history (ever/never), use of oral contraceptives (ever/never), parity (categorical), and smoking history (never, current, former) and were included in the final model if their inclusion changed the exposure estimate by > 10% [301]. None of these altered the estimate by greater than 10%. Final models were adjusted only for 5-year age group.

4.3 Results

4.3.1 Hardy Weinberg Equilibrium

Genotype frequencies for polymorphisms in MSH3 (rs1650663) MLH1 (rs1799977, rs2286940) and MSH2 (rs2303428, rs3732182, rs4583514) are reported in Table 1. All genotypes were consistent with HWE.

4.3.2 Main SNP Effects

There was no evidence of a main effect for any of the individual MMR minor alleles on breast cancer risk (**Table 4.1**), and these findings did not change materially with menopausal status (**Table A.19**). When we restricted these analyses to Whites only, results were similar to those observed among the entire study population. For some variants (MLH1 rs1799977, MLH1 rs2286940, and MSH2 rs2303428) there was suggestion of heterogeneity by HR status (**Table A.21**). We observed moderate risk reductions for the minor alleles of rs1799977, rs2286940, and rs2303428 among HR negative cases (OR; 95% CI = 0.77; 0.52-1.12, 0.69; 0.46-1.01 and 0.71; 0.39-1.27, respectively) and approximately null or slightly increased risk among HR positive cases (OR; 95% CI = 1.11; 0.89-1.37, 0.99; 0.78-1.24 and 1.15; 0.86-1.54, respectively). There was no evidence of tumor heterogeneity by HR status for the remaining MMR polymorphisms (**Table A.21**).

4.3.3 Gene-Gene (GxG) Interactions

We evaluated all potential 2-way multiplicative interactions among MMR genes and found one significant effect between MLH1 (rs1799977) and MSH2 (rs2303428) (p for multiplicative interaction = 0.045). The modest decrease in breast cancer risk among minor allele carriers of the MLH1 (OR=0.81; 95% CI, 0.58-1.14) or MSH2 (OR=0.94; 95% CI, 0.77-1.14) was antagonistically reversed among women who carried at least one minor allele in both genes (OR=1.25; 95% CI, 0.88-1.76) compared with women who were homozygous dominant for both MLH1 and MSH2, although the effect was not statistically significant (**Table A.39**).

We found little or no association with breast cancer upon combining genotypes across the MMR pathway (**Table A.25**).

4.3.4 Gene-Environment (GxE) Interactions

The main effects of RPA [63] and DNA repair [122, 222, 224, 259, 260, 331] genes on breast cancer risk have been published previously in the LIBCSP study population [63],

although no study had previously reported DNA repair SNP-RPA interactions. In the LIBCSP, the effect estimate of moderate postmenopausal RPA (<9.24 hrs/wk) was near null (OR=0.99; 95% CI: 0.77-1.26), while high postmenopausal RPA (≥ 9.24 hrs/wk) was associated with reduced risk of postmenopausal breast cancer (OR=0.77; 95% CI: 0.60-0.99) [63]. When we restricted these analyses to women with data available for genotyping we found similar, but more pronounced, effects for moderate and high postmenopausal RPA. SNP-specific main effects for the genes of interest showed no significant association with breast cancer risk. These data are summarized in **Table 4.2** to facilitate understanding of SNP-RPA interactions shown in **Table 4.3** and reported below.

Our models did not support the presence of additive interactions between RPA and DNA repair genotypes (**Tables A.33-A.34**). However, when we considered multiplicative interactions, significant results (p ≤ 0.05) were observed for lifetime RPA among all women and women carrying variants in MGMT, OGG1, XPC, XPF, and XPG (**Table 4.4**). The interaction between MLH1 and lifetime RPA was near significant (p=0.051). Effect estimates were stronger and more consistent once we restricted our models to postmenopausal participants only. As shown in **Table 4.3**, we found statistically significant multiplicative interactions between postmenopausal RPA and several DNA repair gene polymorphisms including: XPC Ala499Val (rs2228000, p=0.048), XPF Arg415Gln (rs1800067, p=0.012), XPG Asp1104His (rs17655, p=0.022) and MLH1 Ile219Val (rs1799977, p=0.010). Postmenopausal women homozygous for the major alleles of XPC Ala499Val or XPF Arg415Gln who engaged in greater than 9.23 hrs/wk of RPA from menopause to reference date experienced statistically significant reductions in breast cancer risk (OR=0.57; 95% CI, 0.38-0.84 and OR=0.64; 95% CI, 0.46-0.89, respectively) compared to women homozygous for the major alleles who were inactive (<0.01 hrs/wk). Postmenopausal women with minor alleles in XPG Asp1104His and MLH1 Ile219Val who were inactive during the postmenopausal years were at greatest risk of breast cancer risk (OR=1.42; 95% CI, 0.90-

2.25 and OR=1.63; 95% CI, 1.04-2.56, respectively) compared to inactive women who carried no minor alleles. Although involvement in RPA attenuated the increased risk associated with minor allele carriers, the effects were not statistically significant.

4.4 Discussion

In the current study, we found no association between MMR variants and breast cancer risk. There was evidence of a significant gene-gene interaction on the multiplicative scale for MLH1 (rs1799977) and MSH2 (rs2303428); women who harbored minor alleles in both genes were at greater risk of breast cancer than women who carried no minor alleles or minor alleles in only one gene. Given both genes are necessary components for the MMR complex our finding of interaction on the multiplicative scale is biologically plausible [210]. When we assessed the combined effects of MMR genes by counting the number of minor alleles, we found no indication of association with breast cancer risk.

Our study also provides some evidence for modification of DNA repair genotype effect by postmenopausal activity level. Highly active women homozygous for the major allele in XPC Ala499Val and XPF Arg415Gln experienced significant breast cancer risk reductions compared to inactive women with the same genotype. In addition, we observed increased breast cancer risk among inactive women with minor alleles in XPG Asp1104His and MLH1 lle219Val compared to inactive women who were homozygous for the major allele. Our findings suggest that low RPA may augment breast cancer risk among women with variant alleles in DNA repair genes and that high levels of RPA could be particularly beneficial to women who are homozygous for the major allele.

While there are several studies that assess the association between MMR variants and cancer risk, particularly for colorectal cancer, few investigations have been reported for breast cancer. Previous breast cancer studies [225, 228, 233-235] have reported significant associations with minor alleles in MMR SNPs, which may vary by menopausal status. But results are inconsistent, conducted among different international populations with varying

genetic profiles, and based on small, select samples. In our larger population-based study conducted among primarily white women on Long Island, NY, we found little evidence for associations between MMR polymorphisms and breast cancer, even when we considered menopausal status. These findings warrant additional investigation in larger studies with diverse populations as our data were suggestive of differences by HR status.

Although previous studies indicate modest effects of DNA repair variants on breast cancer outcomes, it is possible that associations between some DNA repair variants and cancer risk may be apparent only in the presence of DNA damaging/repairing agents. Several investigations of breast cancer, including our own, have been suggestive of interactions between DNA repair polymorphisms and cigarette smoking, alcohol drinking, radiation exposure, body mass, intake of fruits, vegetables and antioxidant micronutrients [5, 122, 228, 229, 259-264, 266]. Data from animal and clinical studies show that DNA repair can be up-regulated with long term exercise, but to our knowledge our report is the first epidemiologic study to consider interactions between PA, DNA repair and breast cancer risk among women.

PA has consistently been shown to increase the formation of reactive oxygen and nitrogen species [119, 163, 164], which may influence carcinogenesis. Regular exercise training can also result in improvement of damage repair systems [170, 171, 267, 268]. A 2002 study by Radak and colleagues [270] showed that regular exercise decreased the accumulation oxidative stress-induced 8-oxodeoxyguanosine (8-oxo-dG) lesions in the skeletal muscle of exercised rats. These lesions are excised exclusively by 8-oxoguanine DNA glycosylase, the enzyme encoded by OGG1 [180]. Similarly, investigators found that 8-oxo-dG repair, measured by the nicking of a ^{32}P-labeled damaged oligonucleotide, increased in the muscle of exercising animals as did the chymotrypsin-like activity of the proteasome complex, a repair enzyme important in the degradation of proteins modified by oxidative stress [270]. Clinical studies have shown that trained cyclists [271] and marathon

101

runners [272] exhibit up-regulation of DNA excision repair enzymes NESP and RAD23A [271], as well as OGG1 [272].

We hypothesized that up-regulation of DNA repair enzymes might be an important part of the exercise induced adaptation process and that these changes could reduce the likelihood of initiating events in breast carcinogenesis. Our study results lend support to this hypothesis, as we found significant multiplicative interactions between variants in several DNA repair genes and both lifetime and postmenopausal RPA. The current data suggest that women who are homozygous for the major allele and engage in high postmenopausal RPA may be at reduced risk of postmenopausal breast cancer compared to their inactive counterparts. Moreover, women who are inactive and harbor minor DNA repair alleles may be at greater risk of breast cancer. While these findings support our hypothesis and are biologically plausible, the functional significance of many DNA repair gene polymorphisms remains largely unknown [216, 229]. Determining functional status of variants is challenging because of differences in study population characteristics, variations in assay used between studies, and limited ability to detect changes in DNA repair capacity based on single SNP analysis [332]. Both replication and additional mechanistic data are necessary to fully understand the PA, DNA repair, breast cancer association.

Our study has several limitations. In the LIBCSP, blood donation varied by both age and race [18]. While genotype is likely associated with race, given the small number of non-white women included in the study racial variations in blood donation is likely negligible. Moreover, when we restricted our analyses to White women, we observed little change in effect estimates. While our ability to generalize findings to non-white racial groups is limited, we believe that this study may provide clues about the underlying biologic mechanisms of DNA repair and RPA, which likely do not vary by race. Inaccurate recall of exposure variables may have biased our study results. However, neither cases nor controls were aware of their genotype at the time of the interview. Although exposure information on RPA

may be differentially recalled by cases and controls, this misclassification would not likely be differential by genotype. Finally, even with a sample size of 2000+ women, we were limited in our ability to detect modest GxE associations. Our results will therefore need to be confirmed in larger studies with similarly detailed assessment of RPA but expanded genetic data to capture additional variability in the MMR pathway. Multiple comparisons is a consideration in this study, as there were no a priori evidence that an association exists between PA, DNA repair variants and breast cancer. Given we selected genes based on their biologic relevance to breast cancer and PA, we were primarily interested in estimating effect estimates and corresponding 95% confidence intervals without adjusting for multiple comparisons. However, this approach could result in chance findings. The population-based study design, hypothesis-driven DNA repair SNP selection and testing, and detailed exposure assessment, are strengths of this study.

In conclusion, we found little evidence to indicate that any of the six MMR polymorphisms investigated were associated with breast cancer risk. We did find evidence of an interaction between two SNPs in MLH1 (rs1799977) and MSH2 (rs2303428) and breast cancer risk. These genes are central to all mismatch recognition and alterations in them have been shown to be the most common mechanism inducing cancer-related MSI [252-254]. They may therefore play an important role in breast carcinogenesis. We previously reported significant inverse associations between RPA performed after menopause and postmenopausal breast cancer risk [63], and, in this report we find that this protection may be partially dependent upon DNA repair status, particularly XPC 499 C-allele, XPF 415 G-allele, XPG 1104 C-allele, and MLH1 219 G-allele status. Our findings provide clues toward understanding the underlying role of DNA repair in the PA-breast cancer association. For example, our results suggest that exercise could result in up-regulation of certain DNA repair enzymes lowering cancer risk. While the functional significance of many DNA repair polymorphisms with respect to breast cancer remains

largely unknown, the observed associations are consistent across multiple indicators of

physical activity and repair pathways reducing the likelihood that these findings are

attributable to chance. Our results therefore merit further investigation.

TABLE 4.1 Age Adjusted Odds Ratios and 95% Confidence Intervals for the Association between Mismatch Repair Genes and Breast Cancer Risk. The Long Island Breast Cancer Study Project (1996-1997).

Gene (rs)	Genotype	Cases (N=1053)		Controls (N=1102)		OR	95% CI
		N	%	N	%		
MSH3 (rs1650663)	TT	497	49.35%	530	49.58%	1.00	Reference
	CT	412	40.91%	429	40.13%	1.02	(0.85, 1.22)
	CC	98	9.73%	110	10.29%	0.98	(0.72, 1.32)
	CT and CC	510	50.64%	539	50.42%	1.01	(0.85, 1.20)
MLH1 (rs1799977)	AA	503	49.75%	542	50.56%	1.00	Reference
	AG	410	40.55%	443	41.32%	0.98	(0.81, 1.17)
	GG	98	9.69%	87	8.12%	1.17	(0.85, 1.60)
	AG and GG	508	50.24%	530	49.44%	1.01	(0.85, 1.20)
MLH1 (rs2286940)	CC	340	33.83%	342	31.84%	1.00	Reference
	CT	467	46.47%	541	50.37%	0.85	(0.70, 1.04)
	TT	198	19.70%	191	17.87%	1.01	(0.79, 1.30)
	CT and TT	665	66.17%	732	68.24%	0.89	(0.74, 1.08)
MSH2 (rs2303428)	TT	828	83.98%	886	84.14%	1.00	Reference
	CT	150	15.21%	155	14.72%	1.07	(0.83, 1.36)
	CC	8	0.81%	12	1.14%	0.66	(0.27, 1.66)
	CT and CC	158	16.02%	167	15.86%	1.04	(0.81, 1.32)
MSH2 (rs3732182)	GG	542	54.36%	574	53.75%	1.00	Reference
	GT	375	37.61%	411	38.48%	0.98	(0.82, 1.18)
	TT	80	8.02%	83	7.77%	1.04	(0.74, 1.45)
	GT and TT	455	45.63%	494	46.25%	0.99	(0.83, 1.18)
MSH2 (rs4583514)	GG	394	39.01%	404	37.69%	1.00	Reference
	AG	485	48.02%	521	48.60%	0.96	(0.80, 1.16)
	AA	131	12.97%	147	13.71%	0.93	(0.70, 1.23)
	AG and AA	616	60.99%	668	62.31%	0.95	(0.80, 1.14)

OR, odds ratio; 95% CI, 95% confidence interval

TABLE 4.2 Age Adjusted Odds Ratios and 95% Confidence Intervals for the Main Effects of DNA Repair Genes on Postmenopausal Breast Cancer Risk. The Long Island Breast Cancer Study Project (1996-1997).

SNPs for Gene-Environment Interactions	OR	95% CI
Gene Effect (Dominant Models)		
Gene (rs) major/minor alleles		
ERCC1 (rs3212986) C/A[1]	1.16	(0.93, 1.44)
MGMT (rs12917) C/T[2]	1.12	(0.88, 1.43)
MGMT (rs2308321) A/G[2]	0.89	(0.69, 1.15)
MGMT (rs2308327) A/G[2]	0.85	(0.65, 1.11)
OGG1 (rs1052133) C/G[3]	0.95	(0.76, 1.18)
XPA (rs1800975) G/A[1]	1.04	(0.84, 1.29)
XPC (rs2228000) C/T[4]	1.07	(0.86, 1.33)
XPC (rs2228001) A/C[4]	0.99	(0.79, 1.25)
XPD (rs1799793) G/A[1]	1.15	(0.92, 1.43)
XPD (rs13181) A/C[5]	1.16	(0.93, 1.44)
XPF (rs1800067) G/A[1]	1.03	(0.77, 1.40)
XPG (rs17655) G/C[1]	0.98	(0.78, 1.22)
XRCC1 (rs1799782) C/T[6]	1.00	(0.72, 1.38)
XRCC1 (rs25487) G/A[6]	0.99	(0.80, 1.24)
MSH3 (rs1650663) T/C	1.07	(0.86, 1.33)
MLH1 (rs1799977) A/G	1.00	(0.80, 1.24)
MLH1 (rs2286940) C/T	0.83	(0.66, 1.05)
MSH2 (rs2303428) T/C	1.00	(0.73, 1.36)
MSH2 (rs3732182) G/T	1.11	(0.89, 1.38)
MSH2 (rs4583514) A/G	1.00	(0.80, 1.24)

OR, odds ratio; 95% CI, 95% confidence interval

[1] Crew et al. Cancer Epidemiol Biomarkers Prev. (2007)

[2] Shen et al. Carcinogenesis. (2005)

[3] Rossner et al. Cancer Epidemiol Biomarkers Prev. (2006)

[4] Shen et al. Eur J. Cancer. (2008)

[5] Terry et al. Cancer Epidemiol Biomarkers Prev. (2004)

[6] Shen et al. Cancer Epidemiol Biomarkers Prev. (2005)

TABLE 4.3 Age Adjusted Odds Ratios and 95% Confidence Intervals for Multiplicative Effects of DNA Repair SNPs and Recreational Physical Activity on Postmenopausal Breast Cancer Risk. The Long Island Breast Cancer Study Project (1996-1997).

Gene (SNP) major/minor alleles Postmenopausal RPA	Homozygous for major allele			At least one copy of minor allele		
	Ca/Co	OR	(95% CI)	Ca/Co	OR	(95% CI)
ERCC1 (rs3212986) C/A						
<0.01 hrs/wk	93/84	1.00	reference	82/71	1.00	(0.64, 1.55)
0.01-9.23 hrs/wk	128/116	0.96	(0.65, 1.43)	104/82	1.14	(0.75, 1.73)
>9.23 hrs/wk	93/114	0.70	(0.46, 1.05)	92/98	0.78	(0.52, 1.19)
MGMT (rs12917) C/T						
<0.01 hrs/wk	137/113	1.00	reference	40/40	0.83	(0.50, 1.38)
0.01-9.23 hrs/wk	163/141	0.96	(0.68, 1.34)	70/59	0.99	(0.64, 1.53)
>9.23 hrs/wk	128/162	0.64	(0.45, 0.90)	58/50	0.87	(0.55, 1.39)
MGMT (rs2308321) A/G						
<0.01 hrs/wk	141/125	1.00	reference	36/28	1.16	(0.66, 2.03)
0.01-9.23 hrs/wk	175/149	1.04	(0.75, 1.45)	58/50	1.05	(0.67, 1.66)
>9.23 hrs/wk	143/159	0.77	(0.55, 1.07)	43/53	0.68	(0.42, 1.10)
MGMT (rs2308327) A/G						
<0.01 hrs/wk	149/130	1.00	reference	29/25	1.04	(0.57, 1.88)
0.01-9.23 hrs/wk	186/156	1.04	(0.76, 1.44)	47/44	0.96	(0.59, 1.55)
>9.23 hrs/wk	145/163	0.75	(0.54, 1.04)	42/50	0.70	(0.43, 1.13)
OGG1 (rs1052133) C/G						
<0.01 hrs/wk	113/90	1.00	reference	64/61	0.82	(0.52, 1.28)
0.01-9.23 hrs/wk	128/122	0.83	(0.57, 1.21)	94/73	1.02	(0.67, 1.55)
>9.23 hrs/wk	106/118	0.69	(0.47, 1.02)	75/91	0.61	(0.40, 0.92)
XPA (rs1800975) G/A						
<0.01 hrs/wk	84/73	1.00	reference	93/81	0.98	(0.63, 1.53)
0.01-9.23 hrs/wk	110/83	1.15	(0.75, 1.77)	121/112	0.93	(0.62, 1.41)
>9.23 hrs/wk	75/97	0.64	(0.41, 0.99)	112/115	0.80	(0.53, 1.22)

XPC (rs2228000) C/T

<0.01 hrs/wk	93/81	1.00	reference	71/67	0.83	(0.53, 1.30)
0.01-9.23 hrs/wk	121/112	0.83	(0.56, 1.22)	109/82	1.11	(0.74, 1.68)
>9.23 hrs/wk	112/115	0.57	(0.38, 0.84)	91/86	0.82	(0.54, 1.25)

XPC (rs2228001) A/C

<0.01 hrs/wk	67/44	1.00	reference	110/108	0.68	(0.42, 1.09)
0.01-9.23 hrs/wk	75/65	0.79	(0.47, 1.32)	156/131	0.79	(0.50, 1.23)
>9.23 hrs/wk	55/71	0.50	(0.29, 0.84)	131/142	0.58	(0.37, 0.91)

XPD (rs1799793) G/A

<0.01 hrs/wk	72/64	1.00	reference	105/90	1.03	(0.66, 1.62)
0.01-9.23 hrs/wk	92/95	0.87	(0.56, 1.37)	138/101	1.21	(0.79, 1.86)
>9.23 hrs/wk	85/88	0.82	(0.52, 1.29)	101/125	0.68	(0.44, 1.06)

XPD (rs13181) A/C

<0.01 hrs/wk	66/68	1.00	reference	112/85	1.36	(0.87, 2.13)
0.01-9.23 hrs/wk	87/88	1.05	(0.67, 1.66)	141/108	1.33	(0.87, 2.05)
>9.23 hrs/wk	75/82	0.90	(0.56, 1.43)	107/130	0.82	(0.53, 1.27)

XPF (rs1800067) G/A

<0.01 hrs/wk	150/119	1.00	reference	21/27	0.57	(0.30, 1.07)
0.01-9.23 hrs/wk	179/168	0.83	(0.60, 1.15)	43/25	1.38	(0.79, 2.40)
>9.23 hrs/wk	150/175	0.64	(0.46, 0.89)	30/30	0.74	(0.42, 1.31)

XPG (rs17655) G/C

<0.01 hrs/wk	88/89	1.00	reference	80/57	1.42	(0.90, 2.25)
0.01-9.23 hrs/wk	126/111	1.14	(0.77, 1.70)	92/81	1.15	(0.75, 1.76)
>9.23 hrs/wk	113/107	0.98	(0.65, 1.47)	65/96	0.68	(0.44, 1.06)

XRCC1 (rs1799782) C/T

<0.01 hrs/wk	166/135	1.00	reference	12/19	0.56	(0.26, 1.21)
0.01-9.23 hrs/wk	205/177	0.95	(0.70, 1.29)	28/22	1.10	(0.60, 2.03)
>9.23 hrs/wk	160/186	0.69	(0.50, 0.94)	26/27	0.69	(0.38, 1.25)

XRCC1 (rs25487) G/A

<0.01 hrs/wk	64/65	1.00	reference	114/90	1.26	(0.81, 1.98)
0.01-9.23 hrs/wk	86/80	1.05	(0.66, 1.68)	147/120	1.26	(0.82, 1.94)
>9.23 hrs/wk	69/78	0.86	(0.53, 1.40)	118/135	0.83	(0.54, 1.28)

MSH3 (rs1650663) T/C

	Ca/Co	OR	(95% CI)	Ca/Co	OR	(95% CI)
<0.01 hrs/wk	97/79	1.00	reference	74/70	0.87	(0.55, 1.36)
0.01-9.23 hrs/wk	103/109	0.80	(0.53, 1.20)	114/83	1.09	(0.72, 1.65)
>9.23 hrs/wk	83/97	0.67	(0.44, 1.03)	94/113	0.65	(0.43, 0.98)
MLH1 (rs1799977) A/G						
<0.01 hrs/wk	72/81	1.00	reference	101/70	1.63	(1.04, 2.56)
0.01-9.23 hrs/wk	107/93	1.29	(0.84, 1.99)	110/99	1.27	(0.83, 1.94)
>9.23 hrs/wk	93/96	1.09	(0.70, 1.68)	85/114	0.78	(0.51, 1.20)
MLH1 (rs2286940) C/T						
<0.01 hrs/wk	50/47	1.00	reference	123/104	1.16	(0.72, 1.89)
0.01-9.23 hrs/wk	76/55	1.37	(0.80, 2.34)	138/139	0.96	(0.60, 1.54)
>9.23 hrs/wk	62/59	1.00	(0.58, 1.73)	114/150	0.70	(0.43, 1.12)
MSH2 (rs2303428) T/C						
<0.01 hrs/wk	140/125	1.00	reference	25/20	1.07	(0.56, 2.05)
0.01-9.23 hrs/wk	174/160	0.97	(0.70, 1.35)	36/31	1.04	(0.60, 1.79)
>9.23 hrs/wk	152/176	0.73	(0.53, 1.02)	23/31	0.64	(0.35, 1.16)
MSH2 (rs3732182) G/T						
<0.01 hrs/wk	98/82	1.00	reference	75/69	0.92	(0.59, 1.43)
0.01-9.23 hrs/wk	108/102	0.89	(0.60, 1.34)	102/90	0.95	(0.63, 1.44)
>9.23 hrs/wk	95/121	0.61	(0.41, 0.92)	80/88	0.76	(0.49, 1.17)
MSH2 (rs4583514) A/G						
<0.01 hrs/wk	73/61	1.00	reference	100/90	0.96	(0.61, 1.51)
0.01-9.23 hrs/wk	81/70	0.96	(0.60, 1.55)	135/124	0.94	(0.62, 1.44)
>9.23 hrs/wk	71/80	0.71	(0.44, 1.15)	107/129	0.68	(0.44, 1.05)

Ca, case; Co, control; OR, odds ratio; CI, confidence interval

TABLE 4.4 Age Adjusted Odds Ratios and 95% Confidence Intervals for Multiplicative Effects of DNA Repair SNPs and Lifetime Recreational Physical Activity on Breast Cancer Risk. The Long Island Breast Cancer Study Project (1996-1997).

Gene (SNP) major/minor alleles Lifetime RPA	Homozygous for major allele			At least one copy of minor allele		
	Ca/Co	OR	95% CI	Ca/Co	OR	95% CI
ERCC1 (rs3212986) C/A						
<0.01 hrs/wk	119/122	1.00	reference	110/101	1.09	(0.75, 1.58)
0.01-6.35 hrs/wk	227/243	1.02	(0.74, 1.39)	178/161	1.20	(0.86, 1.68)
>6.35 hrs/wk	181/218	0.87	(0.63, 1.20)	193/204	0.99	(0.72, 1.37)
MGMT (rs12917) C/T						
<0.01 hrs/wk	173/155	1.00	reference	58/67	0.78	(0.51, 1.18)
0.01-6.35 hrs/wk	300/296	0.98	(0.75, 1.29)	107/111	0.92	(0.65, 1.31)
>6.35 hrs/wk	276/329	0.75	(0.57, 0.99)	109/95	1.06	(0.74, 1.51)
MGMT (rs2308321) A/G						
<0.01 hrs/wk	180/175	1.00	reference	51/46	1.09	(0.69, 1.71)
0.01-6.35 hrs/wk	321/309	1.10	(0.84, 1.43)	86/96	0.92	(0.64, 1.33)
>6.35 hrs/wk	286/322	0.90	(0.69, 1.17)	90/101	0.90	(0.63, 1.28)
MGMT (rs2308327) A/G						
<0.01 hrs/wk	189/180	1.00	reference	44/43	1.00	(0.63, 1.61)
0.01-6.35 hrs/wk	333/320	1.08	(0.83, 1.40)	74/88	0.86	(0.59, 1.24)
>6.35 hrs/wk	294/336	0.87	(0.67, 1.13)	83/89	0.92	(0.63, 1.32)
OGG1 (rs1052133) C/G						
<0.01 hrs/wk	148/132	1.00	reference	84/89	0.84	(0.57, 1.23)
0.01-6.35 hrs/wk	214/262	0.78	(0.58, 1.05)	182/138	1.28	(0.92, 1.77)
>6.35 hrs/wk	226/231	0.91	(0.68, 1.24)	141/187	0.68	(0.49, 0.95)
XPA (rs1800975) G/A						
<0.01 hrs/wk	118/104	1.00	reference	114/118	0.83	(0.57, 1.21)
0.01-6.35 hrs/wk	185/174	1.00	(0.71, 1.40)	218/228	0.90	(0.65, 1.25)

>6.35 hrs/wk	164/186	0.80	(0.57, 1.13)	211/238	0.79	(0.57, 1.09)

XPC (rs2228000) C/T

<0.01 hrs/wk	129/123	1.00	reference	102/99	0.95	(0.65, 1.38)
0.01-6.35 hrs/wk	251/225	1.12	(0.82, 1.53)	153/179	0.88	(0.63, 1.22)
>6.35 hrs/wk	208/252	0.80	(0.59, 1.10)	168/173	0.94	(0.68, 1.30)

XPC (rs2228001) A/C

<0.01 hrs/wk	88/71	1.00	reference	143/150	0.77	(0.52, 1.14)
0.01-6.35 hrs/wk	124/140	0.77	(0.52, 1.15)	280/263	0.93	(0.65, 1.33)
>6.35 hrs/wk	117/136	0.72	(0.48, 1.08)	258/289	0.74	(0.52, 1.06)

XPD (rs1799793) G/A

<0.01 hrs/wk	90/97	1.00	reference	140/125	1.21	(0.83, 1.76)
0.01-6.35 hrs/wk	166/199	0.98	(0.68, 1.40)	237/204	1.35	(0.95, 1.91)
>6.35 hrs/wk	144/173	0.93	(0.65, 1.35)	230/252	1.02	(0.72, 1.43)

XPD (rs13181) A/C

<0.01 hrs/wk	84/97	1.00	reference	149/124	1.41	(0.97, 2.07)
0.01-6.35 hrs/wk	148/177	1.07	(0.74, 1.54)	250/226	1.39	(0.98, 1.97)
>6.35 hrs/wk	131/155	1.02	(0.70, 1.49)	241/269	1.09	(0.77, 1.54)

XPF (rs1800067) G/A

<0.01 hrs/wk	191/170	1.00	reference	31/41	0.64	(0.38, 1.07)
0.01-6.35 hrs/wk	322/315	0.98	(0.75, 1.27)	65/75	0.83	(0.56, 1.24)
>6.35 hrs/wk	304/358	0.78	(0.60, 1.00)	56/54	0.96	(0.62, 1.48)

XPG (rs17655) G/C

<0.01 hrs/wk	117/117	1.00	reference	102/94	1.14	(0.77, 1.67)
0.01-6.35 hrs/wk	205/211	1.06	(0.77, 1.47)	175/171	1.14	(0.81, 1.59)
>6.35 hrs/wk	214/214	1.04	(0.76, 1.44)	140/191	0.78	(0.55, 1.09)

XRCC1 (rs1799782) C/T

<0.01 hrs/wk	212/196	1.00	reference	21/26	0.76	(0.41, 1.40)
0.01-6.35 hrs/wk	347/349	0.99	(0.77, 1.27)	60/58	1.04	(0.69, 1.58)
>6.35 hrs/wk	334/374	0.86	(0.67, 1.11)	42/51	0.74	(0.47, 1.17)

XRCC1 (rs25487) G/A

<0.01 hrs/wk	95/91	1.00	reference	138/132	0.99	(0.68, 1.45)
0.01-6.35 hrs/wk	162/163	1.02	(0.70, 1.47)	245/245	1.03	(0.73, 1.46)
>6.35 hrs/wk	134/166	0.80	(0.55, 1.16)	243/259	0.93	(0.66, 1.30)

MSH3 (rs1650663) T/C

	Ca/Co	OR	(95% CI)	Ca/Co	OR	(95% CI)
<0.01 hrs/wk	119/115	1.00	reference	103/104	0.97	(0.66, 1.41)
0.01-6.35 hrs/wk	179/199	0.94	(0.68, 1.32)	201/187	1.11	(0.80, 1.55)
>6.35 hrs/wk	175/190	0.92	(0.66, 1.28)	181/222	0.82	(0.59, 1.14)

MLH1 (rs1799977) A/G

<0.01 hrs/wk	106/123	1.00	reference	117/97	1.37	(0.94, 2.00)
0.01-6.35 hrs/wk	204/193	1.31	(0.94, 1.82)	179/195	1.13	(0.81, 1.58)
>6.35 hrs/wk	175/198	1.05	(0.75, 1.47)	181/214	1.00	(0.72, 1.39)

MLH1 (rs2286940) C/T

<0.01 hrs/wk	73/78	1.00	reference	150/142	1.12	(0.76, 1.67)
0.01-6.35 hrs/wk	142/117	1.40	(0.93, 2.10)	239/274	1.00	(0.69, 1.44)
>6.35 hrs/wk	112/127	0.99	(0.65, 1.49)	240/284	0.92	(0.64, 1.33)

MSH2 (rs2303428) T/C

<0.01 hrs/wk	180/186	1.00	reference	35/29	1.24	(0.73, 2.13)
0.01-6.35 hrs/wk	315/322	1.08	(0.83, 1.40)	58/62	1.05	(0.69, 1.59)
>6.35 hrs/wk	293/338	0.92	(0.71, 1.19)	56/65	0.92	(0.61, 1.40)

MSH2 (rs3732182) G/T

<0.01 hrs/wk	122/113	1.00	reference	101/106	0.90	(0.62, 1.31)
0.01-6.35 hrs/wk	209/205	1.02	(0.73, 1.41)	166/181	0.93	(0.66, 1.30)
>6.35 hrs/wk	185/228	0.78	(0.56, 1.07)	166/183	0.88	(0.63, 1.24)

MSH2 (rs4583514) A/G

<0.01 hrs/wk	86/84	1.00	reference	138/136	1.02	(0.70, 1.51)
0.01-6.35 hrs/wk	148/143	1.10	(0.75, 1.62)	233/248	1.01	(0.71, 1.44)
>6.35 hrs/wk	141/157	0.92	(0.63, 1.34)	215/252	0.88	(0.62, 1.26)

Ca, case; Co, control; OR, odds ratio; CI, confidence interval

Chapter V: Discussion

The purpose of this dissertation was to explore whether the association between recreational physical activity (RPA) and breast cancer risk was modified by individual variability in genetic polymorphisms from two closely linked pathways: oxidative stress (Chapter 3) and DNA repair (Chapter 4). To address these novel aims, this dissertation employed a candidate gene approach, focusing on potentially functional polymorphisms in pathways shown to be biologically relevant for both physical activity and breast cancer based on previous work in animal and clinical studies. These analyses were accomplished using data from a large, population-based, study of women with comprehensive exposure assessment and genetic data.

Chapter five provides a detailed discussion of the study strengths, limitations and impact in light of the findings presented in chapters 3 and 4. Specifically, this chapter summarizes the study results (5.1), considers the impact of several sources of bias on the study findings (5.2-5.6) and concludes with directions for future research (5.7) and public health impact (5.8).

5.1 Summary of Results

This investigation began by estimating odds ratios for the association between individual genes in the oxidative stress pathway (*CAT*), the MMR pathway (*MLH1, MSH2, and MSH3*) and breast cancer risk. The majority of single SNP associations were non-significant and close to the null, although some displayed moderate associations by menopausal or hormone receptor status. Notably, one SNP in *CAT* (rs4656146) was associated with a 23% reduced risk of postmenopausal breast cancer. Several variants in the MMR pathway (*MLH1* rs1799977, *MLH1* rs2286940, and *MSH2* rs2303428) were

associated with 20-30% risk reduction among HR negative cases (ER and PR negative), but showed weak positive or approximately null associations among HR positive cases. None of the associations were statistically significant. When I considered effects of multiple genetic variants (either GxG or pathway effects) on breast cancer risk I found a significant multiplicative interaction between *MLH1* (rs1799977) and *MSH2* (rs2303428) as well as a positive association between the number of putative high risk alleles in *CAT* and risk of breast cancer. Women who carried 7-8 high-risk alleles had a 69% increased risk of breast cancer compared with women who had ≤ 2 high-risk alleles. Given lack of association among most single variants, these results support the hypothesis that it is important to consider multiple gene/SNP analyses in uncovering breast cancer etiology.

Statistical interactions between RPA and breast cancer-associated SNPs in oxidative stress and DNA repair pathways were evaluated using both multiplicative and additive models. There was evidence to suggest multiplicative interaction between postmenopausal RPA and SNPs in *CAT, GSTP1*, *XPC*, *XPG*, *XPF*, and *MLH1,* supporting previous work in clinical and rodent studies. These investigations showed that physical activity, especially activity that is strenuous or exhaustive, may result in the generation of ROS [119, 163, 164]. However, regular exercise training results in adaptation of antioxidant capacity [170, 171] and improved damage repair systems [267, 268]. The results of this dissertation show that breast cancer risk reductions are greatest among active women with genotypes related to improved endogenous neutralization of ROS or DNA repair. Women without these genotypes experienced little benefit from physical activity, even if they were highly active. Collectively, these findings suggest that exercise induced ROS are best removed in environments where there are sufficient antioxidant and DNA repair capacity, although other mechanisms of interaction may also be present [74] and should be considered in future studies.

5.2 Generalizability of the Study Sample

Despite the population-based sample selection approach utilized for the Long Island Breast Cancer Study Project (LIBCSP), a limitation of the study was its relatively homogenous population. Because of the racial and economic homogeneity of the source population in the geographic location of the parent study, the current study sample consists primarily of postmenopausal white women with high socioeconomic status. The experiences of the study participants, specifically with regard to RPA involvement or genetic profiles, may therefore not be representative of all US women. While this limits external validity, the internal validity of the study is enhanced and data obtained from these analyses will apply to individuals with the highest risk of developing breast cancer – namely white, postmenopausal women. Moreover, this study may provide some clues about the underlying biology of the physical activity-breast cancer association which is unlikely to vary by race, allele frequency or exposure prevalence.

The racial homogeneity of the study population also restricted the ability to evaluate modification by breast cancer subtypes (e.g. Luminal A, Luminal B, Human epidermal growth factor receptor 2, and Basal-like) which are known to vary by both menopausal status and race [307]. The overwhelming majority of the women included in the parent LIBCSP study were classified as ER/PR positive, which is indicative of the luminal subtype and comprises approximately 60% of breast cancer cases [302]. These findings are therefore applicable to the largest proportion of women diagnosed with breast cancer in the U.S.

5.3 Study Power

Small sample size is an inherent limitation of many molecular epidemiology studies. Given the subtle risk often associated with low penetrant polymorphisms [292], large samples (e.g. 500-2000 case-control pairs) are most likely needed to estimate effects of genes [305, 333], when using a pathway-based approach. This study benefited from the

relatively-large sample size, which increased power to detect main effects of both oxidative stress and DNA repair variants, perform subgroup analyses, and examine joint effects of polymorphisms on breast cancer risk. However, even very large studies assessing main effects of genetic variants can quickly become underpowered when examining gene-environment interactions. These analyses generally require 1500-5000 case control pairs [305, 333], depending on several factors including exposure prevalence, allele frequency, the magnitude of effects and the scale assessed for interaction (i.e. additive or multiplicative). Although the LIBCSP had a sample size of 2000+ women, I was restricted in my ability to detect modest gene-environment associations. This was primarily due to the infrequency of the homozygous recessive genotype, causing the cell sizes of some comparisons to be less than optimal.

5.4 Selection Bias

Another limitation of this dissertation was the potential for selection bias related to study participation. Similar to other population-based case control studies, the parent LIBCSP experienced lower participation rates among controls compared to cases (62.7% vs. 82.1%, respectively). Differences in participation among cases and controls are primarily attributed to poor response among elderly control women ≥ 65 years, where a 43.3% response rate was achieved in comparison to the 71.2% response rate among case women. These differences are less evident among women under 65 years of age with an 88.9% and 76.1% response rate among cases and controls, respectively [18]. For these analyses, the underlying assumption is that genotypes among case study participants are most likely representative of genotypes among all cases sampled in the source population, and control participant genotypes are representative of genotypes among all sampled controls despite the differences in participation by age. When tested empirically, genotype frequencies among women <65 years and ≥ 65 years were approximately equal for both age groups, reducing likelihood of potential biases due to age-related non-participation.

116

In the LIBCSP blood donation not only varied by age, but also by race [18]. While genotype is likely unrelated to age in the LIBCSP, genotype frequencies have been observed to vary by race [334], and thus differences in the availability of blood by race could result in selection bias. Given the very small number of non-white women (6% of cases and 8% of controls) included in the study [18], however, any racial variations in blood donation are likely negligible. In fact, effect estimates obtained in the restricted analyses among Caucasian women did not vary materially from those obtained among analyses based on data among all women.

5.5 Information Bias

Information bias refers to a systematic distortion or error that arises from the procedures used for classification or measurement of the disease, the exposure, or other relevant variables. In most molecular epidemiology studies, information bias may be introduced from genetic or self-reported exposure data. For this dissertation the major sources of bias include those associated with genotyping, including laboratory error and SNP selection, and those associated with RPA assessment, including problems associated with participant recall.

5.5.1 Genotyping

A number of steps were taken to minimize genotyping error in LIBCSP. Samples from matched sets were assayed together in the same batch to ensure that effect estimates did not vary because of inter-assay variability. For quality control purposes 10% duplicates of the samples were distributed throughout DNA samples and laboratory personnel were blinded to case control status. Thus, any laboratory or genotyping error that may have resulted in exposure misclassification is likely non-differential.

The approach used to select for and assess genetic polymorphisms may introduce bias and warrant additional discussion. SNPs were selected using a combination of methods to increase the likelihood of identifying polymorphisms associated with breast cancer. Using

117

a candidate-gene approach I pre-selected regions of the genome that were most likely to be associated with both physical activity and breast cancer based on *a priori* knowledge of exercise pathophysiology and breast cancer etiology. This approach to uncovering the genetic architecture of complex traits capitalizes on both the biological understanding of the phenotype, as well as the increased statistical efficiency of association analysis [279]. While useful, this method lacks the flexibility of GWAS which has facilitated the identification of previously undiscovered regions of DNA shown to be associated with cancer outcomes. However, because genome-wide scanning methods are based primarily on statistical associations, ignoring prior knowledge about disease pathobiology [277, 278], and have limited facility in GxE studies, the candidate gene approach was most appropriate for this dissertation.

The selection of tag SNPs (*CAT* and MMR pathway genes) in addition to potentially functional SNPs (MMR pathway genes) maximized the ability to capture genetic variation across a genomic region and allowed for the identification of previously unreported breast cancer associations. Tag SNPs are unlikely to have a direct effect on gene function, but they may be in linkage disequilibrium with one or more un-typed variants which are functional. Although the tagging procedure used in this analysis was designed to capture most of the variation for the population of interest (CEU), I could not evaluate all potential SNPs for each gene of interest or systematically capture the effects of non-SNP variation (e.g. copy number variants, insertion-deletion polymorphisms, and repeat polymorphisms) which may result in some residual unmeasured genetic variation.

5.5.2 RPA Assessment

Most epidemiologic studies are susceptible to differential recall bias, particularly in case-control studies where participants are asked to self-report environmental exposures from the distant past. Questions related to physical activity could be perceived as sensitive inducing social desirability bias. However, considering RPA levels within genotype during

118

the statistical analysis reduces the impact of potential recall bias, given that women were unaware of their genotype status at the time of the interview. While the exposure information on physical activity may be differentially recalled among diseased women, this misclassification is unlikely to vary by genotype [308]. Thus, any association between RPA and breast cancer observed within genotype cannot be simply a product of potential recall bias.

In addition to biases in recall it is often a concern that physical activity questionnaires lack content validity and reliability, specifically when employed in case-control studies [335], and could result in non-differential recall error. Efforts to reduce this error were undertaken during the design and conduct of the parent study, and included use of visual aids and life course calendars to assist women in recalling specific activities and dates of participation. Similarly, LIBCSP investigators used a comprehensive RPA assessment, developed by Dr. Leslie Bernstein, designed specifically for case control studies of physical activity and breast cancer [274]. This instrument has revealed important relationships between exercise and breast cancer risk in several epidemiologic models [274, 309]. The semi open-ended format of the questions in the RPA assessment strengthens this study, by being sufficiently detailed making it difficult for cases, or controls, to gloss over what they are recalling and by providing a comprehensive assessment of activity type, duration and frequency over broad periods of time. The results obtained in the LIBCSP [63] for the main effect of physical activity among postmenopausal women (28% reduction in risk) is consistent with the 25% risk reduction reported in other studies [44, 46] suggesting that the RPA assessment methods employed in the parent LIBCSP were valid. Although short-term prospective evaluations of physical activity may be better for assessing current activity levels, the use of alternative assessments including activity diaries, accelerometers and pedometers are not feasible in large observational population-based studies, and would fail to capture the etiologically relevant time period for breast cancer incidence.

5.6 Multiple Comparisons

Molecular epidemiology studies with a large number of comparisons are likely to find some statistically significant associations by chance alone. A number of formal approaches have been proposed to account for multiple comparisons [303]; however, there is continuing debate on whether and how to correct for these potential errors [336]. The first approach would be to make no correction for multiple testing. This approach would obligate that the SNPs assessed are not highly correlated. Other approaches, such as Bonferonni correction [311] and false discovery rate [337] adjust the alpha level based on the number of statistical comparisons. These methods are often criticized for being overly conservative [303, 338]. They additionally fail to account for study size and power and inadequately address the issue of false positive associations [339]. Post hoc calculations of the false positive report probability are easy to implement and require fewer assumptions than more complex empirical methods. However, it oversimplifies the assumption for the prior distribution, relies on the study p-value and may not useful for assessing GxE interactions [339, 340]. Bayesian adjustment, works by shrinking the estimates of various SNPs across genes toward the geometric mean of the estimates (empirical Bayes) or some pre-specified value (semi-Bayes). Although these methods can provide more accurate and precise estimates than single level models, they are computationally intensive [287, 310] and may provide no added benefit over more direct approaches. Second stage models (e.g. hierarchical models) similarly shrink the estimates towards their respective group means [341], and have been useful in addressing issues of multiple comparisons and GxE interactions in molecular epidemiology studies [310, 341]. Given the strong biologic rationale of this study and low correlation between SNPs (R^2 range between 0.1-0.6), these dissertation analyses did not warrant the added complexity of Bayesian or hierarchal models. Moreover, other methods to correct for multiple comparisons (e.g. Bonferonni) may be unduly conservative and could increase the type II error rate reducing sensitivity [342]. By not correcting for multiple

comparisons it is possible that I have made incorrect determinations regarding the null hypothesis. However, by not adjusting for multiple comparisons I was able to fully explore my biologically driven hypothesis without penalty and was able to identify potentially important findings which can be verified (or refuted) in future studies.

5.7 Future Directions

The results of this dissertation support two main conclusions: 1) there is evidence that some oxidative stress or DNA repair genotypes, alone or in combination, are associated with breast cancer; and 2) allelic variability in genes related to oxidative stress or DNA repair may modify associations between RPA and breast cancer. While the strengths of this study are plentiful (including pathway-specific selection of potential genotypes, comprehensive assessment of lifecourse RPA, homogenous study sample to maximize the limited study power, and population-based selection of the study sample to increase generalizability of the study findings), there are limitations that warrant consideration in planning for future work in this area.

Replication of SNP associations and gene*gene interactions are needed to make strong conclusions about the validity and strength of the association with respect to breast cancer. Data indicate that the first investigation often suggests a stronger genetic effect than is found by subsequent investigations [343]. This overestimation may be due to systematic biases, including publication bias, or population diversity. Although I identified few biases that would substantially impact the observed associations, it would be useful to consider the effect estimates obtained in LIBCSP in the context of associations estimated in other populations. Replication will provide the most comprehensive, un-biased, assessment of the relationship between these variants and breast cancer occurrence.

In a similar manner, the gene*physical activity study results will need to be confirmed in replication studies. The genotype and physical activity variables were 2-level and 3-level variables, respectively, which may have led to imprecise stratified ORs and/or masked trend

effects. To improve on the precision of this investigation future studies should have: (1) expanded genetic data; and (2) equally detailed RPA assessment as the LIBCSP. Ideally, a prospective cohort would be used to address these questions, although such a design may limit detailed assessment of RPA.

Many of the genetic associations observed in this investigation were among tag SNPs with no known function, suggesting that these polymorphisms are in linkage disequilibrium (LD) with the true causal variant. Alternatively, the typed SNPs may be associated with breast cancer through downstream genetic regulation (e.g. splicing, transcriptional regulation, translational regulation, or regulatory miRNA target sites). Future studies would benefit from comprehensive assessments of genomic elements in regions where the breast cancer-associated SNPs are located. This would aid in identifying causal variants and uncovering biological mechanism through which these variants act.

Future studies may also consider simultaneously assessing genetic variation of all polymorphisms in the same or related biologic pathway. For example, studies assessing three-way interactions between DNA repair, oxidative stress and physical activity on breast incidence my facilitate a more comprehensive understanding of the physical activity-breast cancer association. Second stage models (i.e. hierarchical regression models) may be useful in these efforts. Moreover, these methods may increase precision and reduce the likelihood of false positives among relatively small samples [310]. Despite these advantages, however, an increased sample size of the study population would facilitate implementation of this approach.

5.8 Significance

Breast cancer remains an important public health concern in both the United States and abroad. Although risk factor epidemiology has aided in our understanding of breast cancer etiology, many of the underlying mechanisms that tie these factors to disease risk are not fully understood [5]. It is well established that a large number of risk factors,

specifically those related to hormones, play an important role in breast cancer development [5]. But many of these (e.g., menstrual and reproductive history) are not easily modifiable on an individual-level basis. However, a proportion of breast cancer risk can be attributed to those factors which are amendable to intervention. Physical activity has emerged as an important, potentially modifiable, target for breast cancer risk reduction and may account for a large proportion of breast cancer cases. Identifying pathways through which exercise operates could play an important role in advancing the knowledge of breast cancer etiology, and improving risk reduction strategies for breast cancer. Given the widespread accessibility to physical activity in the US [56], a better understanding of the genetic influences underlying the association with breast cancer has become increasingly important.

The proposed study has widespread implications for breast cancer. First, this study was among the first to systematically evaluate associations between common variation in MLH1, MSH2, MSH3 and CAT using a large population-based sample. Uncovering the role of functional and non-functional SNPs in these pathways is important to understanding the etiology of breast cancer. Additional research may aid in identifying causal variants near these SNPs and could be used to define subgroups of susceptible women who may benefit from increased surveillance.

Second, the novel approach of considering interactions between genotypes and recreational physical activity is an important aspect of investigation and may help define subgroups with elevated or decreased breast cancer risk as well as explain dose response associations. For example, the antagonism detected for *CAT* (rs1001179) and RPA suggests that women who carry the TT genotype and also engage in high levels of RPA would experience greater cases of breast cancer than would be expected by the independent effect of RPA on breast cancer. These observations could indicate that interventions to increase physical activity may not provide the expected breast cancer risk reduction if the TT genotype is highly prevalent in the population. Data from this

investigation show significant multiplicative interaction between RPA and several other polymorphisms including variants in *GSTP1*, *XPC*, *XPF*, *XPG* and *MLH1*. Collectively, these findings suggest that exercise induced ROS are best neutralized in environments where there is sufficient antioxidant and DNA repair capacity via endogenous or exogenous mechanisms.

Studies such as this may help to identify women who are particularly susceptible to the beneficial effects of physical activity based on genetic characteristics. Increasing knowledge of the patterns of biological interaction in the population could aid in the identification of new targets for intervention and inform public health recommendations for lowering breast cancer risk.

5.9 Conclusions

The goal of this dissertation was to examine inter individual variation of genes involved in oxidative stress, DNA repair pathways and RPA to improve our understanding of breast cancer etiology, to explain heterogeneity between studies of physical activity and breast cancer and to identify subgroups of women who may be targeted for tailored public health interventions. Findings from this population-based case-control study of women indicate that most individual polymorphisms in *CAT* and the MMR pathway are not associated with breast cancer risk. However, this investigation is the first to show that when variants in catalase or MMR are considered in combination (gene*gene or pathway approaches) there may be some association with breast cancer occurrence. Further, this was the **first** epidemiologic study to explore interactions between genetic variants, physical activity and cancer risk. Data from this dissertation suggest that active women with genotypes related to enhanced endogenous neutralization of ROS or DNA repair capacity may receive the greatest benefits from exercise.

Future studies should improve on the precision of this investigation by including larger numbers of study participants in addition to assessing genetic variation across

124

multiple pathways implicated in pathobiology of physical activity. Although genotype is non-modifiable, it is encouraging to note that women who were moderately physically active had enhanced risk reduction when they were carriers of alleles related to higher enzymatic activity or DNA repair capacity. While highly active women with risk alleles did not experience the same benefit from exercise as their non-risk alleles carrying counterparts, these women were no more increased risk of breast cancer than non-active women with non-risk alleles.

Appendix I: Tables

TABLE A.1 Risk Factors for Breast Cancer
(Adapted from Hankinson *et al.*, 2004)

Risk Factor*	Direction of Effect †
Family history in first degree relative or genetic predisposition	↑↑
High endogenous estrogen levels	↑↑
Menarche at < 12 years vs. > 14 years	↑
Age at first birth > 30 years vs. < 20 years	↑↑
Parity	↓
Lactation (longer durations)	↓
Menopause at > 54 years vs. at < 45 years	↑↑
Current oral contraceptive use	↑
Postmenopausal hormone use	↑
Ionizing radiation exposure	↑↑
Smoking	↑/↓
PAH exposure	↑
Alcohol use (~1 or more drinks/day)	↑
Diet	↑/↓
High BMI (postmenopausal)	↑
High BMI (premenopausal)	↓
Physical activity	↓

*Among Caucasian populations - there is some evidence that these risk factors may vary by race/ethnicity

† Arrows indicate approximate magnitude of the relationship: ↑ slight to moderate increase in risk; ↑↑ moderate to large increase in risk; ↓ slight to moderate decrease in risk; ↑/↓ no clear direction of effect

TABLE A.2 Summary of DNA Repair Mechanisms (Adapted from Goode *et al.,* 2002)

DNA Repair	Damage	Repair	Genetic Polymorphisms
Base Excision Repair (BER)	Oxidized or reduced bases, nonbulky adducts, and damage produced by methylating agents.	The damaged base is removed by base-specific DNA glycosylases and subsequently restored by endonuclease action, DNA synthesis, and ligation	APEX, APE, polynucleotide kinase, XRCC1, and DNA polymerase
Double Strand Break (DSB)	Replication errors and exogenous agents	*Homologous recombination pathway*: DNA ends are resected and strands extended by DNA polymerase *Nonhomologous end-joining repair pathway*: Direct ligation of the two double-strand-break ends	BRCA1, BRCA2, and XRCC3
Nucleotide Excision Repair (NER)	Bulky lesions such as pyrimidine dimers, other photo-products, larger chemical adducts, and cross-links	The damage is recognized and DNA subsequently unwound. The damaged single strand fragment is removed and a new strand is synthesized by DNA polymerase.	XPC, XPD, XPF, ERCC1 and ERCC4
Mismatch Repair (MMR)	Replication errors (base-base or insertion-deletion mismatched) caused by DNA polymerase errors.	The damage recognized, followed by excision, polymerization, and ligation.	MLH1, MSH2, MSH3, MSH6, and PMS2

TABLE A.3 Response Rates by Study Interview Component and Age at Reference among Respondents. Long Island Breast Cancer Study Project, 1996-1997 (Adapted from Gammon *et al.*, 2002)

Study Interview Component		All	Age at reference				
			<45	45-54	55-64	65-74	75+
Cases							
Questionnaires	Main	1508 (100%)	221 (100%)	397 (100%)	371 (100%)	364 (100%)	155 (100%)
	FFQ	1481 (98.2%)	213 (96.4%)	389 (98.0%)	371 (100%)	357 (98.1%)	151 (97.4%)
Biologic Specimens	Blood	1102 (73.1%)	163 (73.8%)	292 (73.6%)	268 (72.2%)	268 (73.6%)	111 (71.6%)
Medical Records	Signed	1473 (97.7%)	213 (96.4%)	383 (96.5%)	365 (98.4%0	357 (98.1%)	155 (100%)
	Retrieved	1402 (95.2%)	206 (96.7%)	361 (94.3%)	344 (94.2%)	343 (96.1%)	147 (94.8%)
Controls							
Questionnaires	Main	1556 (100%)	298 (100%)	413 (100%)	407 (100%)	308 (100%)	130 (100%)
	FFQ	1518 (97.6%)	292 (98.0%)	405 (98.1%)	401 (98.5%)	300 (97.4%)	120 (92.3%)
Biologic Specimens	Blood	1141 (73.3%)	129 (76.8%)	318 (77.0%)	312 (76.7%)	206 (66.9%)	76 (58.5%)

TABLE A.4 Candidate Genes and SNPs for Gene-Environment Models

Pathway	Gene	RS	Functional Change
Oxidative Stress	CAT	rs1001179	TFBS
	COMT	rs4680	nsSNP
		rs737865	TFBS
	GPX1	rs1050450	nsSNP
	GSTA1	rs3957356	TFBS
	GSTM1	--	Gene Deletion
	GSTP1	rs1695	nsSNP
	GSTT1	--	Gene Deletion
	MnSOD	rs4880/s1799725	nsSNP
	MPO	rs2333227	TFBS
DNA Repair	ERCC1/CD3EAP	rs3212986	nsSNP
	MGMT	rs12917	nsSNP
		rs2308321	nsSNP
		rs2308327	nsSNP
	MLH1	rs1799977	nsSNP
	OGG1	rs1052133	nsSNP
	XPA/XRCC4	rs1800975	TFBS
	XPC	rs222800	nsSNP
		rs2228001	nsSNP
	XPD	rs1799793	nsSNP
		rs 13181	nsSNP
	XPF/ERCC4	rs1800067	nsSNP
	XPG/ERCC5	rs17655	nsSNP
	XRCC1	rs1799782	nsSNP
		rs25487	nsSNP

TABLE A.5 SNP Selection for Main Effects Models

Gene	rs	Position	Allele	Function	Average MAF	Average r2	Size	SNPs captured
CAT	rs480575	34424222	G/A	--	0.3136	0.8742	9	rs11032699, rs11032700, rs480575, rs482322, rs484214, rs524154, rs525938, rs769214, rs7943316
CAT	rs2284365	34441549	T/C	--	0.1887	1	8	rs10836244, rs1408035, rs2284365, rs2284368, rs2284369, rs2420388, rs769217, rs769218
CAT	rs4756146	34420315	C/T	--	0.115	1	5	rs16925614, rs2076556, rs2300182, rs4755374, rs4756146
MLH1	rs1799977	37028572	A/G	nsSNP/ Splicing (ESE or ESS)	0.344	1	1	rs1799977
MLH1	rs2286940	37045110	C/T	--	0.4773	0.9564	23	rs11129748, rs1558528, rs1558529, rs2241031, rs2286939, rs2286940, rs3774335, rs3774338, rs3774339, rs3774341, rs3774343, rs4234259, rs4647215, rs4647222, rs4647269, rs4647277, rs4678925, rs6550445, rs6550447, rs748766, rs9852378, rs9852810, rs9876116
MSH2	rs2303428	47557004	C/T	--	0.101	1	4	rs12999145, rs17036614, rs2042649, rs2303428
MSH2	rs3732182	47547210	G/T	--	0.3352	0.9098	11	rs1981928, rs2059520, rs3732182, rs3732183, rs3764959, rs3771280, rs6726691, rs6757035, rs7565513, rs7602094, rs7607076
MSH2	rs4583514	47557389	A/G	--	0.455	0.8268	21	rs10188090, rs11684661, rs11684737, rs11886591, rs2059520, rs2347794, rs3764960, rs3771274, rs3771275, rs3771276, rs3771280, rs3771281*, rs3821227, rs4583514, rs6544990, rs6711675, rs6724382, rs6724876, rs6726832, rs6729015, rs7607312
MSH3	rs1650663	79998953	T/C	--	0.2256	0.9456	61	rs1382543, rs1478834, rs1628627, rs1643655, rs1650648, rs1650650, rs1650651, rs1650652, rs1650653, rs1650654, rs1650658, rs1650663, rs1650665, rs1650666, rs1650667, rs1650670, rs1650692, rs1650737, rs1677626, rs1677628, rs1677629, rs1677630, rs1677635, rs1677638, rs1677639, rs1677641, rs1677643, rs1677650, rs1677652, rs1677653, rs1677667, rs1677703*, rs1824837, rs1824838, rs1824839, rs245332, rs26267, rs26282, rs28027, rs2897262, rs32959, rs6151614, rs6151615, rs6151618, rs6151619, rs6151620, rs836794, rs836795, rs836801, rs836802, rs836806, rs836808, rs836810, rs836813, rs844369, rs863214

*TAG in CEU population

TABLE A.6 Pairwise LD for all Newly Genotyped *CAT* and Mismatch Repair SNPs. The Long Island Breast Cancer Study Project (1996-1997).

Gene	SNP1	SNP2	R^2
MSH2	rs3732182	rs4583514	34
	rs3732182	rs2303428	34
	rs4583514	rs3732182	0.634
	rs4583514	rs4583514	0.148
	rs2303428	rs2303428	0.234
	rs2303428	rs3732182	0.148
MLH1	rs2286940	rs1799977	0.522
	rs1799977	rs2286940	0.522
CAT	rs1001179	rs480575	0.121
	rs4756146	rs2284365	0.527
	rs4756146	rs480575	0.243
	rs2284365	rs4756146	0.527
	rs2284365	rs480575	0.462
	rs480575	rs2284365	0.462
	rs480575	rs4756146	0.243
	rs480575	rs1001179	0.121

A priori criteria for haplotype analysis $R^2 > 0.70$

TABLE A.7 Hardy Weinberg Equilibrium for Newly Genotyped Catalase SNPs. The Long Island Breast Cancer Study Project (1996-1997).

Gene (rs)	Genotype	Frequency	Chi Squared	P-Value	Decision
CAT (rs480575)	AA	467	1.6225	0.2027	*Fail to reject Null*
	AG	393			
	GG	100			
CAT (rs2284365)	CC	78	6.1010	0.0135	*Reject Null*
	CT	344			
	TT	565			
CAT (rs4756146)	CC	27	5.7089	0.0169	*Reject Null*
	CT	214			
	TT	746			

A priori criteria for rejection of null is $p<0.05$

TABLE A.8 Hardy Weinberg Equilibrium for Newly Genotyped Mismatch Repair SNPs. The Long Island Breast Cancer Study Project (1996-1997).

Gene (rs)	Genotype	Frequency	Chi Squared	P-Value	Decision
MSH3 (rs1650663)	CC	110	2.7663	0.0963	*Fail to reject Null*
	CT	429			
	TT	530			
MLH1 (rs1799977)	AA	542	0.0704	0.7908	*Fail to reject Null*
	AG	443			
	GG	87			
MLH1 (rs2286940)	CC	342	0.8279	0.3629	*Fail to reject Null*
	CT	541			
	TT	191			
MSH2 (rs2303428)	CC	12	3.0300	0.0817	*Fail to reject Null*
	CT	155			
	TT	886			
MSH2 (rs3732182)	GG	574	0.6185	0.4316	*Fail to reject Null*
	GT	411			
	TT	83			
MSH2 (rs4583514)	AA	147	1.0494	0.3056	*Fail to reject Null*
	AG	521			
	GG	404			

A priori criteria for rejection of null is $p<0.05$

TABLE A.9 Distribution of Outcome, Main Exposure, and Key Covariates for *CAT* Analysis. The Long Island Breast Cancer Study Project (1996-1997).

Study Variable		Cases (N=1508)		Controls (N=1556)	
		N	%	N	%
CAT (rs4756146)					
	TT	774	77.87%	809	75.82%
	CT	201	20.22%	229	21.46%
	CC	19	1.91%	29	2.72%
	CT and CC	220	22.13%	258	24.18%
	missing	514	--	489	--
CAT (rs2284365)					
	TT	589	59.43%	610	57.22%
	CT	344	34.71%	371	34.80%
	CC	58	5.85%	85	7.97%
	CT and CC	402	40.56%	456	42.77%
	missing	517	--	490	--
CAT (rs480575)					
	AA	517	52.54%	504	48.60%
	AG	378	38.41%	422	40.69%
	GG	89	9.04%	111	10.70%
	AG and GG	467	47.45%	533	51.39%
	missing	524	--	519	--
Race					
	White	1411	93.75%	1429	91.84%
	Black	69	4.58%	85	5.46%
	Other	25	1.66%	42	2.70%
	missing	3	--	0	--
Family History					
	No	1166	79.81%	1321	87.02%
	Yes	295	20.19%	197	12.98%
	missing	47	--	38	--
Religion					
	None	14	0.93%	12	0.77%
	Protestant	360	23.92%	373	24.02%
	Catholic	859	57.08%	916	58.98%
	Jewish	259	17.21%	239	15.39%
	Other	13	0.86%	13	0.84%
	missing	3	--	3	--

TABLE A.10 Distribution of Outcome, Main Exposure, and Key Covariates for Mismatch Repair Analysis. The Long Island Breast Cancer Study Project (1996-1997).

Study Variable		Cases (N=1508)		Controls (N=1556)	
		N	%	N	%
MSH3 (rs1650663)					
	TT	497	49.35%	530	49.58%
	CT	412	40.91%	429	40.13%
	CC	98	9.73%	110	10.29%
	CT and CC	510	50.64%	539	50.42%
	missing	501	--	487	--
MLH1 (rs1799977)					
	AA	503	49.75%	542	50.56%
	AG	410	40.55%	443	41.32%
	GG	98	9.69%	87	8.12%
	AG and GG	508	50.24%	530	49.44%
	missing	497	--	484	--
MLH1 (rs2286940)					
	CC	340	33.83%	342	31.84%
	CT	467	46.47%	541	50.37%
	TT	198	19.70%	191	17.87%
	CT and TT	665	66.17%	732	68.24%
	missing	503	--	482	--
MSH2 (rs2303428)					
	TT	828	83.98%	886	84.14%
	CT	150	15.21%	155	14.72%
	CC	8	0.81%	12	1.14%
	CT and CC	158	16.02%	167	15.86%
	missing	522	--	503	--
MSH2 (rs3732182)					
	GG	542	54.36%	574	53.75%
	GT	375	37.61%	411	38.48%
	TT	80	8.02%	83	7.77%
	GT and TT	455	45.63%	494	46.25%
	missing	511	--	488	--
MSH2 (rs4583514)					
	GG	394	39.01%	404	37.69%
	AG	485	48.02%	521	48.60%
	AA	131	12.97%	147	13.71%
	AG and AA	616	60.99%	668	62.31%
	missing	498	--	484	--
Race					
	White	1411	93.75%	1429	91.84%

Black	69	4.58%	85	5.46%
Other	25	1.66%	42	2.70%
missing	3	--	0	--
Family History				
No	1166	79.81%	1321	87.02%
Yes	295	20.19%	197	12.98%
missing	47	--	38	--
Religion				
None	14	0.93%	12	0.77%
Protestant	360	23.92%	373	24.02%
Catholic	859	57.08%	916	58.98%
Jewish	259	17.21%	239	15.39%
Other	13	0.86%	13	0.84%
missing	3	--	3	--

TABLE A.11 Distribution of Outcome and Main Exposure for *CAT* Analysis Among Whites. The Long Island Breast Cancer Study Project (1996-1997).

Study Variable		Cases (N=1411)		Controls (N=1429)	
		N	%	N	%
CAT (rs4756146)					
	TT	729	77.72%	746	75.58%
	CT	191	20.36%	214	21.68%
	CC	18	1.92%	27	2.74%
	CT and CC	209	22.28%	241	24.42%
	missing	473	--	442	--
CAT (rs2284365)					
	TT	555	59.36%	565	57.24%
	CT	328	35.08%	344	34.85%
	CC	52	5.56%	78	7.90%
	CT and CC	380	40.64%	422	42.75%
	missing	476	--	442	--
CAT (rs480575)					
	AA	496	53.51%	467	48.65%
	AG	352	37.97%	393	40.94%
	GG	79	8.52%	100	10.42%
	AG and GG	431	46.49%	493	51.36%
	missing	484	--	469	--

TABLE A.12 Distribution of Outcome and Main Exposure for Mismatch Repair Analysis among Whites. The Long Island Breast Cancer Study Project (1996-1997).

Study Variable		Cases (N=1411)		Controls (N=1429)	
		N	%	N	%
MSH3 (rs1650663)					
	TT	473	49.89%	488	49.29%
	CT	385	40.61%	403	40.71%
	CC	90	9.49%	99	10.00%
	CT and CC	475	50.10%	502	50.71%
	missing	463	--	439	--
MLH1 (rs1799977)					
	AA	459	48.21%	481	48.54%
	AG	397	41.70%	424	42.79%
	GG	96	10.08%	86	8.68%
	AG and GG	493	51.78%	510	51.47%
	missing	459	--	438	--
MLH1 (rs2286940)					
	CC	302	31.89%	298	30.01%
	CT	453	47.84%	513	51.66%
	TT	192	20.27%	182	18.33%
	CT and TT	645	68.11%	695	69.99%
	missing	464	--	436	--
MSH2 (rs2303428)					
	TT	778	83.66%	817	83.79%
	CT	144	15.48%	147	15.08%
	CC	8	0.86%	11	1.13%
	CT and CC	152	16.34%	158	16.21%
	missing	481	--	454	--
MSH2 (rs3732182)					
	GG	527	56.12%	554	56.07%
	GT	351	37.38%	370	37.45%
	TT	61	6.50%	64	6.48%
	GT and TT	412	43.88%	434	43.93%
	missing	472	--	441	--
MSH2 (rs4583514)					
	GG	386	40.55%	390	39.35%
	AG	459	48.21%	483	48.74%
	AA	107	11.24%	118	11.91%
	AG and AA	566	59.45%	601	60.65%
	missing	459	--	438	--

TABLE A.13 Age Adjusted Odds Ratios and 95% Confidence Intervals for the Association between *CAT* Genes and Breast Cancer Risk. The Long Island Breast Cancer Study Project (1996-1997).

Gene (rs)		Cases	Controls	Pre- and Postmenopausal women		
				OR	95% CI	CLR
CAT (rs4756146)						
	TT	774	809	1.00	Reference	--
	CT	201	229	0.93	(0.75, 1.16)	1.55
	CC	19	29	0.68	(0.37, 1.22)	3.30
	CT and CC	220	258	0.90	(0.74, 1.11)	1.50
CAT (rs2284365)						
	TT	589	610	1.00	Reference	--
	CT	344	371	0.98	(0.81, 1.18)	1.46
	CC	58	85	0.72	(0.50, 1.03)	2.06
	CT and CC	402	456	0.93	(0.78, 1.11)	1.42
CAT (rs480575)						
	AA	517	504	1.00	Reference	--
	AG	378	422	0.89	(0.74, 1.08)	1.46
	GG	89	111	0.78	(0.58, 1.06)	1.83
	AG and GG	467	533	0.87	(0.73, 1.04)	1.42

OR, odds ratio; 95% CI, 95% confidence interval; CLR, Confidence Limit Ratio

TABLE A.14 Age Adjusted Odds Ratios and 95% Confidence Intervals for the Association between Mismatch Repair Genes and Breast Cancer Risk. The Long Island Breast Cancer Study Project (1996-1997).

Gene (rs)		Cases	Controls	Pre- and Postmenopausal women		
				OR	95% CI	CLR
MSH3 (rs1650663)						
	TT	497	530	1.00	Reference	--
	CT	412	429	1.02	(0.85, 1.22)	1.44
	CC	98	110	0.98	(0.72, 1.32)	1.83
	CT and CC	510	539	1.01	(0.85, 1.20)	1.41
MLH1 (rs1799977)						
	AA	503	542	1.00	Reference	--
	AG	410	443	0.98	(0.81, 1.17)	1.44
	GG	98	87	1.17	(0.85, 1.60)	1.88
	AG and GG	508	530	1.01	(0.85, 1.20)	1.41
MLH1 (rs2286940)						
	CC	340	342	1.00	Reference	--
	CT	467	541	0.85	(0.70, 1.04)	1.49
	TT	198	191	1.01	(0.79, 1.30)	1.65
	CT and TT	665	732	0.89	(0.74, 1.08)	1.46
MSH2 (rs2303428)						
	TT	828	886	1.00	Reference	--
	CT	150	155	1.07	(0.83, 1.36)	1.64
	CC	8	12	0.66	(0.27, 1.66)	6.15
	CT and CC	158	167	1.04	(0.81, 1.32)	1.63
MSH2 (rs3732182)						
	GG	542	574	1.00	Reference	--
	GT	375	411	0.98	(0.82, 1.18)	1.44
	TT	80	83	1.04	(0.74, 1.45)	1.96
	GT and TT	455	494	0.99	(0.83, 1.18)	1.42
MSH2 (rs4583514)						
	GG	394	404	1.00	Reference	--
	AG	485	521	0.96	(0.80, 1.16)	1.45
	AA	131	147	0.93	(0.70, 1.23)	1.76
	AG and AA	616	668	0.95	(0.80, 1.14)	1.43

OR, odds ratio; 95% CI, 95% confidence interval; CLR, Confidence Limit Ratio

TABLE A.15 Age Adjusted Odds Ratios and 95% Confidence Intervals for the Association between *CAT* Genes and Breast Cancer Risk Among Whites. The Long Island Breast Cancer Study Project (1996-1997).

Gene (rs)		Cases	Controls	Pre- and Postmenopausal women		
				OR	95% CI	CLR
CAT (rs4756146)						
	TT	774	809	1.00	Reference	--
	CT	201	229	0.93	(0.75, 1.17)	1.56
	CC	12	29	0.69	(0.37, 1.27)	3.43
	CT and CC	213	258	0.90	(0.73, 1.12)	1.53
CAT (rs2284365)						
	TT	589	610	1.00	Reference	--
	CT	344	371	0.99	(0.81, 1.20)	1.48
	CC	58	85	0.69	(0.48, 1.01)	2.10
	CT and CC	402	456	0.93	(0.78, 1.12)	1.44
CAT (rs480575)						
	AA	517	504	1.00	Reference	--
	AG	378	411	0.86	(0.71, 1.05)	1.48
	GG	89	111	0.75	(0.54, 1.04)	1.93
	AG and GG	467	533	0.84	(0.70, 1.01)	1.44

OR, odds ratio; 95% CI, 95% confidence interval; CLR, Confidence Limit Ratio

TABLE A.16 Age Adjusted Odds Ratios and 95% Confidence Intervals for the Association between Mismatch Repair Genes and Breast Cancer Risk among Whites. The Long Island Breast Cancer Study Project (1996-1997).

Gene (rs)		Cases	Controls	Pre- and Postmenopausal women		
				OR	95% CI	CLR
MSH3 (rs1650663)						
	TT	497	530	1.00	Reference	--
	CT	412	429	0.98	(0.81, 1.18)	1.46
	CC	98	110	0.96	(0.70, 1.32)	1.89
	CT and CC	510	539	1.01	(0.85, 1.20)	1.41
MLH1 (rs1799977)						
	AA	503	542	1.00	Reference	--
	AG	410	443	0.97	(0.80, 1.18)	1.48
	GG	98	87	1.14	(0.83, 1.57)	1.89
	AG and GG	508	530	1.01	(0.85, 1.20)	1.41
MLH1 (rs2286940)						
	CC	340	342	1.00	Reference	--
	CT	467	541	0.87	(0.71, 1.07)	1.51
	TT	198	191	1.03	(0.79, 1.34)	1.70
	CT and TT	665	732	0.89	(0.74, 1.08)	1.46
MSH2 (rs2303428)						
	TT	828	886	1.00	Reference	--
	CT	150	155	1.07	(0.83, 1.37)	1.65
	CC	8	12	0.72	(0.28, 1.84)	6.57
	CT and CC	158	167	1.04	(0.81, 1.32)	1.63
MSH2 (rs3732182)						
	GG	542	574	1.00	Reference	--
	GT	375	411	1.01	(0.84, 1.22)	1.45
	TT	80	83	1.02	(0.67, 1.48)	2.21
	GT and TT	455	494	0.99	(0.83, 1.18)	1.42
MSH2 (rs4583514)						
	GG	394	404	1.00	Reference	--
	AG	485	521	0.96	(0.79, 1.16)	0.08
	AA	131	147	0.94	(0.69, 1.26)	1.83
	AG and AA	616	668	0.95	(0.80, 1.14)	1.43

OR, odds ratio; 95% CI, 95% confidence interval; CLR, Confidence Limit Ratio

TABLE A.17 Comparison of *CAT* Crude and Adjusted Odds Ratios by each Covariate. The Long Island Breast Cancer Study Project (1996-1997).

Gene (rs)		Crude Estimate	Race		Family History		Religion	
			Covariate Adjusted OR	ln[coOR]	Covariate Adjusted OR	ln[coOR]	Covariate Adjusted OR	ln[coOR]
CAT (rs4756146)								
	TT	1.00						
	CT and CC	0.90	0.91	0.0110	0.87	0.0110	0.92	0.0220
CAT (rs2284365)								
	TT	1.00						
	CT and CC	0.93	0.93	0.0000	0.90	0.0000	0.93	0.0000
CAT (rs480575)								
	AA	1.00						
	AG and GG	0.87	0.88	0.0114	0.83	0.0114	0.87	0.0000

OR, odds ratio; ln [coOR] = ln [crude estimate/adjusted estimate]

TABLE A.18 Comparison of Mismatch Repair Crude and Adjusted Odds Ratios by each Covariate. The Long Island Breast Cancer Study Project (1996-1997).

Gene (rs)			Race		Family History		Religion	
In (Crude/Adjusted)		Crude Estimate	Covariate Adjusted OR	In[coOR]	Covariate Adjusted OR	In[coOR]	Covariate Adjusted OR	In[coOR]
MSH3 (rs1650663)								
	TT	*1.00*	1.00		1.00		1.00	
	CT and CC	*1.01*	1.00	0.0109	1.00	0.0159	1.02	0.0030
MLH1 (rs1799977)								
	AA	*1.00*	1.00		1.00		1.00	
	AG and GG	*1.01*	0.99	0.0209	1.02	0.0079	1.01	0.0010
MLH1 (rs2286940)								
	CC	*1.00*	1.00		1.00		1.00	
	CT and TT	*0.89*	0.87	0.0182	0.92	0.0332	0.90	0.0090
MSH2 (rs2303428)								
	TT	*1.00*	1.00		1.00		1.00	
	CT and CC	*1.04*	1.04	0.0019	1.06	0.0190	1.04	0.0019
MSH2 (rs3732182)								
	GG	*1.00*	1.00		1.00		1.00	
	GT and TT	*0.99*	1.01	0.0170	0.99	0.0010	1.00	0.0070
MSH2 (rs4583514)								
	GG	*1.00*	1.00		1.00		1.00	
	AG and AA	*0.95*	0.96	0.0115	0.95	0.0063	0.96	0.0063

OR, odds ratio; In [coOR] = In [crude estimate/adjusted estimate]

144

TABLE A.19 Assessment of Effect Measure Modification in Strata of Covariates for *CAT*. The Long Island Breast Cancer Study Project (1996-1997).

Covariate	Stratum Specific		*Breslow-Day p-value*	Mantel Hanzel	
	OR	95% CI		OR	95% CI
CAT (rs4756146)					
Menopausal Status					
Pre	1.20	(0.84, 1.73)	*0.0419*	0.88	(0.72, 1.09)
Post	0.76	(0.59, 0.98)			
Race					
White	0.89	(0.72, 1.10)	*0.9120*	0.89	(0.73, 1.10)
Black	0.87	(0.23, 3.33)			
Other	1.16	(0.35, 3.84)			
Family History					
No	0.88	(0.70, 1.10)	*0.5913*	0.86	(0.70, 1.06)
Yes	0.75	(0.45, 1.27)			
Religion					
None	1.00	(0.10, 9.61)	*0.8265*	0.90	(0.74, 1.11)
Protestant	1.04	(0.67, 1.63)			
Catholic	0.83	(0.64, 1.08)			
Jewish	1.09	(0.64, 1.85)			
Other	0.60	(0.09, 3.99)			
CAT (rs2284365)					
Menopausal Status					
Pre	1.04	(0.77, 1.42)	*0.3031*	0.91	(0.77, 1.09)
Post	0.86	(0.69, 1.07)			
Race					
White	0.92	(0.76, 1.10)	*0.1699*	0.91	(0.77, 1.09)
Black	0.52	(0.21, 1.31)			
Other	2.23	(0.66, 7.54)			
Family History					
No	0.90	(0.74, 1.10)	*0.6068*	0.88	(0.74, 1.06)
Yes	0.79	(0.51, 1.24)			
Religion					
None	1.67	(0.23, 12.22)	*0.8723*	0.91	(0.76, 1.09)
Protestant	0.90	(0.63, 1.30)			
Catholic	0.90	(0.71, 1.13)			
Jewish	0.91	(0.58, 1.41)			
Other	2.40	(0.30, 19.04)			
CAT (rs480575)					
Menopausal Status					
Pre	0.91	(0.67, 1.24)	*0.5237*	0.84	(0.70, 1.01)

Post	0.81	(0.65, 1.00)			
Race					
White	0.82	(0.69, 0.99)	*0.0962**	0.86	(0.72, 1.03)
Black	1.99	(0.83, 4.74)			
Other	1.70	(0.38, 7.50)			
Family History					
No	0.86	(0.71, 1.05)	*0.1791*	0.82	(0.68, 0.97)
Yes	0.62	(0.40, 0.96)			
Religion					
None	1.00	(0.14, 7.10)	*0.9681*	0.85	(0.72, 1.02)
Protestant	0.93	(0.65, 1.33)			
Catholic	0.82	(0.66, 1.04)			
Jewish	0.82	(0.53, 1.29)			
Other	1.50	(0.11, 20.68)			

A priori, criteria for rejecting the null hypothesis of homogeneity across strata of a covariate is $\alpha < 0.10$
* CLR too wide (few 'black' and 'other' study participants)

TABLE A.20 Assessment of Effect Measure Modification in Strata of Covariates for Mismatch Repair. The Long Island Breast Cancer Study Project (1996-1997).

Covariate	Stratum Specific		Breslow-Day p-value	Mantel Hanzel	
	OR	95% CI		OR	95% CI
MSH3 (rs1650663)					
Menopausal Status					
Pre	0.91	(0.67, 1.23)	*0.3519*	1.02	(0.86, 1.21)
Post	1.08	(0.87, 1.34)			
Race					
White	0.98	(0.82, 1.17)	*0.3879*	1.00	(0.84, 1.19)
Black	1.36	(0.56, 3.29)			
Other	2.04	(0.61, 6.84)			
Family History					
No	0.98	(0.81, 1.19)	*0.6465*	1.00	(0.84, 1.19)
Yes	1.10	(0.71, 1.70)			
Religion					
None	1.67	(0.23, 12.22)	*0.7708*	1.02	(0.85, 1.21)
Protestant	1.02	(0.72, 1.44)			
Catholic	0.97	(0.78, 1.22)			
Jewish	1.21	(0.78, 1.89)			
Other	0.33	(0.03, 4.04)			
MLH1 (rs1799977)					
Menopausal Status					
Pre	1.05	(0.77, 1.41)	*0.8887*	1.03	(0.86, 1.22)
Post	1.02	(0.82, 1.26)			
Race					
White	1.01	(0.85, 1.21)	*0.9653*	1.01	(0.85, 1.20)
Black	0.93	(0.34, 2.49)			
Other	1.15	(0.31, 4.26)			
Family History					
No	1.02	(0.84, 1.23)	*0.5113*	1.04	(0.88, 1.24)
Yes	1.19	(0.77, 1.84)			
Religion					
None	2.33	(0.17, 32.58)	*0.5872*	1.04	(0.87, 1.23)
Protestant	0.94	(0.66, 1.34)			
Catholic	1.15	(0.92, 1.44)			
Jewish	0.80	(0.52, 1.24)			
Other	0.90	(0.13, 6.08)			
MLH1 (rs2286940)					
Menopausal Status					
Pre	1.11	(0.81, 1.52)	*0.1719*	0.93	(0.77, 1.12)
Post	0.84	(0.67, 1.07)			
Race					
White	0.92	(0.76, 1.11)	*0.2987*	0.89	(0.74, 1.07)
Black	0.44	(0.18, 1.10)			

Other	0.91	(0.28, 2.89)			
Family History					
No	0.91	(0.74, 1.12)	*0.5685*	0.94	(0.78, 1.13)
Yes	1.05	(0.67, 1.66)			
Religion					
None	2.33	(0.17, 32.58)	*0.2613*	0.92	(0.76, 1.10)
Protestant	0.80	(0.55, 1.15)			
Catholic	1.04	(0.82, 1.33)			
Jewish	0.65	(0.41, 1.04)			
Other	2.50	(0.37, 16.89)			

MSH2 (rs2303428)
Menopausal Status

Pre	1.07	(0.71, 1.59)	*0.7255*	1.01	(0.79, 1.28)
Post	0.97	(0.72, 1.32)			
Race					
White	1.01	(0.79, 1.29)	*0.5960*	1.01	(0.80, 1.29)
Black	0.42	(0.04, 4.21)			
Other	1.60	(0.41, 6.26)			
Family History					
No	1.02	(0.78, 1.32)	*0.6049*	1.04	(0.82, 1.33)
Yes	1.22	(0.64, 2.36)			
Religion					
None	--				
Protestant	0.83	(0.50, 1.38)	*0.6849*	1.02	(0.80, 1.29)
Catholic	1.13	(0.84, 1.54)			
Jewish	0.93	(0.52, 1.69)			
Other	1.17	(0.12, 10.99)			

MSH2 (rs3732182)
Menopausal Status

Pre	0.85	(0.63, 1.16)	*0.2393*	0.99	(0.83, 1.18)
Post	1.07	(0.86, 1.33)			
Race					
White	1.00	(0.83, 1.19)	*0.6467*	0.99	(0.83, 1.18)
Black	0.64	(0.21, 1.96)			
Other	1.37	(0.41, 4.58)			
Family History					
No	0.99	(0.82, 1.20)	*0.7032*	0.98	(0.82, 1.16)
Yes	0.90	(0.58, 1.40)			
Religion					
None	0.40	(0.05, 3.42)	*0.3763*	0.98	(0.82, 1.16)
Protestant	0.86	(0.60, 1.22)			
Catholic	1.01	(0.81, 1.27)			
Jewish	1.02	(0.65, 1.60)			
Other	7.00	(0.61, 79.87)			

MSH2 (rs4583514)
Menopausal Status

Pre	0.88	(0.65, 1.21)	*0.6385*	*0.94*	*(0.79, 1.13)*

	Post	0.97	(0.78, 1.21)		
Race					
	White	0.95	(0.79, 1.14)	*0.7173*	0.96	(0.80, 1.14)
	Black	0.76	(0.18, 3.24)		
	Other	1.59	(0.42, 6.07)		
Family History					
	No	0.95	(0.78, 1.16)	*0.6465*	0.94	(0.78, 1.12)
	Yes	0.85	(0.55, 1.33)		
Religion					
	None	2.33	(0.17, 32.58)	*0.3257*	0.95	(0.79, 1.13)
	Protestant	0.95	(0.66, 1.37)		
	Catholic	0.87	(0.69, 1.10)		
	Jewish	1.14	(0.74, 1.77)		
	Other	7.00	(0.61, 79.87)		

A priori, criteria for rejecting the null hypothesis of homogeneity across strata of a covariate is α<0.10

TABLE A.21 Age Adjusted Odds Ratios and 95% Confidence Intervals for the Association between *CAT* SNPs and Hormone Receptor Status. The Long Island Breast Cancer Study Project (1996-1997).

Gene (rs)		HR Negative			HR Positive		
		Case/Control	OR	95% CI	Case/Control	OR	95% CI
CAT (rs4756146)							
	TT	100/809	1.00	Reference	398/809	1.00	Reference
	CT and CC	21/258	0.66	(0.40, 1.08)	112/258	0.89	(0.69, 1.15)
CAT (rs2284365)							
	TT	76/610	1.00	Reference	303/610	1.00	Reference
	CT and CC	45/456	0.81	(0.54, 1.20)	206/456	0.94	(0.75, 1.17)
CAT (rs480575)							
	AA	61/504	1.00	Reference	272/504	1.00	Reference
	AG and GG	61/533	0.99	(0.67, 1.44)	237/533	0.84	(0.68, 1.04)

OR, odds ratio; 95% CI, 95% confidence interval; CLR, Confidence Limit Ratio; HR Positive, ER+/PR+ ER+/PR- ER-/PR+; HR Negative, ER-/PR-

TABLE A.22 Age Adjusted Odds Ratios and 95% Confidence Intervals for the Association between Mismatch Repair Genes and Hormone Receptor Status. The Long Island Breast Cancer Study Project (1996-1997).

Gene (rs)		HR Negative			HR Positive		
		Case/Control	OR	95% CI	Case/Control	OR	95% CI
MSH3 (rs1650663)							
	TT	62/530	1.00	Reference	272/530	1.00	Reference
	CT and CC	61/539	0.99	(0.68, 1.44)	247/539	0.89	(0.72, 1.10)
MLH1 (rs1799977)							
	AA	70/542	1.00	Reference	248/542	1.00	Reference
	AG and GG	53/530	0.77	(0.52, 1.12)	273/530	1.11	(0.89, 1.37)
MLH1 (rs2286940)							
	CC	49/342	1.00	Reference	162/342	1.00	Reference
	CT and TT	73/732	0.69	(0.46, 1.01)	356/732	0.99	(0.78, 1.24)
MSH2 (rs2303428)							
	TT	106/886	1.00	Reference	422/886	1.00	Reference
	CT and CC	14/167	0.71	(0.39, 1.27)	87/167	1.15	(0.86, 1.54)
MSH2 (rs3732182)							
	GG	69/574	1.00	Reference	278/574	1.00	Reference
	GT and TT	51/494	0.85	(0.58, 1.25)	235/494	1.00	(0.80, 1.24)
MSH2 (rs4583514)							
	GG	44/404	1.00	Reference	210/404	1.00	Reference
	AG and AA	78/668	1.05	(0.71, 1.57)	311/668	0.91	(0.73, 1.13)

OR, odds ratio; 95% CI, 95% confidence interval; CLR, Confidence Limit Ratio; HR Positive, ER+/PR+ ER+/PR- ER-/PR+; HR Negative, ER-/PR-

TABLE A.23 Likelihood Ratio Test for SNP*SNP Interactions in the *CAT* Gene. The Long Island Breast Cancer Study Project (1996-1997).

	No Interaction Term	Interaction Term Included	Difference (B-C)	p-value
SNP*SNP Interactions				
CAT (rs4756146*rs2284365)	2760.515	2759.047	1.468	0.226
CAT (rs4756146*rs480575)	2718.150	2716.790	1.360	0.244
CAT (rs2284365*rs480575)	2717.824	2714.903	2.921	0.087
CAT (rs1001179*rs4756146)	2739.668	2739.606	0.062	0.803
CAT (rs1001179*rs2284365)	2732.163	2732.146	0.017	0.896
CAT (rs1001179*rs480575)	2688.248	2687.495	0.753	0.386

A priori *p* value for interaction on the multiplicative scale is 0.10

TABLE A.24 Likelihood Ratio Test for Gene*Gene and SNP*SNP Interactions in the Mismatch Repair Pathway. The Long Island Breast Cancer Study Project (1996-1997).

	No Interaction Term	Interaction Term Included	Difference	p-value
*Gene*Gene Interactions*				
MSH3 (rs1650663) * *MLH1* (rs1799977)	2825.194	2823.410	1.784	0.182
MSH3 (rs1650663) * *MLH1* (rs2286940)	2815.901	2812.445	3.456	0.063
MSH3 (rs1650663) * *MSH2* (rs2303428)	2770.021	2769.270	0.751	0.386
MSH3 (rs1650663) * *MSH2* (rs3732182)	2798.508	2798.507	0.001	0.975
MSH3 (rs1650663) * *MSH2* (rs4583514)	2819.465	2819.329	0.136	0.712
MLH1 (rs1799977) * *MSH2* (rs2303428)	2774.930	2770.922	4.008	0.045
MLH1 (rs1799977) * *MSH2* (rs3732182)	2810.043	2808.329	1.714	0.190
MLH1 (rs1799977) * *MSH2* (rs4583514)	2830.626	2829.295	1.331	0.249
MLH1 (rs2286940) * *MSH2* (rs2303428)	2770.263	2770.139	0.124	0.725
MLH1 (rs2286940) * *MSH2* (rs3732182)	2803.660	2802.143	1.517	0.218
MLH1 (rs2286940) * *MSH2* (rs4583514)	2824.145	2822.577	1.568	0.210
*SNP*SNP Interactions*				
MLH1 (rs1799977 * rs2286940)	2825.258	2823.759	1.499	0.221
MSH2 (rs2303428 * rs3732182)	2752.403	2751.024	1.379	0.240
MSH2 (rs2303428 * rs4583514)	2773.357	2771.924	1.433	0.231
MSH2 (rs3732182 * rs4583514)	2809.689	2808.243	1.446	0.229

A priori *p* value for interaction on the multiplicative scale is 0.10

TABLE A.25 Decision for Inclusion - Combined Effects of Polymorphisms in the *CAT* Gene on Breast Cancer Risk. The Long Island Breast Cancer Study Project (1996-1997).

Gene (rs)		Cases	Controls	Pre- and Postmenopausal women		Decision using # at-risk alleles
				OR	95% CI	
CAT (rs4756146)						
	TT	774	809	1.00	Reference	*Include CAT-rs4756146 (wild-type T allele is risk allele)*
	CT	201	229	0.92	(0.74, 1.14)	
	CC	12	29	0.65	(0.36, 1.18)	
	CT and CC	213	258	0.89	(0.73, 1.10)	
CAT (rs2284365)						
	TT	589	610	1.00	Reference	*Include CAT-rs2284365 (wild-type T allele is risk allele)*
	CT	344	371	0.97	(0.80, 1.16)	
	CC	58	85	0.70	(0.49, 1.00)	
	CT and CC	402	456	0.92	(0.77, 1.09)	
CAT (rs480575)						
	AA	517	504	1.00	Reference	*Include CAT-rs480575 (wild-type A allele is risk allele)*
	AG	378	411	0.88	(0.73, 1.06)	
	GG	89	111	0.77	(0.57, 1.05)	
	AG and GG	467	533	0.86	(0.72, 1.02)	
CAT (rs1001179)						
	CC	634	696	1.00	Reference	*Include CAT-rs1001179 (variant T allele is risk allele)*
	CT	356	348	1.12	(0.93, 1.35)	
	TT	46	42	1.20	(0.78, 1.85)	
	CT and TT	402	390	1.13	(0.95, 1.35)	

OR, odds ratio; 95% CI, 95% confidence interval; CLR, Confidence Limit Ratio

154

TABLE A.26 Decision for Inclusion - The Analysis of Combined Effects of Polymorphisms in the MMR Pathway on Breast Cancer Risk. The Long Island Breast Cancer Study Project (1996-1997).

Gene (rs)		Cases	Controls	Pre- and Postmenopausal women		Decision using # variant alleles
				OR	95% CI	
MSH3 (rs1650663)						
	TT	497	530	1.00	Reference	*Include MSH3-rs1650663 (TT=0; CT=1; CC=2)*
	CT	412	429	1.03	(0.86, 1.23)	
	CC	98	110	0.97	(0.72, 1.30)	
	CT and CC	510	539	1.01	(0.85, 1.21)	
MLH1 (rs1799977)						
	AA	503	542	1.00	Reference	*Include MLH1-rs1799977 (AA=0; AG=1; GG=2)*
	AG	410	443	0.98	(0.82, 1.18)	
	GG	98	87	1.18	(0.86, 1.62)	
	AG and GG	508	530	1.01	(0.85, 1.21)	
MLH1 (rs2286940)						
	CC	340	342	1.00	Reference	*Include MLH1-rs2286940 (CC=0; CT=1; TT=2)*
	CT	467	541	0.85	(0.70, 1.03)	
	TT	198	191	1.02	(0.79, 1.31)	
	CT and TT	665	732	0.89	(0.74, 1.07)	
MSH2 (rs2303428)						
	TT	828	886	1.00	Reference	*Include MSH2-rs2303428 (TT=0; CT=1; CC=2)*
	CT	150	155	1.04	(0.81, 1.33)	
	CC	8	12	0.72	(0.29, 1.77)	
	CT and CC	158	167	1.02	(0.80, 1.29)	
MSH2 (rs3732182)						
	GG	542	574	1.00	Reference	*Include MSH2-rs3732182 (GG=0; GT=1; TT=2)*
	GT	375	411	0.97	(0.81, 1.17)	
	TT	80	83	1.02	(0.73, 1.41)	
	GT and TT	455	494	0.98	(0.82, 1.17)	
MSH2 (rs4583514)						
	GG	394	404	1.00	Reference	*Include MSH2-rs4583514 (GG=0; AG=1; AA=2)*
	AG	485	521	0.96	(0.80, 1.16)	
	AA	131	147	0.91	(0.70, 1.20)	
	AG and AA	616	668	0.95	(0.80, 1.14)	

OR, odds ratio; 95% CI, 95% confidence interval; CLR, Confidence Limit Ratio

TABLE A.27 Association between Number of High-Risk Alleles in the *CAT* gene and Breast Cancer Risk. The Long Island Breast Cancer Study Project (1996-1997).

number of at risk alleles*	Cases	Controls	Pre- and Postmenopausal women			Pathway Category		
			OR	95% CI	CLR		OR	95% CI
0	14	23	1.00	Reference	--	*0-2 high-risk alleles*	1.00	Reference
1	28	35	1.35	(0.58, 3.12)	5.38			
2	23	40	0.94	(0.40, 2.19)	5.48			
3	115	125	1.54	(0.75, 3.16)	4.21	*3-6 high-risk alleles*	1.36	(0.97, 1.90)
4	148	164	1.52	(0.75, 3.07)	4.09			
5	93	115	1.34	(0.65, 2.77)	4.26			
6	253	274	1.53	(0.76, 3.05)	4.01			
7	209	186	1.85	(0.92, 3.72)	4.04	*≥7 high-risk alleles*	1.69	(1.18, 2.44)
8	45	39	1.90	(0.86, 4.21)	4.90			

OR, odds ratio (age adjusted); 95% CI, 95% confidence interval; CLR, Confidence Limit Ratio

*Includes rs4756146 (wild-type T allele); rs2284365 (wild-type T allele); rs480575 (wild-type A allele); rs1001179 (variant T allele)

TABLE A.28 Association between Number of Variant Alleles in the Mismatch Repair Pathway and Breast Cancer Risk. The Long Island Breast Cancer Study Project (1996-1997).

# variant alleles	Cases	Controls	Pre- and Postmenopausal women			Pathway Category		
			OR	95% CI	CLR		OR	95% CI
0	55	51	1.00	Reference	--	*0-3 variant alleles*	1.00	Reference
1	106	110	0.89	(0.56, 1.43)	2.55			
2	161	184	0.83	(0.54, 1.29)	2.39			
3	183	205	0.83	(0.54, 1.28)	2.37			
4	169	181	0.88	(0.57, 1.36)	2.39	*4-7 variant alleles*	1.01	(0.85, 1.22)
5	132	143	0.88	(0.56, 1.39)	2.48			
6	81	85	0.86	(0.53, 1.41)	2.66			
7	40	47	0.81	(0.45, 1.44)	3.20			
8	24	17	1.28	(0.61, 2.67)	4.38	*8-11 variant alleles*	1.28	(0.76, 2.16)
9	4	7	0.50	(0.14, 1.83)	13.07			
10	4	3	1.24	(0.26, 5.90)	22.69			
11	2	1	1.93	(0.17, 22.12)	130.12			

OR, odds ratio; 95% CI, 95% confidence interval; CLR, Confidence Limit Ratio

TABLE A.29 Age Adjusted Odds Ratios and 95% Confidence Intervals for the Multiplicative Effect of Oxidative Stress SNPs and Lifetime Recreational Physical Activity on Breast Cancer Risk in the Long Island Breast Cancer Study Project (1996-1997).

Gene (SNP) major/minor alleles Lifetime RPA	Homozygous for major allele			At least one copy of minor allele			*p for interaction*
	Ca/Co	OR	95% CI	Ca/Co	OR	95% CI	
CAT (rs4756146) T/C							
<0.01 hrs/wk	169/161	1.00	reference	48/57	0.75	(0.47, 1.18)	*0.247*
0.01-6.35 hrs/wk	296/290	1.00	reference	82/94	0.86	(0.61, 1.21)	
>6.35 hrs/wk	270/318	1.00	reference	82/95	1.05	(0.74, 1.48)	
CAT (rs2284365) T/C							
<0.01 hrs/wk	131/117	1.00	reference	88/101	0.77	(0.52, 1.14)	*0.311*
0.01-6.35 hrs/wk	229/227	1.00	reference	151/163	0.95	(0.71, 1.27)	
>6.35 hrs/wk	198/223	1.00	reference	145/174	1.01	(0.75, 1.36)	
CAT (rs480575) A/G							
<0.01 hrs/wk	115/100	1.00	reference	102/114	0.77	(0.52, 1.14)	*0.479*
0.01-6.35 hrs/wk	197/190	1.00	reference	176/184	0.93	(0.70, 1.25)	
>6.35 hrs/wk	179/191	1.00	reference	167/206	0.90	(0.67, 1.20)	
CAT (rs1001179) C/T							
<0.01 hrs/wk	133/134	1.00	reference	94/81	1.12	(0.75, 1.66)	*0.141*
0.01-6.35 hrs/wk	254/254	1.00	reference	138/144	0.98	(0.73, 1.31)	
>6.35 hrs/wk	216/275	1.00	reference	152/144	1.30	(0.97, 1.75)	
COMT (rs4680) G/A							
<0.01 hrs/wk	61/58	1.00	reference	167/162	0.95	(0.61, 1.46)	*0.490*
0.01-6.35 hrs/wk	114/109	1.00	reference	284/291	0.96	(0.70, 1.32)	
>6.35 hrs/wk	100/99	1.00	reference	271/321	0.83	(0.60, 1.16)	
COMT (rs737865) T/C							
<0.01 hrs/wk	113/105	1.00	reference	111/116	0.85	(0.58, 1.26)	*0.536*
0.01-6.35 hrs/wk	190/190	1.00	reference	204/207	0.96	(0.73, 1.28)	

>6.35 hrs/wk	168/202	1.00	reference	193/218	1.04	(0.78, 1.39)	
GPX (rs1050450) C/T							
<0.01 hrs/wk	110/106	1.00	reference	116/113	0.99	(0.67, 1.46)	*0.736*
0.01-6.35 hrs/wk	178/184	1.00	reference	219/214	1.04	(0.79, 1.39)	
>6.35 hrs/wk	169/205	1.00	reference	198/213	1.07	(0.80, 1.42)	
GSTA1 (rs3957356) G/A							
<0.01 hrs/wk	70/74	1.00	reference	157/145	1.03	(0.68, 1.55)	*0.779*
0.01-6.35 hrs/wk	136/143	1.00	reference	260/255	1.05	(0.78, 1.41)	
>6.35 hrs/wk	120/149	1.00	reference	247/270	1.14	(0.85, 1.55)	
GSTP1 (rs1695) A/G							
<0.01 hrs/wk	116/107	1.00	reference	111/106	0.86	(0.58, 1.27)	*0.269*
0.01-6.35 hrs/wk	206/184	1.00	reference	183/203	0.80	(0.60, 1.06)	
>6.35 hrs/wk	167/195	1.00	reference	193/221	1.03	(0.77, 1.38)	
GSTM1 (gene deletion)							
<0.01 hrs/wk	109/111	1.00	reference	102/87	1.17	(0.78, 1.76)	*0.381*
0.01-6.35 hrs/wk	182/197	1.00	reference	190/170	1.22	(0.91, 1.64)	
>6.35 hrs/wk	188/215	1.00	reference	158/171	1.05	(0.78, 1.42)	
GSTT1 (gene deletion)							
<0.01 hrs/wk	163/155	1.00	reference	49/46	1.00	(0.62, 1.63)	*0.664*
0.01-6.35 hrs/wk	292/288	1.00	reference	83/85	0.92	(0.65, 1.30)	
>6.35 hrs/wk	287/311	1.00	reference	64/81	0.91	(0.63, 1.32)	
MnSOD (rs4880) T/C							
<0.01 hrs/wk	48/55	1.00	reference	178/160	1.20	(0.76, 1.91)	*0.383*
0.01-6.35 hrs/wk	105/96	1.00	reference	288/302	0.86	(0.62, 1.19)	
>6.35 hrs/wk	100/108	1.00	reference	264/309	0.97	(0.70, 1.34)	
MPO (rs2333227) G/A							
<0.01 hrs/wk	142/127	1.00	reference	86/93	0.84	(0.57, 1.25)	*0.353*
0.01-6.35 hrs/wk	247/232	1.00	reference	151/167	0.87	(0.65, 1.16)	
>6.35 hrs/wk	226/257	1.00	reference	140/161	1.00	(0.74, 1.34)	

OR, odds ratio; 95% CI, 95% confidence interval

A priori, criteria for interaction on the multiplicative scale is α<0.10

TABLE A.30 Age Adjusted Odds Ratios and 95% Confidence Intervals for Multiplicative Effects of Oxidative Stress SNPs and Postmenopausal Recreational Physical Activity on Postmenopausal Breast Cancer Risk. The Long Island Breast Cancer Study Project (1996-1997).

Gene (SNP) major/minor alleles Postmenopausal RPA	Homozygous for major allele			At least one copy of minor allele			p for interaction
	Ca/Co	OR	95% CI	Ca/Co	OR	95% CI	
CAT (rs4756146) T/C							
<0.01 hrs/wk	133/104	1.00	reference	33/44	0.57	(0.33, 0.98)	*0.126*
0.01-9.23 hrs/wk	169/140	1.00	reference	48/51	0.78	(0.49, 1.24)	
>9.23 hrs/wk	137/163	1.00	reference	36/45	1.05	(0.63, 1.76)	
CAT (rs2284365) T/C							
<0.01 hrs/wk	104/80	1.00	reference	64/69	0.70	(0.44, 1.13)	*0.331*
0.01-9.23 hrs/wk	126/109	1.00	reference	92/81	0.97	(0.65, 1.44)	
>9.23 hrs/wk	104/117	1.00	reference	64/89	0.87	(0.57, 1.35)	
CAT (rs480575) A/G							
<0.01 hrs/wk	90/71	1.00	reference	77/76	0.76	(0.48, 1.22)	*0.692*
0.01-9.23 hrs/wk	107/85	1.00	reference	104/98	0.83	(0.55, 1.24)	
>9.23 hrs/wk	93/99	1.00	reference	79/100	0.91	(0.60, 1.39)	
CAT (rs1001179) C/T							
<0.01 hrs/wk	103/95	1.00	reference	70/53	1.36	(0.83, 2.21)	*0.043*
0.01-9.23 hrs/wk	149/126	1.00	reference	75/71	0.89	(0.59, 1.34)	
>9.23 hrs/wk	100/143	1.00	reference	82/69	1.61	(1.06, 2.45)	
COMT (rs4680) G/A							
<0.01 hrs/wk	46/37	1.00	reference	128/116	0.81	(0.47, 1.40)	*0.446*
0.01-9.23 hrs/wk	64/54	1.00	reference	162/142	0.99	(0.64, 1.53)	
>9.23 hrs/wk	56/52	1.00	reference	130/160	0.78	(0.49, 1.23)	
COMT (rs737865) T/C							
<0.01 hrs/wk	89/77	1.00	reference	80/76	0.97	(0.60, 1.56)	*0.439*
0.01-9.23 hrs/wk	109/87	1.00	reference	118/107	0.88	(0.59, 1.30)	

>9.23 hrs/wk	77/98	1.00	reference	102/114	1.11	(0.74, 1.68)	
GPX (rs1050450) C/T							
<0.01 hrs/wk	97/82	1.00	reference	93/70	1.46	(0.92, 2.34)	*0.349*
0.01-9.23 hrs/wk	107/97	1.00	reference	120/96	1.08	(0.73, 1.61)	
>9.23 hrs/wk	79/103	1.00	reference	103/109	1.18	(0.78, 1.79)	
GSTA1 (rs3957356) G/A							
<0.01 hrs/wk	49/48	1.00	reference	124/104	1.11	(0.67, 1.84)	*0.295*
0.01-9.23 hrs/wk	76/68	1.00	reference	151/126	1.07	(0.71, 1.63)	
>9.23 hrs/wk	59/61	1.00	reference	124/152	0.84	(0.54, 1.30)	
GSTP1 (rs1695) A/G							
<0.01 hrs/wk	84/80	1.00	reference	90/68	1.22	(0.77, 1.95)	*0.006*
0.01-9.23 hrs/wk	123/80	1.00	reference	97/109	0.56	(0.38, 0.84)	
>9.23 hrs/wk	76/92	1.00	reference	103/120	1.08	(0.71, 1.64)	
GSTM1 (gene deletion)							
<0.01 hrs/wk	72/73	1.00	reference	86/65	1.40	(0.86, 2.28)	*0.387*
0.01-9.23 hrs/wk	105/99	1.00	reference	111/78	1.38	(0.92, 2.08)	
>9.23 hrs/wk	93/108	1.00	reference	82/88	1.12	(0.73, 1.71)	
GSTT1 (gene deletion)							
<0.01 hrs/wk	124/109	1.00	reference	34/31	0.92	(0.51, 1.64)	*0.526*
0.01-9.23 hrs/wk	175/136	1.00	reference	42/42	0.78	(0.48, 1.28)	
>9.23 hrs/wk	139/153	1.00	reference	36/47	0.96	(0.58, 1.60)	
MnSOD (rs4880) T/C							
<0.01 hrs/wk	46/39	1.00	reference	128/110	0.87	(0.51, 1.48)	*0.219*
0.01-9.23 hrs/wk	59/57	1.00	reference	164/138	1.17	(0.76, 1.81)	
>9.23 hrs/wk	57/51	1.00	reference	125/161	0.74	(0.47, 1.16)	
MPO (rs2333227) G/A							
<0.01 hrs/wk	105/95	1.00	reference	69/57	1.16	(0.72, 1.87)	*0.119*
0.01-9.23 hrs/wk	144/112	1.00	reference	82/84	0.76	(0.51, 1.14)	
>9.23 hrs/wk	109/136	1.00	reference	74/76	1.16	(0.76, 1.76)	

OR, odds ratio; 95% CI, 95% confidence interval

A priori, criteria for interaction on the multiplicative scale is α<0.10

TABLE A.31 Age Adjusted Odds Ratios and 95% Confidence Intervals for Multiplicative Effects of DNA Repair SNPs and Lifetime Recreational Physical Activity on Breast Cancer Risk. The Long Island Breast Cancer Study Project (1996-1997).

Gene (SNP) major/minor alleles Postmenopausal RPA	Homozygous for major allele			At least one copy of minor allele			p for interaction
	Ca/Co	OR	95% CI	Ca/Co	OR	95% CI	
ERCC1 (rs3212986) C/A							
<0.01 hrs/wk	119/122	1.00	reference	110/101	1.06	(0.72, 1.55)	*0.665*
0.01-6.35 hrs/wk	227/243	1.00	reference	178/161	1.16	(0.87, 1.54)	
>6.35 hrs/wk	181/218	1.00	reference	193/204	1.16	(0.87, 1.54)	
MGMT (rs12917) C/T							
<0.01 hrs/wk	173/155	1.00	reference	58/67	0.81	(0.53, 1.23)	*0.022*
0.01-6.35 hrs/wk	300/296	1.00	reference	107/111	0.92	(0.67, 1.26)	
>6.35 hrs/wk	276/329	1.00	reference	109/95	1.40	(1.01, 1.94)	
MGMT (rs2308321) A/G							
<0.01 hrs/wk	180/175	1.00	reference	51/46	1.05	(0.66, 1.68)	*0.324*
0.01-6.35 hrs/wk	321/309	1.00	reference	86/96	0.81	(0.58, 1.13)	
>6.35 hrs/wk	286/322	1.00	reference	90/101	0.98	(0.70, 1.37)	
MGMT (rs2308327) A/G							
<0.01 hrs/wk	189/180	1.00	reference	44/43	0.99	(0.61, 1.60)	*0.217*
0.01-6.35 hrs/wk	333/320	1.00	reference	74/88	0.76	(0.54, 1.08)	
>6.35 hrs/wk	294/336	1.00	reference	83/89	1.05	(0.74, 1.48)	
MLH1 (rs1799977) A/G							
<0.01 hrs/wk	106/123	1.00	reference	117/97	1.37	(0.92, 2.03)	*0.051*
0.01-6.35 hrs/wk	204/193	1.00	reference	179/195	0.84	(0.63, 1.12)	
>6.35 hrs/wk	175/198	1.00	reference	181/214	0.95	(0.71, 1.26)	
MLH1 (rs2286940) C/T							
<0.01 hrs/wk	73/78	1.00	reference	150/142	1.13	(0.75, 1.70)	*0.058*
0.01-6.35 hrs/wk	142/117	1.00	reference	239/274	0.69	(0.51, 0.94)	

	>6.35 hrs/wk	112/127	1.00	reference	240/284	0.92	(0.67, 1.26)	
MSH2 (rs2303428) T/C								
	<0.01 hrs/wk	180/186	1.00	reference	35/29	1.33	(0.76, 2.33)	*0.392*
	0.01-6.35 hrs/wk	315/322	1.00	reference	58/62	0.97	(0.65, 1.45)	
	>6.35 hrs/wk	293/338	1.00	reference	56/65	1.01	(0.68, 1.50)	
MSH2 (rs3732182) G/T								
	<0.01 hrs/wk	122/113	1.00	reference	101/106	0.87	(0.59, 1.29)	*0.212*
	0.01-6.35 hrs/wk	209/205	1.00	reference	166/181	0.92	(0.69, 1.24)	
	>6.35 hrs/wk	185/228	1.00	reference	166/183	1.15	(0.86, 1.54)	
MSH2 (rs4583514) A/G								
	<0.01 hrs/wk	86/84	1.00	reference	138/136	1.01	(0.68, 1.51)	*0.713*
	0.01-6.35 hrs/wk	148/143	1.00	reference	233/248	0.92	(0.68, 1.24)	
	>6.35 hrs/wk	141/157	1.00	reference	215/252	0.96	(0.71, 1.29)	
MSH3 (rs1650663) T/C								
	<0.01 hrs/wk	119/115	1.00	reference	103/104	0.98	(0.67, 1.44)	*0.150*
	0.01-6.35 hrs/wk	179/199	1.00	reference	201/187	1.20	(0.89, 1.60)	
	>6.35 hrs/wk	175/190	1.00	reference	181/222	0.89	(0.67, 1.19)	
OGG1 (rs1052133) C/G								
	<0.01 hrs/wk	148/132	1.00	reference	84/89	0.88	(0.59, 1.30)	*<0.001*
	0.01-6.35 hrs/wk	214/262	1.00	reference	182/138	1.68	(1.26, 2.24)	
	>6.35 hrs/wk	226/231	1.00	reference	141/187	0.73	(0.54, 0.97)	
XPA (rs1800975) G/A								
	<0.01 hrs/wk	118/104	1.00	reference	114/118	0.81	(0.55, 1.19)	*0.458*
	0.01-6.35 hrs/wk	185/174	1.00	reference	218/228	0.91	(0.69, 1.21)	
	>6.35 hrs/wk	164/186	1.00	reference	211/238	0.99	(0.74, 1.32)	
XPC (rs2228000) C/T								
	<0.01 hrs/wk	129/123	1.00	reference	102/99	0.98	(0.66, 1.43)	*0.038*
	0.01-6.35 hrs/wk	251/225	1.00	reference	153/179	0.76	(0.57, 1.00)	
	>6.35 hrs/wk	208/252	1.00	reference	168/173	1.19	(0.90, 1.59)	
XPC (rs2228001) A/C								
	<0.01 hrs/wk	88/71	1.00	reference	143/150	0.76	(0.51, 1.14)	*0.060*

	0.01-6.35 hrs/wk	124/140	1.00	reference	280/263	1.25	(0.93, 1.68)	
	>6.35 hrs/wk	117/136	1.00	reference	258/289	1.01	(0.75, 1.37)	
XPD (rs1799793) G/A								
	<0.01 hrs/wk	90/97	1.00	reference	140/125	1.24	(0.84, 1.82)	*0.242*
	0.01-6.35 hrs/wk	166/199	1.00	reference	237/204	1.38	(1.04, 1.83)	
	>6.35 hrs/wk	144/173	1.00	reference	230/252	1.09	(0.82, 1.46)	
XPD (rs13181) A/C								
	<0.01 hrs/wk	84/97	1.00	reference	149/124	1.42	(0.96, 2.09)	*0.231*
	0.01-6.35 hrs/wk	148/177	1.00	reference	250/226	1.30	(0.98, 1.73)	
	>6.35 hrs/wk	131/155	1.00	reference	241/269	1.07	(0.80, 1.45)	
XPF (rs1800067) G/A								
	<0.01 hrs/wk	191/170	1.00	reference	31/41	0.59	(0.34, 1.01)	*0.046*
	0.01-6.35 hrs/wk	322/315	1.00	reference	65/75	0.85	(0.59, 1.23)	
	>6.35 hrs/wk	304/358	1.00	reference	56/54	1.26	(0.83, 1.90)	
XPG (rs17655) G/C								
	<0.01 hrs/wk	117/117	1.00	reference	102/94	1.14	(0.77, 1.70)	*0.043*
	0.01-6.35 hrs/wk	205/211	1.00	reference	175/171	1.08	(0.81, 1.45)	
	>6.35 hrs/wk	214/214	1.00	reference	140/191	0.76	(0.56, 1.02)	
XRCC1 (rs1799782) C/T								
	<0.01 hrs/wk	212/196	1.00	reference	21/26	0.81	(0.43, 1.52)	*0.340*
	0.01-6.35 hrs/wk	347/349	1.00	reference	60/58	1.06	(0.72, 1.58)	
	>6.35 hrs/wk	334/374	1.00	reference	42/51	0.85	(0.54, 1.33)	
XRCC1 (rs25487) G/A								
	<0.01 hrs/wk	95/91	1.00	reference	138/132	1.01	(0.68, 1.49)	*0.446*
	0.01-6.35 hrs/wk	162/163	1.00	reference	245/245	1.00	(0.75, 1.33)	
	>6.35 hrs/wk	134/166	1.00	reference	243/259	1.16	(0.87, 1.56)	

OR, odds ratio; 95% CI, 95% confidence interval

A priori, criteria for interaction on the multiplicative scale is α<0.10

TABLE A.32 Age Adjusted Odds Ratios and 95% Confidence Intervals for Multiplicative Effects of DNA Repair SNPs and Postmenopausal Recreational Physical Activity on Breast Cancer Risk. The Long Island Breast Cancer Study Project (1996-1997).

Gene (SNP) major/minor alleles Postmenopausal RPA	Homozygous for major allele			At least one copy of minor allele			*p for interaction*
	Ca/Co	OR	95% CI	Ca/Co	OR	95% CI	
ERCC1 (rs3212986) C/A							
<0.01 hrs/wk	93/84	1.00	reference	82/71	0.98	(0.62, 1.56)	*0.466*
0.01-9.23 hrs/wk	128/116	1.00	reference	104/82	1.13	(0.76, 1.68)	
>9.23 hrs/wk	93/114	1.00	reference	92/98	1.14	(0.76, 1.71)	
MGMT (rs12917) C/T							
<0.01 hrs/wk	137/113	1.00	reference	40/40	0.83	(0.48, 1.41)	*0.129*
0.01-9.23 hrs/wk	163/141	1.00	reference	70/59	1.02	(0.67, 1.55)	
>9.23 hrs/wk	128/162	1.00	reference	58/50	1.41	(0.89, 2.23)	
MGMT (rs2308321) A/G							
<0.01 hrs/wk	141/125	1.00	reference	36/28	1.06	(0.60, 1.88)	*0.554*
0.01-9.23 hrs/wk	175/149	1.00	reference	58/50	0.99	(0.63, 1.54)	
>9.23 hrs/wk	143/159	1.00	reference	43/53	0.93	(0.58, 1.51)	
MGMT (rs2308327) A/G							
<0.01 hrs/wk	149/130	1.00	reference	29/25	0.94	(0.51, 1.73)	*0.752*
0.01-9.23 hrs/wk	186/156	1.00	reference	47/44	0.88	(0.55, 1.42)	
>9.23 hrs/wk	145/163	1.00	reference	42/50	0.99	(0.61, 1.61)	
MLH1 (rs1799977) A/G							
<0.01 hrs/wk	72/81	1.00	reference	101/70	1.65	(1.04, 2.64)	*0.010*
0.01-9.23 hrs/wk	107/93	1.00	reference	110/99	0.91	(0.61, 1.36)	
>9.23 hrs/wk	93/96	1.00	reference	85/114	0.68	(0.45, 1.03)	
MLH1 (rs2286940) C/T							
<0.01 hrs/wk	50/47	1.00	reference	123/104	1.19	(0.72, 1.98)	*0.085*
0.01-9.23 hrs/wk	76/55	1.00	reference	138/139	0.69	(0.45, 1.05)	
>9.23 hrs/wk	62/59	1.00	reference	114/150	0.65	(0.42, 1.02)	
MSH2 (rs2303428) T/C							
<0.01 hrs/wk	140/125	1.00	reference	25/20	1.37	(0.66, 2.82)	*0.476*

0.01-9.23 hrs/wk	174/160	1.00	reference	36/31	1.03	(0.61, 1.76)	
>9.23 hrs/wk	152/176	1.00	reference	23/31	0.89	(0.49, 1.63)	
MSH2 (rs3732182) G/T							
<0.01 hrs/wk	98/82	1.00	reference	75/69	0.95	(0.59, 1.51)	*0.329*
0.01-9.23 hrs/wk	108/102	1.00	reference	102/90	1.04	(0.70, 1.55)	
>9.23 hrs/wk	95/121	1.00	reference	80/88	1.26	(0.83, 1.93)	
MSH2 (rs4583514) A/G							
<0.01 hrs/wk	73/61	1.00	reference	100/90	0.98	(0.61, 1.56)	*0.964*
0.01-9.23 hrs/wk	81/70	1.00	reference	135/124	0.94	(0.63, 1.41)	
>9.23 hrs/wk	71/80	1.00	reference	107/129	0.98	(0.64, 1.50)	
MSH3 (rs1650663) T/C							
<0.01 hrs/wk	97/79	1.00	reference	74/70	0.89	(0.56, 1.42)	*0.127*
0.01-9.23 hrs/wk	103/109	1.00	reference	114/83	1.38	(0.93, 2.06)	
>9.23 hrs/wk	83/97	1.00	reference	94/113	0.95	(0.63, 1.44)	
OGG1 (rs1052133) C/G							
<0.01 hrs/wk	113/90	1.00	reference	64/61	0.86	(0.53, 1.38)	*0.177*
0.01-9.23 hrs/wk	128/122	1.00	reference	94/73	1.27	(0.85, 1.89)	
>9.23 hrs/wk	106/118	1.00	reference	75/91	0.80	(0.52, 1.22)	
XPA (rs1800975) G/A							
<0.01 hrs/wk	84/73	1.00	reference	93/81	0.89	(0.56, 1.43)	*0.131*
0.01-9.23 hrs/wk	110/83	1.00	reference	121/112	0.83	(0.56, 1.22)	
>9.23 hrs/wk	75/97	1.00	reference	112/115	1.27	(0.84, 1.91)	
XPC (rs2228000) C/T							
<0.01 hrs/wk	93/81	1.00	reference	71/67	0.84	(0.53, 1.34)	*0.048*
0.01-9.23 hrs/wk	121/112	1.00	reference	109/82	1.29	(0.87, 1.91)	
>9.23 hrs/wk	112/115	1.00	reference	91/86	1.53	(1.01, 2.30)	
XPC (rs2228001) A/C							
<0.01 hrs/wk	67/44	1.00	reference	110/108	0.63	(0.39, 1.04)	*0.077*
0.01-9.23 hrs/wk	75/65	1.00	reference	156/131	1.06	(0.69, 1.60)	
>9.23 hrs/wk	55/71	1.00	reference	131/142	1.20	(0.77, 1.86)	
XPD (rs1799793) G/A							
<0.01 hrs/wk	72/64	1.00	reference	105/90	0.99	(0.62, 1.59)	*0.066*
0.01-9.23 hrs/wk	92/95	1.00	reference	138/101	1.35	(0.92, 2.00)	

>9.23 hrs/wk	85/88	1.00	reference	101/125	0.83	(0.55, 1.24)	
XPD (rs13181) A/C							
<0.01 hrs/wk	66/68	1.00	reference	112/85	1.35	(0.85, 2.15)	*0.173*
0.01-9.23 hrs/wk	87/88	1.00	reference	141/108	1.25	(0.84, 1.85)	
>9.23 hrs/wk	75/82	1.00	reference	107/130	0.92	(0.61, 1.40)	
XPF (rs1800067) G/A							
<0.01 hrs/wk	150/119	1.00	reference	21/27	0.54	(0.28, 1.06)	*0.012*
0.01-9.23 hrs/wk	179/168	1.00	reference	43/25	1.64	(0.95, 2.82)	
>9.23 hrs/wk	150/175	1.00	reference	30/30	1.17	(0.66, 2.06)	
XPG (rs17655) G/C							
<0.01 hrs/wk	88/89	1.00	reference	80/57	1.45	(0.90, 2.34)	*0.022*
0.01-9.23 hrs/wk	126/111	1.00	reference	92/81	0.99	(0.67, 1.48)	
>9.23 hrs/wk	113/107	1.00	reference	65/96	0.71	(0.46, 1.08)	
XRCC1 (rs1799782) C/T							
<0.01 hrs/wk	166/135	1.00	reference	12/19	0.66	(0.29, 1.48)	*0.188*
0.01-9.23 hrs/wk	205/177	1.00	reference	28/22	1.20	(0.65, 2.19)	
>9.23 hrs/wk	160/186	1.00	reference	26/27	0.96	(0.53, 1.77)	
XRCC1 (rs25487) G/A							
<0.01 hrs/wk	64/65	1.00	reference	114/90	1.27	(0.79, 2.04)	*0.343*
0.01-9.23 hrs/wk	86/80	1.00	reference	147/120	1.19	(0.80, 1.76)	
>9.23 hrs/wk	69/78	1.00	reference	118/135	0.96	(0.63, 1.46)	

OR, odds ratio; 95% CI, 95% confidence interval

A priori, criteria for interaction on the multiplicative scale is α<0.10

TABLE A.33 Age Adjusted Odds Ratios and 95% Confidence Intervals for the Additive Effect of Oxidative Stress SNPs and Lifetime Recreational Physical Activity on Breast Cancer Risk in the Long Island Breast Cancer Study Project (1996-1997).

Gene (SNP) major/minor alleles Lifetime RPA	Homozygous for major allele			At least one copy of minor allele			ICR (95% CI)
	Ca/Co	OR	95% CI	Ca/Co	OR	95% CI	
CAT (rs4756146) T/C							
<0.01 hrs/wk	169/161	1.00	reference	48/57	0.80	(0.51, 1.25)	
0.01-6.35 hrs/wk	296/290	1.05	(0.80, 1.38)	82/94	0.90	(0.62, 1.30)	*0.05 (-0.86, 0.96)*
>6.35 hrs/wk	270/318	0.83	(0.63, 1.09)	82/95	0.88	(0.61, 1.27)	*0.25 (-0.61, 1.10)*
CAT (rs2284365) T/C							
<0.01 hrs/wk	131/117	1.00	reference	88/101	0.78	(0.53, 1.14)	
0.01-6.35 hrs/wk	229/227	0.96	(0.70, 1.31)	151/163	0.90	(0.65, 1.27)	*0.17 (-0.70, 1.23)*
>6.35 hrs/wk	198/223	0.78	(0.56, 1.06)	145/174	0.78	(0.56, 1.09)	*0.22 (-0.55, 0.99)*
CAT (rs480575) A/G							
<0.01 hrs/wk	115/100	1.00	reference	102/114	0.77	(0.52, 1.13)	
0.01-6.35 hrs/wk	197/190	0.96	(0.68, 1.34)	176/184	0.90	(0.64, 1.26)	*0.17 (-0.71, 1.06)*
>6.35 hrs/wk	179/191	0.83	(0.59, 1.17)	167/206	0.73	(0.52, 1.03)	*0.13 (-0.66, 0.92)*
CAT (rs1001179) C/T							
<0.01 hrs/wk	133/134	1.00	reference	94/81	1.19	(0.81, 1.74)	
0.01-6.35 hrs/wk	254/254	1.09	(0.80, 1.47)	138/144	1.05	(0.75, 1.48)	*-0.22 (-1.30, 0.86)*
>6.35 hrs/wk	216/275	0.83	(0.61, 1.12)	152/144	1.09	(0.78, 1.52)	*0.08 (-0.96, 1.11)*
COMT (rs4680) G/A							
<0.01 hrs/wk	61/58	1.00	reference	167/162	1.00	(0.65, 1.52)	
0.01-6.35 hrs/wk	114/109	1.06	(0.68, 1.67)	284/291	1.02	(0.68, 1.52)	*-0.04 (-1.33, 1.25)*
>6.35 hrs/wk	100/99	1.00	(0.63, 1.59)	271/321	0.84	(0.56, 1.25)	*-0.16 (-1.35, 1.03)*
COMT (rs737865) T/C							
<0.01 hrs/wk	113/105	1.00	reference	111/116	0.88	(0.61, 1.29)	
0.01-6.35 hrs/wk	190/190	1.01	(0.72, 1.42)	204/207	0.98	(0.70, 1.36)	*0.08 (-0.88, 1.05)*
>6.35 hrs/wk	168/202	0.81	(0.57, 1.13)	193/218	0.84	(0.60, 1.17)	*0.15 (-0.71, 1.00)*

GPX (rs1050450) C/T								
<0.01 hrs/wk	110/106	1.00	reference	116/113	1.01	(0.70, 1.48)	
0.01-6.35 hrs/wk	178/184	1.02	(0.72, 1.43)	219/214	1.07	(0.77, 1.49)	*0.04 (-1.01, 1.09)*
>6.35 hrs/wk	169/205	0.85	(0.60, 1.19)	198/213	0.92	(0.66, 1.28)	*0.05 (-0.89, 1.00)*
GSTA1 (rs3957356) G/A								
<0.01 hrs/wk	70/74	1.00	reference	157/145	1.13	(0.76, 1.68)	
0.01-6.35 hrs/wk	136/143	1.08	(0.72, 1.62)	260/255	1.15	(0.79, 1.67)	*-0.06 (-1.35, 1.23)*
>6.35 hrs/wk	120/149	0.87	(0.58, 1.30)	247/270	0.99	(0.68, 1.43)	*0.00 (-1.15, 1.14)*
GSTP1 (rs1695) A/G								
<0.01 hrs/wk	116/107	1.00	reference	111/106	0.95	(0.65, 1.39)	
0.01-6.35 hrs/wk	206/184	1.10	(0.79, 1.53)	183/203	0.89	(0.64, 1.25)	*-0.16 (-1.14, 0.82)*
>6.35 hrs/wk	167/195	0.80	(0.57, 1.12)	193/221	0.83	(0.59, 1.15)	*0.07 (-0.80, 0.94)*
GSTM1 (gene deletion)								
<0.01 hrs/wk	109/111	1.00	reference	102/87	1.16	(0.78, 1.72)	
0.01-6.35 hrs/wk	182/197	0.99	(0.71, 1.39)	190/170	1.24	(0.88, 1.75)	*0.09 (-1.09, 1.27)*
>6.35 hrs/wk	188/215	0.91	(0.65, 1.27)	158/171	0.96	(0.68, 1.36)	*-0.11 (-1.16, 0.94)*
GSTT1 (gene deletion)								
<0.01 hrs/wk	163/155	1.00	reference	49/46	1.02	(0.64, 1.62)	
0.01-6.35 hrs/wk	292/288	1.05	(0.80, 1.39)	83/85	0.98	(0.67, 1.43)	*-0.09 (-1.15, 0.98)*
>6.35 hrs/wk	287/311	0.90	(0.69, 1.19)	64/81	0.80	(0.54, 1.19)	*-0.12 (-1.08, 0.85)*
MnSOD (rs4880) T/C								
<0.01 hrs/wk	48/55	1.00	reference	178/160	1.26	(0.81, 1.97)	
0.01-6.35 hrs/wk	105/96	1.36	(0.84, 2.20)	288/302	1.16	(0.76, 1.78)	*-0.46 (-2.14, 1.22)*
>6.35 hrs/wk	100/108	1.06	(0.66, 1.71)	264/309	1.01	(0.66, 1.55)	*-0.31 (-1.78, 1.17)*
MPO (rs2333227) G/A								
<0.01 hrs/wk	142/127	1.00	reference	86/93	0.84	(0.58, 1.23)	
0.01-6.35 hrs/wk	247/232	1.03	(0.76, 1.39)	151/167	0.88	(0.63, 1.23)	*0.01 (-0.87, 0.89)*
>6.35 hrs/wk	226/257	0.81	(0.60, 1.10)	140/161	0.82	(0.58, 1.14)	*0.16 (-0.64, 0.96)*

OR, odds ratio; 95% CI, 95% confidence interval; ICR, interaction contrast ratio

TABLE A.34 Age Adjusted Odds Ratios and 95% Confidence Intervals for the Additive Effect of Oxidative Stress SNPs and Postmenopausal Recreational Physical Activity on Postmenopausal Breast Cancer Risk in the Long Island Breast Cancer Study Project (1996-1997).

Gene (SNP) major/minor alleles Postmenopausal RPA	Homozygous for major allele			At least one copy of minor allele			
	Ca/Co	OR	95% CI	Ca/Co	OR	95% CI	ICR (95% CI)
CAT (rs4756146) T/C							
<0.01 hrs/wk	133/104	1.00	reference	33/44	0.59	(0.35, 1.00)	
0.01-9.23 hrs/wk	169/140	0.95	(0.67, 1.34)	48/51	0.74	(0.46, 1.20)	*0.21 (-0.70, 1.11)*
>9.23 hrs/wk	137/163	0.62	(0.44, 0.88)	36/45	0.63	(0.38, 1.06)	*0.42 (-0.36, 1.20)*
CAT (rs2284365) T/C							
<0.01 hrs/wk	104/80	1.00	reference	64/69	0.71	(0.45, 1.11)	
0.01-9.23 hrs/wk	126/109	0.88	(0.59, 1.30)	92/81	0.88	(0.58, 1.35)	*0.30 (-0.69, 1.28)*
>9.23 hrs/wk	104/117	0.63	(0.42, 0.94)	64/89	0.54	(0.35, 0.84)	*0.20 (-0.55, 0.96)*
CAT (rs480575) A/G							
<0.01 hrs/wk	90/71	1.00	reference	77/76	0.77	(0.49, 1.22)	
0.01-9.23 hrs/wk	107/85	0.97	(0.63, 1.50)	104/98	0.84	(0.55, 1.28)	*0.09 (-0.98, 1.16)*
>9.23 hrs/wk	93/99	0.67	(0.44, 1.04)	79/100	0.60	(0.39, 0.93)	*0.15 (-0.70, 1.00)*
CAT (rs1001179) C/T							
<0.01 hrs/wk	103/95	1.00	reference	70/53	1.30	(0.82, 2.06)	
0.01-9.23 hrs/wk	149/126	1.13	(0.78, 1.63)	75/71	0.99	(0.64, 1.53)	*-0.43 (-1.78, 0.92)*
>9.23 hrs/wk	100/143	0.64	(0.43, 0.94)	82/69	1.04	(0.68, 1.60)	*0.11 (-1.28, 1.35)*
COMT (rs4680) G/A							
<0.01 hrs/wk	46/37	1.00	reference	128/116	0.89	(0.54, 1.49)	
0.01-9.23 hrs/wk	64/54	0.95	(0.54, 1.70)	162/142	0.93	(0.56, 1.53)	*0.08 (-1.36, 1.52)*
>9.23 hrs/wk	56/52	0.82	(0.48, 1.48)	130/160	0.63	(0.38, 1.03)	*-0.08 (-1.30, 1.12)*
COMT (rs737865) T/C							
<0.01 hrs/wk	89/77	1.00	reference	80/76	0.89	(0.57, 1.39)	
0.01-9.23 hrs/wk	109/87	1.08	(0.71, 1.65)	118/107	0.94	(0.62, 1.42)	*-0.03 (-1.24,1.15)*
>9.23 hrs/wk	77/98	0.65	(0.42, 1.01)	102/114	0.72	(0.47, 1.08)	*0.18 (-0.75, 1.11)*

		OR	(95% CI)		OR	(95% CI)	ICR
GPX (rs1050450) C/T							
<0.01 hrs/wk	97/82	1.00	reference	93/70	1.44	(0.92, 2.25)	
0.01-9.23 hrs/wk	107/97	1.21	(0.80, 1.85)	120/96	1.28	(0.85, 1.95)	*-0.37 (-1.98, 1.24)*
>9.23 hrs/wk	79/103	0.80	(0.52, 1.24)	103/109	0.92	(0.61, 1.40)	*0.32 (-1.63, 0.99)*
GSTA1 (rs3957356) G/A							
<0.01 hrs/wk	49/48	1.00	reference	124/104	1.19	(0.73, 1.93)	
0.01-9.23 hrs/wk	76/68	1.09	(0.65, 1.85)	151/126	1.20	(0.75, 1.92)	*-0.08 (-1.75, 3.16)*
>9.23 hrs/wk	59/61	0.90	(0.52, 1.56)	124/152	0.77	(0.48, 1.23)	*-0.33 (-1.70, 1.05)*
GSTP1 (rs1695) A/G							
<0.01 hrs/wk	84/80	1.00	reference	90/68	1.24	(0.79, 1.95)	
0.01-9.23 hrs/wk	123/80	1.45	(0.95, 2.22)	97/109	0.84	(0.55, 1.28)	*-0.85 (-2.28, 0.57)*
>9.23 hrs/wk	76/92	0.72	(0.46, 1.12)	103/120	0.78	(0.51, 1.18)	*-0.19(-1.33, 0.95)*
GSTM1 (gene deletion)							
<0.01 hrs/wk	72/73	1.00	reference	86/65	1.34	(0.84, 2.14)	
0.01-9.23 hrs/wk	105/99	1.06	(0.69, 1.63)	111/78	1.48	(0.95, 2.30)	*0.07 (-1.60, 1.75)*
>9.23 hrs/wk	93/108	0.82	(0.53, 1.27)	82/88	0.91	(0.58, 1.44)	*-0.25 (-1.58, 1.08)*
GSTT1 (gene deletion)							
<0.01 hrs/wk	124/109	1.00	reference	34/31	0.94	(0.54, 1.64)	
0.01-9.23 hrs/wk	175/136	1.13	(0.80, 1.60)	42/42	0.86	(0.52, 1.42)	*-0.21 (-1.43, 1,01)*
>9.23 hrs/wk	139/153	0.74	(0.52, 1.06)	36/47	0.68	(0.41, 1.13)	*-0.00 (-1.04, 1.04)*
MnSOD (rs4880) T/C							
<0.01 hrs/wk	46/39	1.00	reference	128/110	0.91	(0.55, 1.51)	
0.01-9.23 hrs/wk	59/57	0.84	(0.47, 1.48)	164/138	0.95	(0.58, 1.56)	*0.20 (-1.18, 1.58)*
>9.23 hrs/wk	57/51	0.82	(0.46, 1.46)	125/161	0.60	(0.37, 0.99)	*-0.13 (-1.32, 1.07)*
MPO (rs2333227) G/A							
<0.01 hrs/wk	105/95	1.00	reference	69/57	1.15	(0.73, 1.82)	
0.01-9.23 hrs/wk	144/112	1.17	(0.80, 1.72)	82/84	0.92	(0.61, 1.40)	*-0.41 (-1.67, 0.86)*
>9.23 hrs/wk	109/136	0.71	(0.48, 1.04)	74/76	0.85	(0.55, 1.30)	*-0.02 (-1.11, 1.08)*

OR, odds ratio; 95% CI, 95% confidence interval; ICR, interaction contrast ratio

TABLE A.35 Age Adjusted Odds Ratios and 95% Confidence Intervals for the Additive Effect of DNA Repair SNPs and Lifetime Recreational Physical Activity on Breast Cancer Risk in the Long Island Breast Cancer Study Project (1996-1997).

Gene (SNP) major/minor alleles Lifetime RPA	Homozygous for major allele			At least one copy of minor allele			ICR (95% CI)
	Ca/Co	OR	95% CI	Ca/Co	OR	95% CI	
ERCC1 (rs3212986) C/A							
<0.01 hrs/wk	119/122	1.00	reference	110/101	1.09	(0.75, 1.58)	
0.01-6.35 hrs/wk	227/243	1.02	(0.74, 1.39)	178/161	1.20	(0.86, 1.68)	*0.10 (-0.99, 1.18)*
>6.35 hrs/wk	181/218	0.87	(0.63, 1.20)	193/204	0.99	(0.72, 1.37)	*0.03 (-0.94, 1.01)*
MGMT (rs12917) C/T							
<0.01 hrs/wk	173/155	1.00	reference	58/67	0.78	(0.51, 1.18)	
0.01-6.35 hrs/wk	300/296	0.98	(0.75, 1.29)	107/111	0.92	(0.65, 1.31)	*0.17 (-0.69, 1.03)*
>6.35 hrs/wk	276/329	0.75	(0.57, 0.99)	109/95	1.06	(0.74, 1.51)	*0.53 (-0.34, 1.39)*
MGMT (rs2308321) A/G							
<0.01 hrs/wk	180/175	1.00	reference	51/46	1.09	(0.69, 1.71)	
0.01-6.35 hrs/wk	321/309	1.10	(0.84, 1.43)	86/96	0.92	(0.64, 1.33)	*-0.26 (-1.31, 0.79)*
>6.35 hrs/wk	286/322	0.90	(0.69, 1.17)	90/101	0.90	(0.63, 1.28)	*-0.09 (-1.09, 0.91)*
MGMT (rs2308327) A/G							
<0.01 hrs/wk	189/180	1.00	reference	44/43	1.00	(0.63, 1.61)	
0.01-6.35 hrs/wk	333/320	1.08	(0.83, 1.40)	74/88	0.86	(0.59, 1.24)	*-0.23 (-1.23, 0.77)*
>6.35 hrs/wk	294/336	0.87	(0.67, 1.13)	83/89	0.92	(0.63, 1.32)	*0.04 (-0.95, 1.03)*
MLH1 (rs1799977) A/G							
<0.01 hrs/wk	106/123	1.00	reference	117/97	1.37	(0.94, 2.00)	
0.01-6.35 hrs/wk	204/193	1.31	(0.94, 1.82)	179/195	1.13	(0.81, 1.58)	*-0.55 (-1.82, 0.72)*
>6.35 hrs/wk	175/198	1.05	(0.75, 1.47)	181/214	1.00	(0.72, 1.39)	*-0.42 (-1.57, 0.73)*
MLH1 (rs2286940) C/T							
<0.01 hrs/wk	73/78	1.00	reference	150/142	1.12	(0.76, 1.67)	
0.01-6.35 hrs/wk	142/117	1.40	(0.93, 2.10)	239/274	1.00	(0.69, 1.44)	*-0.53 (-1.87, 0.82)*
>6.35 hrs/wk	112/127	0.99	(0.65, 1.49)	240/284	0.92	(0.64, 1.33)	*-0.19 (-1.36, 0.97)*

MSH2 (rs2303428) T/C

<0.01 hrs/wk	180/186	1.00	reference	35/29	1.24	(0.73, 2.13)	
0.01-6.35 hrs/wk	315/322	1.08	(0.83, 1.40)	58/62	1.05	(0.69, 1.59)	*-0.28 (-1.60, 1.05)*
>6.35 hrs/wk	293/338	0.92	(0.71, 1.19)	56/65	0.92	(0.61, 1.40)	*-0.24 (-1.48, 1.00)*

MSH2 (rs3732182) G/T

<0.01 hrs/wk	122/113	1.00	reference	101/106	0.90	(0.62, 1.31)	
0.01-6.35 hrs/wk	209/205	1.02	(0.73, 1.41)	166/181	0.93	(0.66, 1.30)	*0.01 (-0.93, 0.95)*
>6.35 hrs/wk	185/228	0.78	(0.56, 1.07)	166/183	0.88	(0.63, 1.24)	*0.21 (-0.64, 1.07)*

MSH2 (rs4583514) A/G

<0.01 hrs/wk	86/84	1.00	reference	138/136	1.02	(0.70, 1.51)	
0.01-6.35 hrs/wk	148/143	1.10	(0.75, 1.62)	233/248	1.01	(0.71, 1.44)	*-0.11 (-1.26, 1.03)*
>6.35 hrs/wk	141/157	0.92	(0.63, 1.34)	215/252	0.88	(0.62, 1.26)	*-0.06 (-1.09, 0.97)*

MSH3 (rs1650663) T/C

<0.01 hrs/wk	119/115	1.00	reference	103/104	0.97	(0.66, 1.41)	
0.01-6.35 hrs/wk	179/199	0.94	(0.68, 1.32)	201/187	1.11	(0.80, 1.55)	*0.20 (-0.81, 1.21)*
>6.35 hrs/wk	175/190	0.92	(0.66, 1.28)	181/222	0.82	(0.59, 1.14)	*-0.07 (-0.96, 0.83)*

OGG1 (rs1052133) C/G

<0.01 hrs/wk	148/132	1.00	reference	84/89	0.84	(0.57, 1.23)	
0.01-6.35 hrs/wk	214/262	0.78	(0.58, 1.05)	182/138	1.28	(0.92, 1.77)	*0.66 (-0.29, 1.61)*
>6.35 hrs/wk	226/231	0.91	(0.68, 1.24)	141/187	0.68	(0.49, 0.95)	*-0.07 (-0.84, 0.70)*

XPA (rs1800975) G/A

<0.01 hrs/wk	118/104	1.00	reference	114/118	0.83	(0.57, 1.21)	
0.01-6.35 hrs/wk	185/174	1.00	(0.71, 1.40)	218/228	0.90	(0.65, 1.25)	*0.08 (-0.83, 0.99)*
>6.35 hrs/wk	164/186	0.80	(0.57, 1.13)	211/238	0.79	(0.57, 1.09)	*0.16 (-0.66, 0.97)*

XPC (rs2228000) C/T

<0.01 hrs/wk	129/123	1.00	reference	102/99	0.95	(0.65, 1.38)	
0.01-6.35 hrs/wk	251/225	1.12	(0.82, 1.53)	153/179	0.88	(0.63, 1.22)	*-0.19 (-1.13, 0.75)*
>6.35 hrs/wk	208/252	0.80	(0.59, 1.10)	168/173	0.94	(0.68, 1.30)	*0.19 (-0.69, 1.07)*

XPC (rs2228001) A/C

<0.01 hrs/wk	88/71	1.00	reference	143/150	0.77	(0.52, 1.14)	
0.01-6.35 hrs/wk	124/140	0.77	(0.52, 1.15)	280/263	0.93	(0.65, 1.33)	*0.39 (-0.54, 1.32)*
>6.35 hrs/wk	117/136	0.72	(0.48, 1.08)	258/289	0.74	(0.52, 1.06)	*0.25 (-0.59, 1.09)*

XPD (rs1799793) G/A

								ICR
<0.01 hrs/wk	90/97	1.00	reference	140/125	1.21	(0.83, 1.76)		
0.01-6.35 hrs/wk	166/199	0.98	(0.68, 1.40)	237/204	1.35	(0.95, 1.91)	*0.16 (-1.08, 1.41)*
>6.35 hrs/wk	144/173	0.93	(0.65, 1.35)	230/252	1.02	(0.72, 1.43)	*-0.13 (-1.24, 0.99)*
XPD (rs13181) A/C							
<0.01 hrs/wk	84/97	1.00	reference	149/124	1.41	(0.97, 2.07)	
0.01-6.35 hrs/wk	148/177	1.07	(0.74, 1.54)	250/226	1.39	(0.98, 1.97)	*-0.09 (-1.47, 1.29)*
>6.35 hrs/wk	131/155	1.02	(0.70, 1.49)	241/269	1.09	(0.77, 1.54)	*-0.34 (-1.60, 0.92)*
XPF (rs1800067) G/A							
<0.01 hrs/wk	191/170	1.00	reference	31/41	0.64	(0.38, 1.07)	
0.01-6.35 hrs/wk	322/315	0.98	(0.75, 1.27)	65/75	0.83	(0.56, 1.24)	*0.22 (-0.63, 1.08)*
>6.35 hrs/wk	304/358	0.78	(0.60, 1.00)	56/54	0.96	(0.62, 1.48)	*0.55 (-0.34, 1.44)*
XPG (rs17655) G/C							
<0.01 hrs/wk	117/117	1.00	reference	102/94	1.14	(0.77, 1.67)	
0.01-6.35 hrs/wk	205/211	1.06	(0.77, 1.47)	175/171	1.14	(0.81, 1.59)	*-0.06 (-1.18, 1.06)*
>6.35 hrs/wk	214/214	1.04	(0.76, 1.44)	140/191	0.78	(0.55, 1.09)	*-0.40 (-1.38, 0.58)*
XRCC1 (rs1799782) C/T							
<0.01 hrs/wk	212/196	1.00	reference	21/26	0.76	(0.41, 1.40)	
0.01-6.35 hrs/wk	347/349	0.99	(0.77, 1.27)	60/58	1.04	(0.69, 1.58)	*0.29 (-0.83, 1.40)*
>6.35 hrs/wk	334/374	0.86	(0.67, 1.11)	42/51	0.74	(0.47, 1.17)	*0.12 (-0.85, 1.08)*
XRCC1 (rs25487) G/A							
<0.01 hrs/wk	95/91	1.00	reference	138/132	0.99	(0.68, 1.45)	
0.01-6.35 hrs/wk	162/163	1.02	(0.70, 1.47)	245/245	1.03	(0.73, 1.46)	*0.02 (-1.05, 1.10)*
>6.35 hrs/wk	134/166	0.80	(0.55, 1.16)	243/259	0.93	(0.66, 1.30)	*0.13 (-0.83, 1.10)*

OR, odds ratio; 95% CI, 95% confidence interval; ICR, Interaction Contrast Ratio

174

TABLE A.36 Age Adjusted Odds Ratios and 95% Confidence Intervals for the Additive Effect of DNA Repair SNPs and Postmenopausal Recreational Physical Activity on Breast Cancer Risk in the Long Island Breast Cancer Study Project (1996-1997).

Gene (SNP) major/minor alleles Postmenopausal RPA	Homozygous for major allele			At least one copy of minor allele			
	Ca/Co	OR	95% CI	Ca/Co	OR	95% CI	ICR (95% CI)
ERCC1 (rs3212986) C/A							
<0.01 hrs/wk	93/84	1.00	reference	82/71	1.00	(0.64, 1.55)	
0.01-9.23 hrs/wk	128/116	0.96	(0.65, 1.43)	104/82	1.14	(0.75, 1.73)	*0.18 (-1.06, 1.41)*
>9.23 hrs/wk	93/114	0.70	(0.46, 1.05)	92/98	0.78	(0.52, 1.19)	*0.09 (-0.91, 1.09)*
MGMT (rs12917) C/T							
<0.01 hrs/wk	137/113	1.00	reference	40/40	0.83	(0.50, 1.38)	
0.01-9.23 hrs/wk	163/141	0.96	(0.68, 1.34)	70/59	0.99	(0.64, 1.53)	*0.21 (-0.89, 1.32)*
>9.23 hrs/wk	128/162	0.64	(0.45, 0.90)	58/50	0.87	(0.55, 1.39)	*0.41 (-0.57, 1.40)*
MGMT (rs2308321) A/G							
<0.01 hrs/wk	141/125	1.00	reference	36/28	1.16	(0.66, 2.03)	
0.01-9.23 hrs/wk	175/149	1.04	(0.75, 1.45)	58/50	1.05	(0.67, 1.66)	*-0.15 (-1.54, 1.23)*
>9.23 hrs/wk	143/159	0.77	(0.55, 1.07)	43/53	0.68	(0.42, 1.10)	*-0.25 (-1.39, 0.89)*
MGMT (rs2308327) A/G							
<0.01 hrs/wk	149/130	1.00	reference	29/25	1.04	(0.57, 1.88)	
0.01-9.23 hrs/wk	186/156	1.04	(0.76, 1.44)	47/44	0.96	(0.59, 1.55)	*-0.12 (-1.44, 1.20)*
>9.23 hrs/wk	145/163	0.75	(0.54, 1.04)	42/50	0.70	(0.43, 1.13)	*-0.09 (1.21, 1.04)*
MLH1 (rs1799977) A/G							
<0.01 hrs/wk	72/81	1.00	reference	101/70	1.63	(1.04, 2.56)	
0.01-9.23 hrs/wk	107/93	1.29	(0.84, 1.99)	110/99	1.27	(0.83, 1.94)	*-0.66 (-2.38, 1.05)*
>9.23 hrs/wk	93/96	1.09	(0.70, 1.68)	85/114	0.78	(0.51, 1.20)	*-0.94 (-2.38, 0.50)*
MLH1 (rs2286940) C/T							
<0.01 hrs/wk	50/47	1.00	reference	123/104	1.16	(0.72, 1.89)	
0.01-9.23 hrs/wk	76/55	1.37	(0.80, 2.34)	138/139	0.96	(0.60, 1.54)	*-0.57 (-2.26,1.12)*

>9.23 hrs/wk	62/59	1.00	(0.58, 1. 73)	114/150	0.70	(0.43, 1.12)	*-0.47 (-1.85, 0.91)*
MSH2 (rs2303428) T/C							
<0.01 hrs/wk	140/125	1.00	reference	25/20	1.07	(0.56, 2.05)	
0.01-9.23 hrs/wk	174/160	0.97	(0.70, 1.35)	36/31	1.04	(0.60, 1.79)	*-0.01 (-1.49, 1.47)*
>9.23 hrs/wk	152/176	0.73	(0.53, 1.02)	23/31	0.64	(0.35, 1.16)	*-1.17 (-1.37, 1.03)*
MSH2 (rs3732182) G/T							
<0.01 hrs/wk	98/82	1.00	reference	75/69	0.92	(0.59, 1.43)	
0.01-9.23 hrs/wk	108/102	0.89	(0.60, 1.34)	102/90	0.95	(0.63, 1.44)	*0.14 (-0.97, 1.25)*
>9.23 hrs/wk	95/121	0.61	(0.41, 0.92)	80/88	0.76	(0.49, 1.17)	*0.23 (-0.70, 1.16)*
MSH2 (rs4583514) A/G							
<0.01 hrs/wk	73/61	1.00	reference	100/90	0.96	(0.61, 1.51)	
0.01-9.23 hrs/wk	81/70	0.96	(0.60, 1.55)	135/124	0.94	(0.62, 1.44)	*0.02 (-1.23, 1.27)*
>9.23 hrs/wk	71/80	0.71	(0.44, 1.15)	107/129	0.68	(0.44, 1.05)	*0.01 (-1.01, 1.03)*
MSH3 (rs1650663) T/C							
<0.01 hrs/wk	97/79	1.00	reference	74/70	0.87	(0.55, 1.36)	
0.01-9.23 hrs/wk	103/109	0.80	(0.53, 1.20)	114/83	1.09	(0.72, 1.65)	*0.42 (-0.70, 1.54)*
>9.23 hrs/wk	83/97	0.67	(0.44, 1.03)	94/113	0.65	(0.43, 0.98)	*0.11 (-0.78, 1.00)*
OGG1 (rs1052133) C/G							
<0.01 hrs/wk	113/90	1.00	reference	64/61	0.82	(0.52, 1.28)	
0.01-9.23 hrs/wk	128/122	0.83	(0.57, 1.21)	94/73	1.02	(0.67, 1.55)	*0.37 (-0.68, 1.42)*
>9.23 hrs/wk	106/118	0.69	(0.47, 1.02)	75/91	0.61	(0.40, 0.92)	*0.10 (-0.74, 0.93)*
XPA (rs1800975) G/A							
<0.01 hrs/wk	84/73	1.00	reference	93/81	0.98	(0.63, 1.53)	
0.01-9.23 hrs/wk	110/83	1.15	(0.75, 1.77)	121/112	0.93	(0.62, 1.41)	*-0.20 (-1.45, 1.06)*
>9.23 hrs/wk	75/97	0.64	(0.41, 0.99)	112/115	0.80	(0.53, 1.22)	*0.18 (-0.83, 1.19)*
XPC (rs2228000) C/T							
<0.01 hrs/wk	93/81	1.00	reference	71/67	0.83	(0.53, 1.30)	
0.01-9.23 hrs/wk	121/112	0.83	(0.56, 1.22)	109/82	1.11	(0.74, 1.68)	*0.46 (-0.65, 1.56)*
>9.23 hrs/wk	112/115	0.57	(0.38, 0.84)	91/86	0.82	(0.54, 1.25)	*0.43 (-0.47, 1.33)*
XPC (rs2228001) A/C							
<0.01 hrs/wk	67/44	1.00	reference	110/108	0.68	(0.42, 1.09)	
0.01-9.23 hrs/wk	75/65	0.79	(0.47, 1.32)	156/131	0.79	(0.50, 1.23)	*0.32 (-0.74, 1.37)*
>9.23 hrs/wk	55/71	0.50	(0.29, 0.84)	131/142	0.58	(0.37, 0.91)	*0.41 (-0.42, 1.23)*

								ICR
XPD (rs1799793) G/A								
	<0.01 hrs/wk	72/64	1.00	reference	105/90	1.03	(0.66, 1.62)	
	0.01-9.23 hrs/wk	92/95	0.87	(0.56, 1.37)	138/101	1.21	(0.79, 1.86)	*0.30 (-1.03, 1.63)*
	>9.23 hrs/wk	85/88	0.82	(0.52, 1.29)	101/125	0.68	(0.44, 1.06)	*-0.17 (-1.24, 0.91)*
XPD (rs13181) A/C								
	<0.01 hrs/wk	66/68	1.00	reference	112/85	1.36	(0.87, 2.13)	
	0.01-9.23 hrs/wk	87/88	1.05	(0.67, 1.66)	141/108	1.33	(0.87, 2.05)	*-0.08 (-1.69, 1.53)*
	>9.23 hrs/wk	75/82	0.90	(0.56, 1.43)	107/130	0.82	(0.53, 1.27)	*-0.44 (-1.75, 0.88)*
XPF (rs1800067) G/A								
	<0.01 hrs/wk	150/119	1.00	reference	21/27	0.57	(0.30, 1.07)	
	0.01-9.23 hrs/wk	179/168	0.83	(0.60, 1.15)	43/25	1.38	(0.79, 2.40)	*0.98 (-0.33, 2.29)*
	>9.23 hrs/wk	150/175	0.64	(0.46, 0.89)	30/30	0.74	(0.42, 1.31)	*0.53 (-0.39, 1.45)*
XPG (rs17655) G/C								
	<0.01 hrs/wk	88/89	1.00	reference	80/57	1.42	(0.90, 2.25)	
	0.01-9.23 hrs/wk	126/111	1.14	(0.77, 1.70)	92/81	1.15	(0.75, 1.76)	*-0.41 (-1.92, 1.09)*
	>9.23 hrs/wk	113/107	0.98	(0.65, 1.47)	65/96	0.68	(0.44, 1.06)	*-0.72 (-1.96, 0.53)*
XRCC1 (rs1799782) C/T								
	<0.01 hrs/wk	166/135	1.00	reference	12/19	0.56	(0.26, 1.21)	
	0.01-9.23 hrs/wk	205/177	0.95	(0.70, 1.29)	28/22	1.10	(0.60, 2.03)	*0.59 (-0.72, 1.90)*
	>9.23 hrs/wk	160/186	0.69	(0.50, 0.94)	26/27	0.69	(0.38, 1.25)	*0.45 (-0.56, 1.45)*
XRCC1 (rs25487) G/A								
	<0.01 hrs/wk	64/65	1.00	reference	114/90	1.26	(0.81, 1.98)	
	0.01-9.23 hrs/wk	86/80	1.05	(0.66, 1.68)	147/120	1.26	(0.82, 1.94)	*-0.06 (-1.62, 1.50)*
	>9.23 hrs/wk	69/78	0.86	(0.53, 1.40)	118/135	0.83	(0.54, 1.28)	*-0.30 (-1.58, 0.99)*

OR, odds ratio (age adjusted); 95% CI, 95% confidence interval; ICR, interaction contrast ratio

TABLE A.37 Distribution and Main Effects of Oxidative Stress Genes for Gene-Environment Interactions among all Women and Postmenopausal Women. The Long Island Breast Cancer Study Project (1996-1997).

Gene (rs)	Genotype	All Women			Postmenopausal Women		
		Ca/Co	OR	95% CI	Ca/Co	OR	95% CI
CAT (rs4756146)	TT	774/809	1.00	Reference	514/491	1.00	Reference
	CT	201/229	0.68	(0.37, 1.22)			
	CC	19/29	0.93	(0.75, 1.16)			
	CT and CC	220/258	0.90	(0.74, 1.11)	139/175	0.77	(0.59, 0.99)
CAT (rs2284365)	TT	589/610	1.00	Reference	390/372	1.00	Reference
	CT	344/371	0.72	(0.50, 1.03)			
	CC	58/85	0.98	(0.81, 1.18)			
	CT and CC	402/456	0.93	(0.78, 1.11)	261/291	0.87	(0.70, 1.09)
CAT (rs480575)	AA	517/504	1.00	Reference	341/306	1.00	Reference
	AG	378/422	0.89	(0.74, 1.08)			
	GG	89/111	0.78	(0.58, 1.06)			
	AG and GG	467/533	0.87	(0.73, 1.04)	306/340	0.82	(0.65, 1.02)
CAT (rs1001179) [1]	CC	634/696	1.00	Reference	416/435	1.00	Reference
	CT	356/348	1.12	(0.93, 1.35)			
	TT	46/42	1.20	(0.77, 1.85)			
	CT and TT	402/390	1.13	(0.95, 1.35)	265/241	1.15	(0.92, 1.43)
COMT (rs4680) [2]	GG	287/277	1.00	Reference	192/176	1.00	Reference
	AG	520/549	0.92	(0.75, 1.13)			
	AA	240/266	0.88	(0.69, 1.13)			
	AG and AA	760/815	0.91	(0.75, 1.10)	496/502	0.90	(0.70, 1.14)
COMT (rs737865) [2]	TT	498/523	1.00	Reference	329/318	1.00	Reference
	CT	430/457	0.97	(0.81, 1.17)			
	CC	101/111	0.93	(0.69, 1.26)			
	CT and CC	531/568	0.97	(0.81, 1.15)	348/359	0.91	(0.73, 1.13)

		Ca/Co	OR	95% CI	Ca/Co	OR	95% CI
GPX (rs1050450) [3]	CC	475/524	1.00	Reference	309/331	1.00	Reference
	CT	456/453	1.09	(0.91, 1.31)			
	TT	109/112	1.04	(0.78, 1.40)			
	CT and TT	565/565	1.08	(0.91, 1.28)	374/345	1.13	(0.91, 1.41)
GSTA1 (rs3957356) [4]	GG	343/386	1.00	Reference	224/229	1.00	Reference
	AG	501/522	1.06	(0.88, 1.29)			
	AA	196/182	1.18	(0.92, 1.51)			
	AG and AA	697/704	1.09	(0.91, 1.31)	461/449	1.04	(0.83, 1.31)
GSTP1 (rs1695) [5]	AA	517/518	1.00	Reference	338/317	1.00	Reference
	AG	413/461	0.89	(0.75, 1.07)			
	GG	96/90	1.09	(0.80, 1.50)			
	AG and GG	509/551	0.93	(0.78, 1.10)	335/345	0.93	(0.74, 1.15)
GSTM1 (Null vs. Present) [5]	0	505/547	1.00	Reference	329/343	1.00	Reference
	1	470/454	1.13	(0.94, 1.35)	315/273	1.21	(0.97, 1.52)
GSTT1 (Null vs. Present) [5]	0	774/795	1.00	Reference	515/487	1.00	Reference
	1	209/221	0.98	(0.79, 1.21)	132/137	0.92	(0.70, 1.21)
MnSOD (rs4880) [6]	TT	270/281	1.00	Reference	184/179	1.00	Reference
	CT	510/539	0.99	(0.78, 1.27)			
	CC	253/264	0.98	(0.80, 1.21)			
	CT and CC	763/803	0.99	(0.81, 1.20)	494/494	0.99	(0.81, 1.21)
MPO (rs2333227) [7]	GG	649/648	1.00	Reference	427/407	1.00	Reference
	AG	331/369	0.82	(0.57, 1.17)			
	AA	62/74	0.91	(0.76, 1.10)			
	AG and AA	393/443	0.90	(0.75, 1.07)	257/272	0.91	(0.73, 1.14)

Ca, cases; Co, controls; OR, odds ratio; 95% CI, 95% confidence interval

[1] Ahn et al. Am J Epidemiol. (2005) *Note: Previous report used recessive model*

[2] Gaudet et al. Breast Cancer Res Treat. (2006)

[3] Ahn et al. Cancer Epidemiol Biomarkers Prev. (2005)

[4] Ahn et al. Carcinogenesis. (2006)

[5] Steck et al. J Nutr. (2007)

[6] Gaudet et al. Cancer Causes Control. (2005)
[7] Ahn et al. Cancer Res. (2004) *Note: Previous report adjusted for age, family history and parity*

TABLE A.38 Age Adjusted Odds Ratios and 95% Confidence Intervals for the Multiplicative Effect of *CAT* SNPs on Breast Cancer Risk. The Long Island Breast Cancer Study Project (1996-1997).

Gene (rs)	Ca/Co	OR	95% CI	Ca/Co	OR	95% CI	*p for interaction*
			CAT (rs2284365)				
CAT (rs480575)		TT			CT and CC		0.087
AA	505/487	1.00	reference	9/14	1.00	reference	
AG and GG	71/100	0.68	(0.49, 0.95)	383/428	1.57	(0.64, 3.84)	

Ca, case; Co, control; OR, odds ratio; 95% CI, 95% confidence interval

TABLE A.39. Age Adjusted Odds Ratios and 95% Confidence Intervals for the Multiplicative Effect of MMR Genes on Breast Cancer Risk. The Long Island Breast Cancer Study Project (1996-1997).

Gene (rs)	Ca/Co	OR	95% CI	Ca/Co	OR	95% CI	*p for interaction*
				MSH3 (rs1650663)			
MLH1 (rs2286940)		TT			CT and CC		*0.063*
CC	157/176	1.00	reference	181/163	1.00	reference	
CT and TT	337/354	1.07	(0.82, 1.40)	324/373	0.75	(0.57, 0.97)	
				MLH1 (rs1799977)			
MSH2 (rs2303428)		AA			AG and GG		*0.045*
TT	417/434	1.00	reference	410/448	1.00	reference	
CT and CC	72/96	0.82	(0.58, 1.15)	86/71	1.35	(0.96, 1.92)	

Ca, case; Co, control; OR, odds ratio; 95% CI, 95% confidence interval

FIGURE A.1 Female Breast Cancer - Incidence and Mortality Rates by Age and Race, US, 2002-2006 (American Cancer Society, Breast Cancer Facts and Figures 2009-2010)
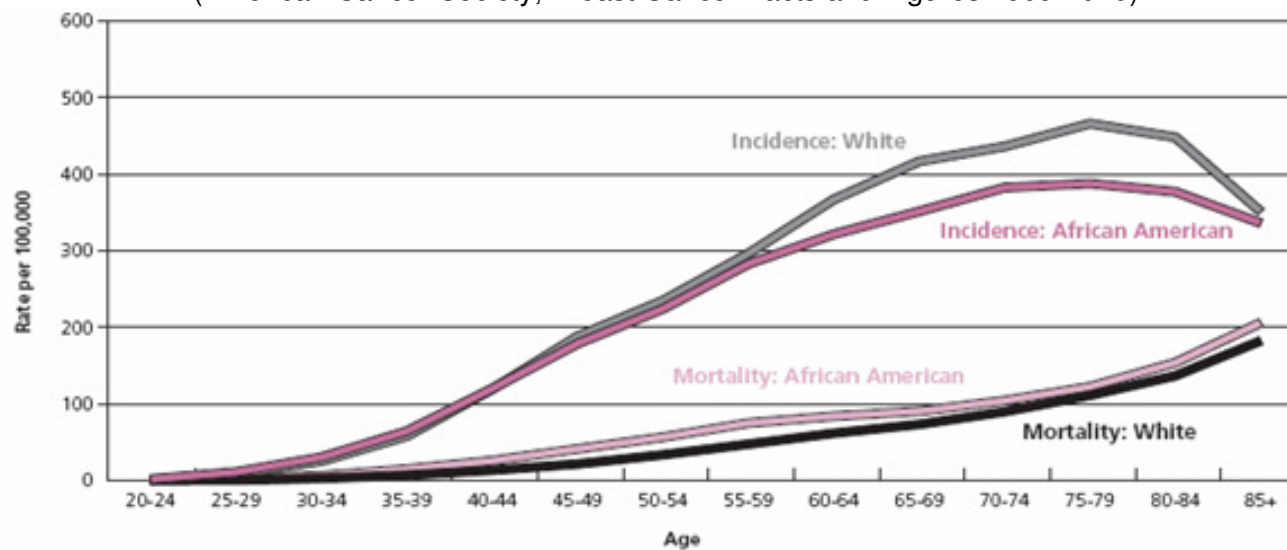
Data sources: Incidence – North American Association of Central Cancer Registries, 2009. Mortality – National Center for Health Statistics, Centers for Disease Control and Prevention, 2009.

FIGURE A.2 Commonly Cited Physical Activity Mechanisms

184
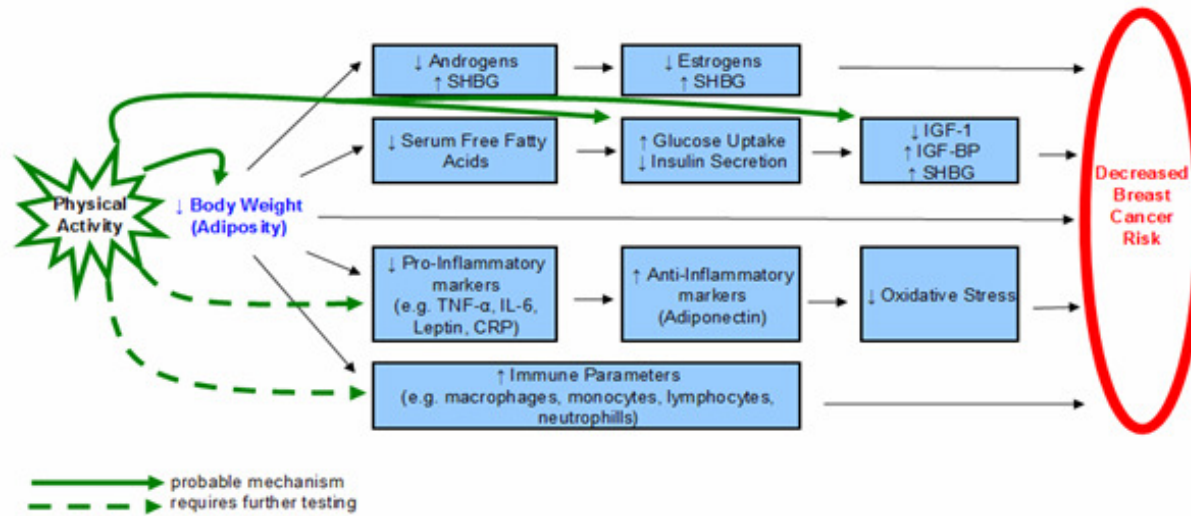
FIGURE A.3 Physical Activity Mechanisms and Carcinogenesis (Adapted from Rundle 2005)
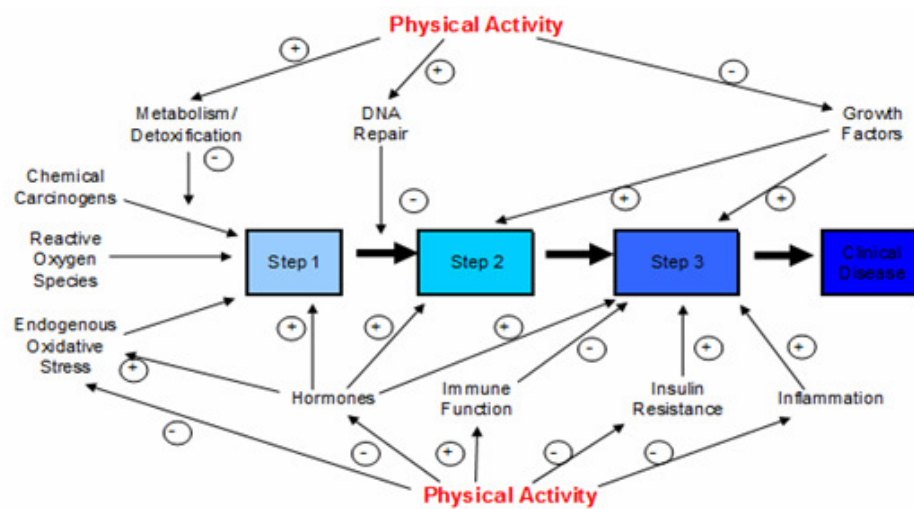
FIGURE A.4 Interactions between complexes of MutS and MutL related proteins during mismatch repair (Adapted from Kolodner and Marsischky 1999)
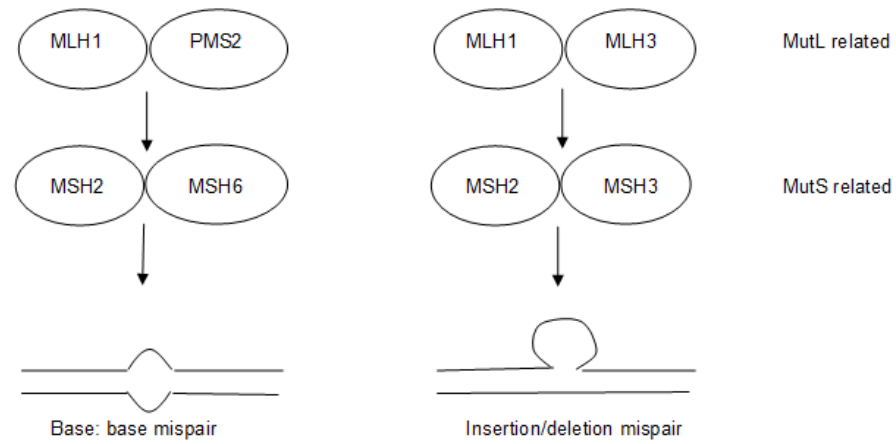
FIGURE A.5 Contact, Cooperation, and Response Rates among case and controls,
Long Island Breast Cancer Study Project. 1996-1997

Cases Sampled (2271) → Ineligible (241), Physician Refusal (193)

Eligible - A (1837) → Unlocatable - B (25), Refusal (217), Too Ill (31), Cognitively Impaired (29), Deceased (16), Other (1), Partial Completion of Main Questionnaire (10)

Completion of Main Questionnaire - C (1508)

Contact Rate = (A-B)/A = D = 98.6%
Cooperation Rate = C/(A-B) = E = 83.2%
Response Rate = D*E = C/A = 82.1%

Controls Sampled (2714) → Ineligible (233)

Eligible - A (2418) → Unlocatable - B (193), Refusal (530), Too Ill (101), Cognitively Impaired (67), Deceased (25), Other (2), Partial Completion of Main Questionnaire (7)

Completion of Main Questionnaire - C (1556)

Contact Rate = (A-B)/A = D = 92.2%
Cooperation Rate = C/(A-B) = E = 68.0%
Response Rate = D*E = C/A = 62.7%

FIGURE A.6 Linkage Disequilibrium Plot *CAT* (CEU HapMap population)

FIGURE A.7 Linkage Disequilibrium Plot *MLH1* (CEU HapMap population)
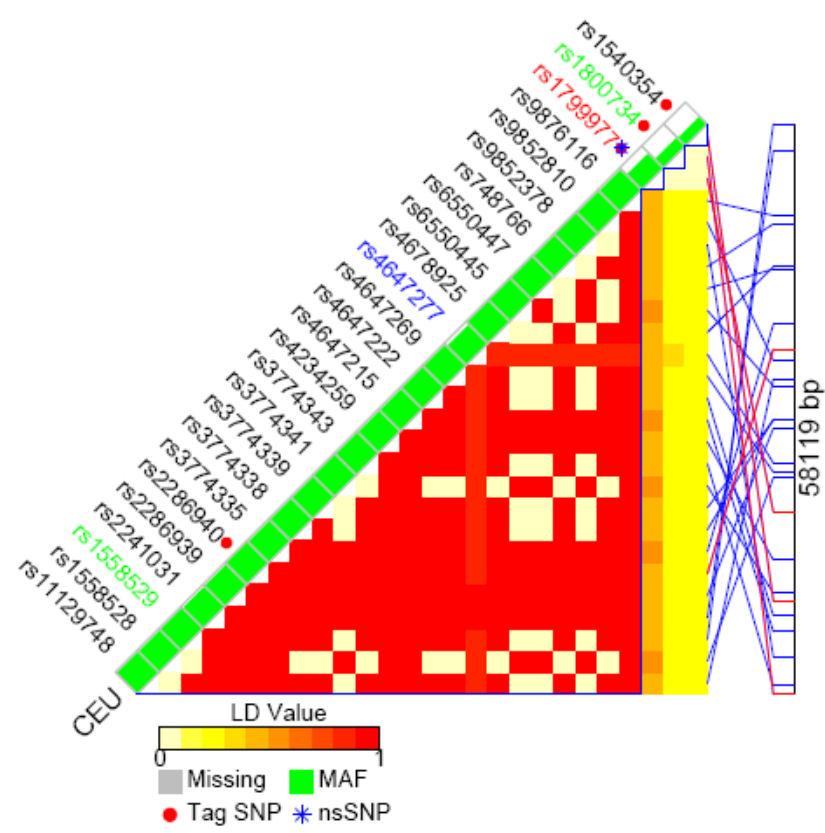
**FIGURE A.8** Linkage Disequilibrium Plot *MSH2* (CEU HapMap population)
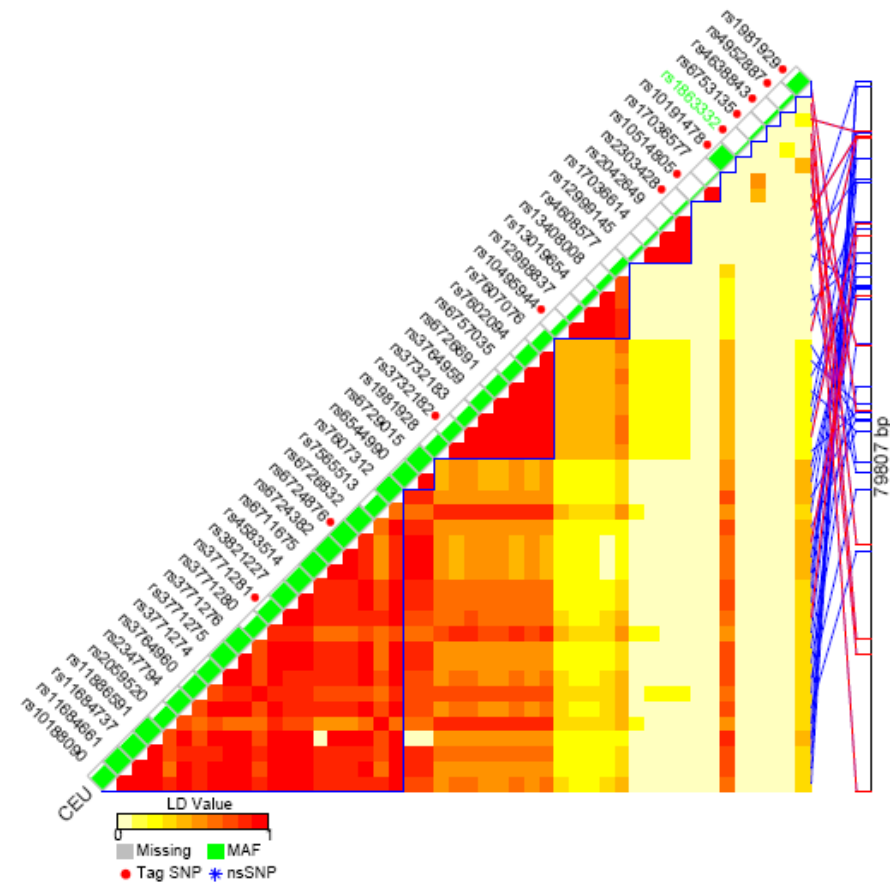
FIGURE A.9 Linkage Disequilibrium Plot *MSH3* (CEU HapMap population)
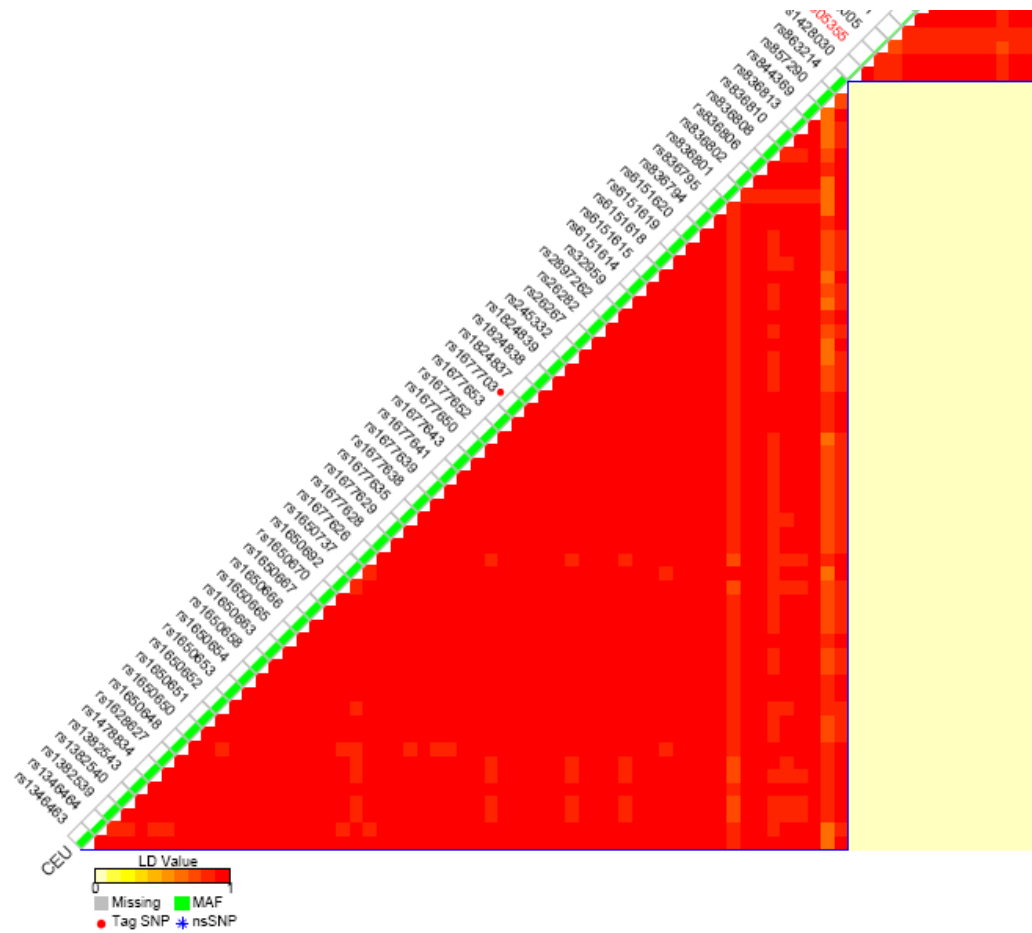
FIGURE A.10 Simple Continuous Plot of Postmenopausal Recreational Physical Activity and Risk of Breast Cancer, Long Island Breast Cancer Study Project. 1996-1997
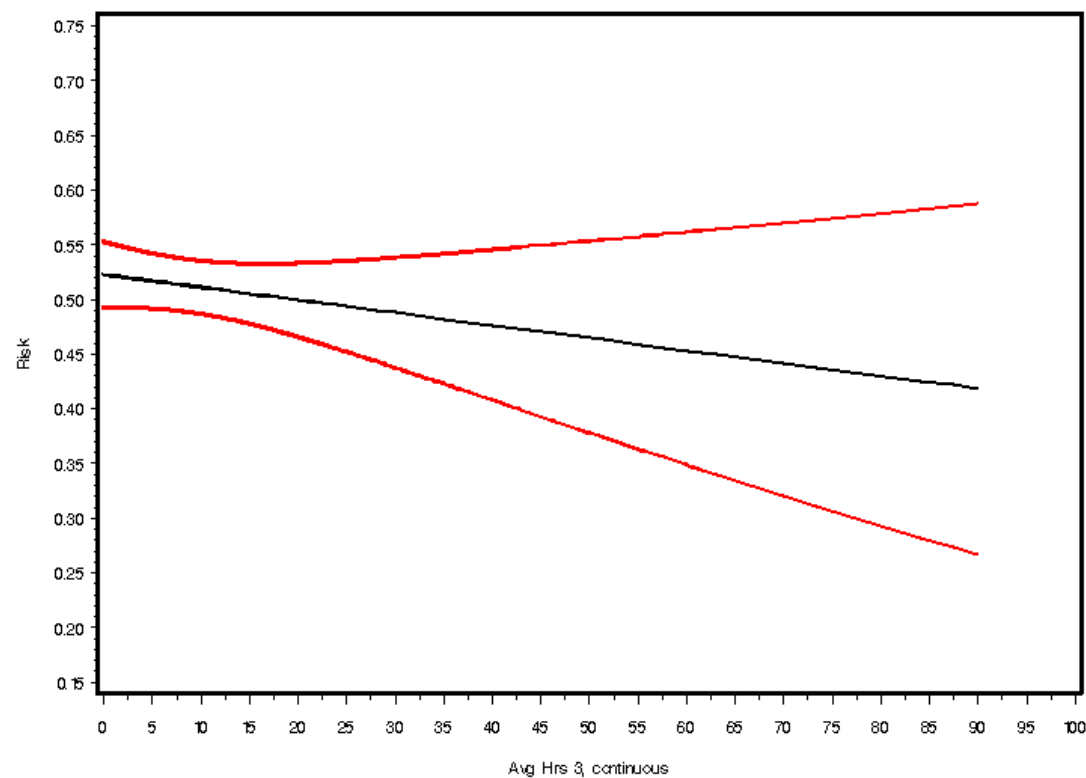
FIGURE A.11 Quadratic Plot of Postmenopausal Recreational Physical Activity and Risk of Breast Cancer, Long Island Breast Cancer Study Project. 1996-1997
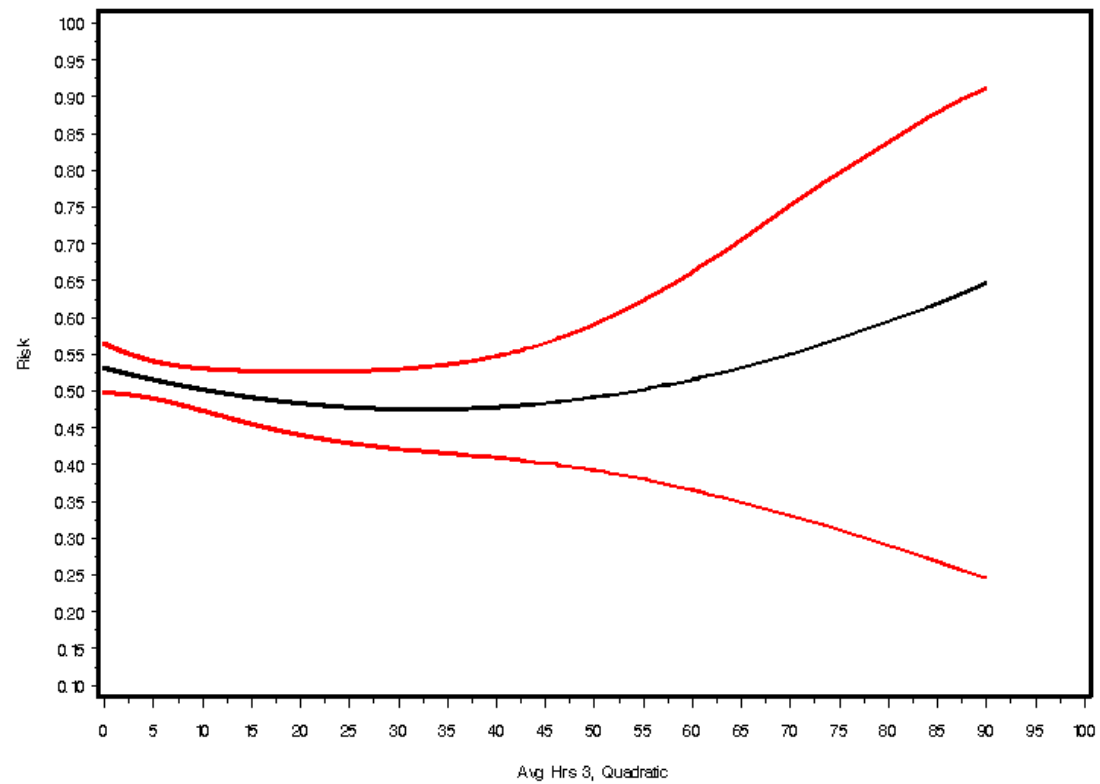
FIGURE A.12 Postmenopausal Recreational Physical Activity and Risk of Breast Cancer Categorized by Deciles among Controls, Long Island Breast Cancer Study Project. 1996-1997
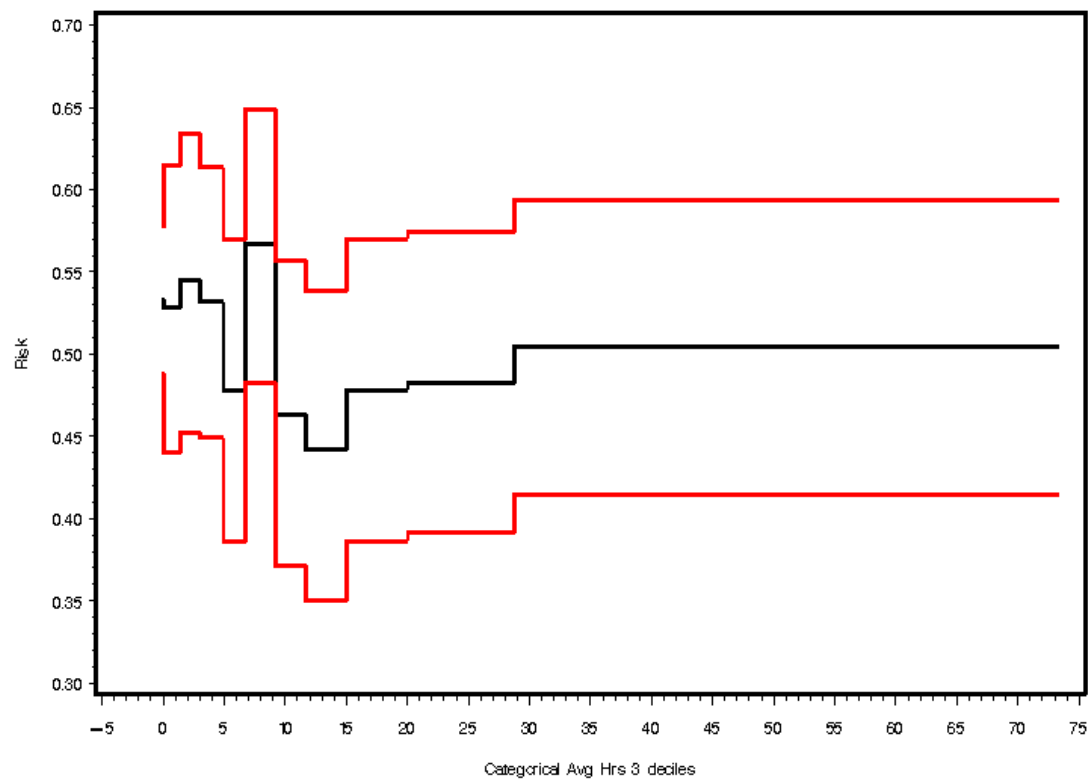
FIGURE A.13 Postmenopausal Recreational Physical Activity and Risk of Breast Cancer Categorized by Quartiles among Controls, Long Island Breast Cancer Study Project. 1996-1997
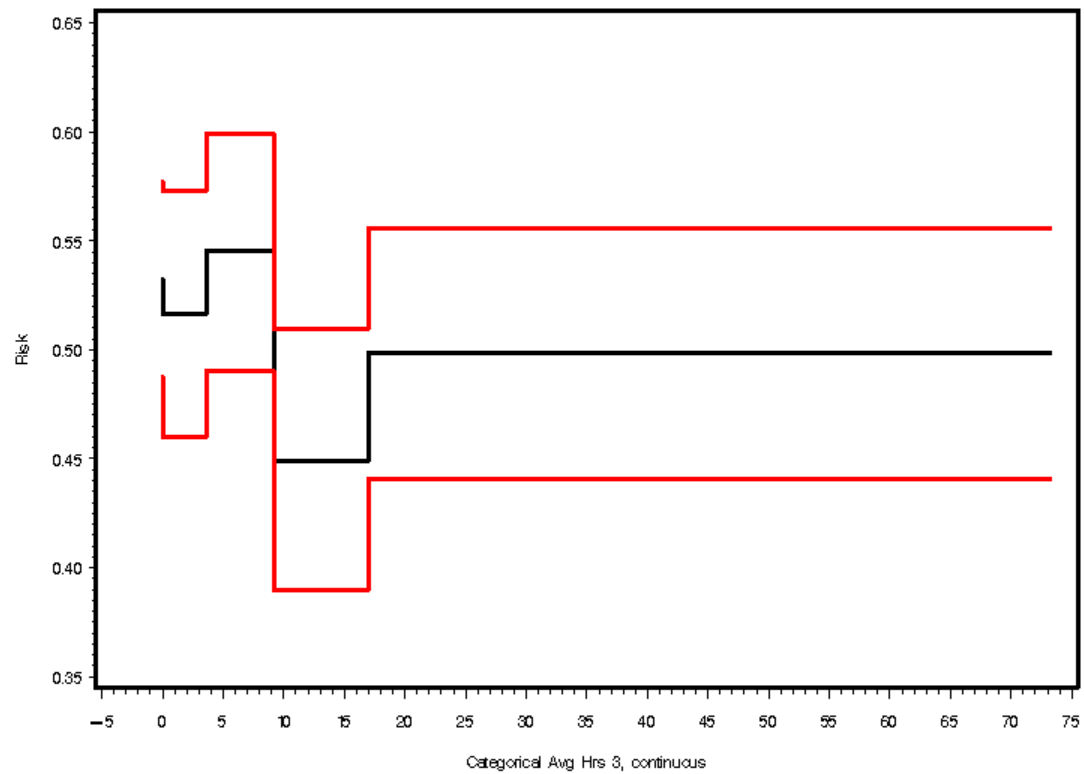
FIGURE A.14 Linear Spline of Postmenopausal Recreational Physical Activity and Risk of Breast Cancer, Long Island Breast Cancer Study Project. 1996-1997
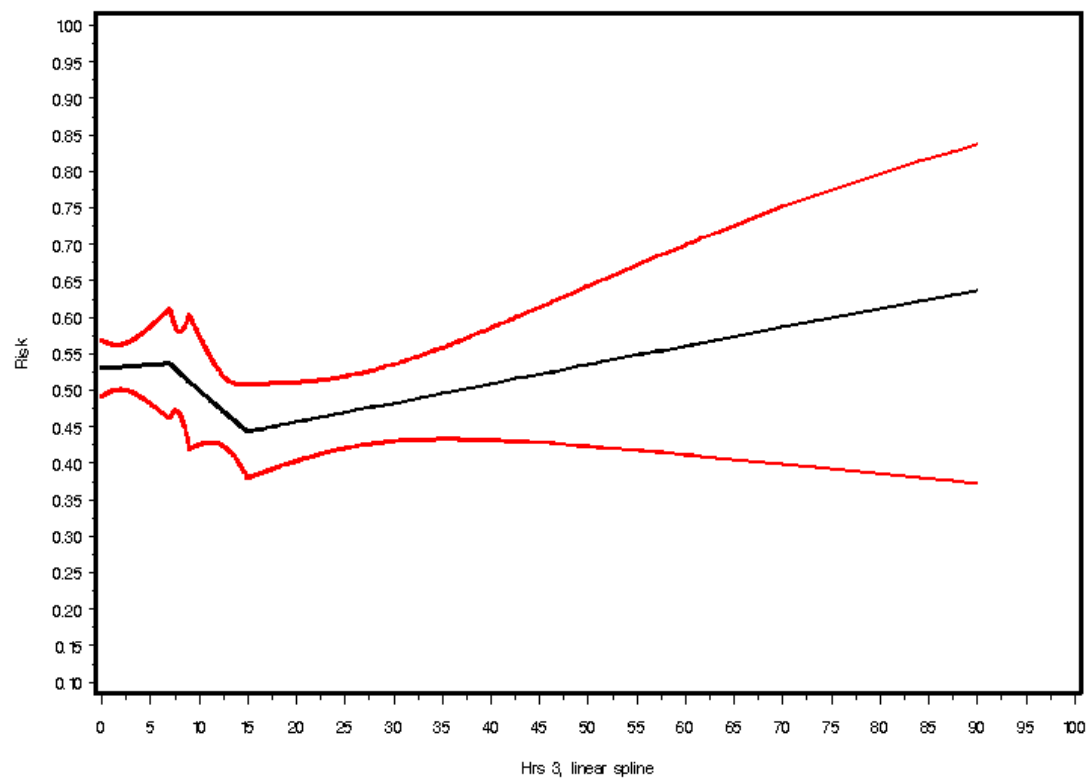
FIGURE A.15 Quadratic Spline of Postmenopausal Recreational Physical Activity and Risk of Breast Cancer, Long Island Breast Cancer Study Project. 1996-1997
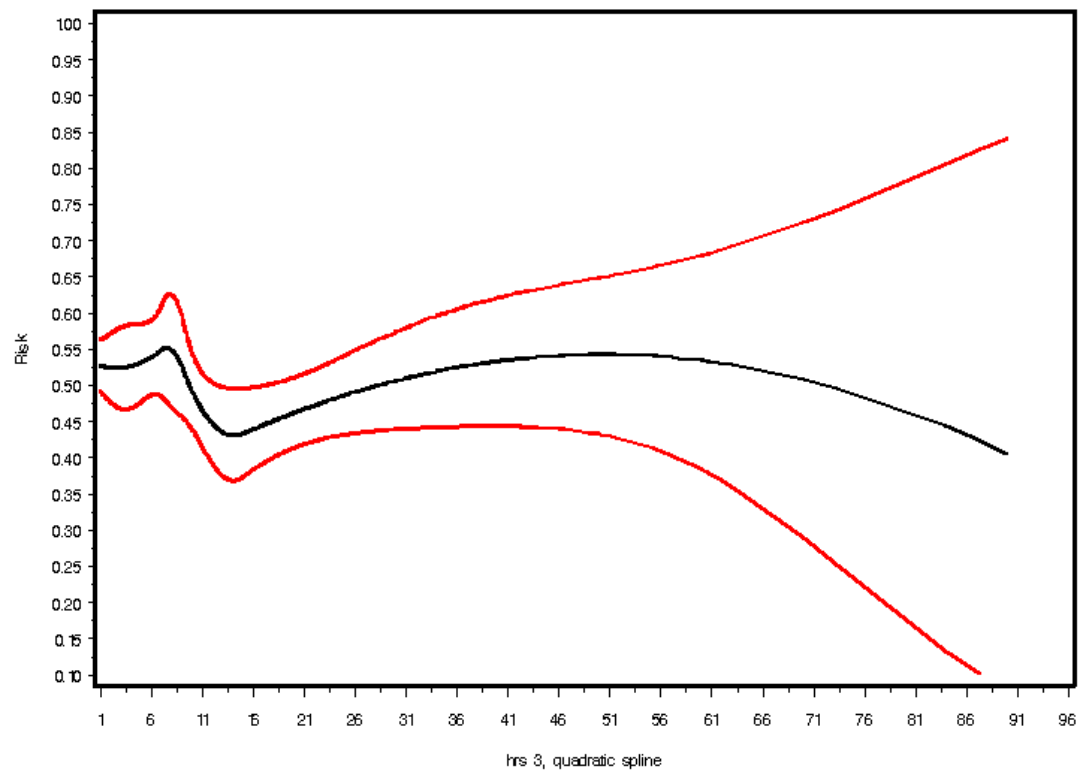
FIGURE A.16 Directed Acyclic Graph for the Association Between Breast Cancer Risk and Genetic Variants.
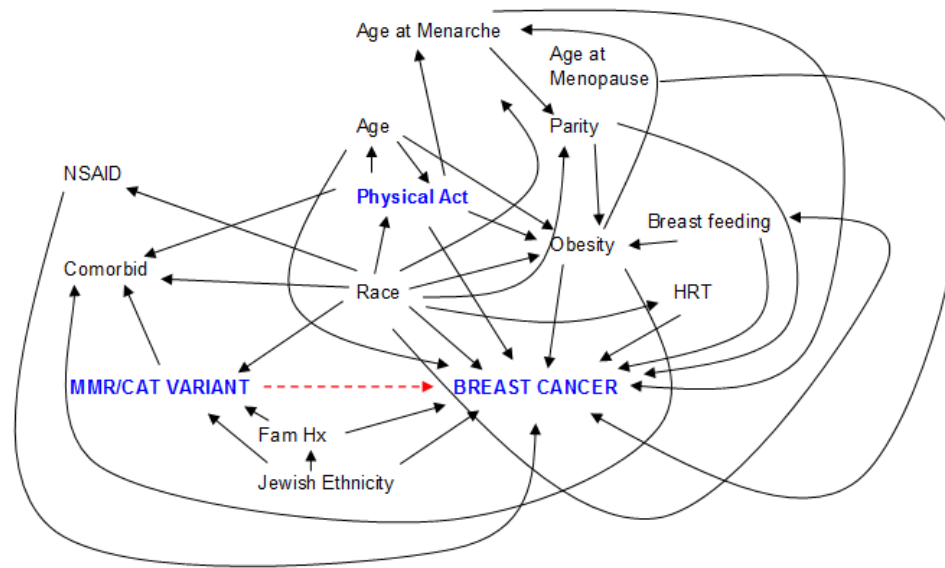
Figure A.17 Power Curves for Main Effects of Genotype and Breast Cancer (All women Combined), Long Island Breast Cancer Study Project. 1996-1997
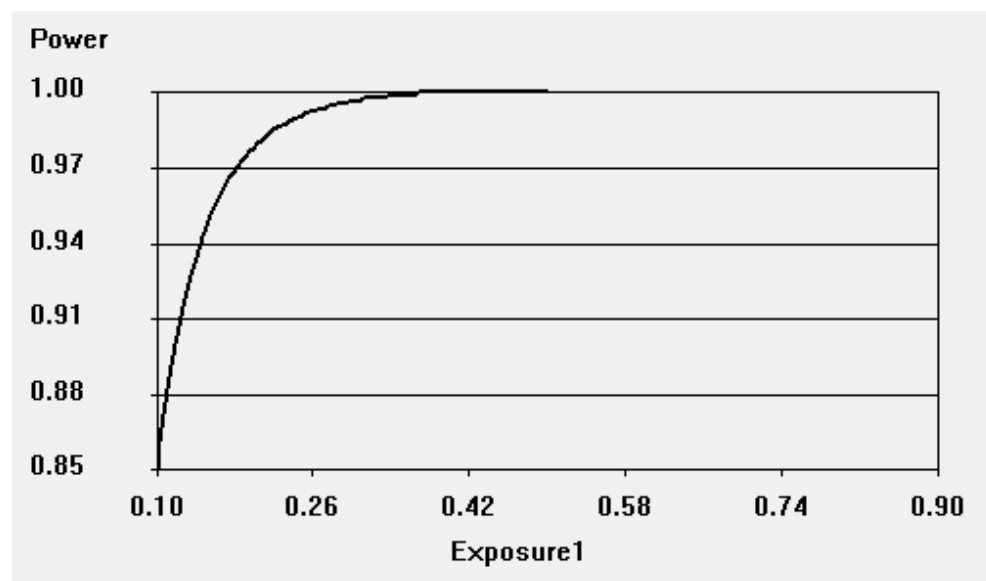
Figure A.18 Power Curves for Multiplicative Interactions and Breast Cancer (All women Combined), Long Island Breast Cancer Study Project. 1996-1997
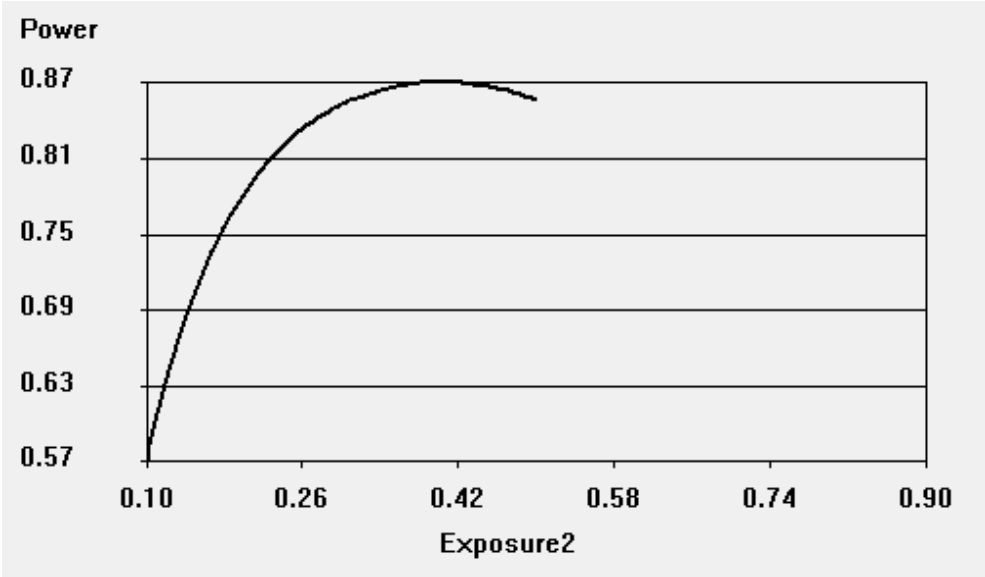
Figure A.19 Power Curves for Additive Interactions and Breast Cancer (All women Combined), Long Island Breast Cancer Study Project. 1996-1997
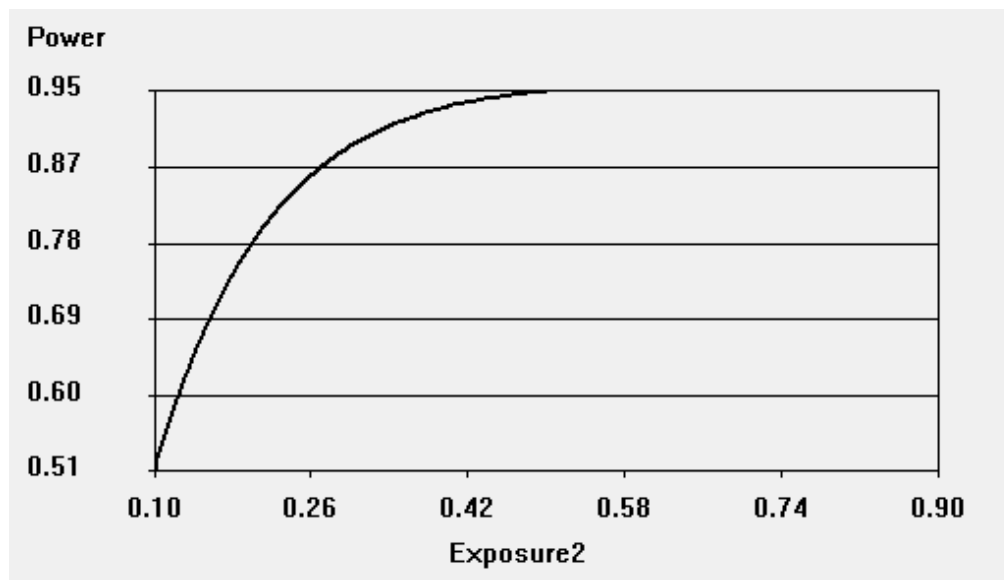
FIGURE A.20(a) Stratum Specific Effects of *CAT* CC genotype using Linear Spline for Postmenopausal Recreational Physical Activity, Long Island Breast Cancer Study Project. 1996-1997
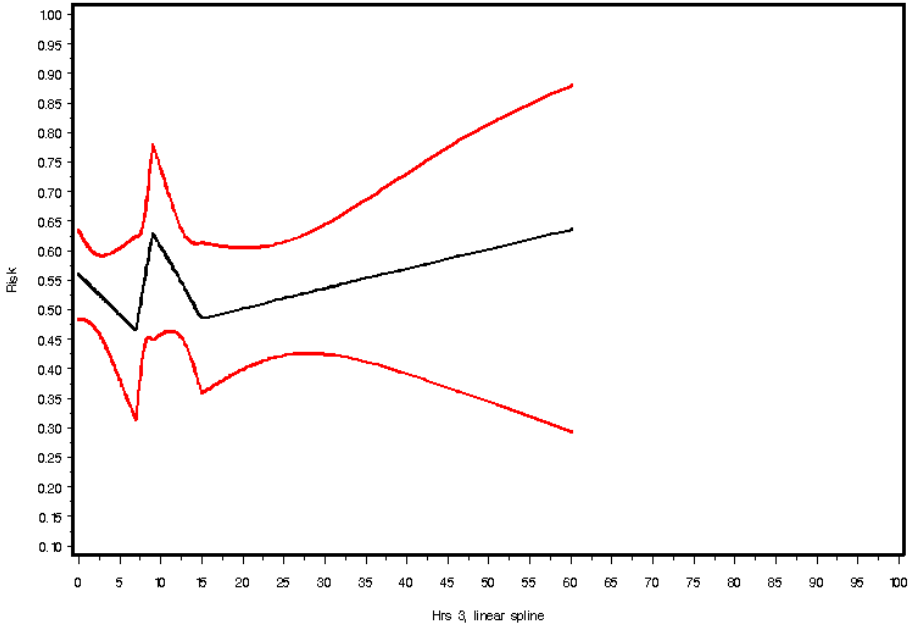
FIGURE A.20(b) Stratum Specific Effects of *CAT* CT or TT genotype using Linear Spline for Postmenopausal Recreational Physical Activity, Long Island Breast Cancer Study Project. 1996-1997
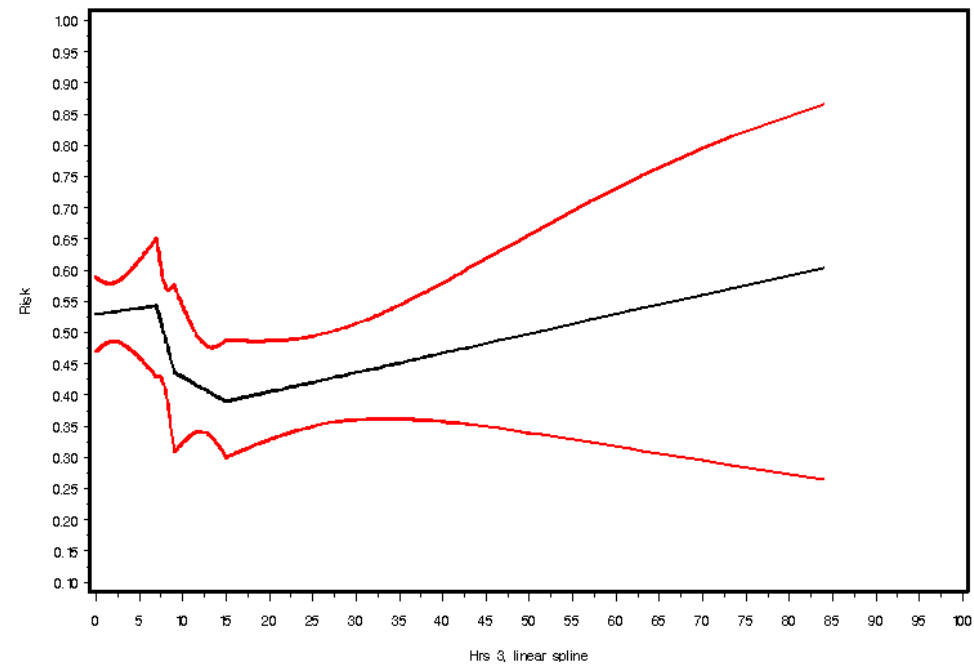
FIGURE A.21(a) Stratum Specific Effects of *GSTP1* AA genotype using Linear Spline for Postmenopausal Recreational Physical Activity, Long Island Breast Cancer Study Project. 1996-1997
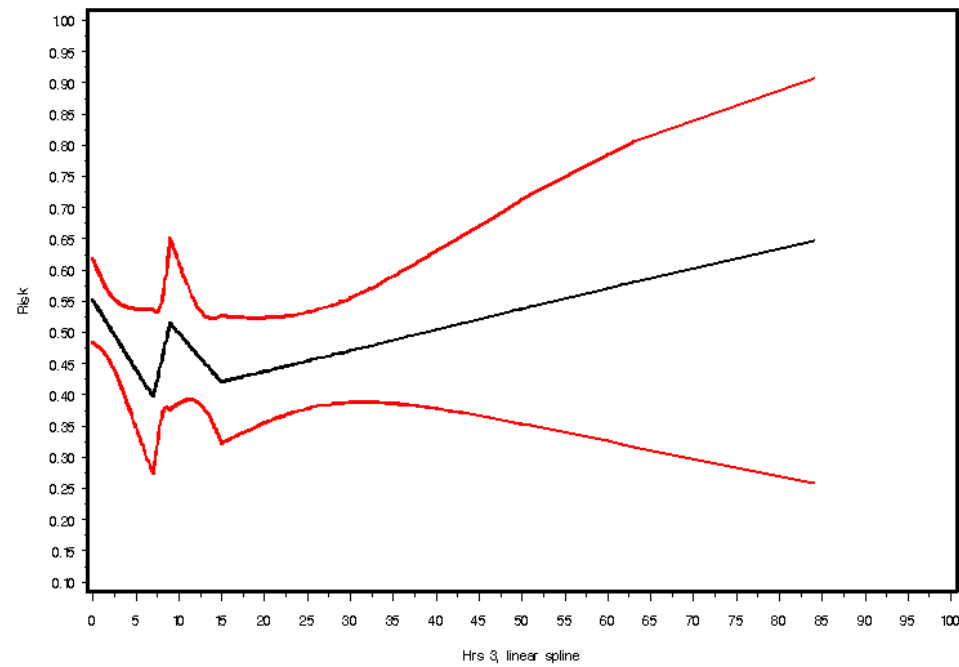
FIGURE A.21(b) Stratum Specific Effects of *GSTP1* AG or GG genotype using Linear Spline for Postmenopausal Recreational Physical Activity, Long Island Breast Cancer Study Project. 1996-1997
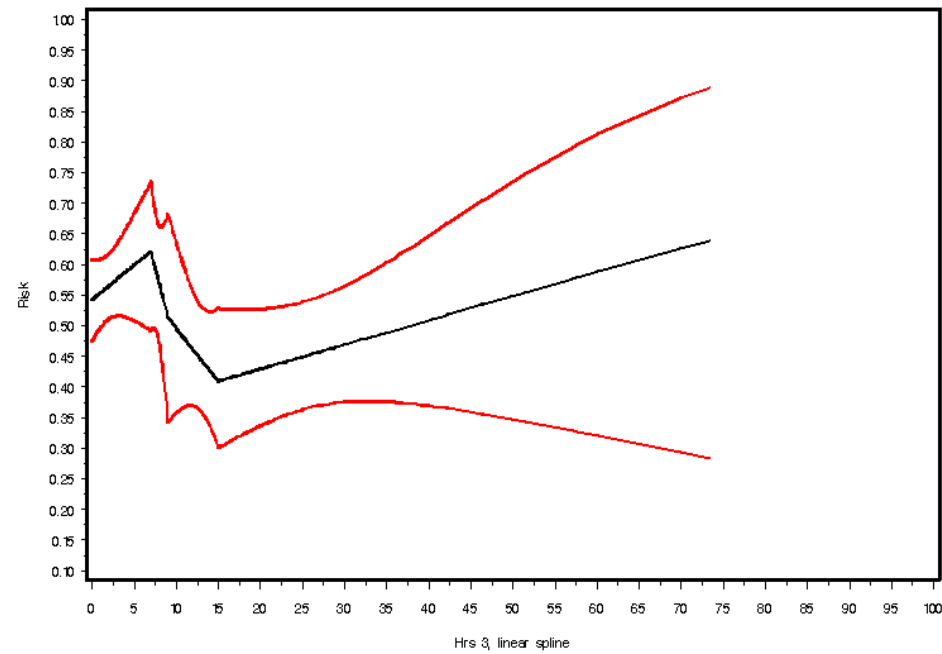
References

1. Kelsey JL, Horn-Ross PL (1993) Breast cancer: magnitude of the problem and descriptive epidemiology. Epidemiol Rev 15:7-16

2. American Cancer Society. (2012). Retrieved 2/20/2012, from http://www.cancer.org/Cancer/BreastCancer/DetailedGuide/breast-cancer-key-statistics

3. American Cancer Society. Breast Cancer Facts & Figures 2007-2008. Atlanta: American Cancer Society, Inc.

4. Singletary SE (2003) Rating the risk factors for breast cancer. Ann Surg 237:474-482. DOI:10.1097/01.SLA.0000059969.64262.87

5. Hankinson SE, Colditz GA, Willett WC (2004) Towards an integrated model for breast cancer etiology: the lifelong interplay of genes, lifestyle, and hormones. Breast Cancer Res 6:213-218. DOI:10.1186/bcr921

6. Clemons M, Goss P (2001) Estrogen and the risk of breast cancer. N Engl J Med 344:276-285. DOI:10.1056/NEJM200101253440407

7. Muti P (2004) The role of endogenous hormones in the etiology and prevention of breast cancer: the epidemiological evidence. Ann N Y Acad Sci 1028:273-282. DOI:10.1196/annals.1322.031

8. Nathanson KL, Wooster R, Weber BL (2001) Breast cancer genetics: what we know and what we need. Nat Med 7:552-556. DOI:10.1038/87876

9. Krainer M, Silva-Arrieta S, FitzGerald MG et al (1997) Differential contributions of BRCA1 and BRCA2 to early-onset breast cancer. N Engl J Med 336:1416-1421. DOI:10.1056/NEJM199705153362003

10. Antoniou A, Pharoah PD, Narod S et al (2003) Average risks of breast and ovarian cancer associated with BRCA1 or BRCA2 mutations detected in case Series unselected for family history: a combined analysis of 22 studies. Am J Hum Genet 72:1117-1130. DOI:10.1086/375033

11. Hirshfield KM, Rebbeck TR, Levine AJ (2010) Germline mutations and polymorphisms in the origins of cancers in women. J Oncol 2010:297671. DOI:10.1155/2010/297671

12. Kelsey JL, Gammon MD, John EM (1993) Reproductive factors and breast cancer. Epidemiol Rev 15:36-47

13. Hsieh CC, Trichopoulos D, Katsouyanni K et al (1990) Age at menarche, age at menopause, height and obesity as risk factors for breast cancer: associations and interactions in an international case-control study. Int J Cancer 46:796-800

14. Lambe M, Hsieh CC, Tsaih SW et al (1998) Parity, age at first birth and the risk of carcinoma in situ of the breast. Int J Cancer 77:330-332

15. Rosner B, Colditz GA, Willett WC (1994) Reproductive risk factors in a prospective study of breast cancer: the Nurses' Health Study. Am J Epidemiol 139:819-835

16. Collaborative Group on Hormonal Factors in Breast Cancer (2002) Breast cancer and breastfeeding: collaborative reanalysis of individual data from 47 epidemiological studies in 30 countries, including 50302 women with breast cancer and 96973 women without the disease. Lancet 360:187-195. DOI:10.1016/S0140-6736(02)09454-0

17. Britt K, Ashworth A, Smalley M (2007) Pregnancy and the risk of breast cancer. Endocr Relat Cancer 14:907-933. DOI:10.1677/ERC-07-0137

18. Gammon MD, Neugut AI, Santella RM et al (2002) The Long Island Breast Cancer Study Project: description of a multi-institutional collaboration to identify environmental risk factors for breast cancer. Breast Cancer Res Treat 74:235-254

19. Casey PM, Cerhan JR, Pruthi S (2008) Oral contraceptive use and risk of breast cancer. Mayo Clin Proc 83:86-90; quiz 90-1

20. Anonymous (1997) Breast cancer and hormone replacement therapy: collaborative reanalysis of data from 51 epidemiological studies of 52,705 women with breast cancer and 108,411 women without breast cancer. Collaborative Group on Hormonal Factors in Breast Cancer. Lancet 350:1047-1059

21. Marchbanks PA, McDonald JA, Wilson HG et al (2002) Oral contraceptives and the risk of breast cancer. N Engl J Med 346:2025-2032. DOI:10.1056/NEJMoa013202

22. Chen FP (2009) Postmenopausal hormone therapy and risk of breast cancer. Chang Gung Med J 32:140-147

23. Chen WY, Manson JE, Hankinson SE et al (2006) Unopposed estrogen therapy and the risk of invasive breast cancer. Arch Intern Med 166:1027-1032. DOI:10.1001/archinte.166.9.1027

24. Shantakumar S, Terry MB, Paykin A et al (2007) Age and menopausal effects of hormonal birth control and hormone replacement therapy in relation to breast cancer risk. Am J Epidemiol 165:1187-1198. DOI:10.1093/aje/kwm006

25. Adami, H.O., et al., Textbook of cancer epidemiology. 2002, Oxford; New York, N.Y.: Oxford University Press. xxiv, 599).

26. Gammon MD, Santella RM, Neugut AI et al (2002) Environmental toxins and breast cancer on Long Island. I. Polycyclic aromatic hydrocarbon DNA adducts. Cancer Epidemiol Biomarkers Prev 11:677-685

27. Gammon MD, Eng SM, Teitelbaum SL et al (2004) Environmental tobacco smoke and breast cancer incidence. Environ Res 96:176-185. DOI:10.1016/j.envres.2003.08.009

28. Singletary KW, Gapstur SM (2001) Alcohol and breast cancer: review of epidemiologic and experimental evidence and potential mechanisms. JAMA 286:2143-2151

29. Fan S, Meng Q, Gao B et al (2000) Alcohol stimulates estrogen receptor signaling in human breast cancer cell lines. Cancer Res 60:5635-5639

30. Key J, Hodgson S, Omar RZ et al (2006) Meta-analysis of studies of alcohol and breast cancer with consideration of the methodological issues. Cancer Causes Control 17:759-770. DOI:10.1007/s10552-006-0011-0

31. Michels KB, Mohllajee AP, Roset-Bahmanyar E et al (2007) Diet and breast cancer: a review of the prospective observational studies. Cancer 109:2712-2749. DOI:10.1002/cncr.22654

32. Brennan SF, Cantwell MM, Cardwell CR et al (2010) Dietary patterns and breast cancer risk: a systematic review and meta-analysis. Am J Clin Nutr 91:1294-1302. DOI:10.3945/ajcn.2009.28796

33. Lof M, Weiderpass E (2009) Impact of diet on breast cancer risk. Curr Opin Obstet Gynecol 21:80-85. DOI:10.1097/GCO.0b013e32831d7f22

34. Friedenreich CM (2001) Review of anthropometric factors and breast cancer risk. Eur J Cancer Prev 10:15-32

35. Pichard C, Plu-Bureau G, Neves-E Castro M et al (2008) Insulin resistance, obesity and breast cancer risk. Maturitas 60:19-30. DOI:10.1016/j.maturitas.2008.03.002

36. Key TJ, Allen NE, Verkasalo PK et al (2001) Energy balance and cancer: the role of sex hormones. Proc Nutr Soc 60:81-89

37. Gammon MD, John EM, Britton JA (1998) Recreational and occupational physical activities and risk of breast cancer. J Natl Cancer Inst 90:100-117

38. Monninkhof EM, Elias SG, Vlems FA et al (2007) Physical activity and breast cancer: a systematic review. Epidemiology 18:137-157. DOI:10.1097/01.ede.0000251167.75581.98

39. Friedenreich CM, Rohan TE (1995) A review of physical activity and breast cancer. Epidemiology 6:311-317

40. Friedenreich CM, Orenstein MR (2002) Physical activity and cancer prevention: etiologic evidence and biological mechanisms. J Nutr 132:3456S-3464S

41. Friedenreich CM (2004) Physical activity and breast cancer risk: the effect of menopausal status. Exerc Sport Sci Rev 32:180-184

42. Friedenreich CM, Cust AE (2008) Physical activity and breast cancer risk: impact of timing, type and dose of activity and population subgroup effects. Br J Sports Med 42:636-647. DOI:10.1136/bjsm.2006.029132

43. Thune I, Furberg AS (2001) Physical activity and cancer risk: dose-response and cancer, all sites and site-specific. Med Sci Sports Exerc 33:S530-50; discussion S609-10

44. McTiernan A (2008) Mechanisms linking physical activity with cancer. Nat Rev Cancer 8:205-211. DOI:10.1038/nrc2325

45. Latikka P, Pukkala E, Vihko V (1998) Relationship between the risk of breast cancer and physical activity. An epidemiological perspective. Sports Med 26:133-143

46. Tehard B, Friedenreich CM, Oppert JM et al (2006) Effect of physical activity on women at increased risk of breast cancer: results from the E3N cohort study. Cancer Epidemiol Biomarkers Prev 15:57-64. DOI:10.1158/1055-9965.EPI-05-0603

47. Mittendorf R, Longnecker MP, Newcomb PA et al (1995) Strenuous physical activity in young adulthood and risk of breast cancer (United States). Cancer Causes Control 6:347-353

48. Thune I, Brenn T, Lund E et al (1997) Physical activity and the risk of breast cancer. N Engl J Med 336:1269-1275. DOI:10.1056/NEJM199705013361801

49. Batty GD (2002) Physical activity and coronary heart disease in older adults. A systematic review of epidemiological studies. Eur J Public Health 12:171-176

50. Elsawy B, Higgins KE (2010) Physical activity guidelines for older adults. Am Fam Physician 81:55-59

51. LaMonte MJ, Blair SN, Church TS (2005) Physical activity and diabetes prevention. J Appl Physiol 99:1205-1213. DOI:10.1152/japplphysiol.00193.2005

52. Schmitt NM, Schmitt J, Doren M (2009) The role of physical activity in the prevention of osteoporosis in postmenopausal women-An update. Maturitas 63:34-38. DOI:10.1016/j.maturitas.2009.03.002

53. Vainio H, Kaaks R, Bianchini F (2002) Weight control and physical activity in cancer prevention: international evaluation of the evidence. Eur J Cancer Prev 11 Suppl 2:S94-100

54. Caspersen CJ (1989) Physical activity epidemiology: concepts, methods, and applications to exercise science. Exerc Sport Sci Rev 17:423-473

55. Newton RU, Galvao DA (2008) Exercise in prevention and management of cancer. Curr Treat Options Oncol 9:135-146. DOI:10.1007/s11864-008-0065-1

56. Centers for Disease Control - Office of Surveillance, Epidemiology, and Laboratory Services (2010) Prevalence and Trend Data - Exercise 2010. 2011

57. Ainsworth BE, Haskell WL, Leon AS et al (1993) Compendium of physical activities: classification of energy costs of human physical activities. Med Sci Sports Exerc 25:71-80

58. Pate RR, Pratt M, Blair SN et al (1995) Physical activity and public health. A recommendation from the Centers for Disease Control and Prevention and the American College of Sports Medicine. JAMA 273:402-407

59. McTiernan A, Ulrich C, Slate S et al (1998) Physical activity and cancer etiology: associations and mechanisms. Cancer Causes Control 9:487-509

60. Maruti SS, Willett WC, Feskanich D et al (2008) A prospective study of age-specific physical activity and premenopausal breast cancer. J Natl Cancer Inst 100:728-737. DOI:10.1093/jnci/djn135

61. Dorn J, Vena J, Brasure J et al (2003) Lifetime physical activity and breast cancer risk in pre- and postmenopausal women. Med Sci Sports Exerc 35:278-285. DOI:10.1249/01.MSS.0000048835.59454.8D

62. Peters TM, Moore SC, Gierach GL et al (2009) Intensity and timing of physical activity in relation to postmenopausal breast cancer risk: the prospective NIH-AARP diet and health study. BMC Cancer 9:349. DOI:10.1186/1471-2407-9-349

63. McCullough LE, Eng SM, Bradshaw PT et al (2012) Fat or fit: The joint effects of physical activity, weight gain, and body size on breast cancer risk. Cancer 118:4860-4868. DOI:10.1002/cncr.27433; 10.1002/cncr.27433

64. Bernstein L, Patel AV, Ursin G et al (2005) Lifetime recreational exercise activity and breast cancer risk among black women and white women. J Natl Cancer Inst 97:1671-1679. DOI:10.1093/jnci/dji374

65. Sesso HD, Paffenbarger RS,Jr, Lee IM (1998) Physical activity and breast cancer risk in the College Alumni Health Study (United States). Cancer Causes Control 9:433-439

66. Breslow RA, Ballard-Barbash R, Munoz K et al (2001) Long-term recreational physical activity and breast cancer in the National Health and Nutrition Examination Survey I epidemiologic follow-up study. Cancer Epidemiol Biomarkers Prev 10:805-808

67. Gilliland FD, Li YF, Baumgartner K et al (2001) Physical activity and breast cancer risk in hispanic and non-hispanic white women. Am J Epidemiol 154:442-450

68. Dallal CM, Sullivan-Halley J, Ross RK et al (2007) Long-term recreational physical activity and risk of invasive and in situ breast cancer: the California teachers study. Arch Intern Med 167:408-415. DOI:10.1001/archinte.167.4.408

69. Schmidt ME, Chang-Claude J, Slanger T et al (2009) Physical activity and postmenopausal breast cancer: effect modification by other breast cancer risk factors. Methods Inf Med 48:444-450. DOI:10.3414/ME9239

70. Sprague BL, Trentham-Dietz A, Newcomb PA et al (2007) Lifetime recreational and occupational physical activity and risk of in situ and invasive breast cancer. Cancer Epidemiol Biomarkers Prev 16:236-243. DOI:10.1158/1055-9965.EPI-06-0713

71. Pijpe A, Manders P, Brohet RM et al (2010) Physical activity and the risk of breast cancer in BRCA1/2 mutation carriers. Breast Cancer Res Treat 120:235-244. DOI:10.1007/s10549-009-0476-0

72. King MC, Marks JH, Mandell JB et al (2003) Breast and ovarian cancer risks due to inherited mutations in BRCA1 and BRCA2. Science 302:643-646. DOI:10.1126/science.1088759

73. Bardia A, Hartmann LC, Vachon CM et al (2006) Recreational physical activity and risk of postmenopausal breast cancer based on hormone receptor status. Arch Intern Med 166:2478-2483. DOI:10.1001/archinte.166.22.2478

74. Rundle A (2005) Molecular epidemiology of physical activity and cancer. Cancer Epidemiol Biomarkers Prev 14:227-236

75. Neilson HK, Friedenreich CM, Brockton NT et al (2009) Physical activity and postmenopausal breast cancer: proposed biologic mechanisms and areas for future research. Cancer Epidemiol Biomarkers Prev 18:11-27. DOI:10.1158/1055-9965.EPI-08-0756

76. Lorincz AM, Sukumar S (2006) Molecular links between obesity and breast cancer. Endocr Relat Cancer 13:279-292. DOI:10.1677/erc.1.00729

77. Calle EE, Kaaks R (2004) Overweight, obesity and cancer: epidemiological evidence and proposed mechanisms. Nat Rev Cancer 4:579-591. DOI:10.1038/nrc1408

78. Rogers CJ, Colbert LH, Greiner JW et al (2008) Physical activity and cancer prevention : pathways and targets for intervention. Sports Med 38:271-296

79. Matsuzawa Y, Shimomura I, Nakamura T et al (1995) Pathophysiology and pathogenesis of visceral fat obesity. Obes Res 3 Suppl 2:187S-194S

80. Bernstein L, Ross RK, Lobo RA et al (1987) The effects of moderate physical activity on menstrual cycle patterns in adolescence: implications for breast cancer prevention. Br J Cancer 55:681-685

81. Harlow SD, Matanoski GM (1991) The association between weight, physical activity, and stress and variation in the length of the menstrual cycle. Am J Epidemiol 133:38-49

82. Cooper GS, Sandler DP, Whelan EA et al (1996) Association of physical and behavioral characteristics with menstrual cycle patterns in women age 29-31 years. Epidemiology 7:624-628

83. Hoffman-Goetz L, Apter D, Demark-Wahnefried W et al (1998) Possible mechanisms mediating an association between physical activity and breast cancer. Cancer 83:621-628

84. Loucks AB (2003) Energy availability, not body fatness, regulates reproductive function in women. Exerc Sport Sci Rev 31:144-148

85. Forney JP, Milewich L, Chen GT et al (1981) Aromatization of androstenedione to estrone by human adipose tissue in vitro. Correlation with adipose tissue mass, age, and endometrial neoplasia. J Clin Endocrinol Metab 53:192-199

86. Dorgan JF, Longcope C, Stephenson HE,Jr et al (1996) Relation of prediagnostic serum estrogen and androgen levels to breast cancer risk. Cancer Epidemiol Biomarkers Prev 5:533-539

87. Shephard RJ, Shek PN (1998) Associations between physical activity and susceptibility to cancer: possible mechanisms. Sports Med 26:293-315

88. Matthews CE, Fowke JH, Dai Q et al (2004) Physical activity, body size, and estrogen metabolism in women. Cancer Causes Control 15:473-481. DOI:10.1023/B:CACO.0000036445.04238.87

89. Haffner SM, Katz MS, Dunn JF (1991) Increased upper body and overall adiposity is associated with decreased sex hormone binding globulin in postmenopausal women. Int J Obes 15:471-478

90. Kramer MM, Wells CL (1996) Does physical activity reduce risk of estrogen-dependent cancer in women? Med Sci Sports Exerc 28:322-334

91. Friedenreich CM, Woolcott CG, McTiernan A et al (2010) Alberta physical activity and breast cancer prevention trial: sex hormone changes in a year-long exercise intervention among postmenopausal women. J Clin Oncol 28:1458-1466. DOI:10.1200/JCO.2009.24.9557

92. Bertone-Johnson ER, Tworoger SS, Hankinson SE (2009) Recreational physical activity and steroid hormone levels in postmenopausal women. Am J Epidemiol 170:1095-1104. DOI:10.1093/aje/kwp254

93. Chan MF, Dowsett M, Folkerd E et al (2007) Usual physical activity and endogenous sex hormones in postmenopausal women: the European prospective investigation into cancer-norfolk population study. Cancer Epidemiol Biomarkers Prev 16:900-905. DOI:10.1158/1055-9965.EPI-06-0745

94. Gunter MJ, Hoover DR, Yu H et al (2009) Insulin, insulin-like growth factor-I, and risk of breast cancer in postmenopausal women. J Natl Cancer Inst 101:48-60. DOI:10.1093/jnci/djn415

95. Giovannucci E (2001) Insulin, insulin-like growth factors and colon cancer: a review of the evidence. J Nutr 131:3109S-20S

96. Kaaks R (1996) Nutrition, hormones, and breast cancer: is insulin the missing link? Cancer Causes Control 7:605-625

97. Harris NS, Winter WE (2004) The chemical pathology of insulin resistance and the metabolic syndrome. MLO Med Lab Obs 36:20, 22-5

98. Kaaks R, Lukanova A (2001) Energy balance and cancer: the role of insulin and insulin-like growth factor-I. Proc Nutr Soc 60:91-106

99. Hursting SD, Lashinger LM, Colbert LH et al (2007) Energy balance and carcinogenesis: underlying pathways and targets for intervention. Curr Cancer Drug Targets 7:484-491

100. Boule NG, Haddad E, Kenny GP et al (2001) Effects of exercise on glycemic control and body mass in type 2 diabetes mellitus: a meta-analysis of controlled clinical trials. JAMA 286:1218-1227

101. Ross R, Janssen I, Dawson J et al (2004) Exercise-induced reduction in obesity and insulin resistance in women: a randomized controlled trial. Obes Res 12:789-798. DOI:10.1038/oby.2004.95

102. Yu H, Rohan T (2000) Role of the insulin-like growth factor family in cancer development and progression. J Natl Cancer Inst 92:1472-1489

103. Friedenreich CM, Neilson HK, Woolcott CG et al (2011) Changes in insulin resistance indicators, IGFs, and adipokines in a year-long trial of aerobic exercise in postmenopausal women. Endocr Relat Cancer 18:357-369. DOI:10.1530/ERC-10-0303

104. Trayhurn P, Bing C, Wood IS (2006) Adipose tissue and adipokines--energy regulation from the human perspective. J Nutr 136:1935S-1939S

105. Arita Y, Kihara S, Ouchi N et al (1999) Paradoxical decrease of an adipose-specific protein, adiponectin, in obesity. Biochem Biophys Res Commun 257:79-83

106. Das UN (2004) Metabolic syndrome X: an inflammatory condition? Curr Hypertens Rep 6:66-73

107. Mishima Y, Kuyama A, Tada A et al (2001) Relationship between serum tumor necrosis factor-alpha and insulin resistance in obese men with Type 2 diabetes mellitus. Diabetes Res Clin Pract 52:119-123

108. Il'yasova D, Colbert LH, Harris TB et al (2005) Circulating levels of inflammatory markers and cancer risk in the health aging and body composition cohort. Cancer Epidemiol Biomarkers Prev 14:2413-2418. DOI:10.1158/1055-9965.EPI-05-0316

109. Aggarwal BB, Shishodia S, Sandur SK et al (2006) Inflammation and cancer: how hot is the link? Biochem Pharmacol 72:1605-1621. DOI:10.1016/j.bcp.2006.06.029

110. Coussens LM, Werb Z (2002) Inflammation and cancer. Nature 420:860-867. DOI:10.1038/nature01322

111. Mattusch F, Dufaux B, Heine O et al (2000) Reduction of the plasma concentration of C-reactive protein following nine months of endurance training. Int J Sports Med 21:21-24. DOI:10.1055/s-2000-8852

112. Starkie R, Ostrowski SR, Jauffred S et al (2003) Exercise and IL-6 infusion inhibit endotoxin-induced TNF-alpha production in humans. FASEB J 17:884-886. DOI:10.1096/fj.02-0670fje

113. Ostrowski K, Rohde T, Asp S et al (1999) Pro- and anti-inflammatory cytokine balance in strenuous exercise in humans. J Physiol 515 (Pt 1):287-291

114. Jakobisiak M, Lasek W, Golab J (2003) Natural mechanisms protecting against cancer. Immunol Lett 90:103-122

115. Shephard RJ, Shek PN (1995) Cancer, immune function, and physical activity. Can J Appl Physiol 20:1-25

116. Nieman DC (1997) Exercise immunology: practical applications. Int J Sports Med 18 Suppl 1:S91-100. DOI:10.1055/s-2007-972705

117. Karihtala P, Winqvist R, Syvaoja JE et al (2006) Increasing oxidative damage and loss of mismatch repair enzymes during breast carcinogenesis. Eur J Cancer 42:2653-2659. DOI:10.1016/j.ejca.2006.05.037

118. Gago-Dominguez M, Jiang X, Esteban Castelao J (2007) Lipid peroxidation and the protective effect of physical exercise on breast cancer. Med Hypotheses 68:1138-1143. DOI:10.1016/j.mehy.2006.09.026

119. Clarkson PM, Thompson HS (2000) Antioxidants: what role do they play in physical activity and health? Am J Clin Nutr 72:637S-46S

120. Millikan RC, Player J, de Cotret AR et al (2004) Manganese superoxide dismutase Ala-9Val polymorphism and risk of breast cancer in a population-based case-control study of African Americans and whites. Breast Cancer Res 6:R264-74. DOI:10.1186/bcr786

121. Klaunig JE, Kamendulis LM (2004) The role of oxidative stress in carcinogenesis. Annu Rev Pharmacol Toxicol 44:239-267. DOI:10.1146/annurev.pharmtox.44.101802.121851

122. Rossner P,Jr, Gammon MD, Terry MB et al (2006) Relationship between urinary 15-F2t-isoprostane and 8-oxodeoxyguanosine levels and breast cancer risk. Cancer Epidemiol Biomarkers Prev 15:639-644. DOI:10.1158/1055-9965.EPI-05-0554

123. Martin KR, Barrett JC (2002) Reactive oxygen species as double-edged swords in cellular processes: low-dose cell signaling versus high-dose toxicity. Hum Exp Toxicol 21:71-75

124. Marnett LJ (2000) Oxyradicals and DNA damage. Carcinogenesis 21:361-370

125. Cooke MS, Evans MD, Dizdaroglu M et al (2003) Oxidative DNA damage: mechanisms, mutation, and disease. FASEB J 17:1195-1214. DOI:10.1096/fj.02-0752rev

126. Tas F, Hansel H, Belce A et al (2005) Oxidative stress in breast cancer. Med Oncol 22:11-15. DOI:10.1385/MO:22:1:011

127. Ambrosone CB (2000) Oxidants and antioxidants in breast cancer. Antioxid Redox Signal 2:903-917

128. Datta K, Sinha S, Chattopadhyay P (2000) Reactive oxygen species in health and disease. Natl Med J India 13:304-310

129. Mates JM, Perez-Gomez C, Nunez de Castro I (1999) Antioxidant enzymes and human diseases. Clin Biochem 32:595-603

130. Abou Ghalia AH, Fouad IM (2000) Glutathione and its metabolizing enzymes in patients with different benign and malignant diseases. Clin Biochem 33:657-662

131. Ambrosone CB, Ahn J, Singh KK et al (2005) Polymorphisms in genes related to oxidative stress (MPO, MnSOD, CAT) and survival after treatment for breast cancer. Cancer Res 65:1105-1111

132. Ahn J, Ambrosone CB, Kanetsky PA et al (2006) Polymorphisms in genes related to oxidative stress (CAT, MnSOD, MPO, and eNOS) and acute toxicities from radiation therapy following lumpectomy for breast cancer. Clin Cancer Res 12:7063-7070. DOI:10.1158/1078-0432.CCR-06-0039

133. Klebanoff SJ (1980) Oxygen metabolism and the toxic properties of phagocytes. Ann Intern Med 93:480-489

134. Martin JH, Alalami O, van den Berg HW (1999) Reduced expression of endothelial and inducible nitric oxide synthase in a human breast cancer cell line which has acquired estrogen independence. Cancer Lett 144:65-74

135. Feig DI, Reid TM, Loeb LA (1994) Reactive oxygen species in tumorigenesis. Cancer Res 54:1890s-1894s

136. Halliwell B (1987) Oxidants and human disease: some new concepts. FASEB J 1:358-364

137. Cross CE, Halliwell B, Borish ET et al (1987) Oxygen radicals and human disease. Ann Intern Med 107:526-545

138. Forsberg L, de Faire U, Morgenstern R (2001) Oxidative stress, human genetic variation, and disease. Arch Biochem Biophys 389:84-93. DOI:10.1006/abbi.2001.2295

139. Kang DH (2002) Oxidative stress, DNA damage, and breast cancer. AACN Clin Issues 13:540-549

140. Behrend L, Henderson G, Zwacka RM (2003) Reactive oxygen species in oncogenic transformation. Biochem Soc Trans 31:1441-1444. DOI:10.1042/

141. Caporaso N (2003) The molecular epidemiology of oxidative damage to DNA and cancer. J Natl Cancer Inst 95:1263-1265

142. Li D, Zhang W, Sahin AA et al (1999) DNA adducts in normal tissue adjacent to breast cancer: a review. Cancer Detect Prev 23:454-462

143. Hristozov D, Gadjeva V, Vlaykova T et al (2001) Evaluation of oxidative stress in patients with cancer. Arch Physiol Biochem 109:331-336. DOI:10.1076/apab.109.4.331.4248

144. Kumaraguruparan R, Kabalimoorthy J, Nagini S (2005) Correlation of tissue lipid peroxidation and antioxidants with clinical stage and menopausal status in patients with adenocarcinoma of the breast. Clin Biochem 38:154-158. DOI:10.1016/j.clinbiochem.2004.10.012

145. Rajneesh CP, Manimaran A, Sasikala KR et al (2008) Lipid peroxidation and antioxidant status in patients with breast cancer. Singapore Med J 49:640-643

146. Kumaraguruparan R, Subapriya R, Viswanathan P et al (2002) Tissue lipid peroxidation and antioxidant status in patients with adenocarcinoma of the breast. Clin Chim Acta 325:165-170

147. Punnonen K, Ahotupa M, Asaishi K et al (1994) Antioxidant enzyme activities and oxidative stress in human breast cancer. J Cancer Res Clin Oncol 120:374-377

148. Kasapovic J, Pejic S, Todorovic A et al (2008) Antioxidant status and lipid peroxidation in the blood of breast cancer patients of different ages. Cell Biochem Funct 26:723-730. DOI:10.1002/cbf.1499

149. Ishii K, Zhen LX, Wang DH et al (1996) Prevention of mammary tumorigenesis in acatalasemic mice by vitamin E supplementation. Jpn J Cancer Res 87:680-684

150. Forsberg L, Lyrenas L, de Faire U et al (2001) 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 30:500-505

151. Zhou XF, Cui J, DeStefano AL et al (2005) Polymorphisms in the promoter region of catalase gene and essential hypertension. Dis Markers 21:3-7

152. Casp CB, She JX, McCormack WT (2002) Genetic association of the catalase gene (CAT) with vitiligo susceptibility. Pigment Cell Res 15:62-66

153. Ahn J, Nowell S, McCann SE et al (2006) Associations between catalase phenotype and genotype: modification by epidemiologic factors. Cancer Epidemiol Biomarkers Prev 15:1217-1222. DOI:10.1158/1055-9965.EPI-06-0104

154. Nadif R, Mintz M, Jedlicka A et al (2005) Association of CAT polymorphisms with catalase activity and exposure to environmental oxidative stimuli. Free Radic Res 39:1345-1350. DOI:10.1080/10715760500306711

155. Bastaki M, Huen K, Manzanillo P et al (2006) Genotype-activity relationship for Mn-superoxide dismutase, glutathione peroxidase 1 and catalase in humans. Pharmacogenet Genomics 16:279-286. DOI:10.1097/01.fpc.0000199498.08725.9c

156. Ahn J, Gammon MD, Santella RM et al (2005) Associations between breast cancer risk and the catalase genotype, fruit and vegetable consumption, and supplement use. Am J Epidemiol 162:943-952. DOI:10.1093/aje/kwi306

157. Quick SK, Shields PG, Nie J et al (2008) Effect modification by catalase genotype suggests a role for oxidative stress in the association of hormone replacement therapy with postmenopausal breast cancer risk. Cancer Epidemiol Biomarkers Prev 17:1082-1087. DOI:10.1158/1055-9965.EPI-07-2755

158. Li Y, Ambrosone CB, McCullough MJ et al (2009) Oxidative stress-related genotypes, fruit and vegetable consumption and breast cancer risk. Carcinogenesis 30:777-784. DOI:10.1093/carcin/bgp053

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

160. Brooks PJ (1997) DNA damage, DNA repair, and alcohol toxicity--a review. Alcohol Clin Exp Res 21:1073-1082

161. Yager JD (2000) Endogenous estrogens as carcinogens through metabolic activation. J Natl Cancer Inst Monogr (27):67-73

162. Ahn J, Gammon MD, Santella RM et al (2006) Effects of glutathione S-transferase A1 (GSTA1) genotype and potential modifiers on breast cancer risk. Carcinogenesis 27:1876-1882. DOI:10.1093/carcin/bgl038

163. Kanter MM (1994) Free radicals, exercise, and antioxidant supplementation. Int J Sport Nutr 4:205-220

164. Ayres S, Baer J, Subbiah MT (1998) Exercised-induced increase in lipid peroxidation parameters in amenorrheic female athletes. Fertil Steril 69:73-77

165. Singh VN (1992) A current perspective on nutrition and exercise. J Nutr 122:760-765

166. Guerra A, Rego C, Laires MJ et al (2001) Lipid profile and redox status in high performance rhythmic female teenagers gymnasts. J Sports Med Phys Fitness 41:505-512

167. Guerra A, Rego C, Castro E et al (2000) LDL peroxidation in adolescent female gymnasts. Rev Port Cardiol 19:1129-1140

168. McBride JM, Kraemer WJ, Triplett-McBride T et al (1998) Effect of resistance exercise on free radical production. Med Sci Sports Exerc 30:67-72

169. Packer L (1997) Oxidants, antioxidant nutrients and the athlete. J Sports Sci 15:353-363

170. Vani M, Reddy GP, Reddy GR et al (1990) Glutathione-S-transferase, superoxide dismutase, xanthine oxidase, catalase, glutathione peroxidase and lipid peroxidation in the liver of exercised rats. Biochem Int 21:17-26

171. Evelo CT, Palmen NG, Artur Y et al (1992) Changes in blood glutathione concentrations, and in erythrocyte glutathione reductase and glutathione S-transferase activity after running training and after participation in contests. Eur J Appl Physiol Occup Physiol 64:354-358

172. Halliwell B (2000) The antioxidant paradox. Lancet 355:1179-1180. DOI:10.1016/S0140-6736(00)02075-4

173. Miyazaki H, Oh-ishi S, Ookawara T et al (2001) Strenuous endurance training in humans reduces oxidative stress following exhausting exercise. Eur J Appl Physiol 84:1-6

174. Ji LL, Stratman FW, Lardy HA (1988) Antioxidant enzyme systems in rat liver and skeletal muscle. Influences of selenium deficiency, chronic training, and acute exercise. Arch Biochem Biophys 263:150-160

175. Singal PK, Dhalla AK, Hill M et al (1993) Endogenous antioxidant changes in the myocardium in response to acute and chronic stress conditions. Mol Cell Biochem 129:179-186

176. Kanter MM, Hamlin RL, Unverferth DV et al (1985) Effect of exercise training on antioxidant enzymes and cardiotoxicity of doxorubicin. J Appl Physiol 59:1298-1303

177. Robertson JD, Maughan RJ, Duthie GG et al (1991) Increased blood antioxidant systems of runners in response to training load. Clin Sci (Lond) 80:611-618

178. Phaneuf S, Leeuwenburgh C (2001) Apoptosis and exercise. Med Sci Sports Exerc 33:393-396

179. Cadenas E (1989) Biochemistry of oxygen toxicity. Annu Rev Biochem 58:79-110. DOI:10.1146/annurev.bi.58.070189.000455

180. Boiteux S, Radicella JP (1999) Base excision repair of 8-hydroxyguanine protects DNA from endogenous oxidative stress. Biochimie 81:59-67

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

182. Braithwaite E, Wu X, Wang Z (1999) Repair of DNA lesions: mechanisms and relative repair efficiencies. Mutat Res 424:207-219

183. Berwick M, Vineis P (2000) Markers of DNA repair and susceptibility to cancer in humans: an epidemiologic review. J Natl Cancer Inst 92:874-897

184. Gammon MD, Santella RM (2008) PAH, genetic susceptibility and breast cancer risk: an update from the Long Island Breast Cancer Study Project. Eur J Cancer 44:636-640. DOI:10.1016/j.ejca.2008.01.026

185. Christmann M, Tomicic MT, Roos WP et al (2003) Mechanisms of human DNA repair: an update. Toxicology 193:3-34

186. Ferguson LR (2010) Chronic inflammation and mutagenesis. Mutat Res 690:3-11. DOI:10.1016/j.mrfmmm.2010.03.007

187. Chang CL, Marra G, Chauhan DP et al (2002) Oxidative stress inactivates the human DNA mismatch repair system. Am J Physiol Cell Physiol 283:C148-54. DOI:10.1152/ajpcell.00422.2001

188. Khanna KK, Jackson SP (2001) DNA double-strand breaks: signaling, repair and the cancer connection. Nat Genet 27:247-254. DOI:10.1038/85798

189. Sancar A, Lindsey-Boltz LA, Unsal-Kacmaz K et al (2004) Molecular mechanisms of mammalian DNA repair and the DNA damage checkpoints. Annu Rev Biochem 73:39-85. DOI:10.1146/annurev.biochem.73.011303.073723

190. Cahill D, Connor B, Carney JP (2006) Mechanisms of eukaryotic DNA double strand break repair. Front Biosci 11:1958-1976

191. Fattah F, Lee EH, Weisensel N et al (2010) Ku regulates the non-homologous end joining pathway choice of DNA double-strand break repair in human somatic cells. PLoS Genet 6:e1000855. DOI:10.1371/journal.pgen.1000855

192. Lu AL, Li X, Gu Y et al (2001) Repair of oxidative DNA damage: mechanisms and functions. Cell Biochem Biophys 35:141-170. DOI:10.1385/CBB:35:2:141

193. Liu Y, Prasad R, Beard WA et al (2007) Coordination of steps in single-nucleotide base excision repair mediated by apurinic/apyrimidinic endonuclease 1 and DNA polymerase beta. J Biol Chem 282:13532-13541. DOI:10.1074/jbc.M611295200

194. Robertson AB, Klungland A, Rognes T et al (2009) DNA repair in mammalian cells: Base excision repair: the long and short of it. Cell Mol Life Sci 66:981-993. DOI:10.1007/s00018-009-8736-z

195. Pascucci B, Stucki M, Jonsson ZO et al (1999) Long patch base excision repair with purified human proteins. DNA ligase I as patch size mediator for DNA polymerases delta and epsilon. J Biol Chem 274:33696-33702

196. Reardon JT, Bessho T, Kung HC et al (1997) In vitro repair of oxidative DNA damage by human nucleotide excision repair system: possible explanation for neurodegeneration in xeroderma pigmentosum patients. Proc Natl Acad Sci U S A 94:9463-9468

197. Braithwaite E, Wu X, Wang Z (1998) Repair of DNA lesions induced by polycyclic aromatic hydrocarbons in human cell-free extracts: involvement of two excision repair mechanisms in vitro. Carcinogenesis 19:1239-1246

198. Eveno E, Bourre F, Quilliet X et al (1995) Different removal of ultraviolet photoproducts in genetically related xeroderma pigmentosum and trichothiodystrophy diseases. Cancer Res 55:4325-4332

199. Tuteja N, Tuteja R (2001) Unraveling DNA repair in human: molecular mechanisms and consequences of repair defect. Crit Rev Biochem Mol Biol 36:261-290. DOI:10.1080/20014091074192

200. Friedberg EC (2001) How nucleotide excision repair protects against cancer. Nat Rev Cancer 1:22-33. DOI:10.1038/35094000

201. Aquilina G, Bignami M (2001) Mismatch repair in correction of replication errors and processing of DNA damage. J Cell Physiol 187:145-154. DOI:10.1002/jcp.1067

202. Chintamani, Jha BP, Bhandari V et al (2007) The expression of mismatched repair genes and their correlation with clinicopathological parameters and response to neo-adjuvant chemotherapy in breast cancer. Int Semin Surg Oncol 4:5. DOI:10.1186/1477-7800-4-5

203. Fritzell JA, Narayanan L, Baker SM et al (1997) Role of DNA mismatch repair in the cytotoxicity of ionizing radiation. Cancer Res 57:5143-5147

204. Peters AC, Young LC, Maeda T et al (2003) Mammalian DNA mismatch repair protects cells from UVB-induced DNA damage by facilitating apoptosis and p53 activation. DNA Repair (Amst) 2:427-435

205. Aquilina G, Crescenzi M, Bignami M (1999) Mismatch repair, G(2)/M cell cycle arrest and lethality after DNA damage. Carcinogenesis 20:2317-2326

206. Hickman MJ, Samson LD (1999) Role of DNA mismatch repair and p53 in signaling induction of apoptosis by alkylating agents. Proc Natl Acad Sci U S A 96:10764-10769

207. Wu J, Gu L, Wang H et al (1999) Mismatch repair processing of carcinogen-DNA adducts triggers apoptosis. Mol Cell Biol 19:8292-8301

208. Sia EA, Kokoska RJ, Dominska M et al (1997) Microsatellite instability in yeast: dependence on repeat unit size and DNA mismatch repair genes. Mol Cell Biol 17:2851-2858

209. Lengauer C, Kinzler KW, Vogelstein B (1998) Genetic instabilities in human cancers. Nature 396:643-649. DOI:10.1038/25292

210. Kolodner RD, Marsischky GT (1999) Eukaryotic DNA mismatch repair. Curr Opin Genet Dev 9:89-96

211. Wood RD, Mitchell M, Sgouros J et al (2001) Human DNA repair genes. Science 291:1284-1289. DOI:10.1126/science.1056154

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

213. Han J, Hankinson SE, Ranu H et al (2004) Polymorphisms in DNA double-strand break repair genes and breast cancer risk in the Nurses' Health Study. Carcinogenesis 25:189-195. DOI:10.1093/carcin/bgh002

214. Han S, Zhang HT, Wang Z et al (2006) DNA repair gene XRCC3 polymorphisms and cancer risk: a meta-analysis of 48 case-control studies. Eur J Hum Genet 14:1136-1144. DOI:10.1038/sj.ejhg.5201681

215. Garcia-Closas M, Egan KM, Newcomb PA et al (2006) Polymorphisms in DNA double-strand break repair genes and risk of breast cancer: two population-based studies in USA and Poland, and meta-analyses. Hum Genet 119:376-388. DOI:10.1007/s00439-006-0135-z

216. Smith TR, Levine EA, Perrier ND et al (2003) DNA-repair genetic polymorphisms and breast cancer risk. Cancer Epidemiol Biomarkers Prev 12:1200-1204

217. Zhang Y, Newcomb PA, Egan KM et al (2006) Genetic polymorphisms in base-excision repair pathway genes and risk of breast cancer. Cancer Epidemiol Biomarkers Prev 15:353-358. DOI:10.1158/1055-9965.EPI-05-0653

218. Thyagarajan B, Anderson KE, Folsom AR et al (2006) No association between XRCC1 and XRCC3 gene polymorphisms and breast cancer risk: Iowa Women's Health Study. Cancer Detect Prev 30:313-321. DOI:10.1016/j.cdp.2006.07.002

219. Sangrajrang S, Schmezer P, Burkholder I et al (2008) Polymorphisms in three base excision repair genes and breast cancer risk in Thai women. Breast Cancer Res Treat 111:279-288. DOI:10.1007/s10549-007-9773-7

220. Chacko P, Rajan B, Joseph T et al (2005) Polymorphisms in DNA repair gene XRCC1 and increased genetic susceptibility to breast cancer. Breast Cancer Res Treat 89:15-21. DOI:10.1007/s10549-004-1004-x

221. Mitra AK, Singh N, Singh A et al (2008) Association of polymorphisms in base excision repair genes with the risk of breast cancer: a case-control study in North Indian women. Oncol Res 17:127-135

222. Crew KD, Gammon MD, Terry MB et al (2007) Polymorphisms in nucleotide excision repair genes, polycyclic aromatic hydrocarbon-DNA adducts, and breast cancer risk. Cancer Epidemiol Biomarkers Prev 16:2033-2041. DOI:10.1158/1055-9965.EPI-07-0096

223. Kumar R, Hoglund L, Zhao C et al (2003) Single nucleotide polymorphisms in the XPG gene: determination of role in DNA repair and breast cancer risk. Int J Cancer 103:671-675. DOI:10.1002/ijc.10870

224. Terry MB, Gammon MD, Zhang FF et al (2004) Polymorphism in the DNA repair gene XPD, polycyclic aromatic hydrocarbon-DNA adducts, cigarette smoking, and breast cancer risk. Cancer Epidemiol Biomarkers Prev 13:2053-2058

225. Lee KM, Choi JY, Kang C et al (2005) Genetic polymorphisms of selected DNA repair genes, estrogen and progesterone receptor status, and breast cancer risk. Clin Cancer Res 11:4620-4626. DOI:10.1158/1078-0432.CCR-04-2534

226. Milne RL, Ribas G, Gonzalez-Neira A et al (2006) ERCC4 associated with breast cancer risk: a two-stage case-control study using high-throughput genotyping. Cancer Res 66:9420-9427. DOI:10.1158/0008-5472.CAN-06-1418

227. Jorgensen TJ, Visvanathan K, Ruczinski I et al (2007) Breast cancer risk is not associated with polymorphic forms of xeroderma pigmentosum genes in a cohort of women from Washington County, Maryland. Breast Cancer Res Treat 101:65-71. DOI:10.1007/s10549-006-9263-3

228. Smith TR, Levine EA, Freimanis RI et al (2008) Polygenic model of DNA repair genetic polymorphisms in human breast cancer risk. Carcinogenesis 29:2132-2138. DOI:10.1093/carcin/bgn193

229. Mechanic LE, Millikan RC, Player J et al (2006) Polymorphisms in nucleotide excision repair genes, smoking and breast cancer in African Americans and whites: a population-based case-control study. Carcinogenesis 27:1377-1385. DOI:10.1093/carcin/bgi330

230. Manuguerra M, Saletta F, Karagas MR et al (2006) XRCC3 and XPD/ERCC2 single nucleotide polymorphisms and the risk of cancer: a HuGE review. Am J Epidemiol 164:297-302. DOI:10.1093/aje/kwj189

231. Jiang Z, Li C, Xu Y et al (2010) Associations between XPD polymorphisms and risk of breast cancer: a meta-analysis. Breast Cancer Res Treat 123:203-212. DOI:10.1007/s10549-010-0751-0

232. Pabalan N, Francisco-Pabalan O, Sung L et al (2010) Meta-analysis of two ERCC2 (XPD) polymorphisms, Asp312Asn and Lys751Gln, in breast cancer. Breast Cancer Res Treat 124:531-541. DOI:10.1007/s10549-010-0863-6

233. Poplawski T, Zadrozny M, Kolacinska A et al (2005) Polymorphisms of the DNA mismatch repair gene HMSH2 in breast cancer occurence and progression. Breast Cancer Res Treat 94:199-204. DOI:10.1007/s10549-005-4793-7

234. Conde J, Silva SN, Azevedo AP et al (2009) Association of common variants in mismatch repair genes and breast cancer susceptibility: a multigene study. BMC Cancer 9:344. DOI:10.1186/1471-2407-9-344

235. Monsees GM, Kraft P, Chanock SJ et al (2011) Comprehensive screen of genetic variation in DNA repair pathway genes and postmenopausal breast cancer risk. Breast Cancer Res Treat 125:207-214. DOI:10.1007/s10549-010-0947-3

236. Yee CJ, Roodi N, Verrier CS et al (1994) Microsatellite instability and loss of heterozygosity in breast cancer. Cancer Res 54:1641-1644

237. Shaw JA, Walsh T, Chappell SA et al (1996) Microsatellite instability in early sporadic breast cancer. Br J Cancer 73:1393-1397

238. Walsh T, Chappell SA, Shaw JA et al (1998) Microsatellite instability in ductal carcinoma in situ of the breast. J Pathol 185:18-24. DOI:2-G

239. Siah SP, Quinn DM, Bennett GD et al (2000) Microsatellite instability markers in breast cancer: a review and study showing MSI was not detected at 'BAT 25' and 'BAT 26' microsatellite markers in early-onset breast cancer. Breast Cancer Res Treat 60:135-142

240. Shen KL, Yang LS, Hsieh HF et al (2000) Microsatellite alterations on human chromosome 11 in in situ and invasive breast cancer: a microdissection microsatellite analysis and correlation with p53, ER (estrogen receptor), and PR (progesterone receptor) protein immunoreactivity. J Surg Oncol 74:100-107. DOI:2-O

241. Strand M, Prolla TA, Liskay RM et al (1993) Destabilization of tracts of simple repetitive DNA in yeast by mutations affecting DNA mismatch repair. Nature 365:274-276. DOI:10.1038/365274a0

242. Thibodeau SN, French AJ, Cunningham JM et al (1998) Microsatellite instability in colorectal cancer: different mutator phenotypes and the principal involvement of hMLH1. Cancer Res 58:1713-1718

243. Thibodeau SN, Bren G, Schaid D (1993) Microsatellite instability in cancer of the proximal colon. Science 260:816-819

244. Chiaravalli AM, Furlan D, Facco C et al (2001) Immunohistochemical pattern of hMSH2/hMLH1 in familial and sporadic colorectal, gastric, endometrial and ovarian carcinomas with instability in microsatellite sequences. Virchows Arch 438:39-48

245. Lynch HT, de la Chapelle A (1999) Genetic susceptibility to non-polyposis colorectal cancer. J Med Genet 36:801-818

246. Fleisher AS, Esteller M, Wang S et al (1999) Hypermethylation of the hMLH1 gene promoter in human gastric cancers with microsatellite instability. Cancer Res 59:1090-1095

247. Gurin CC, Federici MG, Kang L et al (1999) Causes and consequences of microsatellite instability in endometrial carcinoma. Cancer Res 59:462-466

248. Salvesen HB, MacDonald N, Ryan A et al (2000) Methylation of hMLH1 in a population-based series of endometrial carcinomas. Clin Cancer Res 6:3607-3613

249. Baylin SB, Herman JG (2000) DNA hypermethylation in tumorigenesis: epigenetics joins genetics. Trends Genet 16:168-174

250. Herman JG, Baylin SB (2000) Promoter-region hypermethylation and gene silencing in human cancer. Curr Top Microbiol Immunol 249:35-54

251. Schofield MJ, Hsieh P (2003) DNA mismatch repair: molecular mechanisms and biological function. Annu Rev Microbiol 57:579-608. DOI:10.1146/annurev.micro.57.030502.090847

252. Kinzler KW, Vogelstein B (1996) Lessons from hereditary colorectal cancer. Cell 87:159-170

253. Spagnoletti I, Pizzi C, Galietta A et al (2004) Loss of hMSH2 expression in primary breast cancer with p53 alterations. Oncol Rep 11:845-851

254. Cai KQ, Albarracin C, Rosen D et al (2004) Microsatellite instability and alteration of the expression of hMLH1 and hMSH2 in ovarian clear cell carcinoma. Hum Pathol 35:552-559

255. Murata H, Khattar NH, Kang Y et al (2002) Genetic and epigenetic modification of mismatch repair genes hMSH2 and hMLH1 in sporadic breast cancer with microsatellite instability. Oncogene 21:5696-5703. DOI:10.1038/sj.onc.1205683

256. Seitz S, Wassmuth P, Plaschke J et al (2003) Identification of microsatellite instability and mismatch repair gene mutations in breast cancer cell lines. Genes Chromosomes Cancer 37:29-35. DOI:10.1002/gcc.10196

257. Koster F, Schroer A, Fischer D et al (2007) Immunohistochemistry of proteins for DNA mismatch repair in correlation to prognostic factors of mammarian cancer. Oncol Rep 17:1223-1227

258. Khilko N, Bourne P, Qi Y et al (2007) Mismatch repair genes hMLH1 and hMSH2 may not play an essential role in breast carcinogenesis. Int J Surg Pathol 15:233-241. DOI:10.1177/1066896907302116

259. Shen J, Gammon MD, Terry MB et al (2005) Polymorphisms in XRCC1 modify the association between polycyclic aromatic hydrocarbon-DNA adducts, cigarette smoking, dietary antioxidants, and breast cancer risk. Cancer Epidemiol Biomarkers Prev 14:336-342. DOI:10.1158/1055-9965.EPI-04-0414

260. Shen J, Terry MB, Gammon MD et al (2005) MGMT genotype modulates the associations between cigarette smoking, dietary antioxidants and breast cancer risk. Carcinogenesis 26:2131-2137. DOI:10.1093/carcin/bgi179

261. Metsola K, Kataja V, Sillanpaa P et al (2005) XRCC1 and XPD genetic polymorphisms, smoking and breast cancer risk in a Finnish case-control study. Breast Cancer Res 7:R987-97. DOI:10.1186/bcr1333

262. Pachkowski BF, Winkel S, Kubota Y et al (2006) XRCC1 genotype and breast cancer: functional studies and epidemiologic data show interactions between XRCC1 codon 280 His and smoking. Cancer Res 66:2860-2868. DOI:10.1158/0008-5472.CAN-05-3388

263. Rajaraman P, Bhatti P, Doody MM et al (2008) Nucleotide excision repair polymorphisms may modify ionizing radiation-related breast cancer risk in US radiologic technologists. Int J Cancer 123:2713-2716. DOI:10.1002/ijc.23779

264. Millikan RC, Player JS, Decotret AR et al (2005) Polymorphisms in DNA repair genes, medical exposure to ionizing radiation, and breast cancer risk. Cancer Epidemiol Biomarkers Prev 14:2326-2334. DOI:10.1158/1055-9965.EPI-05-0186

265. Han J, Hankinson SE, Zhang SM et al (2004) Interaction between genetic variations in DNA repair genes and plasma folate on breast cancer risk. Cancer Epidemiol Biomarkers Prev 13:520-524

266. Shrubsole MJ, Gao YT, Cai Q et al (2004) MTHFR polymorphisms, dietary folate intake, and breast cancer risk: results from the Shanghai Breast Cancer Study. Cancer Epidemiol Biomarkers Prev 13:190-196

267. Radak Z, Taylor AW, Ohno H et al (2001) Adaptation to exercise-induced oxidative stress: from muscle to brain. Exerc Immunol Rev 7:90-107

268. Sato Y, Nanri H, Ohta M et al (2003) Increase of human MTH1 and decrease of 8-hydroxydeoxyguanosine in leukocyte DNA by acute and chronic exercise in healthy male subjects. Biochem Biophys Res Commun 305:333-338

269. Radak Z, Kaneko T, Tahara S et al (1999) The effect of exercise training on oxidative damage of lipids, proteins, and DNA in rat skeletal muscle: evidence for beneficial outcomes. Free Radic Biol Med 27:69-74

270. Radak Z, Naito H, Kaneko T et al (2002) Exercise training decreases DNA damage and increases DNA repair and resistance against oxidative stress of proteins in aged rat skeletal muscle. Pflugers Arch 445:273-278. DOI:10.1007/s00424-002-0918-6

271. Wittwer M, Billeter R, Hoppeler H et al (2004) Regulatory gene expression in skeletal muscle of highly endurance-trained humans. Acta Physiol Scand 180:217-227

272. Radak Z, Apor P, Pucsok J et al (2003) Marathon running alters the DNA base excision repair in human skeletal muscle. Life Sci 72:1627-1633

273. Waksberg J. 1978. Sampling methods for random digit dialing. J Amer Statistic Assoc 73:40-46.

274. Bernstein L, Henderson BE, Hanisch R et al (1994) Physical exercise and reduced risk of breast cancer in young women. J Natl Cancer Inst 86:1403-1408

275. Rothman N, Stewart WF, Schulte PA (1995) Incorporating biomarkers into cancer epidemiology: a matrix of biomarker and study design categories. Cancer Epidemiol Biomarkers Prev 4:301-311

276. Donnelly P (2008) Progress and challenges in genome-wide association studies in humans. Nature 456:728-731. DOI:10.1038/nature07631

277. Moore JH, Asselbergs FW, Williams SM (2010) Bioinformatics challenges for genome-wide association studies. Bioinformatics 26:445-455. DOI:10.1093/bioinformatics/btp713

278. Zhu M, Zhao S (2007) Candidate gene identification approach: progress and challenges. Int J Biol Sci 3:420-427

279. Tabor HK, Risch NJ, Myers RM (2002) Candidate-gene approaches for studying complex genetic traits: practical considerations. Nat Rev Genet 3:391-397. DOI:10.1038/nrg796

280. Mottagui-Tabar S, Faghihi MA, Mizuno Y et al (2005) Identification of functional SNPs in the 5-prime flanking sequences of human genes. BMC Genomics 6:18. DOI:10.1186/1471-2164-6-18

281. Johnson GC, Esposito L, Barratt BJ et al (2001) Haplotype tagging for the identification of common disease genes. Nat Genet 29:233-237. DOI:10.1038/ng1001-233

282. de Bakker PI, Yelensky R, Pe'er I et al (2005) Efficiency and power in genetic association studies. Nat Genet 37:1217-1223. DOI:10.1038/ng1669

283. Barrett JC, Fry B, Maller J et al (2005) Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics 21:263-265. DOI:10.1093/bioinformatics/bth457

284. Zongli X, Taylor J SNPinfo: Integrating GWAS and Candidate Gene Information into Functional SNP Selection for Genetic Association Studies 2009

285. International HapMap Consortium (2003) The International HapMap Project. Nature 426:789-796. DOI:10.1038/nature02168

286. Carlson CS, Eberle MA, Rieder MJ et al (2004) Selecting a maximally informative set of single-nucleotide polymorphisms for association analyses using linkage disequilibrium. Am J Hum Genet 74:106-120. DOI:10.1086/381000

287. Lin DY, Zeng D, Millikan R (2005) Maximum likelihood estimation of haplotype effects and haplotype-environment interactions in association studies. Genet Epidemiol 29:299-312. DOI:10.1002/gepi.20098

288. Collins FS, Guyer MS, Charkravarti A (1997) Variations on a theme: cataloging human DNA sequence variation. Science 278:1580-1581

289. Fallin D, Cohen A, Essioux L et al (2001) Genetic analysis of case/control data using estimated haplotype frequencies: application to APOE locus variation and Alzheimer's disease. Genome Res 11:143-151

290. Zaykin DV, Westfall PH, Young SS et al (2002) Testing association of statistically inferred haplotypes with discrete and continuous traits in samples of unrelated individuals. Hum Hered 53:79-91

291. Botstein D, Risch N (2003) Discovering genotypes underlying human phenotypes: past successes for mendelian disease, future approaches for complex disease. Nat Genet 33 Suppl:228-237. DOI:10.1038/ng1090

292. Ziegler A, Konig I (2006) A statistical approach to genetic epidemiology. John Wiley & Sons, New York

293. Hosking L, Lumsden S, Lewis K et al (2004) Detection of genotyping errors by Hardy-Weinberg equilibrium testing. Eur J Hum Genet 12:395-399. DOI:10.1038/sj.ejhg.5201164

294. Bustin SA (2000) Absolute quantification of mRNA using real-time reverse transcription polymerase chain reaction assays. J Mol Endocrinol 25:169-193

295. Greenland S, Brumback B (2002) An overview of relations among causal modelling methods. Int J Epidemiol 31:1030-1037

296. Lewis CM (2002) Genetic association studies: design, analysis and interpretation. Brief Bioinform 3:146-153

297. Lettre G, Lange C, Hirschhorn JN (2007) Genetic model testing and statistical power in population-based association studies of quantitative traits. Genet Epidemiol 31:358-362. DOI:10.1002/gepi.20217

298. Eng SM (2002). *A case-control study of lifetime participation in recreational physical activity and changes in body size in relation to breast cancer risk.* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses. (Accession Order No. [3048125])

299. Ahn J, Gammon MD, Santella RM et al (2004) Myeloperoxidase genotype, fruit and vegetable consumption, and breast cancer risk. Cancer Res 64:7634-7639. DOI:10.1158/0008-5472.CAN-04-1843

300. Kleinbaum DG, Klein M (2002) Logistic Regression: A Self-Learning Text. 2nd edition. Springer, New York

301. Greenland S (1989) Modeling and variable selection in epidemiologic analysis. Am J Public Health 79:340-349

302. Yang XR, Chang-Claude J, Goode EL et al (2011) Associations of breast cancer risk factors with tumor subtypes: a pooled analysis from the Breast Cancer Association Consortium studies. J Natl Cancer Inst 103:250-263. DOI:10.1093/jnci/djq526

303. Rothman K, Greenland S (1998) Modern Epidemiology. 2nd edition. Maple Press, Philadelphia

304. Assmann SF, Hosmer DW, Lemeshow S et al (1996) Confidence intervals for measures of interaction. Epidemiology 7:286-290

305. Garcia-Closas M, Lubin JH (1999) Power and sample size calculations in case-control studies of gene-environment interactions: comments on different approaches. Am J Epidemiol 149:689-692

306. Lubin JH, Gail MH (1990) On power and sample size for studying features of the relative odds of disease. Am J Epidemiol 131:552-566

307. Carey LA, Perou CM, Livasy CA et al (2006) Race, breast cancer subtypes, and survival in the Carolina Breast Cancer Study. JAMA 295:2492-2502. DOI:10.1001/jama.295.21.2492

308. Garcia-Closas M, Thompson WD, Robins JM (1998) Differential misclassification and the assessment of gene-environment interactions in case-control studies. Am J Epidemiol 147:426-433

309. Carpenter CL, Ross RK, Paganini-Hill A et al (1999) Lifetime exercise activity and breast cancer risk among post-menopausal women. Br J Cancer 80:1852-1858. DOI:10.1038/sj.bjc.6690610

310. Hung RJ, Brennan P, Malaveille C et al (2004) Using hierarchical modeling in genetic association studies with multiple markers: application to a case-control study of bladder cancer. Cancer Epidemiol Biomarkers Prev 13:1013-1021

311. Hochberg Y, Benjamini Y (1990) More powerful procedures for multiple significance testing. Stat Med 9:811-818

312. Hunter DJ (2005) Gene-environment interactions in human diseases. Nat Rev Genet 6:287-298. DOI:10.1038/nrg1578

313. Ahn J, Gammon MD, Santella RM et al (2005) No association between glutathione peroxidase Pro198Leu polymorphism and breast cancer risk. Cancer Epidemiol Biomarkers Prev 14:2459-2461. DOI:10.1158/1055-9965.EPI-05-0459

314. Gaudet MM, Bensen JT, Schroeder J et al (2006) Catechol-O-methyltransferase haplotypes and breast cancer among women on Long Island, New York. Breast Cancer Res Treat 99:235-240. DOI:10.1007/s10549-006-9205-0

315. Gaudet MM, Gammon MD, Santella RM et al (2005) MnSOD Val-9Ala genotype, pro- and anti-oxidant environmental modifiers, and breast cancer among women on Long Island, New York. Cancer Causes Control 16:1225-1234. DOI:10.1007/s10552-005-0375-6

316. Steck SE, Gaudet MM, Britton JA et al (2007) Interactions among GSTM1, GSTT1 and GSTP1 polymorphisms, cruciferous vegetable intake and breast cancer risk. Carcinogenesis 28:1954-1959. DOI:10.1093/carcin/bgm141

317. Breslow NE, Day NE (1980) Statistical methods in cancer research. Volume I - The analysis of case-control studies. IARC Sci Publ (32):5-338

318. Powers SK, Ji LL, Leeuwenburgh C (1999) Exercise training-induced alterations in skeletal muscle antioxidant capacity: a brief review. Med Sci Sports Exerc 31:987-997

319. Radak Z, Chung HY, Goto S (2008) Systemic adaptation to oxidative challenge induced by regular exercise. Free Radic Biol Med 44:153-159. DOI:10.1016/j.freeradbiomed.2007.01.029

320. Hoffman-Goetz L, Pervaiz N, Guan J (2009) Voluntary exercise training in mice increases the expression of antioxidant enzymes and decreases the expression of TNF-alpha in intestinal lymphocytes. Brain Behav Immun 23:498-506. DOI:10.1016/j.bbi.2009.01.015

321. Siu PM, Pei XM, Teng BT et al (2011) Habitual exercise increases resistance of lymphocytes to oxidant-induced DNA damage by upregulating expression of antioxidant and DNA repairing enzymes. Exp Physiol 96:889-906. DOI:10.1113/expphysiol.2011.058396

322. Hu X, Ji X, Srivastava SK et al (1997) Mechanism of differential catalytic efficiency of two polymorphic forms of human glutathione S-transferase P1-1 in the glutathione conjugation of carcinogenic diol epoxide of chrysene. Arch Biochem Biophys 345:32-38. DOI:10.1006/abbi.1997.0269

323. Sundberg K, Johansson AS, Stenberg G et al (1998) Differences in the catalytic efficiencies of allelic variants of glutathione transferase P1-1 towards carcinogenic diol epoxides of polycyclic aromatic hydrocarbons. Carcinogenesis 19:433-436

324. Mao GE, Morris G, Lu QY et al (2004) Glutathione S-transferase P1 lle105Val polymorphism, cigarette smoking and prostate cancer. Cancer Detect Prev 28:368-374. DOI:10.1016/j.cdp.2004.07.003

325. Schaid DJ, Jacobsen SJ (1999) Biased tests of association: comparisons of allele frequencies when departing from Hardy-Weinberg proportions. Am J Epidemiol 149:706-711

326. Li GM (2008) Mechanisms and functions of DNA mismatch repair. Cell Res 18:85-98. DOI:10.1038/cr.2007.115

327. Moinfar F, Beham A, Friedrich G et al (2008) Macro-environment of breast carcinoma: frequent genetic alterations in the normal appearing skins of patients with breast cancer. Mod Pathol 21:639-646. DOI:10.1038/modpathol.2008.28

328. Radak Z, Pucsuk J, Boros S et al (2000) Changes in urine 8-hydroxydeoxyguanosine levels of super-marathon runners during a four-day race period. Life Sci 66:1763-1767

329. Chen X, Levine L, Kwok PY (1999) Fluorescence polarization in homogeneous nucleic acid analysis. Genome Res 9:492-498

330. Mohrenweiser HW, Wilson DM,3rd, Jones IM (2003) Challenges and complexities in estimating both the functional impact and the disease risk associated with the extensive genetic variation in human DNA repair genes. Mutat Res 526:93-125

331. Shen J, Gammon MD, Terry MB et al (2008) Xeroderma pigmentosum complementation group C genotypes/diplotypes play no independent or interaction role with polycyclic aromatic hydrocarbons-DNA adducts for breast cancer risk. Eur J Cancer 44:710-717. DOI:10.1016/j.ejca.2007.10.027

332. Shen J, Desai M, Agrawal M et al (2006) Polymorphisms in nucleotide excision repair genes and DNA repair capacity phenotype in sisters discordant for breast cancer. Cancer Epidemiol Biomarkers Prev 15:1614-1619. DOI:10.1158/1055-9965.EPI-06-0218

333. Zondervan KT, Cardon LR (2004) The complex interplay among factors that influence allelic association. Nat Rev Genet 5:89-100. DOI:10.1038/nrg1270

334. Tian C, Gregersen PK, Seldin MF (2008) Accounting for ancestry: population substructure and genome-wide association studies. Hum Mol Genet 17:R143-50. DOI:10.1093/hmg/ddn268

335. Lee IM. (2009). Epidemiologic methods in physical activity studies: Oxford University Press. New York.

336. Michels KB, Rosner BA (1996) Data trawling: to fish or not to fish. Lancet 348:1152-1153. DOI:10.1016/S0140-6736(96)05418-9

337. Sabatti C, Service S, Freimer N (2003) False discovery rate in linkage and association genome screens for complex disorders. Genetics 164:829-833

338. Savitz DA, Olshan AF (1995) Multiple comparisons and related issues in the interpretation of epidemiologic data. Am J Epidemiol 142:904-908

339. Wacholder S, Chanock S, Garcia-Closas M et al (2004) Assessing the probability that a positive report is false: an approach for molecular epidemiology studies. J Natl Cancer Inst 96:434-442

340. Thomas DC, Clayton DG (2004) Betting odds and genetic associations. J Natl Cancer Inst 96:421-423

341. Witte JS (1997) Genetic analysis with hierarchical models. Genet Epidemiol 14:1137-1142. DOI:2-H

342. Rothman KJ (1990) No adjustments are needed for multiple comparisons. Epidemiology 1:43-46

343. Ioannidis JP, Ntzani EE, Trikalinos TA et al (2001) Replication validity of genetic association studies. Nat Genet 29:306-309. DOI:10.1038/ng749