Modifying Effects of DNA Repair and Oxidative Stress Variants on Physical Activity and Breast Cancer Risk

Lauren E. McCullough
September 27, 2010
Bios Conference Room

1.1 SPECIFIC AIMS

Breast cancer is the leading cause of global cancer incidence and mortality among women. Among U.S. women it is the primary cause of 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. Physical activity 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 alterations in these pathways. Physical activity may work through a DNA repair pathway by increasing repair activity or similarly through an oxidative stress pathway by up-regulating antioxidant enzymes. The overall effect would be diminished DNA damage and lower likelihood of initiating events.

This study proposes to assess whether the effect of physical activity on breast cancer risk is modified by individual variability in the genetic variants of two different pathways: (1) DNA repair

and (2) oxidative stress. Specifically, this study will assess whether genetic polymorphisms in *MLH1*, *MSH2*, *MSH3,* and *CAT* are associated with breast cancer incidence as well as their potential interaction with physical activity to impact disease development. The study aims are defined by the pathway under investigation:

AIM 1. DNA Repair

AIM 1A: to determine the main effect of select SNPs in three genes (*MLH1, MSH2, and MSH3*) of the MMR pathway on breast cancer risk.

AIM 1B: to evaluate interactions between polymorphisms in DNA repair genes (MMR, BER, and NER pathways) and self-reported lifetime physical activity on breast cancer risk (gene-environment interactions)

AIM 1C: to explore potential interactions between SNPs in genes from three DNA repair pathways (MMR, BER and NER) on breast cancer risk (gene-gene interactions).

AIM 2. Oxidative Stress

AIM 2A: to determine the main effect of select SNPs in *CAT* on breast cancer risk.

AIM 2B: to evaluate interactions between polymorphisms in oxidative stress genes and self-reported lifetime physical activity on breast cancer risk (gene-environment interactions)

AIM 2C: to explore the potential interactions between SNPs in genes from the oxidative stress pathway on breast cancer risk (gene-gene interactions).

The following sections summarize the breast cancer and physical activity literature, details the physical activity-breast cancer mechanisms described to date, and proposes two alternative pathways that may drive the inverse association.

1.2 BREAST CANCER

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. It is well established that a large number of non-modifiable risks, specifically those related to hormones, play an important role in breast cancer development. However, a proportion of breast cancer risk can be attributed to those factors which are amendable. The case for these modifiable or lifestyle risk factors are most apparently revealed by the large differences in breast cancer incidence between countries (Kelsey 1993) and the monotonic increase in risk among immigrants across generations. Excess body weight and low physical activity combined are proposed to account for up to one-third of breast cancer cases (IARC Working Group 2002). While these risks may not be as predictive as those that are non-modifiable, they can ultimately be controlled, aiding in the reduction of breast malignancy development. Identifying pathways through which these factors operate could play an important role in advancing the knowledge of breast cancer etiology. The following section outlines the trends in breast cancer incidence, potential mechanisms, and risk factors for the disease.

1.2.1 TRENDS IN BREAST CANCER INCIDENCE

Breast cancer is the leading cause of global cancer incidence and mortality among women. Among U.S. women, breast cancer is the primary cause of cancer related illness with an estimated 194,280 new cases in 2009 and incidence rate of 123.6 per 100,000 person-years. There were an estimated 40,610 deaths attributable to breast cancer in 2009 yielding a mortality rate of 25.0 per 100,000 person-years (American Cancer Society 2009). Breast cancer differentially affects women by age and race. From 2000-2004 94% of incident cases in the U.S. occurred in women over the age of 40. Most women were of European decent (American

Cancer Society 2007-2008). 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 (American Cancer Society 2007-2008).

1.2.2 MECHANISMS OF BREAST CANCER

Prolonged exposure to circulating estrogens 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 (Singletary 2003). Increased cellular proliferation is likely to result in a greater number of deleterious mutations that, if un-repaired, could result in breast malignancy (Hankinson 2004). 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 (Clemons 2001). Closely linked to estrogen are insulin and insulin-like growth factor (IGF)-1. They have been proposed to work with estrogen to influence breast cancer risk by both increasing cell proliferation and preventing apoptosis (Muti 2004). Insulin is additionally known to reduce synthesis of sex-hormone-binding globulin (SHBG), a transporter of testosterone and estradiol, in the liver (Muti 2004). 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 *BRCA1* and *BRCA2* which account for 15-20% of ancestral breast

cancer clustering (Nathanson 2001), and are commonly associated with early-onset breast cancer (Krainer 1997). A 2003 pooled analysis reports an average cumulative risk of 65% among *BRCA1* carriers by age 70. The reported cumulative risk among *BRCA2* carriers is 45% (Antoniou 2003). 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 (Hirshfield 2010). 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 and increased mutation.

1.2.3 BREAST CANCER RISK FACTOR EPIDEMIOLOGY

Reproductive Factors

Unlike many cancers breast cancer has a number of well established risk factors. As previously described, cumulative exposure to estrogens appear 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 estrogen levels (Kelsey, 1993). Early age at menarche (Hsieh 1990, Kelsey 1993), age at first birth (Lambe 1998), parity (Lambe 1998, Rosner 1994), lactation history (Collaborative Group on Hormonal Factors in Breast Cancer 2002), and late age at menopause (Kelsey 1993) 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 (Hankinson 2004). 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 (Britt 2007). Findings in the Long Island Breast Cancer Study Project (LIBCSP) (Gammon, 2002) indicate that reproductive practices have an important role in breast cancer etiology. Investigators report 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 (EE) and are thus a frequent source of exposure to exogenous hormones in premenopausal women (Casey 2008). 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 (Collaborative Group on Hormonal Factors in Breast Cancer 1997). A risk reduction was observed after stopping OC use (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) (Collaborative Group on Hormonal Factors in Breast Cancer 1997). A more recent 2002 study (Marchbanks 2002) reports 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 reports that while combined estrogen-progestin therapy (EPT) moderately increases the risk of breast cancer (20-40%), the current evidence for unopposed estrogen showed no increase in risk (Chen 2009). 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 (Chen 2006). 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 (Shantakumar 2007). 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 (IARC 2000). Women exposed to the atomic bomb in Nagasaki and Hiroshima, Japan have experienced as much as a 9-fold increase in breast cancer risk (Adami 2002).

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 (Adami 2002). 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 (Gammon 2002). 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) (Gammon 2004).

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 (Gammon 2002). LIBCSP investigators reported a 50% increased risk (95% CI: 1.04, 2.20) among women with highest PAH quintile compared to women in the lowest quartile (Gammon 2002).

Alcohol

The consumption of alcohol has been associated with a modest increased risk of breast cancer (Singletary 2003). 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 (Fan 2000). Other mechanisms for the alcohol-breast cancer relationship have been suggested. Such mechanisms include an increase in reactive oxygen species, hydroxyl radicals, and DNA modification (Singletary 2001). Reports from a 2006 meta-analyses of high quality studies show a 22% (95% CI: 1.09, 1.37) increased risk of breast cancer among women classified as drinkers, compared to abstainers (Key 2006). Contrary to these results Long Island investigators found no association between ever alcohol use and never users (Gammon 2002).

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 (Michels 2007). Studies that have assessed the association vary in both the magnitude and direction of effect, and have typically been shown to only slightly modify risk. There are several rationales for null or weak associations in the diet-breast cancer literature. Lof and Widerpass (2009) suggests 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 (Lof and Widerpass 2009). 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 pre-menopausal women while increasing risk among post-menopausal women (Friedenreich 2001). Pichard and colleagues report a positive relationship between body mass index (BMI, the ratio of weight in kg squared to height in meters) and breast cancer with relative risk (RR) ranging from 1.26 to 2.52 among post-menopausal women (Pichard, 2008). The positive association between obesity and post-menopausal 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 (Key 2001, Muti 2004). Compared to ovarian estrogen production among pre-menopausal women, adipose mediated estrogen production is highly unregulated (Key 2001). 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 (Key 2001).

Physical Activity

There has been an overwhelming amount of epidemiological evidence identifying the beneficial effects of exercise and physical activity in the reduction of breast cancer incidence (Gammon 1998 and Vainio 2002). Study reports show 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 (Tehard 2006 and McTiernan 2008). The most apparent 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) (Friedenreich 2004). However, early epidemiologic studies evaluating the role of physical activity on breast cancer risk show that upon control for body weight, an effect for physical activity persists (Mittendorf 1995 and Thune 1997). These results indicate that in addition to an obesity-mediated pathway, physical activity is likely to influence breast cancer carcinogenesis through independent mechanisms.

1.2.4 CONCLUSIONS

Breast cancer remains an important public health concern in both the United States 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 may of the established risk factors have been linked to reproductive events and estrogen pathways they are primarily non-modifiable. There are a number of modifiable risks known to play a role in breast cancer development. While these risks may not be as predictive as those that are non-modifiable, 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 post-menopausal 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 will focus on pathways relevant to breast cancer and physical activity, specifically DNA repair and oxidative stress. In the following sections I will describe the current state of knowledge of the inverse association observed between physical activity and breast cancer risk. I will supply a summary of 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 will highlight several genes that have been investigated in connection to breast cancer along the DNA repair and oxidative stress pathways and provide a rationale for examining these pathways in the molecular epidemiology of physical activity.

1.3 RECREATIONAL PHYSICAL ACTIVITY

Physical activity has been associated with reduced incidence of a number of chronic diseases including heart disease (Batty 2002 and Elsawy 2010), diabetes (LaMonte 2005, Elsawy 2010, and Quin 2010), stroke (Elsawy 2010), osteoporosis (Schmitt 2009) and cancer (Elsawy 2010). 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 (Gammon 1998 and Vainio 2002). 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.

1.3.1 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 (Caspersen 1989). 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 (Newton 2008). All individuals are exposed to physical activity in several domains across the life

span. While the recommendations for regular physical activity vary by age and health status the CDC's analysis from the Behavioral Risk Factor Surveillance System (BRFSS) indicates that the prevalence of regular physical activity among U.S. women in 2007 was 47.5% (CDC, 2007). With a nationwide prevalence of approximately 50% 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 (Gammon 1998, Friedenreich 2002). 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 facet 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 (Ainsworth 1993). 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 (Pate 1995). 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. Also important is the source of activity. 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 shorter 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 (Newton 2008). While physiological measures of physical activity (e.g. resting heart rate and aerobic capacity) perform particularly well and are commonly regarded as 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 (McTiernan 1998). Questionnaire based assessments are common practice in observational epidemiology and often query participants on the source and all three dimensions of physical activity. While self reported data may only provide crude categorizations of physical activity levels these methods are known have relatively good validity for both high intensity and sedentary activities (Jacobs 1993). Adjustment for BMI may further aid in reducing variation that occur because of differences in race, age, and sex (Gammon 1998). Although there are qualitative differences in physical activity assessment across studies, the physical activity-breast cancer literature overwhelmingly supports an inverse association.

1.3.2 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 (Gammon 1998, Monninkhof 2007, Friednenreich 1995, Thune 2001, Friedenreich 2002, Friedenreich 2004, McTiernan 2008, Frienenreich 2008, and Latikka 1998). Risk reductions reported in these studies range from 20-40% among active women (Thune 2001, Friedenreich 2002, Friedenreich 2004, and McTiernan 2008), though slightly stronger associations have been found in case-control compared with cohort studies (Gammon 1998 and Frienenreich 2008). While early reports show risk reductions for both occupational and recreational activity a comprehensive 2008 review of physical activity parameters and breast cancer risk indicates that the greatest risk reductions are observed for

recreational physical activity (20% risk reduction). Activity related to occupation, transportation, and daily living each resulted in approximately 14% reduced risk (Friedenreich 2008). The same review reports similar risk reductions from vigorous activities (average 26%) and moderate intensity activities (average 22%).

Determining the time at 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 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 reported inverse associations (~82%). 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 (Latikka 1998). 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. Although most of the CIs included the null, 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 (Dorn 2003). A recent prospective study found moderate-to-vigorous activity during the past 10 years to be associated with postmenopausal breast cancer risk (RR=0.84; 95%CI=0.76, 0.93). Activities during other periods of life were not statistically significant and lifetime activity was not assessed (Peters 2009). Although the optimal timing of physical activity for breast cancer protection remains to be resolved, lifetime physical activity was shown to provide the greatest risk reductions in the Long Island study population (Eng 2002). In the only other study known to use this comprehensive physical activity assessment lifetime activity was reported (Bernstein, 2005).

There is some evidence for a dose response relationship between increasing activity levels and decreasing breast cancer risk. Thune et al. (2001) observed graded dose-response relationships in 57% of studies evaluated (N=28). The proportion is as high as 87% in some reviews (Friedenreich 2002). Evidence of dose-response relationships are more frequently observed in case-control compared to cohort studies (Thune 2001, Friedenreich 2008). One review reported that 47% of case-control studies and 39% of cohort studies found linear tends for decreasing risk with increasing activity (Friedenreich 2008). 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 (Monninkhof 2007).

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 physical activity-breast cancer association.

Menopausal Status

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 (Sesso 1998, Breslow 2001, Gilliland 2001, Dorn 2003), although CIs overlap in many studies. A 2004 review study (Friedenreich, 2004) showed that among 26 studies examining the association between physical activity and premenopausal breast cancer risk 50% found no significant association. Thirteen studies reported risk reductions, with seven being statistically significant.

Among the 27 studies conducted in postmenopausal women 22 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 observe similar trends among pre and post-menopausal women (Monninkhof et al. 2007). While it is clear that an effect of physical activity persist in both pre and post-menopausal strata, risk reductions are greater in magnitude for postmenopausal women (40% average risk reduction) than observed for premenopausal women (33% average risk reduction) (Friedenreich 2008).

BMI

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 (Dallal 2007). Investigators of the E3N cohort report no effect modification by BMI on the physical activity-breast cancer association (Tehard 2006) as did investigators of the MARIE study; even upon stratifying by menopausal status (Schmidt 2009). Some review studies have drawn similar conclusions (McTiernan 1998, Monninkhof 2007). 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 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 (Friedenreich 2008).

Family History

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. reports 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) (Dallal 2007). Similar associations for family history have been observed in other studies (Sprague 2007, Bernstein 2005) and the average risk reduction is estimated to be 6% among women without a family history, compared to an average increased risk of 20% for women with a family history of breast malignancy (Friedenreich 2008).

Energy Intake

Energy intake and energy expenditure together determine overall energy balance which greatly affects adiposity and breast cancer risk. It is therefore important to determine if energy intake confounds 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 (Friedenreich 2008 and Mcterrian 1998). An average of risk reduction of 21% was observed for both groups (Friedenreich 2008).

Hormone Receptor Status

While hormone receptor status is not an effect modifier, 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 (2007) report significant decreases in breast cancer risk with increasing levels of both moderate (P trend=0.003) and strenuous (P trend=0.003) recreational physical activity 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 (Bernstein 2005), and others showing stronger associations for ER+ tumors (Bardia 2006). Friedenreich and colleges reported risk reductions of 29% and 14% for hormone receptor negative and positive tumors, respectively (Friedenreich 2008).

1.3.3 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 (McTiernan 2008, Rundle 2005, Neilson 2009). The lack of information on the molecular epidemiology of physical activity and the invariable presence of 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 (Rundle 2005, Neilson 2009, Lorincz 2006, Calle 2004, Rogers 2008).

Bodyweight

The most commonly proposed mechanism for the physical activity breast cancer association is through a reduction in bodyweight. Increased bodyweight is an independent risk factor for postmenopausal breast cancer and its effects on endogenous 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 (Matsuzawa 1995). Physical activity is known to preferentially reduce central obesity (McTiernan 1998) 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 post-menopausal 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 (Bernstein 1987, Harlow 1991, Cooper 1996, Hoffman-Goetz 1998), primarily as a result of luteal phase inadequacy (Frienenerich 1995). While these data suggest that exercise reduces cumulative exposure to sex steroid hormones thereby deceasing the risk of breast cancer (Bernstein 1987, Latikka 1998), more recent research indicates that beyond its effect on energy availability exercise has little disruptive effect on the hormonal milieu of premenopausal women (Loucks 2003). Moreover, there is little evidence of a direct association between disturbances in menstrual characteristics and breast cancer outcomes.

Among premenopausal women the ovaries are the principal source of estrogens (primarily estraidol). 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 (Forney 1981). The P450 enzyme aromatase is known to be involved in the biosynthesis of estrogens from androgenic precursors (Calle 2004) which have been correlated with increased breast cancer risk (Dorgan 1996). In both pre- and postmenopausal women estradiol forms a reversible redox reaction with estrone which is subsequently metabolized along one of two pathways: 2-hydroxyestrone (2HE) or 16α-hydroxyestrone (16HE) (Shepard 1998). Adipose mediated estrogen production is preferential to the bioactive 16HE pathway and is associated with decreased levels of the less bioactive 2HE (Matthews 2004). Among obese postmenopausal women the increased propensity for the 16HE pathway results in an even greater amount of circulating estrogen. In contrast, pre-menopausal women experience consistently high levels of estrogen from menarche, regardless of weight, so any additional exposure from the aromatization of androgens or 16HE metabolizing pathway would not greatly impact overall estrogen levels.

Bioavailability of sex steroids also increases after menopause as the levels of sex hormone binding globulin (SHBG), the predominant carrier of estradiol, decreases (Haffner 1991). 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 (Kramer 1996). Physical activity has been shown to reduce serum concentrations of both estrogens and androgens in postmenopausal women (Friedenreich 2010, Bertone-Johnson 2009, Van Gils 2009, Hoffman-Goetz 1998, and Chan 2007), although there may be the possibility of a U-shaped relation between activity and hormone levels (Bertone-Johnson, 2009). 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

(Gunter 2008) and influenced by both central adiposity and physical activity (Giovannucci 2001). 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 effort to maintain glucose homeostasis (Kaaks 1996, Lorincz 2006, Harris 2004, Calle 2004). 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 IGF-1, and (3) reduced availability of SHBG (Calle 2004, Kaaks 2001). Increases in both insulin and 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 (Hurstings 2007). Reduced SHBG increases the fraction of bioavailable estradiol and testosterone. Collectively, these mechanisms provide an opportunity for tumor development and progression.

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 (Boule 2001, Ross 2004). However, intervention studies show comparable improvements in insulin resistance for diet-induced or exercised-induced weight loss (Ross 2004) suggesting that an overall reduction in adiposity, irrespective of the mechanism, improves 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 (Yu 2000). The relation between physical activity and IGF-1 is not clear however; neither observational nor experimental studies show an effect of exercise on circulating IGF-1 levels (McTiernan 2005, Orenstein MR, Friedenreich 1994). Similarly, animal models indicate that exercise training does not affect basal levels of IGF-1 (Rogers 2008). The evidence is more consistent that physical activity increases levels of IGF-BP resulting in a decrease in overall IGF-1 bioavailability and activity (Hurstings 2007, Yu 2000). 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 (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) (Trayhurn 2006). 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 (IL-1ra) and soluble TNF- α receptors (sTNF-R) (Hursting 2007). 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 (Arita 1999). Chronic inflammation is not only a risk factor for obesity, but has been associated with several other chronic conditions including metabolic syndrome (Das 2004), type 2 diabetes (Mishima 2001), and some cancers (Il'yasova 2005, Aggarwal 2006). Increases in cancer risk due to chronic inflammation may occur because of changes in the microenvironment, increased proliferative activity, and oxidative stress (Coussens 2002). 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 (Mattusch 2000, Starkie 2003, McTiernan 2008). One study reports that while there are small increases of pro-inflammatory markers, a surge of cytokine inhibitors are initiated following exercise (Ostrowski 1999). 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 (McTiernan 2008).

Immune Function

It is hypothesized that the immune system aids in cancer risk reduction by recognizing and eliminating abnormal cells (Jakóbisiak 2003). 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 (Shepard 1998). 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 (Rodgers 2008, Shephard 1995, and Nieman 1997). These effects are transient, with levels of immune parameters dropping below pre-exercise levels following activity (Nieman 1997). 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 (Shepard 1998 and McTiernan, 2008). Depression of immune function may be induced by excessive exercise at high intensity levels (McTiernan, 2008). 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 (Mittendorf 1995 and Thune 1997). Similarly, animal models have shown that the negative energy balance induced by exercise does not, alone, explain the cancer preventive effects of physical activity (Rodgers 2008). 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 1). 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.

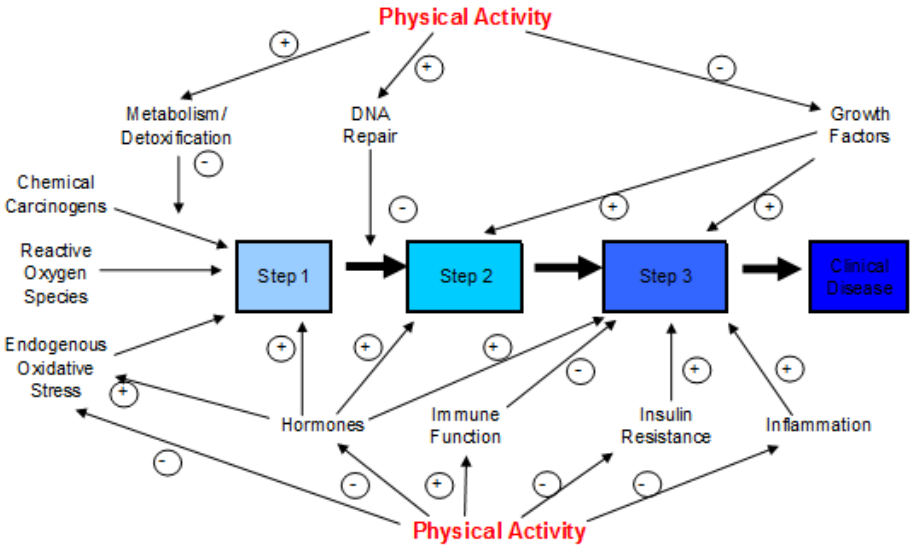

FIGURE 1. Mechanisms of Physical Activity. Adapted from Rundle 2005

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 downstream including pathways related to early stages of malignant transformation (Rundle 2005 and Neilson 2009). Markers associated with DNA repair and oxidative stress have been shown to be pertinent to breast cancer, but they may also be modifiable by exercise (Hoffman-Goetz 1998, Rundle 2005, Friedenreich 2008). To date there are no studies which have evaluated the biologic plausibility of these mechanisms.

1.3.4 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 (Friedenreich 2008). 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 bodyweight – 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 DNA repair and oxidative stress pathways in the physical activity-breast cancer association

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. DNA damage can occur through a variety of endogenous and exogenous processes. In human cells base loss (hydrolysis) is the most frequent type of DNA damage with up to 10,000 abastic sites (apurine/apyrimidine sites resulting form the loss of a purine or pyrimidine residue) generated daily (Lindahl, 1993). DNA integrity may be compromised by oxidative damage *(Further discussed in section 1.5),* which occurs when a cell is exposed to increased amounts of reactive oxygen species (ROS) (Cadenas, 1989). If these compounds are not neutralized by plasma antioxidants they have the capacity to react with biomolecules (e.g. DNA, lipids, and proteins) causing DNA damage. DNA methylation (the addition of a methyl group to DNA) is another route to DNA damage. Although these changes are reversible and occur without modifying the DNA sequence some lesions (i.e. 3-methyladenine) promote mutagenesis if left unrepaired (Frosina, 2000). DNA damage commonly manifests as “non-bulky” DNA adducts or single base modifications (Yu, 1999) and if unrepaired, may be passed on to successive generations during cell division. Damage caused by bulky adducts can result in helical distortions that may inhibit transcription (Balajee 2000).

The integrity of DNA is primarily maintained by DNA 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 that operate on damaged DNA: base excision repair (BER), double strand break (DSB) repair, nucleotide excision repair (NER) and mismatch repair (MMR) (Goode 2002).

1.4.1 DNA REPAIR PATHWAYS

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 (Goode 2002). Repair is initiated by DNA glycosylases, which recognize and remove damaged nucleotides by inverting the double helix and cleaving the N-glycosylic bond generating an abastic (AP) site (Lu 2001). 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) (Lui 2007 and Robertson 2009). Each pathway necessitates its on set of enzymes. The short-patch BER pathway requires four proteins: AP endonuclease (APE1), DNA polymerase β (POLβ), and DNA ligase III/XRCC1 heterodimer. The long-patch BER pathway involves six: APE1, replication factor C (RFC), proliferating cell nuclear antigen (PCNA), flap endonuclease 1 (FEN1), DNA polymerases δ/ε (POLδ/ε), and DNA ligase I (Pascucci 1999, Lu 2001, and Robertson 2009).

Double Strand Break Repair

Double-strand breaks can be produced by exogenous agents, chemotherapeutic agents, endogenously generated reactive oxygen species (ROS) and replication errors (Khanna 2001). Double strand breaks are among the most dangerous type of DNA damage, as the unavailability of a damage-free template makes repair more difficult (Goode 2002). There are two main DSB repair pathways: homologous recombination (HR) and non-homologous end joining (NHEJ)

(Sancar 2004). The HR pathway involves greater than 15 molecules including breast and ovarian cancer genes *BRCA1* and *BRCA2*, as well as X-ray repair complementing defective repair in Chinese hamster cells 1 (*XCCR1)*, *XCCR2*, and *ATM*. DNA-dependent protein kinase A is an essential component of the NHEJ pathway, although other molecules (e.g. Ligase IV) are important (Khanna 2001). In HR the damaged ends are resected 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 – potentially resulting in more error (Cahill 2006). NHEJ additionally has two known sub-pathways: classic NHEJ (described above) and alternative NHEJ (Fattah 2010). Little is known about the mechanisms and factors involved in A-NHEJ.

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 (Reardon 1997, Braithwaite 1998, Goode 2002, Sancar 2004). 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 (Eveno 1995, Tuteja 2001). 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 Sancar 2004). 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 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 (Friedberg 2001, Sancar 2004).

Mismatch Repair

The MMR genes repair incorrect pairings of nucleotide bases (base-base or insertion-deletion mismatched) which may occur as a result of genetic recombination, replicative errors in DNA polymerase, deamination of 5-methylcytosine to thymine, or environmental mutagens (Aquilina 2001 and Chintamani 2007). MMR is essential for maintaining genomic stability as its proteins have been shown to encourage cytotoxicity (Fritzell 1997), p53 phosphorylation (Peters 2003), cell-cycle arrest (Aquilina 1999), and cell death (Hickman 1999, Wu 1999, and Peters 2003) 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 micro satellites 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 (Sia 1997 and Lengauer 1998). Loss of MMR function may occur because of mutations in one of six genes associated with MMR: *MLH1, MLH3, MSH2, MSH3, MSH6* and *PMS2*. These genes make up the MutS and MutL homologue proteins involved in MMR. 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 (Kolodner 1999). 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 ADP from 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 30->50 exonuclease and re-synthesized by DNA polymerase γ and PCNA (Aquilina 2001).

1.4.2 EPIDEMIOLOGY OF DNA REPAIR GENES AND BREAST CANCER

Approximately 130 human genes are involved in the four DNA repair pathways described above, each playing an important role in the maintenance of genomic integrity (Wood 2001). 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 (Berwick 2000 and Goode 2002). 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 (Berwick 2000). 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 (Goode 2002). 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); *ERCC1, XPC, XPD,* and *XPF* (NER genes); and *BRCA2* and *XRCC3* (DSB repair genes) (Goode 2002). 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 R194W polymorphism in *XRCC3* showed a reduced risk of breast cancer associated with the W allele (OR=0.7, 95%CI: 0.4, 1.3 for RW/WW vs. RR) (Duell 2001). The N372H polymorphism of *BRCA2* was associated with breast cancer risk in several

studies. The combined odds of breast cancer among women with the HH genotype was 1.3 times the odds (95% CI: 1.1, 1.6) of breast cancer among women with NN the genotype (Goode 2002).

Following the Goode 2002 review, a number of studies have reported on variants in DNA repair and breast cancer risk in several study populations. Epidemiologic studies conducted among Caucasian women have not supported a role for genetic variations in the BER pathway and carcinogenesis (Zhang 2006 and Thyagarajan 2006). Study results from both Asian (Zhang 2006 and Sangrajrang 2008) and Indian (Chacko 2005 and Mitra 2008) populations are suggestive of an association, specifically among XRCC1 polymorphisms (Li, 2009). While some studies show mixed results for the function of DSB repair genes in the etiology of breast cancer (Han 2004), the vast majority of the evidence is supportative of an association (Han 2006, García-Closas 2006, and Smith 2003). Among all DSB polymorphisms examined *XRCC3* is the most consistently associated with breast cancer susceptibility, although its effects are likely small (García-Closas 2006). Variants of the NER pathway have been evaluated in multiple studies. Several study reports show that genetic polymorphisms in *XPD, XPF, ERCC1*, and *ERCC4* may be associated with increased breast cancer risk (Smith 2003, Crew 2007, Lee 2005, and Terry 2004). There remains a dearth of literature examining polymorphisms in MMR and breast cancer outcomes.

To date there have been only four published studies to estimate the association between MMR variants and breast cancer risk (Poplawaski 2005, Lee 2005, Smith 2008, Conde 2009). Poplawski and colleagues examined two common polymorphisms of the MSH2 gene: an A → G transition at 127 position producing an Asn → Ser substitution at codon 127 (the Asn127Ser polymorphism) and a G → A transition at 1032 position resulting in a Gly → Asp change at codon 322 (the Gly322Asp polymorphism). Both polymorphisms are sense mutations capable of

changing the biological properties, structure, and function of the MSH2 protein (Poplawaski 2005). 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) (Poplawaski 2005). Using a hospital based sample of Korean women Lee and colleagues revealed that the risk of breast cancer increased in a dose response manner with the number of *MLH1* – 93 G alleles (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 (Lee 2005). A 2008 case control study reported a decrease in breast cancer risk with variant alleles in the 219II/IV polymorphism of MLH1 (OR=0.49 95% CI=0.29-0.85) (Smith 2008). While several other polymorphisms were examined (MSH3: R940Q, T1036A and MSH6: G39E) no other statistically significant associations emerged among Caucasian women. The most recent study (Conde 2009), conducted in a hospital setting among Portuguese women, estimated the effects of polymorphism in several MMR genes (*MSH3, MSH4, MSH6, MLH1, MLH3, PMS1 and MUTYH*). The L844P, G>A polymorphism of *MLH3* was found to be associated with breast cancer incidence. The OR's (95% CI's) for GA and AA genotypes were 0.65 (0.45-0.95) and OR = 0.62 (0.41-0.94), respectively. 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 MISMATCH REPAIR AND BREAST CANCER

Available evidence indicates that that a considerable fraction of breast tumors show instability in sequence motifs of mono- and di-nucleotide repeats (Yee 1994, Shaw 1996, Walsh 1998, Shen 2000, and Siah 2000). This phenonema is known as microsatellite instability (MSI) and has

been shown to be closely associated with MMR deficiency (Strand 1993 and Goode 2002). These deficiencies occur frequently in hereditary non-polyposis colorectal cancers (HNPCC) (Thibodeau 1993, Thibodeau 1998, Lynch 1999, and Chiaravall 2001) as well as other MSI related cancers including some sporadic colorectal tumors, hematological malignancies, endometrial, prostatic, and gastric cancers (Fleisher 1999, Gurin 1999, Salvesen 2000, Chintamani 2007, and Cohen 2008). 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 (Baylin 2000, Herman 2000). These changes may lead to increased mutations of oncogenes, tumor suppressor genes, and loss of DNA damage-induced apoptosis, therefore facilitating carcinogenesis (Schofield 2003). While several MMR genes are associated with cancer predisposition, the MSH2 and MLH1 genes are central to all mismatch recognition and have been shown to be the most common mechanism inducing cancer-related MSI (Kinzler 1996, Spagnoletti 2004, Cai 2004). 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), MLH1 (N=5), or both (N=7) and approximately 50% had genetic alterations is both genes (Murata 2002). 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 (Shen 2000, Murata 2002, Seitz 2003). Chintamani and colleagues report a range from 5-30% (Chintamani 2007). This variation is likely due to differences in microsatellie markers used for analysis but may also indicate that progression differs, mechanistically, in primary breast cancer compared to HNPCC (Chintamani 2007). For example, the lack of 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 genes are related to tumor progression and invasion (Block 2000, Murata 2002, Köster 2007, Chintamani 2007) although some studies report null effects (Köster 2007, Khilko 2007). Although modest, the evidence to date indicates that breast cancer is associated with MSI, MSI is primarily caused by variations in MMR genes, polymorphisms in MMR genes may be related to breast cancer risk, and expression of these genes are potentially important modulators of breast cancer progression.

1.4.3 DNA REPAIR AND PHYSICAL ACTIVITY

While many epidemiologic studies report modest effects of DNA repair variants on breast cancer outcomes a number of investigations have found that these polymorphisms may more profoundly influence carcinogenesis by modifying the effect of environmental exposures on cancer risk. The effects of smoking (Shen [MGMT] 2005, Shen [XRCC1] 2005, Metsola 2005, Pachkowski 2006, Mechanic 2006, Smith 2008); PHA-DNA adducts (Crew 2007); radiation exposure (Millikan 2005, Rajaraman 2008); and dietary factors (Han 2004, Shen [MGMT] 2005, Shen [XRCC1] 2005, Shrubsole 2004) have each been shown to be modified by SNPs in DNA repair genes. To date there are no studies which evaluate these associations with physical activity.

Repair enzymes have been shown to be up-regulated with long term exercise in several studies. As early as 1999, Radak and colleagues (Radak 1999) assessed whether regular exercise

decreased the degree of oxidative damage in lipids, proteins, and DNA (namely 8-hydroxydeoxyguanosine – 8-OH-dG). Oxidative modified DNA in the form of 8-OH-dG can be quantified to indicate the extent of DNA damage (Wu 2004). Radak showed that regular exercise decreased the accumulation of nuclear 8-hydroxydeoxyguanosine (8-OH-dG) in the skeletal muscle of exercised rats (Radak 1999). The investigators hypothesize that the observed reduction could be attributed to the increased regulation of repair systems and later proposed that regular exercise causes adaptation of both antioxidant and repair systems resulting in increased resistance to oxidative damage (Radak 2001). A 2002 study of aged rats (20-30 months old) confirmed this, as regular exercise attenuated the age-associated increase in 8-OH-dG in skeletal muscle as well as increased activity of the proteasome complex, a repair enzyme important in the degradation of proteins modified by oxidative stress (Radak 2002). Sato and colleagues extended these observations to exercise in healthy male participants. They found that the basal 8-OH-dG levels of physically active men were significantly lower than those of sedentary men. The 8-OH-dG levels of active men remained unchanged post exercise and 8-OH-dG levels of the sedentary participants decreased following 30 min of mild exercise (Sato 2003). Similarly, studies of trained cyclist (Wittwer 2004) and marathon runners (Radak 2003) show up-regulation of DNA excision repair enzymes NESP and RAD23A (Wittwer 2004), as well as HOGG1 (Radak 2003). Collectively, these data suggest that regular exercise may reduce genomic damage by increasing any number of DNA repair mechanisms.

1.4.4 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. To date there are only four documented analyses of the association. Although these studies showed modest to strong associations with breast cancer risk, odds ratios ranging from 1.2 - 8.0, they generally

relied on non-representative samples and lacked statistical power. Further investigations of these genes are warranted 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, polycyclic aromatic hydrocarbon-DNA adducts, dietary antioxidants, and fruit and vegetable consumption (Shen [MGMT] 2005, Shen [XRCC1] 2005, Metsola 2005, Pachkowski 2006, Mechanic 2006, Smith 2008, Crew 2007, Millikan 2005, Rajaraman 2008 Han 2004, Shrubsole 2004). While it has been recognized that physical activity may exert its effects via DNA repair, no study has considered a possible interaction with physical activity levels. Animal and clinical studies show that DNA repair enzymes are up regulated with physical activity which is likely to result in diminished DNA damage. An epidemiologic model could be tested by assessing the joint effects of low physical activity and reduced DNA repair capacity on breast cancer outcomes.

1.5 OXIDATIVE STRESS

I previously highlighted several factors that could initiate DNA damage *(Section 1.4).* Among them are reactive oxygen species (ROS), which may be generated through any number of endogenous or exogenous processes. The term reactive oxygen species (ROS) is commonly used to describe certain reactive oxygen and nitrogen metabolites containing unpaired electrons in their outer orbit (Karihtala 2006 and Gago Dominguiez 2007). ROS includes free radical derivatives of molecular oxygen (e.g. superoxide radical $O_2^{\bullet -}$, hydrogen peroxide H_2O_2, hydroxyl radical $^{\bullet}OH$) and are continuously generated within a cell as a result of oxidative metabolism (Clarkson 2000 and Millikan, 2004). ROS may be produced by other endogenous processes such as estrogen metabolism, peroxisomes, or inflammatory cell activation. ROS and ROS-generating compounds are ubiquitous in the environment commonly found in inhaled smoke, alcohol, and ingested goods (Klaunig 2004 and Rossner 2006). While modest levels of ROS are useful for cell signaling processes (Martin 2002) excess ROS may result in DNA damage, lipid peroxidation, and protein modification (Marnett 2000, Cooke 2003, Klaunig 2004, and Tas 2005). 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 (Ambrosome 2000).

1.5.1 ENDOGENOUS RESPONSES TO OXIDATIVE STRESS

When endogenous or exogenous ROS production occurs in an environment with sufficient plasma antioxidants to scavenge the ROS, there are seemingly few harmful effects. When there is excess ROS production or insufficient *in vivo* 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), superoxide dismutase (SOD), glutathione peroxidase (GPx), as well as glutathione S-transferases (GST) (Ambrosome 2000). CAT, SOD, and GPx enzymes form the first line of defense against superoxide and hydrogen peroxide. Secondary defenses include reduced GPx and GST (Datta 2000 and Mates 1999). These enzymes play a central role in the defense against free radicals, peroxides, as well as a wide array of xenobiotics and carcinogens (Abou Ghalia 2000). 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) (Ambrsome 2005). MPO generates ROS endogenously by performing as an antimicrobial enzyme, catalyzing a reaction between H2O2 and chloride to generate hypochlorous acid (Ahn 2006). Hypochlorous acid further reacts with other biological molecules to generate secondary radicals (Klebanoff 1980). Endothelial nitric oxide synthase (eNOS) is also responsible for the generation of ROS, catalyzing the production of the NO radical (Martin 1999). The levels of potentially cytotoxic ROS may ultimately depend on the balance between endogenous pro- and anti-oxidants.

1.5.2 EPIDEMIOLOGY OF OXIDATIVE STRESS GENES AND BREAST CANCER

Oxidative stress may play an important role in the risk of many chronic diseases including cardiovascular disease, aging, and human cancer (Halliwell 1987, Cross 1987, Feig 1994, and Forsberg 2001). 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 (Ambrosone 2000, Kang 2002, Behrend 2003, and Caporaso 2003). Oxidative damage has been suggested to contribute to the formation of DNA adducts (Li 1999) and has been implicated in neoplastic transformation (Hristozov 2001). Oxidative damage has frequently been reported to be higher among women with breast cancer, compared to controls (Ambrosone, 2000). Similarly, studies have observed enhanced lipid

peroxidation in breast tumor tissue compared to uninvolved adjacent tissue as well as elevations in both enzymatic (SOD, CAT, GPx, GST) and nonenzymatic (GSH) antioxidants of tumor cell lines (Kumaraguruparan 2005, Tas 2005, Rajneesh 2008, and Kumaraguruparan, 2002). In some studies, however, higher SOD and GPx activities were accompanied with lower CAT activity in cancerous tissue compared to nonmalignant tissues (Tas 2005, Punnonen 1994).

CAT is a heme enzyme that has a primary role in neutralizing ROS by converting H2O2 into H2O and O2 (Ambrosone, 2000). Activity levels of the CAT enzymes are likely affected by functional polymorphisms in the genes encoding them. A common catalase-262 C/T polymorphism (rs1001179) has been identified in the promoter region of the human CAT gene (Forsberg, 2001). It is plausible that the endogenous variability associated with this SNP plays a role in the host response to oxidative stress. Studies confirm this, as the variant CAT allele (T) has been associated with both hypertension (Zhou 2005) and vitiligo (Casp 2002); conditions related to oxidative stress. This polymorphism has been shown to affect the transcriptional activity of the promoter (Forsberg 2001) and is thought to result in reduced activity (Ahn 2006, Nadif 2005, and Bastaki 2006). Using a sample of 420 controls Ahn and colleagues examined the functional effects of this variant on 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 (Ahn 2006). These associations were also observed in a smaller, yet independent, study population (Ahn, 2005). This evidence lends support for the potential role of Catalase in breast cancer etiology.

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 (Kasapovic´ 2008). Genetic variants that influence pro or anti-oxidant mechanisms may therefore play an important role in breast cancer carcinogenesis (Ambrosone, 2000). Animal models have shown that acatalasemic mice, with approximately one tenth the CAT blood and tissue levels of normal mice, were more susceptible to mammary carcinoma (Ishii 1996). While these models 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. A few epidemiologic studies have examined the association between rs1001179 and breast cancer risk. In 2005, Ahn reported results from the Long Island Breast Cancer Study Project indicating that women with the common CC genotype had a 17% reduction in risk of breast cancer (95% CI: 0.69, 1.00) compared with those with at least one T allele (CT and TT genotypes) (Ahn, 2005). Follow-up investigations found approximately null associations between CAT genotypes and breast cancer risk (Quick 2008; Li 2009), although there were less participants in both studies. To date no other Catalase variants have been reported in the epidemiologic literature.

1.5.3 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 (Pryor, 1993; Brooks 1997; Yager, 2000; Kumaraguruparan, 2005) while others (i.e. consumption of fruits vegetables and antioxidants) are suggested to oppose ROS formation (Ambrosone, 2000) 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 “at risk” genotypes relating to 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 genotypes (Ahn 2006). Among ever users of HRT the increased risk of breast cancer was more pronounced among women with variant CT or TT CAT genotype (OR=1.88; 95% CI=1.29-2.75) than among women with CC genotype (OR=1.15; 95% CI=0.86-1.54) although CAT genotype alone was not associated with breast cancer risk (Quick 2008). Factors suggested to oppose ROS generation have been shown to obliterate the increased risk associated with genotype. Among women with at risk B/B GSTA1 genotypes, Ahn et al. observed a inverse trend between cruciferous vegetable consumption and breast cancer risk (P for trend = 0.05) (Ahn, April 2006). 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 consumers/variant genotype had OR's (95%CI's) = 0.94 (0.65-1.37) and 1.06 (0.66-1.73), respectively (Ahn, 2005). These data indicate that genotypes related 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 (Clarkson 2000, Kanter 1994, Ayres 1998). While the dimensions of exercise induced ROS production are unknown, there is some evidence that strenuous activity (Singh 1992, Guerra 2000, Guerra 2001), endurance and resistance training (McBride 1998) increase lipid peroxidation. Three mechanisms of exercise induced ROS production have been proposed (Packer, 1997). 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 radical. 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 known to stimulate endogenous antioxidants as a physiological response to the oxidative stress induced by exercise (Vani 1990 and Evelo 1992). Up-regulation of antioxidant enzymes may render cells more resistant to subsequent oxidative insult (Halliwell 2000) thereby neutralizing the potentially mutagenic effects of lipid peroxidation (Miyazaki 2001). Both animal and human studies support changes in antioxidant enzymes GST, SOD, and GPx following exercise. Rodent studies show increased levels of hepatic CAT (Ji 1988, Singal 1993, Kanter 1985) with exercise training. Amplified levels of hepatic SOD (Singal 1993, Kanter 1985) and cytosolic SOD activity (Ji 1988) have also been documented. These studies translate to observations among humans, as there is evidence of a positive correlation between training and enthrocyte GST (Evelo 1992), GPx and CAT activity (Robertson 1991) among runners. In addition to changes in antioxidant enzymes, exercise induced oxidative damage may also lead to increased cellular apoptosis providing yet another mechanism for the inverse association between physical activity and breast cancer (Phaneuf 2001).

1.5.4 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 (Behrend 2003, Kang 2002) 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 (Ambresome 2000). 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 variants in CAT and breast cancer risk, as studies have shown a dose-response reduction in activity with the 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 the 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 (i.e. CAT) modify the effect of physical activity on breast cancer risk.

1.6 CONCLUSIONS AND INNOVATION OF STUDY AIMS

While much attention has been given to obesity related mechanisms in the attempt to understand the molecular effects of physical on breast cancer risk, studies have been unsuccessful. It is well documented that exercise may work through obesity related pathways but independent and review studies alike show that these mechanisms are probable at best. They additionally fail to explain how the effect for physical activity remains after controlling for obesity. We must strengthen the evidence for an independent effect of physical activity on breast cancer incidence in order to refute the claim that the observed association is due to a healthy person effect.

To my knowledge the proposed study will be the first 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 environmental modulators, their impact on physical activity has not been evaluated in the epidemiologic literature. The aim of this ancillary study is therefore two-fold. I would like to estimate the main effect of understudied DNA repair and oxidative stress variants on breast cancer risk by assaying new SNPs in the LIBCSP. Understanding how specific genetic changes in these pathways impact breast cancer incidence will contribute to the current knowledge of disease etiology. Additionally, it is my aim to better understand the molecular epidemiology of physical activity. Pathways outside the obesity paradigm will be explored to establish an independent mechanism for the physical activity-breast cancer association.

AIM 1. DNA Repair

AIM 1A: to determine the main effect of select SNPs in three genes (*MLH1, MSH2, and MSH3*) of the MMR pathway on breast cancer risk.

AIM 1B: to evaluate interactions between polymorphisms in DNA repair genes (MMR, BER, and NER pathways) and self-reported lifetime physical activity on breast cancer risk (gene-environment interactions)

AIM 1C: to explore potential interactions between SNPs in genes from three DNA repair pathways (MMR, BER and NER) on breast cancer risk (gene-gene interactions).

AIM 2. Oxidative Stress

AIM 2A: to determine the main effect of select SNPs in *CAT* on breast cancer risk.

AIM 2B: to evaluate interactions between polymorphisms in oxidative stress genes and self-reported lifetime physical activity on breast cancer risk (gene-environment interactions)

AIM 2C: to explore the potential interactions between SNPs in genes from the oxidative stress pathway on breast cancer risk (gene-gene interactions).

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

The proposed study has widespread implications for breast cancer. First, 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. The proposed study would be the first to systematically evaluate these associations in a large population-base sample. Second, Catalase plays an important role in neutralizing

oxygen derivatives produced by endogenous and exogenous sources. While one SNP in CAT has been found to be associated with breast cancer risk (in the Long Island study population) other variants should be examined. Finally, there is consensus in the importance of understanding the mechanism through which physical activity exerts its protective effects on breast cancer risk. However, little research has been conducted outside the obesity related pathways despite the strong biologic plausibility of DNA repair and oxidative stress 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.

REFERENCES

Abou Ghalia AH, Fouad IM. Glutathione and its metabolizing enzymes in patients with different benign and malignant diseases. Clin Biochem. 2000 Nov;33(8):657-62. PubMed PMID: 11166013.

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

Aggarwal BB, Shishodia S, Sandur SK, Pandey MK, Sethi G. Inflammation and cancer: how hot is the link? Biochem Pharmacol. 2006 Nov 30;72(11):1605-21. Epub 2006 Aug 4. Review. PubMed PMID: 16889756.

Ahn J, Ambrosone CB, Kanetsky PA, Tian C, Lehman TA, Kropp S, Helmbold I, von Fournier D, Haase W, Sautter-Bihl ML, Wenz F, Chang-Claude J. 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. 2006 Dec 1;12(23):7063-70. PubMed PMID: 17145829.

Ahn J, Gammon MD, Santella RM, Gaudet MM, Britton JA, Teitelbaum SL, Terry MB, Neugut AI, Eng SM, Zhang Y, Garza C, Ambrosone CB. Effects of glutathione S-transferase A1 (GSTA1) genotype and potential modifiers on breast cancer risk. Carcinogenesis. 2006 Sep;27(9):1876-82. Epub 2006 Apr 19. PubMed PMID: 16624829.

Ahn J, Nowell S, McCann SE, Yu J, Carter L, Lang NP, Kadlubar FF, Ratnasinghe LD, Ambrosone CB. Associations between catalase phenotype and genotype: modification by epidemiologic factors. Cancer Epidemiol Biomarkers Prev. 2006 Jun;15(6):1217-22. PubMed PMID: 16775184.

Ahn J, Gammon MD, Santella RM, Gaudet MM, Britton JA, Teitelbaum SL, Terry MB, Nowell S, Davis W, Garza C, Neugut AI, Ambrosone CB. Associations between breast cancer risk and the catalase genotype, fruit and vegetable consumption, and supplement use. Am J Epidemiol. 2005 Nov 15;162(10):943-52. Epub 2005 Sep 28. PubMed PMID: 16192345.

Ainsworth BE, Haskell WL, Leon AS, Jacobs DR Jr, Montoye HJ, Sallis JF, Paffenbarger RS Jr. Compendium of physical activities: classification of energy costs of human physical activities. Med Sci Sports Exerc. 1993 Jan;25(1):71-80. PubMed PMID: 8292105.

Ambrosone CB, Ahn J, Singh KK, Rezaishiraz H, Furberg H, Sweeney C, Coles B, Trovato A. Polymorphisms in genes related to oxidative stress (MPO, MnSOD, CAT) and survival after treatment for breast cancer. Cancer Res. 2005 Feb

1;65(3):1105-11. PubMed PMID: 15705913.

Ambrosone CB. Oxidants and antioxidants in breast cancer. Antioxid Redox Signal. 2000 Winter;2(4):903-17. Review. PubMed PMID: 11213491.

American Cancer Society, *Breast Cancer Facts and Figures 2007-2008*. Retrieved 2010, American Cancer Society: Atlanta, GA

American Cancer Society, *Cancer Facts & Figures 2009*. Retrieved 2010, American Cancer Society: Atlanta, GA

Antoniou A, Pharoah PD, Narod S, Risch HA, Eyfjord JE, Hopper JL, Loman N, Olsson H, Johannsson O, Borg A, Pasini B, Radice P, Manoukian S, Eccles DM, Tang N, Olah E, Anton-Culver H, Warner E, Lubinski J, Gronwald J, Gorski B, Tulinius H, Thorlacius S, Eerola H, Nevanlinna H, Syrjäkoski K, Kallioniemi OP, Thompson D, Evans C, Peto J, Lalloo F, Evans DG, Easton DF. 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. 2003 May;72(5):1117-30. Epub 2003 Apr 3. Erratum in: Am J Hum Genet. 2003 Sep;73(3):709. PubMed PMID: 12677558; PubMed Central PMCID: PMC1180265.

Aquilina G, Bignami M. Mismatch repair in correction of replication errors and processing of DNA damage. J Cell Physiol. 2001 May;187(2):145-54. Review. PubMed PMID: 11267994.

Aquilina G, Crescenzi M, Bignami M. Mismatch repair, G(2)/M cell cycle arrest and lethality after DNA damage. Carcinogenesis. 1999 Dec;20(12):2317-26. PubMed PMID: 10590226.

Arita Y, Kihara S, Ouchi N, Takahashi M, Maeda K, Miyagawa J, Hotta K, Shimomura I, Nakamura T, Miyaoka K, Kuriyama H, Nishida M, Yamashita S, Okubo K, Matsubara K, Muraguchi M, Ohmoto Y, Funahashi T, Matsuzawa Y. Paradoxical decrease of an adipose-specific protein, adiponectin, in obesity. Biochem Biophys Res Commun. 1999 Apr 2;257(1):79-83. PubMed PMID: 10092513.

Ayres S, Baer J, Subbiah MT. Exercised-induced increase in lipid peroxidation parameters in amenorrheic female athletes. Fertil Steril. 1998 Jan;69(1):73-7. PubMed PMID: 9457937.

Balajee AS, Bohr VA. Genomic heterogeneity of nucleotide excision repair. Gene. 2000 May 30;250(1-2):15-30. Review. PubMed PMID: 10854775.

Bardia A, Hartmann LC, Vachon CM, Vierkant RA, Wang AH, Olson JE, Sellers TA, Cerhan JR. Recreational physical activity and risk of postmenopausal breast cancer based on hormone receptor status. Arch Intern Med. 2006 Dec 11-25;166(22):2478-83. PubMed PMID: 17159013.

Bastaki M, Huen K, Manzanillo P, Chande N, Chen C, Balmes JR, Tager IB, Holland N. Genotype-activity relationship for Mn-superoxide dismutase, glutathione peroxidase 1 and catalase in humans. Pharmacogenet Genomics. 2006 Apr;16(4):279-86. PubMed PMID: 16538174.

Batty GD. Physical activity and coronary heart disease in older adults. A systematic review of epidemiological studies. Eur J Public Health. 2002 Sep;12(3):171-6. Review. PubMed PMID: 12232954.

Baylin SB, Herman JG. DNA hypermethylation in tumorigenesis: epigenetics joins genetics. Trends Genet. 2000 Apr;16(4):168-74. Review. PubMed PMID: 10729832.

Behrend L, Henderson G, Zwacka RM. Reactive oxygen species in oncogenic transformation. Biochem Soc Trans. 2003 Dec;31(Pt 6):1441-4. Review. PubMed PMID: 14641084.

Bernstein L, Patel AV, Ursin G, Sullivan-Halley J, Press MF, Deapen D, Berlin JA, Daling JR, McDonald JA, Norman SA, Malone KE, Strom BL, Liff J, Folger SG, Simon MS, Burkman RT, Marchbanks PA, Weiss LK, Spirtas R. Lifetime recreational exercise activity and breast cancer risk among black women and white women. J Natl Cancer Inst. 2005 Nov 16;97(22):1671-9. PubMed PMID: 16288120.

Bernstein L, Ross RK, Lobo RA, Hanisch R, Krailo MD, Henderson BE. The effects of moderate physical activity on menstrual cycle patterns in adolescence: implications for breast cancer prevention. Br J Cancer. 1987 Jun;55(6):681-5. PubMed PMID: 3620313; PubMed Central PMCID: PMC2002035.

Bertone-Johnson ER, Tworoger SS, Hankinson SE. Recreational physical activity and steroid hormone levels in postmenopausal women. Am J Epidemiol. 2009 Nov 1;170(9):1095-104. Epub 2009 Sep 25. PubMed PMID: 19783585; PubMed Central PMCID: PMC2781741.

Berwick M, Vineis P. Markers of DNA repair and susceptibility to cancer in humans: an epidemiologic review. J Natl Cancer Inst. 2000 Jun 7;92(11):874-97. Review. PubMed PMID: 10841823.

Bock N, Meden H, Regenbrecht M, Jünemann B, Wangerin J, Marx D. Expression of the mismatch repair protein hMSH2 in carcinoma in situ and invasive cancer of the breast. Anticancer Res. 2000 Jan-Feb;20(1A):119-24. PubMed PMID: 10769643.

Boulé NG, Haddad E, Kenny GP, Wells GA, Sigal RJ. Effects of exercise on glycemic control and body mass in type 2 diabetes mellitus: a meta-analysis of controlled clinical trials. JAMA. 2001 Sep 12;286(10):1218-27. Review. PubMed PMID: 11559268.

Braithwaite E, Wu X, Wang Z. Repair of DNA lesions induced by polycyclic aromatic hydrocarbons in human cell-free extracts: involvement of two excision repair mechanisms in vitro. Carcinogenesis. 1998 Jul;19(7):1239-46. PubMed PMID: 9683183.

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. 1997 Oct 11;350(9084):1047-59. Erratum in: Lancet 1997 Nov 15;350(9089):1484. PubMed PMID: 10213546.

Breslow RA, Ballard-Barbash R, Munoz K, Graubard BI. 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. 2001 Jul;10(7):805-8. PubMed PMID: 11440967.

Britt K, Ashworth A, Smalley M. Pregnancy and the risk of breast cancer. Endocr Relat Cancer. 2007 Dec;14(4):907-33. Review. PubMed PMID: 18045947.

Brooks PJ. DNA damage, DNA repair, and alcohol toxicity--a review. Alcohol Clin Exp Res. 1997 Sep;21(6):1073-82. Review. PubMed PMID: 9309320.

Cadenas E. Biochemistry of oxygen toxicity. Annu Rev Biochem. 1989;58:79-110. Review. PubMed PMID: 2673022.

Cahill D, Connor B, Carney JP. Mechanisms of eukaryotic DNA double strand break repair. Front Biosci. 2006 May 1;11:1958-76. Review. PubMed PMID: 16368571.

Cai KQ, Albarracin C, Rosen D, Zhong R, Zheng W, Luthra R, Broaddus R, Liu J. Microsatellite instability and alteration of the expression of hMLH1 and hMSH2 in ovarian clear cell carcinoma. Hum Pathol. 2004 May;35(5):552-9. PubMed PMID: 15138928.

Calle EE, Kaaks R. Overweight, obesity and cancer: epidemiological evidence and proposed mechanisms. Nat Rev Cancer. 2004 Aug;4(8):579-91. Review. PubMed PMID: 15286738.

Caporaso N. The molecular epidemiology of oxidative damage to DNA and cancer. J Natl Cancer Inst. 2003 Sep 3;95(17):1263-5. PubMed PMID: 12953074.

Casey PM, Cerhan JR, Pruthi S. Oral contraceptive use and risk of breast cancer. Mayo Clin Proc. 2008 Jan;83(1):86-90; quiz 90-1. Review. PubMed PMID: 18174010.

Casp CB, She JX, McCormack WT. Genetic association of the catalase gene (CAT) with vitiligo susceptibility. Pigment Cell Res. 2002 Feb;15(1):62-6. PubMed PMID: 11837458.

Caspersen CJ. Physical activity epidemiology: concepts, methods, and applications to exercise science. Exerc Sport Sci Rev. 1989;17:423-73. Review. PubMed PMID: 2676554.

CDC http://apps.nccd.cdc.gov/BRFSS/sex.asp?cat=PA&yr=2007&qkey=4418&state= UB

Chacko P, Rajan B, Joseph T, Mathew BS, Pillai MR. Polymorphisms in DNA repair gene XRCC1 and increased genetic susceptibility to breast cancer. Breast Cancer Res Treat. 2005 Jan;89(1):15-21. PubMed PMID: 15666192.

Chan MF, Dowsett M, Folkerd E, Bingham S, Wareham N, Luben R, Welch A, Khaw KT. Usual physical activity and endogenous sex hormones in postmenopausal women: the European prospective investigation into cancer-norfolk population study. Cancer Epidemiol Biomarkers Prev. 2007 May;16(5):900-5. PubMed PMID: 17507613.

Chen FP. Postmenopausal hormone therapy and risk of breast cancer. Chang Gung Med J. 2009 Mar-Apr;32(2):140-7. Review. PubMed PMID: 19403003.

Chen WY, Manson JE, Hankinson SE, Rosner B, Holmes MD, Willett WC, Colditz GA. Unopposed estrogen therapy and the risk of invasive breast cancer. Arch Intern Med. 2006 May 8;166(9):1027-32. PubMed PMID: 16682578.

Chiaravalli AM, Furlan D, Facco C, Tibiletti MG, Dionigi A, Casati B, Albarello L, Riva C, Capella C. Immunohistochemical pattern of hMSH2/hMLH1 in

familial and sporadic colorectal, gastric, endometrial and ovarian carcinomas with instability in microsatellite sequences. Virchows Arch. 2001 Jan;438(1):39-48. PubMed PMID: 11213834.

Chintamani, Jha BP, Bhandari V, Bansal A, Saxena S, Bhatnagar D. 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. 2007 Feb 14;4:5. PubMed PMID: 17394628; PubMed Central PMCID: PMC1802872.

Clarkson PM, Thompson HS. Antioxidants: what role do they play in physical activity and health? Am J Clin Nutr. 2000 Aug;72(2 Suppl):637S-46S. Review. PubMed PMID: 10919970.

Clemons M, Goss P. Estrogen and the risk of breast cancer. N Engl J Med. 2001 Jan 25;344(4):276-85. Review. Erratum in: N Engl J Med 2001 Jun 7;344(23):1804. PubMed PMID: 11172156.

Cohn DE, Pavelka JC, Frankel WL, Morrison CD, Hampel H, Copeland LJ, Fowler JM. Correlation between patient weight and defects in DNA mismatch repair: is this the link between an increased risk of previous cancer in thinner women with endometrial cancer? Int J Gynecol Cancer. 2008 Jan-Feb;18(1):136-40. Epub 2007 Apr 26. PubMed PMID: 17466051.

Collaborative Group on Hormonal Factors in Breast Cancer. 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. 2002 Jul 20;360(9328):187-95. PubMed PMID: 12133652.

Conde J, Silva SN, Azevedo AP, Teixeira V, Pina JE, Rueff J, Gaspar JF. Association of common variants in mismatch repair genes and breast cancer susceptibility: a multigene study. BMC Cancer. 2009 Sep 25;9:344. PubMed PMID: 19781088; PubMed Central PMCID: PMC2761943.

Cooke MS, Evans MD, Dizdaroglu M, Lunec J. Oxidative DNA damage: mechanisms, mutation, and disease. FASEB J. 2003 Jul;17(10):1195-214. Review. PubMed PMID: 12832285.

Cooper GS, Sandler DP, Whelan EA, Smith KR. Association of physical and behavioral characteristics with menstrual cycle patterns in women age 29-31 years. Epidemiology. 1996 Nov;7(6):624-8. PubMed PMID: 8899389.

Coussens LM, Werb Z. Inflammation and cancer. Nature. 2002 Dec 19-26;420(6917):860-7. Review. PubMed PMID: 12490959; PubMed Central PMCID: PMC2803035.

Crew KD, Gammon MD, Terry MB, Zhang FF, Zablotska LB, Agrawal M, Shen J, Long CM, Eng SM, Sagiv SK, Teitelbaum SL, Neugut AI, Santella RM. Polymorphisms in nucleotide excision repair genes, polycyclic aromatic hydrocarbon-DNA adducts, and breast cancer risk. Cancer Epidemiol Biomarkers Prev. 2007 Oct;16(10):2033-41. PubMed PMID: 17932351.

Cross CE, Halliwell B, Borish ET, Pryor WA, Ames BN, Saul RL, McCord JM, Harman D. Oxygen radicals and human disease. Ann Intern Med. 1987 Oct;107(4):526-45. Review. PubMed PMID: 3307585.

Dallal CM, Sullivan-Halley J, Ross RK, Wang Y, Deapen D, Horn-Ross PL, Reynolds P, Stram DO, Clarke CA, Anton-Culver H, Ziogas A, Peel D, West DW, Wright W, Bernstein L. Long-term recreational physical activity and risk of invasive and in situ breast cancer: the California teachers study. Arch Intern Med. 2007 Feb 26;167(4):408-15. PubMed PMID: 17325304.

Das UN. Metabolic syndrome X: an inflammatory condition? Curr Hypertens Rep. 2004 Feb;6(1):66-73. Review. PubMed PMID: 14972097.

Datta K, Sinha S, Chattopadhyay P. Reactive oxygen species in health and disease. Natl Med J India. 2000 Nov-Dec;13(6):304-10. Review. PubMed PMID: 11209486.

Dorgan JF, Longcope C, Stephenson HE Jr, Falk RT, Miller R, Franz C, Kahle L, Campbell WS, Tangrea JA, Schatzkin A. Relation of prediagnostic serum estrogen and androgen levels to breast cancer risk. Cancer Epidemiol Biomarkers Prev. 1996 Jul;5(7):533-9. PubMed PMID: 8827358.

Dorn J, Vena J, Brasure J, Freudenheim J, Graham S. Lifetime physical activity and breast cancer risk in pre- and postmenopausal women. Med Sci Sports Exerc. 2003 Feb;35(2):278-85. PubMed PMID: 12569217.

Duell EJ, Millikan RC, Pittman GS, Winkel S, Lunn RM, Tse CK, Eaton A, Mohrenweiser HW, Newman B, Bell DA. Polymorphisms in the DNA repair gene XRCC1 and breast cancer. Cancer Epidemiol Biomarkers Prev. 2001 Mar;10(3):217-22. PubMed PMID: 11303590.

Elsawy B, Higgins KE. Physical activity guidelines for older adults. Am Fam Physician. 2010 Jan 1;81(1):55-9. PubMed PMID: 20052963.

Eng, SM. Lifetime participation in recreational physical activity in relation to breast cancer among women on Long Island, New York. Unpublished Data. April 2002.

Evelo CT, Palmen NG, Artur Y, Janssen GM. 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. 1992;64(4):354-8. PubMed PMID: 1592062.

Eveno E, Bourre F, Quilliet X, Chevallier-Lagente O, Roza L, Eker AP, Kleijer WJ, Nikaido O, Stefanini M, Hoeijmakers JH, et al. Different removal of ultraviolet photoproducts in genetically related xeroderma pigmentosum and trichothiodystrophy diseases. Cancer Res. 1995 Oct 1;55(19):4325-32. PubMed PMID: 7671243.

Fan S, Meng Q, Gao B, Grossman J, Yadegari M, Goldberg ID, Rosen EM. Alcohol stimulates estrogen receptor signaling in human breast cancer cell lines. Cancer Res. 2000 Oct 15;60(20):5635-9. PubMed PMID: 11059753.

Fattah F, Lee EH, Weisensel N, Wang Y, Lichter N, Hendrickson EA. Ku regulates the non-homologous end joining pathway choice of DNA double-strand break repair in human somatic cells. PLoS Genet. 2010 Feb 26;6(2):e1000855. PubMed PMID: 20195511; PubMed Central PMCID: PMC2829059.

Feig DI, Reid TM, Loeb LA. Reactive oxygen species in tumorigenesis. Cancer Res. 1994 Apr 1;54(7 Suppl):1890s-1894s. Review. PubMed PMID: 8137306.

Fleisher AS, Esteller M, Wang S, Tamura G, Suzuki H, Yin J, Zou TT, Abraham JM, Kong D, Smolinski KN, Shi YQ, Rhyu MG, Powell SM, James SP, Wilson KT, Herman JG, Meltzer SJ. Hypermethylation of the hMLH1 gene promoter in human gastric cancers with microsatellite instability. Cancer Res. 1999 Mar 1;59(5):1090-5. PubMed PMID: 10070967.

Forney JP, Milewich L, Chen GT, Garlock JL, Schwarz BE, Edman CD, MacDonald PC. Aromatization of androstenedione to estrone by human adipose tissue in vitro. Correlation with adipose tissue mass, age, and endometrial neoplasia. J Clin Endocrinol Metab. 1981 Jul;53(1):192-9. PubMed PMID: 7240376.

Forsberg L, de Faire U, Morgenstern R. Oxidative stress, human genetic

variation, and disease. Arch Biochem Biophys. 2001 May 1;389(1):84-93. Review. PubMed PMID: 11370676.

Forsberg L, Lyrenäs L, de Faire U, Morgenstern R. A common functional C-T substitution polymorphism in the promoter region of the human catalase gene influences transcription factor binding, reporter gene transcription and is correlated to blood catalase levels. Free Radic Biol Med. 2001 Mar 1;30(5):500-5. PubMed PMID: 11182520.

Friedberg EC. How nucleotide excision repair protects against cancer. Nat Rev Cancer. 2001 Oct;1(1):22-33. Review. PubMed PMID: 11900249.

Friedenreich CM, Woolcott CG, McTiernan A, Ballard-Barbash R, Brant RF, Stanczyk FZ, Terry T, Boyd NF, Yaffe MJ, Irwin ML, Jones CA, Yasui Y, Campbell KL, McNeely ML, Karvinen KH, Wang Q, Courneya KS. Alberta physical activity and breast cancer prevention trial: sex hormone changes in a year-long exercise intervention among postmenopausal women. J Clin Oncol. 2010 Mar 20;28(9):1458-66. Epub 2010 Feb 16. PubMed PMID: 20159820; PubMed Central PMCID: PMC2849767.

Friedenreich CM, Cust AE. Physical activity and breast cancer risk: impact of timing, type and dose of activity and population subgroup effects. Br J Sports Med. 2008 Aug;42(8):636-47. Epub 2008 May 16. Review. PubMed PMID: 18487249.

Friedenreich CM. Physical activity and breast cancer risk: the effect of menopausal status. Exerc Sport Sci Rev. 2004 Oct;32(4):180-4. Review. PubMed PMID: 15604938.

Friedenreich CM. Review of physical activity and the IGF family. J Physi Activ Health 2004;1:291-320

Friedenreich CM, Orenstein MR. Physical activity and cancer prevention: etiologic evidence and biological mechanisms. J Nutr. 2002 Nov;132(11 Suppl):3456S-3464S. Review. PubMed PMID: 12421870.

Friedenreich CM. Review of anthropometric factors and breast cancer risk. Eur J Cancer Prev. 2001 Feb;10(1):15-32. Review. PubMed PMID: 11263588.

Friedenreich CM, Rohan TE. A review of physical activity and breast cancer. Epidemiology. 1995 May;6(3):311-7. Review. PubMed PMID: 7619942.

Fritzell JA, Narayanan L, Baker SM, Bronner CE, Andrew SE, Prolla TA, Bradley A, Jirik FR, Liskay RM, Glazer PM. Role of DNA mismatch repair in the

cytotoxicity of ionizing radiation. Cancer Res. 1997 Nov 15;57(22):5143-7. PubMed PMID: 9371516.

Frosina G. Overexpression of enzymes that repair endogenous damage to DNA. Eur J Biochem. 2000 Apr;267(8):2135-49. Review. PubMed PMID: 10759836.

Gago-Dominguez M, Castelao JE, Pike MC, Sevanian A, Haile RW. Role of lipid peroxidation in the epidemiology and prevention of breast cancer. Cancer Epidemiol Biomarkers Prev. 2005 Dec;14(12):2829-39. Review. PubMed PMID: 16364997.

Gammon MD, Eng SM, Teitelbaum SL, Britton JA, Kabat GC, Hatch M, Paykin AB, Neugut AI, Santella RM. Environmental tobacco smoke and breast cancer incidence. Environ Res. 2004 Oct;96(2):176-85. PubMed PMID: 15325878.

Gammon MD, Santella RM, Neugut AI, Eng SM, Teitelbaum SL, Paykin A, Levin B, Terry MB, Young TL, Wang LW, Wang Q, Britton JA, Wolff MS, Stellman SD, Hatch M, Kabat GC, Senie R, Garbowski G, Maffeo C, Montalvan P, Berkowitz G, Kemeny M, Citron M, Schnabel F, Schuss A, Hajdu S, Vinceguerra V. Environmental toxins and breast cancer on Long Island. I. Polycyclic aromatic hydrocarbon DNA adducts. Cancer Epidemiol Biomarkers Prev. 2002 Aug;11(8):677-85. PubMed PMID: 12163319.

Gammon MD, Neugut AI, Santella RM, Teitelbaum SL, Britton JA, Terry MB, Eng SM, Wolff MS, Stellman SD, Kabat GC, Levin B, Bradlow HL, Hatch M, Beyea J, Camann D, Trent M, Senie RT, Garbowski GC, Maffeo C, Montalvan P, Berkowitz GS, Kemeny M, Citron M, Schnabe F, Schuss A, Hajdu S, Vincguerra V, Collman GW, Obrams GI. 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. 2002 Jun;74(3):235-54. PubMed PMID: 12206514.

Gammon MD, John EM, Britton JA. Recreational and occupational physical activities and risk of breast cancer. J Natl Cancer Inst. 1998 Jan 21;90(2):100-17. Review. PubMed PMID: 9450570.

García-Closas M, Egan KM, Newcomb PA, Brinton LA, Titus-Ernstoff L, Chanock S, Welch R, Lissowska J, Peplonska B, Szeszenia-Dabrowska N, Zatonski W, Bardin-Mikolajczak A, Struewing JP. 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. 2006 May;119(4):376-88. Epub 2006 Feb 17. PubMed PMID: 16485136.

Gilliland FD, Li YF, Baumgartner K, Crumley D, Samet JM. Physical activity

and breast cancer risk in hispanic and non-hispanic white women. Am J Epidemiol. 2001 Sep 1;154(5):442-50. PubMed PMID: 11532786.

Giovannucci E. Insulin, insulin-like growth factors and colon cancer: a review of the evidence. J Nutr. 2001 Nov;131(11 Suppl):3109S-20S. Review. PubMed PMID: 11694656.

Goode EL, Ulrich CM, Potter JD. Polymorphisms in DNA repair genes and associations with cancer risk. Cancer Epidemiol Biomarkers Prev. 2002 Dec;11(12):1513-30. Review. Erratum in: Cancer Epidemiol Biomarkers Prev. 2003 Oct;12(10):1119. PubMed PMID: 12496039.

Guerra A, Rego C, Laires MJ, Castro EM, Silva D, Monteiro C, Silva Z, Lebre E, Bicho M. Lipid profile and redox status in high performance rhythmic female teenagers gymnasts. J Sports Med Phys Fitness. 2001 Dec;41(4):505-12. PubMed PMID: 11687771.

Guerra A, Rego C, Castro E, Laires MJ, Silva D, Sinde S, Monteiro C, Silva Z, Lebre E, Bicho M. LDL peroxidation in adolescent female gymnasts. Rev Port Cardiol. 2000 Nov;19(11):1129-40. English, Portuguese. PubMed PMID: 11201629.

Gunter MJ, Hoover DR, Yu H, Wassertheil-Smoller S, Rohan TE, Manson JE, Li J, Ho GY, Xue X, Anderson GL, Kaplan RC, Harris TG, Howard BV, Wylie-Rosett J, Burk RD, Strickler HD. Insulin, insulin-like growth factor-I, and risk of breast cancer in postmenopausal women. J Natl Cancer Inst. 2009 Jan 7;101(1):48-60. Epub 2008 Dec 30. PubMed PMID: 19116382; PubMed Central PMCID: PMC2639294.

Gurin CC, Federici MG, Kang L, Boyd J. Causes and consequences of microsatellite instability in endometrial carcinoma. Cancer Res. 1999 Jan 15;59(2):462-6. PubMed PMID: 9927063.

Haffner SM, Katz MS, Dunn JF. Increased upper body and overall adiposity is associated with decreased sex hormone binding globulin in postmenopausal women. Int J Obes. 1991 Jul;15(7):471-8. PubMed PMID: 1894424.

Halliwell B. The antioxidant paradox. Lancet. 2000 Apr 1;355(9210):1179-80. Review. PubMed PMID: 10791396.

Halliwell B. Oxidants and human disease: some new concepts. FASEB J. 1987 Nov;1(5):358-64. Review. PubMed PMID: 2824268.

Han J, Hankinson SE, Zhang SM, De Vivo I, Hunter DJ. Interaction between genetic variations in DNA repair genes and plasma folate on breast cancer risk. Cancer Epidemiol Biomarkers Prev. 2004 Apr;13(4):520-4. PubMed PMID: 15066914.

Han J, Hankinson SE, Ranu H, De Vivo I, Hunter DJ. Polymorphisms in DNA double-strand break repair genes and breast cancer risk in the Nurses' Health Study. Carcinogenesis. 2004 Feb;25(2):189-95. Epub 2003 Oct 24. PubMed PMID: 14578164.

Han S, Zhang HT, Wang Z, Xie Y, Tang R, Mao Y, Li Y. DNA repair gene XRCC3 polymorphisms and cancer risk: a meta-analysis of 48 case-control studies. Eur J Hum Genet. 2006 Oct;14(10):1136-44. Epub 2006 Jun 21. PubMed PMID: 16791138.

Hankinson SE, Colditz GA, Willett WC. Towards an integrated model for breast cancer etiology: the lifelong interplay of genes, lifestyle, and hormones. Breast Cancer Res. 2004;6(5):213-8. Epub 2004 Aug 6. Review. PubMed PMID: 15318928; PubMed Central PMCID: PMC549181.

Harlow SD, Matanoski GM. The association between weight, physical activity, and stress and variation in the length of the menstrual cycle. Am J Epidemiol. 1991 Jan;133(1):38-49. PubMed PMID: 1983897.

Harris NS, Winter WE. The chemical pathology of insulin resistance and the metabolic syndrome. MLO Med Lab Obs. 2004 Oct;36(10):20, 22-5. Review. PubMed PMID: 15536804.

Herman JG, Baylin SB. Promoter-region hypermethylation and gene silencing in human cancer. Curr Top Microbiol Immunol. 2000;249:35-54. Review. PubMed PMID: 10802937.

Hickman MJ, Samson LD. Role of DNA mismatch repair and p53 in signaling induction of apoptosis by alkylating agents. Proc Natl Acad Sci U S A. 1999 Sep 14;96(19):10764-9. PubMed PMID: 10485900; PubMed Central PMCID: PMC17957.

Hoffman-Goetz L, Apter D, Demark-Wahnefried W, Goran MI, McTiernan A, Reichman ME. Possible mechanisms mediating an association between physical activity and breast cancer. Cancer. 1998 Aug 1;83(3 Suppl):621-8. Review. PubMed PMID: 9690525.

Hristozov D, Gadjeva V, Vlaykova T, Dimitrov G. Evaluation of oxidative stress in patients with cancer. Arch Physiol Biochem. 2001 Oct;109(4):331-6. PubMed PMID: 11935368.

Hsieh CC, Trichopoulos D, Katsouyanni K, Yuasa S. 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. 1990 Nov 15;46(5):796-800. PubMed PMID: 2228308.

Hursting SD, Lashinger LM, Colbert LH, Rogers CJ, Wheatley KW, Nunez NP, Mahabir S, Barrett JC, Forman MR, Perkins SN. Energy balance and carcinogenesis: underlying pathways and targets for intervention. Curr Cancer Drug Targets. 2007 Aug;7(5):484-91. Review. PubMed PMID: 17691908.

Il'yasova D, Colbert LH, Harris TB, Newman AB, Bauer DC, Satterfield S, Kritchevsky SB. Circulating levels of inflammatory markers and cancer risk in the health aging and body composition cohort. Cancer Epidemiol Biomarkers Prev. 2005 Oct;14(10):2413-8. PubMed PMID: 16214925.

IARC (2000). Monographs on the evaluation of carcinogenic risks to humans. Vol. 75: Ionizing Radiation, Part 1: X- and Gamma (g)-Radiation, and Neutrons. IARC: Lyon, France

IARC Working Group. IARC handbook of cancer prevention, volume 6: weight control and physical activity. Lyon: IARC 2002

Ishii K, Zhen LX, Wang DH, Funamori Y, Ogawa K, Taketa K. Prevention of mammary tumorigenesis in acatalasemic mice by vitamin E supplementation. Jpn J Cancer Res. 1996 Jul;87(7):680-4. PubMed PMID: 8698615.

Jacobs DR Jr, Ainsworth BE, Hartman TJ, Leon AS. A simultaneous evaluation of 10 commonly used physical activity questionnaires. Med Sci Sports Exerc. 1993 Jan;25(1):81-91. PubMed PMID: 8423759.

Jakóbisiak M, Lasek W, Gołab J. Natural mechanisms protecting against cancer. Immunol Lett. 2003 Dec 15;90(2-3):103-22. Review. Erratum in: Immunol Lett. 2004 Feb 15;91(2-3):255. PubMed PMID: 14687712.

Ji LL, Stratman FW, Lardy HA. Antioxidant enzyme systems in rat liver and skeletal muscle. Influences of selenium deficiency, chronic training, and acute exercise. Arch Biochem Biophys. 1988 May 15;263(1):150-60. PubMed PMID: 3369860.

Kaaks R, Lukanova A. Energy balance and cancer: the role of insulin and insulin-like growth factor-I. Proc Nutr Soc. 2001 Feb;60(1):91-106. Review. PubMed PMID: 11310428.

Kaaks R. Nutrition, hormones, and breast cancer: is insulin the missing link? Cancer Causes Control. 1996 Nov;7(6):605-25. Review. PubMed PMID: 8932921.

Kang DH. Oxidative stress, DNA damage, and breast cancer. AACN Clin Issues. 2002 Nov;13(4):540-9. Review. PubMed PMID: 12473916.

Kanter MM. Free radicals, exercise, and antioxidant supplementation. Int J Sport Nutr. 1994 Sep;4(3):205-20. Review. PubMed PMID: 7987357.

Kanter MM, Hamlin RL, Unverferth DV, Davis HW, Merola AJ. Effect of exercise training on antioxidant enzymes and cardiotoxicity of doxorubicin. J Appl Physiol. 1985 Oct;59(4):1298-303. PubMed PMID: 4055607.

Karihtala P, Winqvist R, Syväoja JE, Kinnula VL, Soini Y. Increasing oxidative damage and loss of mismatch repair enzymes during breast carcinogenesis. Eur J Cancer. 2006 Oct;42(15):2653-9. Epub 2006 Sep 22. PubMed PMID: 16996262.

Kasapović J, Pejić S, Todorović A, Stojiljković V, Pajović SB. Antioxidant status and lipid peroxidation in the blood of breast cancer patients of different ages. Cell Biochem Funct. 2008 Aug;26(6):723-30. PubMed PMID: 18636415.

Kelsey JL, Horn-Ross PL. Breast cancer: magnitude of the problem and descriptive epidemiology. Epidemiol Rev. 1993;15(1):7-16. Review. PubMed PMID: 8405214.

Kelsey JL, Gammon MD, John EM. Reproductive factors and breast cancer. Epidemiol Rev. 1993;15(1):36-47. Review. PubMed PMID: 8405211.

Key J, Hodgson S, Omar RZ, Jensen TK, Thompson SG, Boobis AR, Davies DS, Elliott P. Meta-analysis of studies of alcohol and breast cancer with consideration of the methodological issues. Cancer Causes Control. 2006 Aug;17(6):759-70. PubMed PMID: 16783604.

Key TJ, Allen NE, Verkasalo PK, Banks E. Energy balance and cancer: the role of sex hormones. Proc Nutr Soc. 2001 Feb;60(1):81-9. Review. PubMed PMID: 11310427.

Khanna KK, Jackson SP. DNA double-strand breaks: signaling, repair and the cancer connection. Nat Genet. 2001 Mar;27(3):247-54. Review. PubMed PMID:

11242102.

Khilko N, Bourne P, Qi Yang, Ping Tang. Mismatch repair genes hMLH1 and hMSH2 may not play an essential role in breast carcinogenesis. Int J Surg Pathol. 2007 Jul;15(3):233-41. PubMed PMID: 17652529.

Kinzler KW, Vogelstein B. Lessons from hereditary colorectal cancer. Cell. 1996 Oct 18;87(2):159-70. Review. PubMed PMID: 8861899.

Klaunig JE, Kamendulis LM. The role of oxidative stress in carcinogenesis. Annu Rev Pharmacol Toxicol. 2004;44:239-67. Review. PubMed PMID: 14744246.

Klebanoff SJ. Oxygen metabolism and the toxic properties of phagocytes. Ann Intern Med. 1980 Sep;93(3):480-9. Review. PubMed PMID: 6254418.

Kolodner RD, Marsischky GT. Eukaryotic DNA mismatch repair. Curr Opin Genet Dev. 1999 Feb;9(1):89-96. Review. PubMed PMID: 10072354.

Krainer M, Silva-Arrieta S, FitzGerald MG, Shimada A, Ishioka C, Kanamaru R, MacDonald DJ, Unsal H, Finkelstein DM, Bowcock A, Isselbacher KJ, Haber DA. Differential contributions of BRCA1 and BRCA2 to early-onset breast cancer. N Engl J Med. 1997 May 15;336(20):1416-21. PubMed PMID: 9145678.

Kramer MM, Wells CL. Does physical activity reduce risk of estrogen-dependent cancer in women? Med Sci Sports Exerc. 1996 Mar;28(3):322-34. Review. PubMed PMID: 8776221.

Kumaraguruparan R, Kabalimoorthy J, Nagini S. Correlation of tissue lipid peroxidation and antioxidants with clinical stage and menopausal status in patients with adenocarcinoma of the breast. Clin Biochem. 2005 Feb;38(2):154-8. PubMed PMID: 15642278.

Kumaraguruparan R, Subapriya R, Viswanathan P, Nagini S. Tissue lipid peroxidation and antioxidant status in patients with adenocarcinoma of the breast. Clin Chim Acta. 2002 Nov;325(1-2):165-70. PubMed PMID: 12367782.

Köster F, Schröer A, Fischer D, Horn AK, Diedrich K, Friedrich M. Immunohistochemistry of proteins for DNA mismatch repair in correlation to prognostic factors of mammarian cancer. Oncol Rep. 2007 May;17(5):1223-7. PubMed PMID: 17390069.

Lambe M, Hsieh CC, Tsaih SW, Ekbom A, Trichopoulos D, Adami HO. Parity, age at first birth and the risk of carcinoma in situ of the breast. Int J Cancer. 1998 Jul 29;77(3):330-2. PubMed PMID: 9663590.

LaMonte MJ, Blair SN, Church TS. Physical activity and diabetes prevention. J Appl Physiol. 2005 Sep;99(3):1205-13. Review. PubMed PMID: 16103523.

Latikka P, Pukkala E, Vihko V. Relationship between the risk of breast cancer and physical activity. An epidemiological perspective. Sports Med. 1998 Sep;26(3):133-43. Review. PubMed PMID: 9802171.

Lee KM, Choi JY, Kang C, Kang CP, Park SK, Cho H, Cho DY, Yoo KY, Noh DY, Ahn SH, Park CG, Wei Q, Kang D. Genetic polymorphisms of selected DNA repair genes, estrogen and progesterone receptor status, and breast cancer risk. Clin Cancer Res. 2005 Jun 15;11(12):4620-6. PubMed PMID: 15958648.

Lengauer C, Kinzler KW, Vogelstein B. Genetic instabilities in human cancers. Nature. 1998 Dec 17;396(6712):643-9. Review. PubMed PMID: 9872311.

Li D, Zhang W, Sahin AA, Hittelman WN. DNA adducts in normal tissue adjacent to breast cancer: a review. Cancer Detect Prev. 1999;23(6):454-62. Review. PubMed PMID: 10571655.

Li H, Ha TC, Tai BC. XRCC1 gene polymorphisms and breast cancer risk in different populations: a meta-analysis. Breast. 2009 Jun;18(3):183-91. Epub 2009 May 14. Review. PubMed PMID: 19446452.

Li Y, Ambrosone CB, McCullough MJ, Ahn J, Stevens VL, Thun MJ, Hong CC. Oxidative stress-related genotypes, fruit and vegetable consumption and breast cancer risk. Carcinogenesis. 2009 May;30(5):777-84. Epub 2009 Mar 2. PubMed PMID: 19255063.

Lindahl T. Instability and decay of the primary structure of DNA. Nature. 1993 Apr 22;362(6422):709-15. Review. PubMed PMID: 8469282.

Liu Y, Prasad R, Beard WA, Kedar PS, Hou EW, Shock DD, Wilson SH. Coordination of steps in single-nucleotide base excision repair mediated by apurinic/apyrimidinic endonuclease 1 and DNA polymerase beta. J Biol Chem. 2007 May 4;282(18):13532-41. Epub 2007 Mar 12. PubMed PMID: 17355977; PubMed Central PMCID: PMC2366199.

Lof M, Weiderpass E. Impact of diet on breast cancer risk. Curr Opin Obstet Gynecol. 2009 Feb;21(1):80-5. Review. PubMed PMID: 19125007.

Lorincz AM, Sukumar S. Molecular links between obesity and breast cancer. Endocr Relat Cancer. 2006 Jun;13(2):279-92. Review. PubMed PMID: 16728564.

Loucks AB. Energy availability, not body fatness, regulates reproductive function in women. Exerc Sport Sci Rev. 2003 Jul;31(3):144-8. Review. PubMed PMID: 12882481.

Lu AL, Li X, Gu Y, Wright PM, Chang DY. Repair of oxidative DNA damage: mechanisms and functions. Cell Biochem Biophys. 2001;35(2):141-70. Review. PubMed PMID: 11892789.

Lynch HT, de la Chapelle A. Genetic susceptibility to non-polyposis colorectal cancer. J Med Genet. 1999 Nov;36(11):801-18. Review. PubMed PMID: 10544223; PubMed Central PMCID: PMC1734258.

Marchbanks PA, McDonald JA, Wilson HG, Folger SG, Mandel MG, Daling JR, Bernstein L, Malone KE, Ursin G, Strom BL, Norman SA, Wingo PA, Burkman RT, Berlin JA, Simon MS, Spirtas R, Weiss LK. Oral contraceptives and the risk of breast cancer. N Engl J Med. 2002 Jun 27;346(26):2025-32. PubMed PMID: 12087137.

Marnett LJ. Oxyradicals and DNA damage. Carcinogenesis. 2000 Mar;21(3):361-70. Review. PubMed PMID: 10688856.

Martin JH, Alalami O, van den Berg HW. Reduced expression of endothelial and inducible nitric oxide synthase in a human breast cancer cell line which has acquired estrogen independence. Cancer Lett. 1999 Sep 20;144(1):65-74. PubMed PMID: 10503879.

Martin KR, Barrett JC. Reactive oxygen species as double-edged swords in cellular processes: low-dose cell signaling versus high-dose toxicity. Hum Exp Toxicol. 2002 Feb;21(2):71-5. PubMed PMID: 12102499.

Matsuzawa Y, Shimomura I, Nakamura T, Keno Y, Kotani K, Tokunaga K. Pathophysiology and pathogenesis of visceral fat obesity. Obes Res. 1995 Sep;3 Suppl 2:187S-194S. Review. PubMed PMID: 8581775.

Matthews CE, Fowke JH, Dai Q, Leon Bradlow H, Jin F, Shu XO, Gao YT,

Longcope C, Hebert JR, Zheng W. Physical activity, body size, and estrogen metabolism in women. Cancer Causes Control. 2004 Jun;15(5):473-81. PubMed PMID: 15286467.

Mattusch F, Dufaux B, Heine O, Mertens I, Rost R. Reduction of the plasma concentration of C-reactive protein following nine months of endurance training. Int J Sports Med. 2000 Jan;21(1):21-4. PubMed PMID: 10683094.

Matés JM, Pérez-Gómez C, Núñez de Castro I. Antioxidant enzymes and human diseases. Clin Biochem. 1999 Nov;32(8):595-603. Review. PubMed PMID: 10638941.

McBride JM, Kraemer WJ, Triplett-McBride T, Sebastianelli W. Effect of resistance exercise on free radical production. Med Sci Sports Exerc. 1998 Jan;30(1):67-72. PubMed PMID: 9475646.

McTiernan A. Mechanisms linking physical activity with cancer. Nat Rev Cancer. 2008 Mar;8(3):205-11. Review. PubMed PMID: 18235448.

McTiernan 2005, Orenstein MR, Friedenreich CM. Review of physical activity and the IGF family. J Physi Activ Health 2004;1:291-320

McTiernan A, Ulrich C, Slate S, Potter J. Physical activity and cancer etiology: associations and mechanisms. Cancer Causes Control. 1998 Oct;9(5):487-509. Review. PubMed PMID: 9934715.

Mechanic LE, Millikan RC, Player J, de Cotret AR, Winkel S, Worley K, Heard K, Heard K, Tse CK, Keku T. Polymorphisms in nucleotide excision repair genes, smoking and breast cancer in African Americans and whites: a population-based case-control study. Carcinogenesis. 2006 Jul;27(7):1377-85. Epub 2006 Jan 6. PubMed PMID: 16399771.

Metsola K, Kataja V, Sillanpää P, Siivola P, Heikinheimo L, Eskelinen M, Kosma VM, Uusitupa M, Hirvonen A. XRCC1 and XPD genetic polymorphisms, smoking and breast cancer risk in a Finnish case-control study. Breast Cancer Res. 2005;7(6):R987-97. Epub 2005 Oct 11. PubMed PMID: 16280050; PubMed Central PMCID: PMC1410770.

Michels KB, Mohllajee AP, Roset-Bahmanyar E, Beehler GP, Moysich KB. Diet and breast cancer: a review of the prospective observational studies. Cancer. 2007 Jun 15;109(12 Suppl):2712-49. Review. PubMed PMID: 17503428.

Millikan RC, Player JS, Decotret AR, Tse CK, Keku T. Polymorphisms in DNA repair genes, medical exposure to ionizing radiation, and breast cancer risk. Cancer Epidemiol Biomarkers Prev. 2005 Oct;14(10):2326-34. PubMed PMID: 16214912.

Millikan RC, Player J, de Cotret AR, Moorman P, Pittman G, Vannappagari V, Tse CK, Keku T. 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. 2004;6(4):R264-74. Epub 2004 Apr 7. PubMed PMID: 15217492; PubMed Central PMCID: PMC468618.

Mishima Y, Kuyama A, Tada A, Takahashi K, Ishioka T, Kibata M. Relationship between serum tumor necrosis factor-alpha and insulin resistance in obese men with Type 2 diabetes mellitus. Diabetes Res Clin Pract. 2001 May;52(2):119-23. PubMed PMID: 11311966.

Mitra AK, Singh N, Singh A, Garg VK, Agarwal A, Sharma M, Chaturvedi R, Rath SK. Association of polymorphisms in base excision repair genes with the risk of breast cancer: a case-control study in North Indian women. Oncol Res. 2008;17(3):127-35. PubMed PMID: 18669164.

Mittendorf R, Longnecker MP, Newcomb PA, Dietz AT, Greenberg ER, Bogdan GF, Clapp RW, Willett WC. Strenuous physical activity in young adulthood and risk of breast cancer (United States). Cancer Causes Control. 1995 Jul;6(4):347-53. PubMed PMID: 7548722.

Miyazaki H, Oh-ishi S, Ookawara T, Kizaki T, Toshinai K, Ha S, Haga S, Ji LL, Ohno H. Strenuous endurance training in humans reduces oxidative stress following exhausting exercise. Eur J Appl Physiol. 2001 Jan-Feb;84(1-2):1-6. PubMed PMID: 11394236.

Monninkhof EM, Elias SG, Vlems FA, van der Tweel I, Schuit AJ, Voskuil DW, van Leeuwen FE; TFPAC. Physical activity and breast cancer: a systematic review. Epidemiology. 2007 Jan;18(1):137-57. Review. PubMed PMID: 17130685.

Murata H, Khattar NH, Kang Y, Gu L, Li GM. Genetic and epigenetic modification of mismatch repair genes hMSH2 and hMLH1 in sporadic breast cancer with microsatellite instability. Oncogene. 2002 Aug 22;21(37):5696-703. PubMed PMID: 12173039.

Muti P. The role of endogenous hormones in the etiology and prevention of breast cancer: the epidemiological evidence. Ann N Y Acad Sci. 2004 Dec;1028:273-82. Review. PubMed PMID: 15650252.

Nadif R, Mintz M, Jedlicka A, Bertrand JP, Kleeberger SR, Kauffmann F. Association of CAT polymorphisms with catalase activity and exposure to environmental oxidative stimuli. Free Radic Res. 2005 Dec;39(12):1345-50. PubMed PMID: 16298864; PubMed Central PMCID: PMC1877822.

Nathanson KL, Wooster R, Weber BL. Breast cancer genetics: what we know and what we need. Nat Med. 2001 May;7(5):552-6. Review. Erratum in: Nat Med 2001 Jun;7(6):749. Nathanson KN [corrected to Nathanson KL]. PubMed PMID: 11329055.

Neilson HK, Friedenreich CM, Brockton NT, Millikan RC. Physical activity and postmenopausal breast cancer: proposed biologic mechanisms and areas for future research. Cancer Epidemiol Biomarkers Prev. 2009 Jan;18(1):11-27. Review. PubMed PMID: 19124476.

Newton RU, Galvão DA. Exercise in prevention and management of cancer. Curr Treat Options Oncol. 2008 Jun;9(2-3):135-46. Epub 2008 Aug 13. Review. PubMed PMID: 18704691.

Nieman DC. Exercise immunology: practical applications. Int J Sports Med. 1997 Mar;18 Suppl 1:S91-100. Review. PubMed PMID: 9129268.

Ostrowski K, Rohde T, Asp S, Schjerling P, Pedersen BK. Pro- and anti-inflammatory cytokine balance in strenuous exercise in humans. J Physiol. 1999 Feb 15;515 (Pt 1):287-91. PubMed PMID: 9925898; PubMed Central PMCID: PMC2269132.

Pachkowski BF, Winkel S, Kubota Y, Swenberg JA, Millikan RC, Nakamura J. XRCC1 genotype and breast cancer: functional studies and epidemiologic data show interactions between XRCC1 codon 280 His and smoking. Cancer Res. 2006 Mar 1;66(5):2860-8. PubMed PMID: 16510609.

Packer L. Oxidants, antioxidant nutrients and the athlete. J Sports Sci. 1997 Jun;15(3):353-63. PubMed PMID: 9232561.

Pascucci B, Stucki M, Jónsson ZO, Dogliotti E, Hübscher U. 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. 1999 Nov 19;274(47):33696-702. PubMed PMID: 10559260.

Pate RR, Pratt M, Blair SN, Haskell WL, Macera CA, Bouchard C, Buchner D, Ettinger W, Heath GW, King AC, et al. Physical activity and public health. A

recommendation from the Centers for Disease Control and Prevention and the American College of Sports Medicine. JAMA. 1995 Feb 1;273(5):402-7. Review. PubMed PMID: 7823386.

Peters AC, Young LC, Maeda T, Tron VA, Andrew SE. Mammalian DNA mismatch repair protects cells from UVB-induced DNA damage by facilitating apoptosis and p53 activation. DNA Repair (Amst). 2003 Apr 2;2(4):427-35. PubMed PMID: 12606123.

Peters TM, Moore SC, Gierach GL, Wareham NJ, Ekelund U, Hollenbeck AR, Schatzkin A, Leitzmann MF. Intensity and timing of physical activity in relation to postmenopausal breast cancer risk: the prospective NIH-AARP diet and health study. BMC Cancer. 2009 Oct 1;9:349. PubMed PMID: 19796379; PubMed Central PMCID: PMC2768744.

Phaneuf S, Leeuwenburgh C. Apoptosis and exercise. Med Sci Sports Exerc. 2001 Mar;33(3):393-6. Review. PubMed PMID: 11252065.

Pichard C, Plu-Bureau G, Neves-E Castro M, Gompel A. Insulin resistance, obesity and breast cancer risk. Maturitas. 2008 May 20;60(1):19-30. Epub 2008 May 16. Review. PubMed PMID: 18485631.

Poplawski T, Zadrozny M, Kolacinska A, Rykala J, Morawiec Z, Blasiak J. Polymorphisms of the DNA mismatch repair gene HMSH2 in breast cancer occurence and progression. Breast Cancer Res Treat. 2005 Dec;94(3):199-204. PubMed PMID: 16252083.

Pryor WA, Stone K. Oxidants in cigarette smoke. Radicals, hydrogen peroxide, peroxynitrate, and peroxynitrite. Ann N Y Acad Sci. 1993 May 28;686:12-27; discussion 27-8. Review. PubMed PMID: 8512242.

Punnonen K, Ahotupa M, Asaishi K, Hyöty M, Kudo R, Punnonen R. Antioxidant enzyme activities and oxidative stress in human breast cancer. J Cancer Res Clin Oncol. 1994;120(6):374-7. PubMed PMID: 8138563.

Qin L, Knol MJ, Corpeleijn E, Stolk RP. Does physical activity modify the risk of obesity for type 2 diabetes: a review of epidemiological data. Eur J Epidemiol. 2010;25(1):5-12. Epub 2009 Oct 22. Review. PubMed PMID: 19847656; PubMed Central PMCID: PMC2807936.

Quick SK, Shields PG, Nie J, Platek ME, McCann SE, Hutson AD, Trevisan M, Vito D, Modali R, Lehman TA, Seddon M, Edge SB, Marian C, Muti P, Freudenheim JL. 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. 2008 May;17(5):1082-7. PubMed PMID: 18483329.

Radak Z, Taylor AW, Ohno H, Goto S. Adaptation to exercise-induced oxidative stress: from muscle to brain. Exerc Immunol Rev. 2001;7:90-107. Review. PubMed PMID: 11579750.

Radák Z, Apor P, Pucsok J, Berkes I, Ogonovszky H, Pavlik G, Nakamoto H, Goto S. Marathon running alters the DNA base excision repair in human skeletal muscle. Life Sci. 2003 Feb 21;72(14):1627-33. PubMed PMID: 12551751.

Radák Z, Naito H, Kaneko T, Tahara S, Nakamoto H, Takahashi R, Cardozo-Pelaez F, Goto S. Exercise training decreases DNA damage and increases DNA repair and resistance against oxidative stress of proteins in aged rat skeletal muscle. Pflugers Arch. 2002 Nov;445(2):273-8. Epub 2002 Sep 13. PubMed PMID: 12457248.

Radák Z, Kaneko T, Tahara S, Nakamoto H, Ohno H, Sasvári M, Nyakas C, Goto S. 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. 1999 Jul;27(1-2):69-74. PubMed PMID: 10443921.

Rajaraman P, Bhatti P, Doody MM, Simon SL, Weinstock RM, Linet MS, Rosenstein M, Stovall M, Alexander BH, Preston DL, Sigurdson AJ. Nucleotide excision repair polymorphisms may modify ionizing radiation-related breast cancer risk in US radiologic technologists. Int J Cancer. 2008 Dec 1;123(11):2713-6. PubMed PMID: 18767034.

Rajneesh CP, Manimaran A, Sasikala KR, Adaikappan P. Lipid peroxidation and antioxidant status in patients with breast cancer. Singapore Med J. 2008 Aug;49(8):640-3. PubMed PMID: 18756349.

Reardon JT, Bessho T, Kung HC, Bolton PH, Sancar A. 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. 1997 Aug 19;94(17):9463-8. PubMed PMID: 9256505; PubMed Central PMCID: PMC23224.

Robertson AB, Klungland A, Rognes T, Leiros I. DNA repair in mammalian cells: Base excision repair: the long and short of it. Cell Mol Life Sci. 2009 Mar;66(6):981-93. Review. PubMed PMID: 19153658.

Robertson JD, Maughan RJ, Duthie GG, Morrice PC. Increased blood antioxidant systems of runners in response to training load. Clin Sci (Lond). 1991 Jun;80(6):611-8. PubMed PMID: 1647925.

Rogers CJ, Colbert LH, Greiner JW, Perkins SN, Hursting SD. Physical activity and cancer prevention : pathways and targets for intervention. Sports Med. 2008;38(4):271-96. Review. PubMed PMID: 18348589.

Rose DP, Haffner SM, Baillargeon J. Adiposity, the metabolic syndrome, and breast cancer in African-American and white American women. Endocr Rev. 2007 Dec;28(7):763-77. Epub 2007 Nov 2. Review. PubMed PMID: 17981890.

Rosner B, Colditz GA, Willett WC. Reproductive risk factors in a prospective study of breast cancer: the Nurses' Health Study. Am J Epidemiol. 1994 Apr 15;139(8):819-35. PubMed PMID: 8178795.

Ross R, Janssen I, Dawson J, Kungl AM, Kuk JL, Wong SL, Nguyen-Duy TB, Lee S, Kilpatrick K, Hudson R. Exercise-induced reduction in obesity and insulin resistance in women: a randomized controlled trial. Obes Res. 2004 May;12(5):789-98. PubMed PMID: 15166299.

Rossner P Jr, Gammon MD, Terry MB, Agrawal M, Zhang FF, Teitelbaum SL, Eng SM, Gaudet MM, Neugut AI, Santella RM. Relationship between urinary 15-F2t-isoprostane and 8-oxodeoxyguanosine levels and breast cancer risk. Cancer Epidemiol Biomarkers Prev. 2006 Apr;15(4):639-44. PubMed PMID: 16614103.

Rundle A. Molecular epidemiology of physical activity and cancer. Cancer Epidemiol Biomarkers Prev. 2005 Jan;14(1):227-36. Review. PubMed PMID: 15668499.

Salvesen HB, MacDonald N, Ryan A, Iversen OE, Jacobs IJ, Akslen LA, Das S. Methylation of hMLH1 in a population-based series of endometrial carcinomas. Clin Cancer Res. 2000 Sep;6(9):3607-13. PubMed PMID: 10999752.

Sancar A, Lindsey-Boltz LA, Unsal-Kaçmaz K, Linn S. Molecular mechanisms of mammalian DNA repair and the DNA damage checkpoints. Annu Rev Biochem. 2004;73:39-85. Review. PubMed PMID: 15189136.

Sangrajrang S, Schmezer P, Burkholder I, Waas P, Boffetta P, Brennan P, Bartsch H, Wiangnon S, Popanda O. Polymorphisms in three base excision repair genes and breast cancer risk in Thai women. Breast Cancer Res Treat. 2008 Sep;111(2):279-88. Epub 2007 Oct 6. PubMed PMID: 17922186.

Sato Y, Nanri H, Ohta M, Kasai H, Ikeda M. 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. 2003 May 30;305(2):333-8. PubMed PMID: 12745079.

Schmidt ME, Chang-Claude J, Slanger T, Obi N, Flesch-Janys D, Steindorf K. Physical activity and postmenopausal breast cancer: effect modification by other breast cancer risk factors. Methods Inf Med. 2009;48(5):444-50. Epub 2009 Aug 20. PubMed PMID: 19696953.

Schmitt NM, Schmitt J, Dören M. The role of physical activity in the prevention of osteoporosis in postmenopausal women-An update. Maturitas. 2009 May 20;63(1):34-8. Epub 2009 Apr 7. Review. PubMed PMID: 19356867.

Schofield MJ, Hsieh P. DNA mismatch repair: molecular mechanisms and biological function. Annu Rev Microbiol. 2003;57:579-608. Review. PubMed PMID: 14527292.

Seitz S, Wassmuth P, Plaschke J, Schackert HK, Karsten U, Santibanez-Koref MF, Schlag PM, Scherneck S. Identification of microsatellite instability and mismatch repair gene mutations in breast cancer cell lines. Genes Chromosomes Cancer. 2003 May;37(1):29-35. PubMed PMID: 12661003.

Sesso HD, Paffenbarger RS Jr, Lee IM. Physical activity and breast cancer risk in the College Alumni Health Study (United States). Cancer Causes Control. 1998 Aug;9(4):433-9. PubMed PMID: 9794176.

Shantakumar S, Terry MB, Paykin A, Teitelbaum SL, Britton JA, Moorman PG, Kritchevsky SB, Neugut AI, Gammon MD. Age and menopausal effects of hormonal birth control and hormone replacement therapy in relation to breast cancer risk. Am J Epidemiol. 2007 May 15;165(10):1187-98. Epub 2007 Mar 3. PubMed PMID: 17337757.

Shaw JA, Walsh T, Chappell SA, Carey N, Johnson K, Walker RA. Microsatellite instability in early sporadic breast cancer. Br J Cancer. 1996 Jun;73(11):1393-7. PubMed PMID: 8645585; PubMed Central PMCID: PMC2074483.

Shen J, Terry MB, Gammon MD, Gaudet MM, Teitelbaum SL, Eng SM, Sagiv SK, Neugut AI, Santella RM. MGMT genotype modulates the associations between cigarette smoking, dietary antioxidants and breast cancer risk. Carcinogenesis. 2005 Dec;26(12):2131-7. Epub 2005 Jul 13. Erratum in: Carcinogenesis. 2007

May;28(5):1131. PubMed PMID: 16014702.

Shen J, Gammon MD, Terry MB, Wang L, Wang Q, Zhang F, Teitelbaum SL, Eng SM, Sagiv SK, Gaudet MM, Neugut AI, Santella RM. Polymorphisms in XRCC1 modify the association between polycyclic aromatic hydrocarbon-DNA adducts, cigarette smoking, dietary antioxidants, and breast cancer risk. Cancer Epidemiol Biomarkers Prev. 2005 Feb;14(2):336-42. PubMed PMID: 15734955.

Shen KL, Yang LS, Hsieh HF, Chen CJ, Yu JC, Tsai NM, Harn HJ. 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. 2000 Jun;74(2):100-7. PubMed PMID: 10914818.

Shephard RJ, Shek PN. Associations between physical activity and susceptibility to cancer: possible mechanisms. Sports Med. 1998 Nov;26(5):293-315. Review. PubMed PMID: 9858394.

Shephard RJ, Shek PN. Cancer, immune function, and physical activity. Can J Appl Physiol. 1995 Mar;20(1):1-25. Review. PubMed PMID: 7742765.

Shrubsole MJ, Gao YT, Cai Q, Shu XO, Dai Q, Hébert JR, Jin F, Zheng W. MTHFR polymorphisms, dietary folate intake, and breast cancer risk: results from the Shanghai Breast Cancer Study. Cancer Epidemiol Biomarkers Prev. 2004 Feb;13(2):190-6. PubMed PMID: 14973091.

Sia EA, Kokoska RJ, Dominska M, Greenwell P, Petes TD. Microsatellite instability in yeast: dependence on repeat unit size and DNA mismatch repair genes. Mol Cell Biol. 1997 May;17(5):2851-8. PubMed PMID: 9111357; PubMed Central PMCID: PMC232137.

Siah SP, Quinn DM, Bennett GD, Casey G, Flower RL, Suthers G, Rudzki Z. 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. 2000 Mar;60(2):135-42. Review. PubMed PMID: 10845276.

Singal PK, Dhalla AK, Hill M, Thomas TP. Endogenous antioxidant changes in the myocardium in response to acute and chronic stress conditions. Mol Cell Biochem. 1993 Dec 22;129(2):179-86. Review. PubMed PMID: 8177240.

Singh VN. A current perspective on nutrition and exercise. J Nutr. 1992

Mar;122(3 Suppl):760-5. Review. PubMed PMID: 1542045.

Singletary KW, Gapstur SM. Alcohol and breast cancer: review of epidemiologic and experimental evidence and potential mechanisms. JAMA. 2001 Nov 7;286(17):2143-51. Review. PubMed PMID: 11694156.

Singletary SE. Rating the risk factors for breast cancer. Ann Surg. 2003 Apr;237(4):474-82. Review. PubMed PMID: 12677142; PubMed Central PMCID: PMC1514477.

Smith TR, Levine EA, Freimanis RI, Akman SA, Allen GO, Hoang KN, Liu-Mares W, Hu JJ. Polygenic model of DNA repair genetic polymorphisms in human breast cancer risk. Carcinogenesis. 2008 Nov;29(11):2132-8. Epub 2008 Aug 13. PubMed PMID: 18701435; PubMed Central PMCID: PMC2722862.

Smith TR, Levine EA, Perrier ND, Miller MS, Freimanis RI, Lohman K, Case LD, Xu J, Mohrenweiser HW, Hu JJ. DNA-repair genetic polymorphisms and breast cancer risk. Cancer Epidemiol Biomarkers Prev. 2003 Nov;12(11 Pt 1):1200-4. PubMed PMID: 14652281.

Spagnoletti I, Pizzi C, Galietta A, Di Maio M, Mastranzo P, Daniele S, Limite G, Pettinato G, Contegiacomo A. Loss of hMSH2 expression in primary breast cancer with p53 alterations. Oncol Rep. 2004 Apr;11(4):845-51. PubMed PMID: 15010883.

Sprague BL, Trentham-Dietz A, Newcomb PA, Titus-Ernstoff L, Hampton JM, Egan KM. Lifetime recreational and occupational physical activity and risk of in situ and invasive breast cancer. Cancer Epidemiol Biomarkers Prev. 2007 Feb;16(2):236-43. PubMed PMID: 17301255.

Starkie R, Ostrowski SR, Jauffred S, Febbraio M, Pedersen BK. Exercise and IL-6 infusion inhibit endotoxin-induced TNF-alpha production in humans. FASEB J. 2003 May;17(8):884-6. Epub 2003 Mar 5. PubMed PMID: 12626436.

Strand M, Prolla TA, Liskay RM, Petes TD. Destabilization of tracts of simple repetitive DNA in yeast by mutations affecting DNA mismatch repair. Nature. 1993 Sep 16;365(6443):274-6. Erratum in: Nature 1994 Apr 7;368(6471);569. PubMed PMID: 8371783.

Tas F, Hansel H, Belce A, Ilvan S, Argon A, Camlica H, Topuz E. Oxidative stress in breast cancer. Med Oncol. 2005;22(1):11-5. PubMed PMID: 15750191.

Tehard B, Friedenreich CM, Oppert JM, Clavel-Chapelon F. Effect of physical activity on women at increased risk of breast cancer: results from the E3N cohort study. Cancer Epidemiol Biomarkers Prev. 2006 Jan;15(1):57-64. PubMed PMID: 16434587.

Terry MB, Gammon MD, Zhang FF, Eng SM, Sagiv SK, Paykin AB, Wang Q, Hayes S, Teitelbaum SL, Neugut AI, Santella RM. Polymorphism in the DNA repair gene XPD, polycyclic aromatic hydrocarbon-DNA adducts, cigarette smoking, and breast cancer risk. Cancer Epidemiol Biomarkers Prev. 2004 Dec;13(12):2053-8. PubMed PMID: 15598760.

Thibodeau SN, French AJ, Cunningham JM, Tester D, Burgart LJ, Roche PC, McDonnell SK, Schaid DJ, Vockley CW, Michels VV, Farr GH Jr, O'Connell MJ. Microsatellite instability in colorectal cancer: different mutator phenotypes and the principal involvement of hMLH1. Cancer Res. 1998 Apr 15;58(8):1713-8. PubMed PMID: 9563488.

Thibodeau SN, Bren G, Schaid D. Microsatellite instability in cancer of the proximal colon. Science. 1993 May 7;260(5109):816-9. PubMed PMID: 8484122.

Thune I, Furberg AS. Physical activity and cancer risk: dose-response and cancer, all sites and site-specific. Med Sci Sports Exerc. 2001 Jun;33(6 Suppl):S530-50; discussion S609-10. Review. PubMed PMID: 11427781.

Thune I, Brenn T, Lund E, Gaard M. Physical activity and the risk of breast cancer. N Engl J Med. 1997 May 1;336(18):1269-75. PubMed PMID: 9113929.

Thyagarajan B, Anderson KE, Folsom AR, Jacobs DR Jr, Lynch CF, Bargaje A, Khaliq W, Gross MD. No association between XRCC1 and XRCC3 gene polymorphisms and breast cancer risk: Iowa Women's Health Study. Cancer Detect Prev. 2006;30(4):313-21. Epub 2006 Sep 11. PubMed PMID: 16963196.

Trayhurn P, Bing C, Wood IS. Adipose tissue and adipokines--energy regulation from the human perspective. J Nutr. 2006 Jul;136(7 Suppl):1935S-1939S. Review. PubMed PMID: 16772463.

Tuteja N, Tuteja R. Unraveling DNA repair in human: molecular mechanisms and consequences of repair defect. Crit Rev Biochem Mol Biol. 2001;36(3):261-90. Review. PubMed PMID: 11450971.

Vainio H, Kaaks R, Bianchini F. Weight control and physical activity in

cancer prevention: international evaluation of the evidence. Eur J Cancer Prev. 2002 Aug;11 Suppl 2:S94-100. Review. PubMed PMID: 12570341.

van Gils CH, Peeters PH, Schoenmakers MC, Nijmeijer RM, Onland-Moret NC, van der Schouw YT, Monninkhof EM. Physical activity and endogenous sex hormone levels in postmenopausal women: a cross-sectional study in the Prospect-EPIC Cohort. Cancer Epidemiol Biomarkers Prev. 2009 Feb;18(2):377-83. Epub 2009 Feb 3. PubMed PMID: 19190137.

Vani M, Reddy GP, Reddy GR, Thyagaraju K, Reddanna P. Glutathione-S-transferase, superoxide dismutase, xanthine oxidase, catalase, glutathione peroxidase and lipid peroxidation in the liver of exercised rats. Biochem Int. 1990;21(1):17-26. PubMed PMID: 2386536.

Walsh T, Chappell SA, Shaw JA, Walker RA. Microsatellite instability in ductal carcinoma in situ of the breast. J Pathol. 1998 May;185(1):18-24. PubMed PMID: 9713355.

Wittwer M, Billeter R, Hoppeler H, Flück M. Regulatory gene expression in skeletal muscle of highly endurance-trained humans. Acta Physiol Scand. 2004 Feb;180(2):217-27. PubMed PMID: 14738480.

Wood RD, Mitchell M, Sgouros J, Lindahl T. Human DNA repair genes. Science. 2001 Feb 16;291(5507):1284-9. PubMed PMID: 11181991.

Wu J, Gu L, Wang H, Geacintov NE, Li GM. Mismatch repair processing of carcinogen-DNA adducts triggers apoptosis. Mol Cell Biol. 1999 Dec;19(12):8292-301. PubMed PMID: 10567554; PubMed Central PMCID: PMC84913.

Wu LL, Chiou CC, Chang PY, Wu JT. Urinary 8-OHdG: a marker of oxidative stress to DNA and a risk factor for cancer, atherosclerosis and diabetics. Clin Chim Acta. 2004 Jan;339(1-2):1-9. Review. PubMed PMID: 14687888.

Yager JD. Endogenous estrogens as carcinogens through metabolic activation. J Natl Cancer Inst Monogr. 2000;(27):67-73. Review. PubMed PMID: 10963620.

Yee CJ, Roodi N, Verrier CS, Parl FF. Microsatellite instability and loss of heterozygosity in breast cancer. Cancer Res. 1994 Apr 1;54(7):1641-4. PubMed PMID: 8137273.

Yu H, Rohan T. Role of the insulin-like growth factor family in cancer

development and progression. J Natl Cancer Inst. 2000 Sep 20;92(18):1472-89. Review. PubMed PMID: 10995803.

Yu Z, Chen J, Ford BN, Brackley ME, Glickman BW. Human DNA repair systems: an overview. Environ Mol Mutagen. 1999;33(1):3-20. Review. PubMed PMID: 10037319.

Zhang Y, Newcomb PA, Egan KM, Titus-Ernstoff L, Chanock S, Welch R, Brinton LA, Lissowska J, Bardin-Mikolajczak A, Peplonska B, Szeszenia-Dabrowska N, Zatonski W, Garcia-Closas M. Genetic polymorphisms in base-excision repair pathway genes and risk of breast cancer. Cancer Epidemiol Biomarkers Prev. 2006 Feb;15(2):353-8. PubMed PMID: 16492928.

Zhou XF, Cui J, DeStefano AL, Chazaro I, Farrer LA, Manolis AJ, Gavras H, Baldwin CT. Polymorphisms in the promoter region of catalase gene and essential hypertension. Dis Markers. 2005;21(1):3-7. PubMed PMID: 15735318.

2.1 STUDY OVERVIEW

The numerous physiological consequences of physical activity necessitate a more complete understanding of the mechanisms driving the inverse association. Inherited genetic variants at one or more loci may affect disease susceptibility or modify the effects of known risk factors. This study proposes to assess whether the effect of physical activity on breast cancer risk is modified by individual variability in the genetic variants of two different pathways: (1) DNA repair and (2) oxidative stress. I propose to examine three MMR (*MLH1, MSH2, MSH3*) and one oxidative stress (*CAT*) gene and their association with breast cancer risk. I will then assess physical activity-gene interactions among the aforementioned genes as well as others in the same pathways. The study aims are as follows:

AIM 1. DNA Repair

AIM 1A: to determine the main effect of select SNPs in three genes (*MLH1, MSH2, and MSH3*) of the MMR pathway on breast cancer risk.

AIM 1B: to evaluate interactions between polymorphisms in DNA repair genes (MMR, BER, and NER pathways) and self-reported lifetime physical activity on breast cancer risk (gene-environment interactions)

AIM 1C: to explore potential interactions between SNPs in genes from three DNA repair pathways (MMR, BER and NER) on breast cancer risk (gene-gene interactions).

AIM 2. Oxidative Stress

AIM 2A: to determine the main effect of select SNPs in *CAT* on breast cancer risk.

AIM 2B: to evaluate interactions between polymorphisms in oxidative stress genes and self-reported lifetime physical activity on breast cancer risk (gene-environment interactions)

AIM 2C: to explore the potential interactions between SNPs in genes from the oxidative stress pathway on breast cancer risk (gene-gene interactions).

These aims will be accomplished through the analysis of extant data from the Long Island Breast Cancer Study Project (LIBCSP); a large a population-based case-control study rich in measures of lifetime physical activity and biomarker samples (Gammon, 2002). This ancillary study will employ 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 ancillary study.

2.2 LONG ISLAND BREAST CANCER STUDY PROJECT

2.2.1 ELIGIBILITY

The Long Island Breast Cancer Study Project (LIBCSP) is a federally-funded population based study conducted among adult English-speaking female residents of Nassau and Suffolk counties, Long Island, NY. 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) (Waksberg, 1978). 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 and 1,556 (62.7%) of eligible control women (**Figure 2.1**). 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 (76 (4.1%) and 193 (7.8%)), and non-

locatable, moved out of area, or other (26 (1.4%) and 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 (FFQ) 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 (Bernstein, 1994). A recreational physical activity (RPA) screener was used to query participant's regular involvement in physical activity or exercise during any period in their life. 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 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 2.1**). 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 anticipate approximately 1053 cases and 1102 controls with blood available for genotyping (Terry, 2004). The final sample size is 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 2.1**).

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 (Gammon, 2002). 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 SUMMARY

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 (Hoffman Goetz 1998, McTiernan 1998, Rundle 2005). 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 (Rothman 1995). A general lack of biomarkers of exposure impedes the ability to establish a true causal association between physical activity and chronic disease outcomes (Rundle 2005). 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 (Rothman 1995). Correctly specifying what biomarkers influence or modify the effect of physical activity on breast cancer outcomes will aid in strengthening the argument for a causal association.

2.3.1 SNP SELECTION

To study associations between inter-individual variation of MMR, oxidative stress variants and breast cancer, I propose to examine polymorphisms in three MMR genes (*MLH1, MSH2, MSH3*) and one oxidative stress gene (*CAT*) using a candidate gene approach. These genes were chosen for this project because they play a critical role in the DNA mismatch repair process and against lipid peroxidation, respectively. I did not attempt to capture all genetic variability within

the *MLH1, MSH2, MSH3* and *CAT* genes. Rather, targeted SNP selection was informed by functional data, association studies in the breast cancer literature, and patterns of linkage disequilibrium (LD) within each gene.

Identifying functional SNPs are important because these SNPs are likely to play an essential role in gene expression and therefore cell phenotype (Mottagui-Tabar, 2005). 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 conformtion, 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).

A total of 6 SNPs in the genes of interest have been evaluated in the breast cancer literature. Four of these SNPs (rs1001179, rs1799977, rs1800743, and rs4987188) were associated with breast cancer risk in at least one study. The CAT SNP (rs1001179) was previously genotyped as apart of the LIBCSP. Both rs1799977 and rs1800734 are located in the MLH1 gene. 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 are polymorphisms that are highly correlated with other SNPs in a gene (Johnson, 2001) which, upon genotyping, can be used to infer characteristics of untyped SNPs. This method is based on the degree of LD between the tag and untyped 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 (De Bakker, 2005; Barrett, 2005) and the SNPinfo web server from the National Institute of Environmental Health Service (Zongli, 2009). For both programs tagSNPs were selected using data from phase II of the International HapMap Project database (The International HapMap Consortium, 2003). Given the racial homogeneity of the LIBCSP population with DNA available for the proposed analyses (93.4% White and 6.6% Non-White [Terry, 2004]), the CEU population (30 Utah trios with ancestry from northern and western Europe) was used as the reference panel for SNP selection**.** In addition to the CEU reference panel tagSNPs in the dbSNP European population were examined using the SNPinfo program. In two cases the European tag was selected.

Both programs use pairwise tagging methods to select a maximally informative set of common SNPs incorporating LD information based on the r^2 statistic (Carlson, 2004). 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 (Carlson, 2004). 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. 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 *MLH1, MSH2, MSH3* and *CAT.* 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 sequencing in the four genes under study (**Table 2.2**). **Figures 2.2-2.5** show the LD plots for each gene in the CEU population.

Tests of Hardy-Weinberg equilibrium (HWE) will be conducted for candidate gene SNPs to ensure assumptions of independent inheritance in the LIBCSP source population are upheld. Departures from HWE will be assessed among controls, as this group serves as a good representation for the source population (Ziegler, 2006). 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 an association between the putative at risk allele and disease (Hosking, 2004). 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: $p2 + 2pq + q2 = 1$ (Ziegler, 2006). This equation will be 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. The analysis of HWE among CAT SNPs is currently in progress.

2.3.2 ASSESSMENT OF 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 (Lin, 2005). 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 (Collins, 1997; Fallin, 2001; Zaykin, 2002; Botstein and Schaid, 2004). 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. I therefore anticipated little LD between SNPs. Using plink v1.06 I calculated pairwise LD between all geontyped SNPs and found none to be in high LD. For practical experience, however, I will examine haplotypes in the two SNPs that were most highly correlated: rs3732182 and rs4583514 (r^2 =0.634).

2.3.3 GENOTYPING METHODS AND QUALITY CONTROL

LIBCSP genotyping was conducted by Dr. Regina Santella's laboratory at Columbia University, New York, NY. All DNA samples are available on 96 well master plates. Controls for genotype at each locus and two no-DNA controls were included on each plate. Any samples that were outside the variables defined by the controls were identified as noninformative and were retested. 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. 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 7500 Real Time PCR system.

2.4 VAIABLE CONSTRUCTION AND COVARIATES

2.4.1 OVERVIEW

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 2.2.5 describes the data collection process for the recreational physical activity measurement. Confounders will be selected *a priori* using directed acyclic graphs (DAG) (Greenland, 2002) on the basis of subject matter knowledge. Final selection of confounders will be based on modeling strategies discussed in section 2.5.

2.4.2 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 *MLH1*, *MSH2, MSH3* and *CAT* 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 (Lewis, 2002). All SNP associations will 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 (Lettre, 2007). When the true model form is unknown however, the general model is more powerful than choosing any incorrectly specified model form. In the proposed study the mode of inheritance for most SNPs is not known; the general model is therefore most appropriate to estimate genotype effects. In some loci there may be too few minor allele homozygotes in the study population to employ a general model. In these instances a dominant model will be used.

For each analysis the SNPs will be modeled categorically using indicator variables. Two indicator variables will be created, one for the heterozygote genotype and another for the minor allele homozygote genotype. Estimated ORs will contrast the effect of heterozygote vs. major allele homozygote (referent group) and minor allele homozygote vs. major allele homozygote. If these analyses show that a particular locus has a specific mode of inheritance (dominant, recessive, or additive), an alternative model will be used specifying the correct genetic model.

2.4.3 EFFECT MODIFIER VARIABLE CONSTRUCTION

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

investigate. These genes will be evaluated in addition to the DNA repair and oxidative stress genes already ascertained in the Long Island study population. 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 (Ainsworth, 1993). 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 (Eng, 2002). Results from the logistic models of the three datasets generated by the different imputations were not materially different from the complete case analysis.

The lifetime recreational physical activity measure will be modeled several ways. I will initially model the variable as a categorical variable using high (greater than or equal to control median) and low (less than control median) average hours or MET-hours per week of RPA. An additional category of physical activity includes women who were classified as having no physical activity (based on the RPA screener). I will also explore other models for physical activity including categorization based on control tertiles and quartiles. Physical activity categories I will broadly classify in order to maintain reasonable cell sizes for gene*environment interactions on an additive scale. Selection of the index and referent groups for physical activity will vary by study hypothesis. Among DNA repair SNPs the primary hypothesis is that “at risk” genotypes could act synergistically with low levels of physical activity to increase the risk of breast cancer greater than would be expected by their individual effects. In these analyses the highest physical activity group will serve as the referent. Among oxidative stress SNPs I hypothesize that “at risk” genotypes could have an antagonistic effect on the benefits of high physical activity. The lowest physical activity group (no activity) will be the referent in these analyses.

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 DNA repair or oxidative stress variants within this study population (Terry, 2004; Ahn, 2004; Ahn, 2005; Shen, 2005; Ahn, 2006) these analyses are considered beyond the scope of the project, as the primary aims are to assess the interaction between *MLH1, MSH2, MSH3,* and *CAT* variants with physical activity. Similarly, menopausal status, hormone receptor status, and obesity have been shown in some studies (Friedenreich, 2001; Barida, 2006; Dallal, 2007) to modify the effects of physical activity on breast cancer outcomes. The assessment of three-way interactions will be exploratory in nature as there is both a lack of statistical power and I am primarily interested in the two-way interaction of genetic variants and recreational physical activity.

2.4.4 CONFOUNDERS

I expect minimal confounding of the association between breast cancer and *MLH1, MSH2, MSH3,* and *CAT*. Based on the directed acyclic graph (DAG) in **Figure 2.6**, a minimally sufficient adjustment set include 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 identify as Jewish in this study population (Gammon, 2002).

2.5 DATA ANALYSIS

2.5.1 OVERVIEW

The primary objective of the proposed study is to elucidate the mechanism through which physical activity may act on breast cancer risk. This will be accomplished by assessing potential interactions with both DNA repair variants of the MMR pathway (AIM 1) and oxidative stress variants in Catalase (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 interactions between variants and self-reported lifetime physical activity on breast cancer risk (**gene-environment interactions**), and (C) evaluation of 2-way interactions between SNPs within a pathway and breast cancer outcomes (**gene-gene interactions**). I will additionally estimate haplotypes effects in the MSH2 gene.

2.5.2 MAIN EFFECTS

Genotype Effects

Unconditional logistic regression models will be 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 (Kleinbaum, 2002). The binary logistic model function is: Logit $[D=1|X=x] = \alpha + \beta1X1 + \beta2X2$, where

α = model intercept
X1 = presence or absence of heterozygote genotype
X2 = presence or absence of minor allele homozygote genotype
$\beta1$ = regression coefficient corresponding to heterozygote genotype
$\beta2$ = regression coefficient corresponding to minor allele homozygote genotype
D = case (1) or control (0) status

Continuous covariates will be analyzed using indicator variables, parametric specifications (e.g., logit, linear), and flexible modeling (e.g, quadratic, higher order polynomials, or splines). The

function which best describes the shape of the data using the fewest number of parameters will be included in the final model. A covariate will be considered a confounder if it results in at least a 10% change in estimate when added to the model, compared to a model without the covariate (Greenland, 1989). Covariates that qualify as confounders will then be included in a single model to establish a fully-adjusted estimate of the association between the selected variant and breast cancer. In order to identify the most parsimonious model that will return a point estimate and corresponding 95% CI similar to that observed in the fully-adjusted model, backward elimination techniques will be employed to identify a subset of confounding variables to be selected (Agresti, 1990). Beginning with the fully-adjusted model, covariates which contribute the least to the overall fit of the model, measured by the –2 log likelihood ratio test, will be omitted (Kleinbaum, 2002). This process will be repeated until all covariates remaining in the model have a *p*-value of less than 0.05. In addition to estimating the main effect of the SNPs I will assess potential heterogeneity of effects across strata of menopausal status.

2.5.3 INTERACTIONS

A statistical interaction occurs when there is departure from additivity of effects on one chosen outcome scale (Rothman, 1998). 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 (i.e. ICR) may better reveal biological interactions (Rothman, 1998).

Gene-Environment Interaction

To assess potential modification by genetic polymorphisms on the association between physical activity and breast cancer risk I will examine both multiplicative and additive interactions. I will include a multiplicative interaction term in the regression model and assess departures from the

multiplicative null using the likelihood ratio test (LRT). 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. If there is evidence for interaction (p-values ≤ 0.05) stratified OR's will be reported. The basic logistic regression model allowing for interaction is: logit $[D=1|X=x] = \alpha + \beta1X1 + \beta2X2 + \beta3(X1)(X2)$

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 will be evaluated by the interaction contrast ratio (ICR) (Rothman, 1998). I will use indicator terms for participants with the genotype only, exposure only, and both the genotype and exposure of interest. The magnitude of an additive interaction effect between SNPs and physical activity will be 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 SE (ICR) (Assmann, 1996). ICRs less than zero indicate less than additive effects, ICRs of zero suggest no interaction on the additive scale, and ICRs greater than zero imply superadditivity (Rothman, 1998).

Gene-Gene Interactions

The possible interaction between MMR genotypes in breast cancer development will be 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. I will then collapse the variant alleles across selected genes and calculate adjusted ORs for breast cancer by stratifying on the number of 'high-risk' alleles in the pathway. Low-, intermediate-, and high-risk categories will be defined based on similar point estimates of breast cancer risk for each

number of putative high-risk alleles. All analyses will be conducted using SAS version 9.1 (Cary, NC).

2.5.4. HAPLOTYPE ANALYSIS

Haplotypes can be described as closely linked genetic markers present on one chromosome which tend to be inherited together. When simultaneously examining alleles on the same chromosome unknown phase is a problem among individuals who are heterozygous at more than one loci. Given the unavailability of family genotype information in LIBCSP haplotypes must be inferred based upon the frequency of observed genotypes in the study population. Several methods have been described for inferring haplotypes including: Clark's algorithm, the expectation-maximization (EM) algorithm, coalescence-based algorithms, and the partition-ligation method (Niu, 2004). However, these methods do not account for the uncertainty in haplotypes assignment potentially leading to biased estimates. While recent haplotypes reconstruction methods have improved on the techniques described above, there is still no gold standard. The efficacy of each method depends on the degree of ethnic diversity, number of SNPs, and number of missing genotypes (Niu, 2004). I will use SAS genetics to calculate the haplotypes frequencies among both cases and controls. Odds ratios and standard errors will be estimated using two separate programs in order to compare the usefulness of each one in the Long Island study population. The HAPLOSTAT program is based on score equations for generalized linear models (Schaid 2002). While it allows for specification of haplotype effects (additive, dominant, recessive) and can fit haplotype*covariate interactions it is limited by the assumptions of HWE and at best can be used for significance testing. The HAPSTAT program estimates the probability of a given haplotype using maximum likelihood estimation (Lin, 2005). This program is capable of effect estimation, can accommodate departures from HWE, as well as estimate gene*gene and gene*environment interactions using haplotypes (Lin, 2005). The computational efficiency of this program is limited by both the size and number of haplotypes.

This, however, should not present a problem in MSH2 given the haplotype analysis has been restricted to two loci.

2.5.4 SAMPLE SIZE AND POWER

Genotyping data is estimated to be available for 1053 cases and 1102 controls. For these analyses, the estimates for study power will vary with the genetic model selected, prevalence 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 are 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 et al. (1999).

The study will have more than adequate power to detect even modest associations for main gene effects (using an additive, dominant, or recessive model). Power is 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 2.7** shows the expected study power for varying prevalence's of the at-risk genotype and for pre and postmenopausal women combined. To examine OR effect modification (Lubin, 1990; García-Closas, 1999) of most gene*gene or gene*physical activity combinations 80% power is expected to detect substantial interactions of OR $\geq$ 4.5 and OR $\geq$ 5 multiplicative (**Figure 2.8**) and additive models (**Figure 2.9**), respectively. These calculations assume a 2-level genetic model, 3-level physical activity categorization, and an at risk genotype prevalence of at least 25%. The power decreases 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 is its large population-based design. This will aid in both generalizability of study findings and allow sufficient statistical power to examine gene-gene/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 are less evident among women under 65 years with an 88.9% and 76.1% response rate among cases and controls, respectively (Gammon, 2002).

Despite the large overall sample size 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, Human epidermal growth factor receptor 2, and Basal-like) which are known to vary by both menopausal status and race (Cary, 2006). 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 (Cary, 2006). Given the narrow range of breast cancer subtypes I will have limited 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 (Rundle, 2005). 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 (Gammon, 2002). Nevertheless, differential recall of physical activity among cases and controls could potentially bias results. In these analyses such errors are potentially superfluous because I am interested in estimating the effect of gene*environment interactions whereby there is no differential reporting of physical activity based on genotype. 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 (Bernstein, 1994; Carpenter, 1999). Similarly, the results obtained in the Long Island study data (Eng, 2002) for the main effect of physical activity among post menopausal women (27% reduction in risk) is consistent with the 25% risk reduction reported in other independent and review studies (Tehard, 2006; McTiernan, 2008).

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. In this analysis I will compare demographic and physical activity characteristics of participants with successful genotyping data to those without genotyping data to verify there are no significant differences between the two groups that could bias the results.

Small sample size is an inherent limitation of most molecular epidemiology studies chiefly because of the infrequency of some minor alleles. The LIBCSP is well powered to investigate 2-way interactions (gene*environment and gene*gene), but is underpowered to precisely estimate the effects of 3-way interactions among genes or physical activity. Such analyses would be exploratory, at best.

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. Among individuals who were unable or refused to donate samples available demographic and questionnaire data will allow me to assess whether there are any differences between them and their contributing counterparts.

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 DNA repair/oxidative stress variants, physical activity, and breast cancer. In order to correct for multiple testing I will employ the Bonferroni correction (Hochberg, 1989) and focus the analyses on estimating the main effects of genetic variants and joint effects of variants and physical activity rather than testing for statistical significance. Further, specific aims 2 and 3 will require multiple level stratification to investigate potential gene*gene and gene*environment interactions. Increasing the number of strata will invariably create small cells particularly within strata of variant alleles. I recognize that there will be less power to estimate precise effects in these strata and will use caution when interpreting these findings.

2.7 SUMMARY

The proposed study has 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. The proposed study would be the first to systematically evaluate these associations in a large population-base sample. Similarly, it has been shown across study populations 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 DNA repair and oxidative stress 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.

REFERENCES

Agresti A. Categorical Data Analysis. New York: John Wiley & Sons, 1990.

Ahn J, Gammon MD, Santella RM, Gaudet MM, Britton JA, Teitelbaum SL, Terry MB, Neugut AI, Eng SM, Zhang Y, Garza C, Ambrosone CB. Effects of glutathione S-transferase A1 (GSTA1) genotype and potential modifiers on breast cancer risk. Carcinogenesis. 2006 Sep;27(9):1876-82. Epub 2006 Apr 19. PubMed PMID: 16624829.

Ahn J, Gammon MD, Santella RM, Gaudet MM, Britton JA, Teitelbaum SL, Terry MB, Nowell S, Davis W, Garza C, Neugut AI, Ambrosone CB. Associations between breast cancer risk and the catalase genotype, fruit and vegetable consumption, and supplement use. Am J Epidemiol. 2005 Nov 15;162(10):943-52. Epub 2005 Sep 28. PubMed PMID: 16192345.

Ahn J, Gammon MD, Santella RM, Gaudet MM, Britton JA, Teitelbaum SL, Terry MB, Neugut AI, Josephy PD, Ambrosone CB. Myeloperoxidase genotype, fruit and vegetable consumption, and breast cancer risk. Cancer Res. 2004 Oct 15;64(20):7634-9. PubMed PMID: 15492293.

Ainsworth BE, Haskell WL, Leon AS, Jacobs DR Jr, Montoye HJ, Sallis JF, Paffenbarger RS Jr. Compendium of physical activities: classification of energy costs of human physical activities. Med Sci Sports Exerc. 1993 Jan;25(1):71-80. PubMed PMID: 8292105.

Assmann SF, Hosmer DW, Lemeshow S, Mundt KA. Confidence intervals for measures of interaction. Epidemiology. 1996 May;7(3):286-90. PubMed PMID: 8728443.

Bardia A, Hartmann LC, Vachon CM, Vierkant RA, Wang AH, Olson JE, Sellers TA, Cerhan JR. Recreational physical activity and risk of postmenopausal breast cancer based on hormone receptor status. Arch Intern Med. 2006 Dec 11-25;166(22):2478-83. PubMed PMID: 17159013.

Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics. 2005 Jan 15;21(2):263-5. Epub 2004 Aug 5. PubMed PMID: 15297300.

Bernstein L, Henderson BE, Hanisch R, Sullivan-Halley J, Ross RK. Physical exercise and reduced risk of breast cancer in young women. J Natl Cancer Inst. 1994 Sep 21;86(18):1403-8. PubMed PMID: 8072034.

Block G, Woods M, Potosky A, Clifford C. Validation of a self-administered diet history questionnaire using multiple diet records. J Clin Epidemiol. 1990;43(12):1327-35. PubMed PMID: 2254769.

Block G, Hartman AM, Dresser CM, Carroll MD, Gannon J, Gardner L. A data-based approach to diet questionnaire design and testing. Am J Epidemiol. 1986 Sep;124(3):453-69. PubMed PMID: 3740045.

Botstein D, Risch N. Discovering genotypes underlying human phenotypes: past successes for mendelian disease, future approaches for complex disease. Nat

Genet. 2003 Mar;33 Suppl:228-37. Review. PubMed PMID: 12610532.

Carey LA, Perou CM, Livasy CA, Dressler LG, Cowan D, Conway K, Karaca G, Troester MA, Tse CK, Edmiston S, Deming SL, Geradts J, Cheang MC, Nielsen TO, Moorman PG, Earp HS, Millikan RC. Race, breast cancer subtypes, and survival in the Carolina Breast Cancer Study. JAMA. 2006 Jun 7;295(21):2492-502. PubMed PMID: 16757721.

Carlson CS, Eberle MA, Rieder MJ, Yi Q, Kruglyak L, Nickerson DA. Selecting a maximally informative set of single-nucleotide polymorphisms for association analyses using linkage disequilibrium. Am J Hum Genet. 2004 Jan;74(1):106-20. Epub 2003 Dec 15. PubMed PMID: 14681826; PubMed Central PMCID: PMC1181897.

Carpenter CL, Ross RK, Paganini-Hill A, Bernstein L. Lifetime exercise activity and breast cancer risk among post-menopausal women. Br J Cancer. 1999 Aug;80(11):1852-8. PubMed PMID: 10468309; PubMed Central PMCID: PMC2374273.

Collins FS, Guyer MS, Charkravarti A. Variations on a theme: cataloging human DNA sequence variation. Science. 1997 Nov 28;278(5343):1580-1. PubMed PMID: 9411782.

Dallal CM, Sullivan-Halley J, Ross RK, Wang Y, Deapen D, Horn-Ross PL, Reynolds P, Stram DO, Clarke CA, Anton-Culver H, Ziogas A, Peel D, West DW, Wright W, Bernstein L. Long-term recreational physical activity and risk of invasive and in situ breast cancer: the California teachers study. Arch Intern Med. 2007 Feb 26;167(4):408-15. PubMed PMID: 17325304.

de Bakker PI, Yelensky R, Pe'er I, Gabriel SB, Daly MJ, Altshuler D. Efficiency and power in genetic association studies. Nat Genet. 2005 Nov;37(11):1217-23. Epub 2005 Oct 23. PubMed PMID: 16244653.

Eng, SM. Lifetime participation in recreational physical activity in relation to breast cancer among women on Long Island, New York. Unpublished Data. April 2002.

Fallin D, Cohen A, Essioux L, Chumakov I, Blumenfeld M, Cohen D, Schork NJ. Genetic analysis of case/control data using estimated haplotype frequencies: application to APOE locus variation and Alzheimer's disease. Genome Res. 2001 Jan;11(1):143-51. PubMed PMID: 11156623; PubMed Central PMCID: PMC311030.

Friedenreich CM, Bryant HE, Courneya KS. Case-control study of lifetime physical activity and breast cancer risk. Am J Epidemiol. 2001 Aug 15;154(4):336-47. PubMed PMID: 11495857.

Gammon MD, Neugut AI, Santella RM, Teitelbaum SL, Britton JA, Terry MB, Eng SM, Wolff MS, Stellman SD, Kabat GC, Levin B, Bradlow HL, Hatch M, Beyea J, Camann D, Trent M, Senie RT, Garbowski GC, Maffeo C, Montalvan P, Berkowitz GS, Kemeny M, Citron M, Schnabe F, Schuss A, Hajdu S, Vincguerra V, Collman GW, Obrams GI. 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. 2002 Jun;74(3):235-54. PubMed PMID: 12206514.

García-Closas M, Lubin JH. Power and sample size calculations in case-control studies of gene-environment interactions: comments on different approaches. Am J Epidemiol. 1999 Apr 15;149(8):689-92. PubMed PMID: 10206617.

Greenland S, Brumback B. An overview of relations among causal modelling methods. Int J Epidemiol. 2002 Oct;31(5):1030-7. Review. PubMed PMID: 12435780.

Greenland S. Modeling and variable selection in epidemiologic analysis. Am J Public Health. 1989 Mar;79(3):340-9. PubMed PMID: 2916724; PubMed Central PMCID: PMC1349563.

Hoffman-Goetz L, Apter D, Demark-Wahnefried W, Goran MI, McTiernan A, Reichman ME. Possible mechanisms mediating an association between physical activity and breast cancer. Cancer. 1998 Aug 1;83(3 Suppl):621-8. Review. PubMed PMID: 9690525.

Hochberg Y., Tamhane A.C. Multiple comparison procedures. Wiley, New York, USA, 1989.

Hosking L, Lumsden S, Lewis K, Yeo A, McCarthy L, Bansal A, Riley J, Purvis I, Xu CF. Detection of genotyping errors by Hardy-Weinberg equilibrium testing. Eur J Hum Genet. 2004 May;12(5):395-9. PubMed PMID: 14872201.

International HapMap Consortium. The International HapMap Project. Nature. 2003 Dec 18;426(6968):789-96. PubMed PMID: 14685227.

Johnson GC, Esposito L, Barratt BJ, Smith AN, Heward J, Di Genova G, Ueda H, Cordell HJ, Eaves IA, Dudbridge F, Twells RC, Payne F, Hughes W, Nutland S, Stevens H, Carr P, Tuomilehto-Wolf E, Tuomilehto J, Gough SC, Clayton DG, Todd JA. Haplotype tagging for the identification of common disease genes. Nat Genet. 2001 Oct;29(2):233-7. PubMed PMID: 11586306.

Kleinbaum, D.G. and M. Klein, *Logistic Regression: A self-learning text*. 2nd ed. Statistics for Biology and Health, ed. K. Dietz, et al. 2002, New York: Springer.

Lettre G, Lange C, Hirschhorn JN. Genetic model testing and statistical power in population-based association studies of quantitative traits. Genet Epidemiol. 2007 May;31(4):358-62. PubMed PMID: 17352422.

Lewis CM. Genetic association studies: design, analysis and interpretation. Brief Bioinform. 2002 Jun;3(2):146-53. Review. PubMed PMID: 12139434.

Lin DY, Zeng D, Millikan R. Maximum likelihood estimation of haplotype effects and haplotype-environment interactions in association studies. Genet Epidemiol. 2005 Dec;29(4):299-312. PubMed PMID: 16240443.

Lubin JH, Gail MH. On power and sample size for studying features of the relative odds of disease. Am J Epidemiol. 1990 Mar;131(3):552-66. PubMed PMID: 2301364.

McTiernan A. Mechanisms linking physical activity with cancer. Nat Rev

Cancer. 2008 Mar;8(3):205-11. Review. PubMed PMID: 18235448.

McTiernan A, Ulrich C, Slate S, Potter J. Physical activity and cancer etiology: associations and mechanisms. Cancer Causes Control. 1998 Oct;9(5):487-509. Review. PubMed PMID: 9934715.

Mottagui-Tabar S, Faghihi MA, Mizuno Y, Engström PG, Lenhard B, Wasserman WW, Wahlestedt C. Identification of functional SNPs in the 5-prime flanking sequences of human genes. BMC Genomics. 2005 Feb 17;6(1):18. PubMed PMID: 15717931; PubMed Central PMCID: PMC550646.

Niu T. Algorithms for inferring haplotypes. Genet Epidemiol. 2004 Dec;27(4):334-47. Review. PubMed PMID: 15368348.

Potischman N, Swanson CA, Coates RJ, Weiss HA, Brogan DR, Stanford JL, Schoenberg JB, Gammon MD, Brinton LA. Dietary relationships with early onset (under age 45) breast cancer in a case-control study in the United States: influence of chemotherapy treatment. Cancer Causes Control. 1997 Sep;8(5):713-21. PubMed PMID: 9328193.

Rothman K. Greenland S. Modern Epidemiology. 2nd Ed. Philadelphia: Maple Press, 1998.

Rothman N, Stewart WF, Schulte PA. Incorporating biomarkers into cancer epidemiology: a matrix of biomarker and study design categories. Cancer Epidemiol Biomarkers Prev. 1995 Jun;4(4):301-11. Review. PubMed PMID: 7655323.

Rundle A. Molecular epidemiology of physical activity and cancer. Cancer Epidemiol Biomarkers Prev. 2005 Jan;14(1):227-36. Review. PubMed PMID: 15668499.

Schaid DJ. Evaluating associations of haplotypes with traits. Genet Epidemiol. 2004 Dec;27(4):348-64. Review. PubMed PMID: 15543638.

Schaid DJ, Rowland CM, Tines DE, Jacobson RM, Poland GA. Score tests for association between traits and haplotypes when linkage phase is ambiguous. Am J Hum Genet. 2002 Feb;70(2):425-34. Epub 2001 Dec 27. PubMed PMID: 11791212; PubMed Central PMCID: PMC384917.

Shen J, Terry MB, Gammon MD, Gaudet MM, Teitelbaum SL, Eng SM, Sagiv SK, Neugut AI, Santella RM. MGMT genotype modulates the associations between cigarette smoking, dietary antioxidants and breast cancer risk. Carcinogenesis. 2005 Dec;26(12):2131-7. Epub 2005 Jul 13. Erratum in: Carcinogenesis. 2007 May;28(5):1131. PubMed PMID: 16014702.

Tehard B, Friedenreich CM, Oppert JM, Clavel-Chapelon F. Effect of physical activity on women at increased risk of breast cancer: results from the E3N cohort study. Cancer Epidemiol Biomarkers Prev. 2006 Jan;15(1):57-64. PubMed PMID: 16434587.

Terry MB, Gammon MD, Zhang FF, Eng SM, Sagiv SK, Paykin AB, Wang Q, Hayes S, Teitelbaum SL, Neugut AI, Santella RM. Polymorphism in the DNA repair gene XPD, polycyclic aromatic hydrocarbon-DNA adducts, cigarette smoking, and breast cancer

risk. Cancer Epidemiol Biomarkers Prev. 2004 Dec;13(12):2053-8. PubMed PMID: 15598760.

Waksberg, J., Sampling methods for random digit dialing. J Amer Statistic Assoc, 1978. 73: p. 40-46.

Zaykin DV, Westfall PH, Young SS, Karnoub MA, Wagner MJ, Ehm MG. Testing association of statistically inferred haplotypes with discrete and continuous traits in samples of unrelated individuals. Hum Hered. 2002;53(2):79-91. PubMed PMID: 12037407.

Ziegler A. and Konig I. A Statistical Approach to Genetic Epidemiology. New York: John Wiley & Sons, 2006

Zongli Xu, Jack A. Taylor (2009). SNPinfo: Integrating GWAS and Candidate Gene Information into Functional SNP Selection for Genetic Association Studies. Nucleic Acids Research http://www.niehs.nih.gov/snpinfo

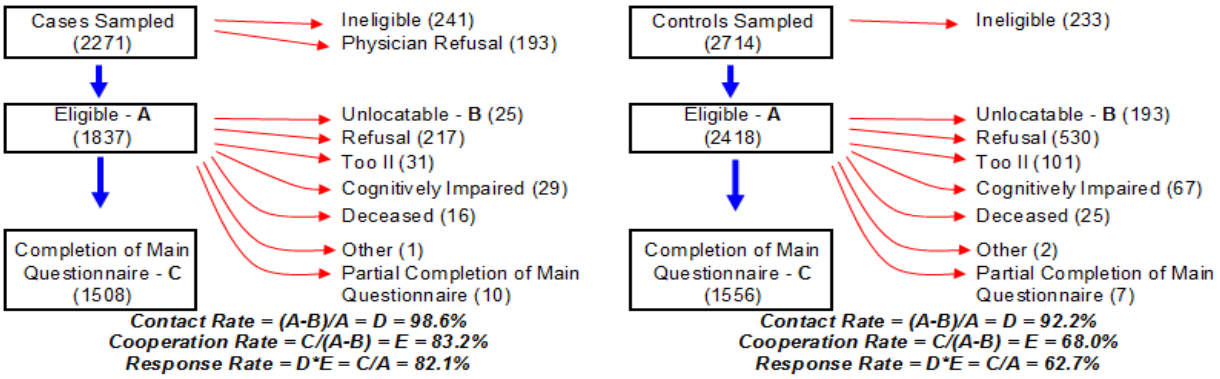

FIGURE 2.1 Contact, Cooperation, and Response Rates among case and controls, Long Island Breast Cancer Study Project. 1996-1997

Table 2.1 Response rated 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 2.2 SNP selection for the proposed study

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

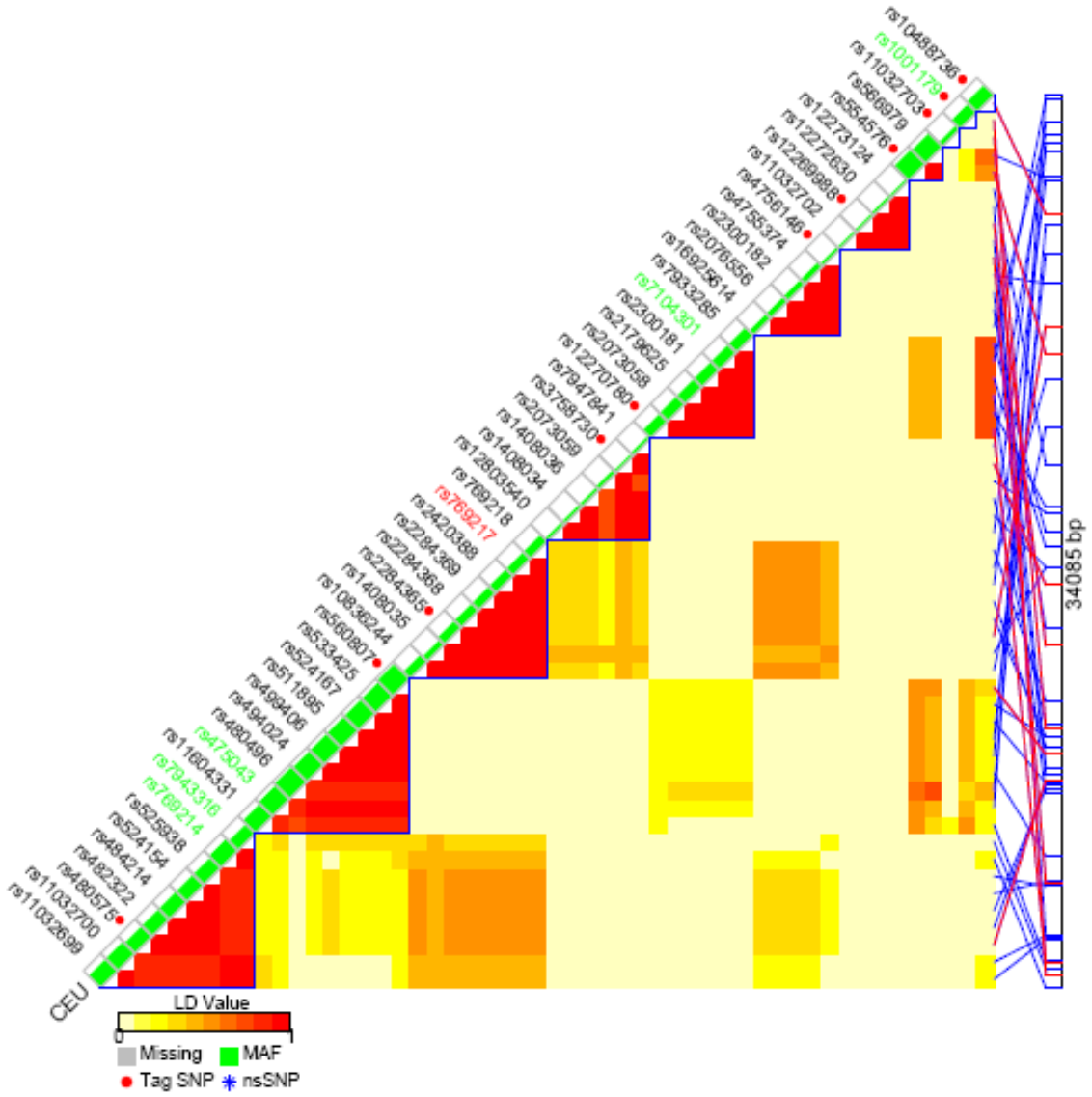

FIGURE 2.2 LD Plot CAT (CEU HapMap population)

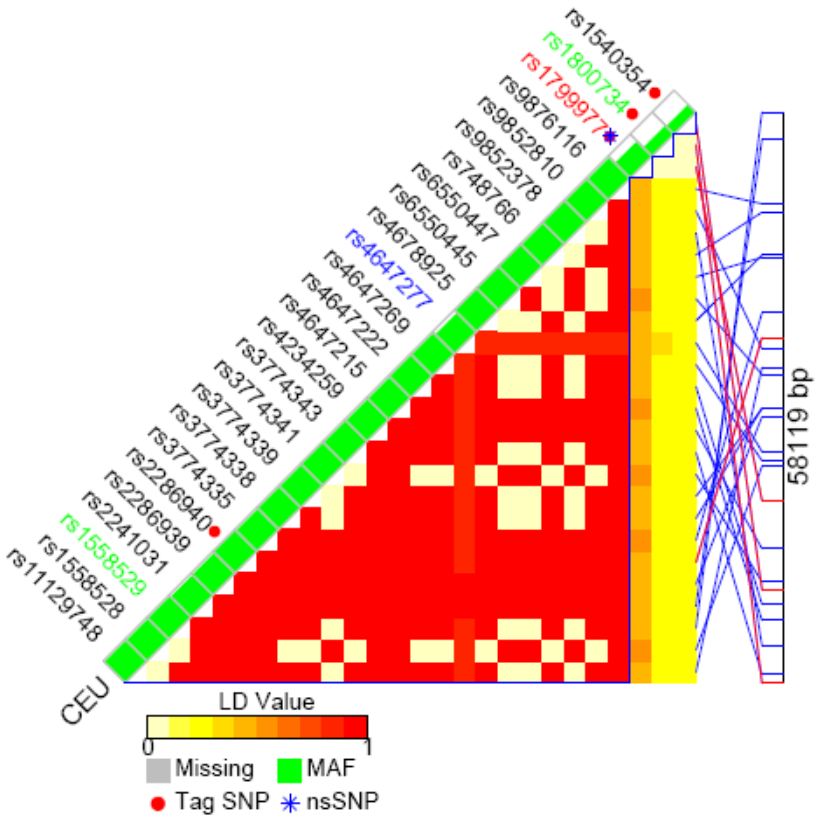

FIGURE 2.3 LD Plot MLH1 (CEU HapMap population)

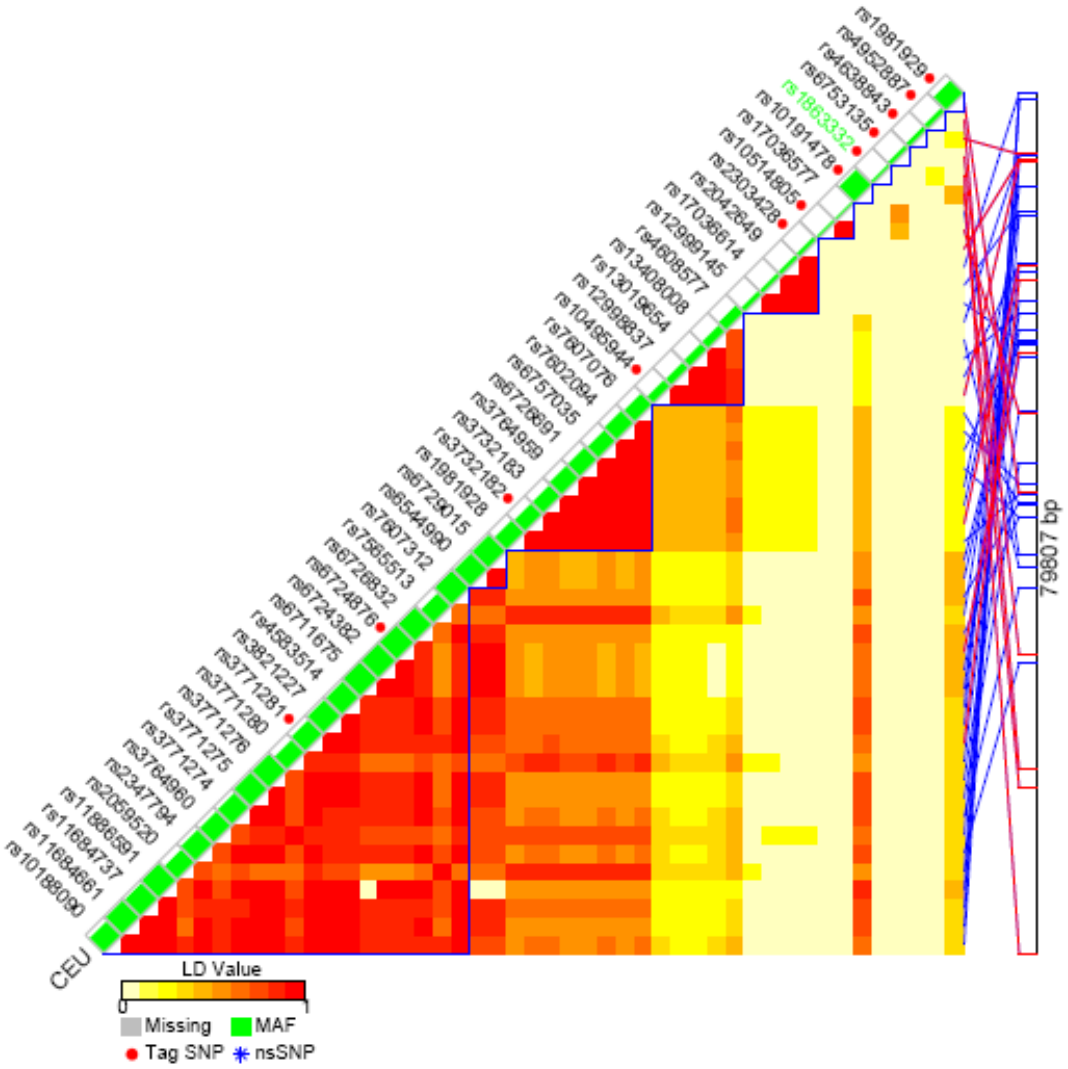

FIGURE 2.4 LD Plot MSH2 (CEU HapMap population)

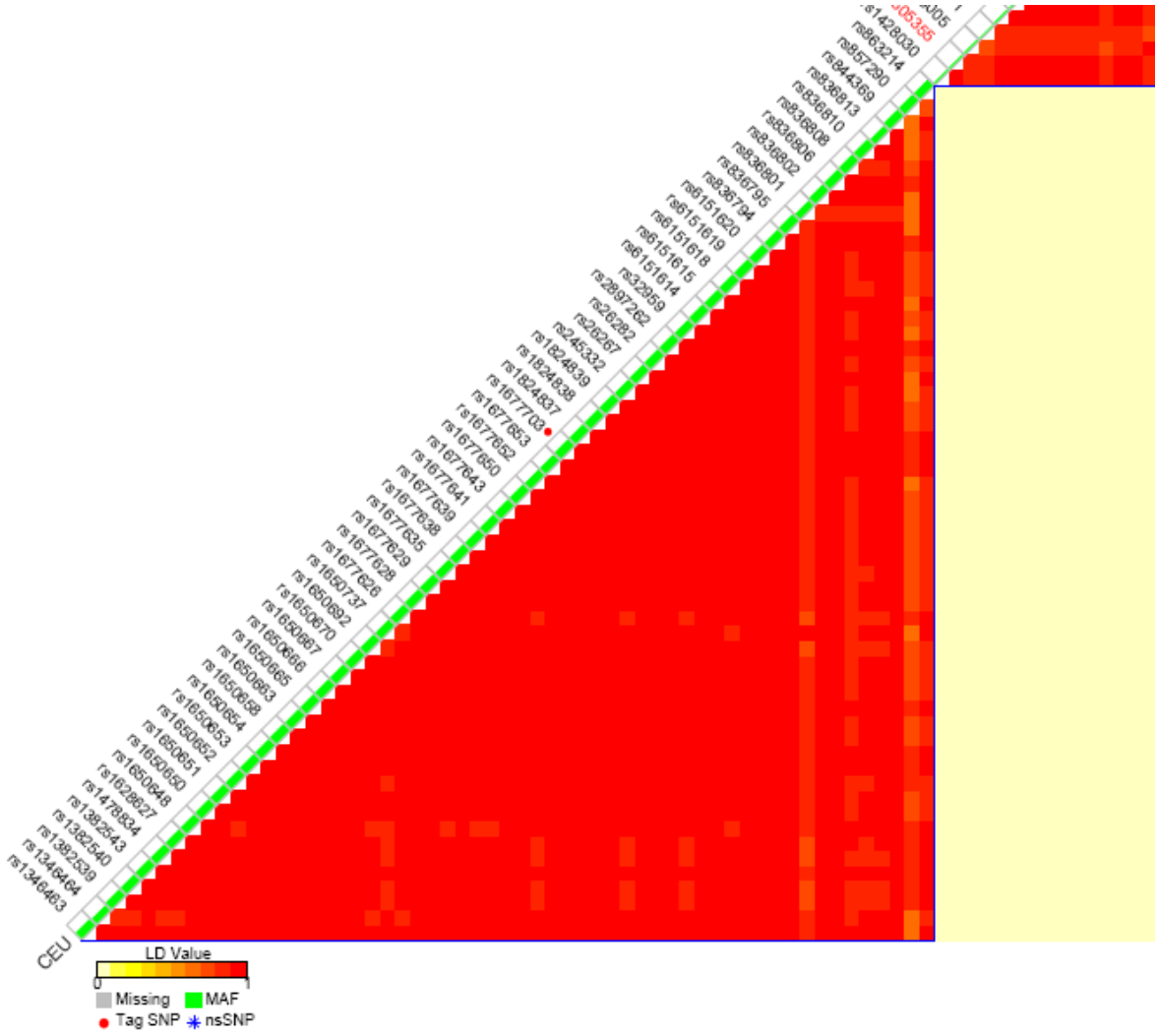

FIGURE 2.5 LD Plot MSH3 (CEU HapMap population)

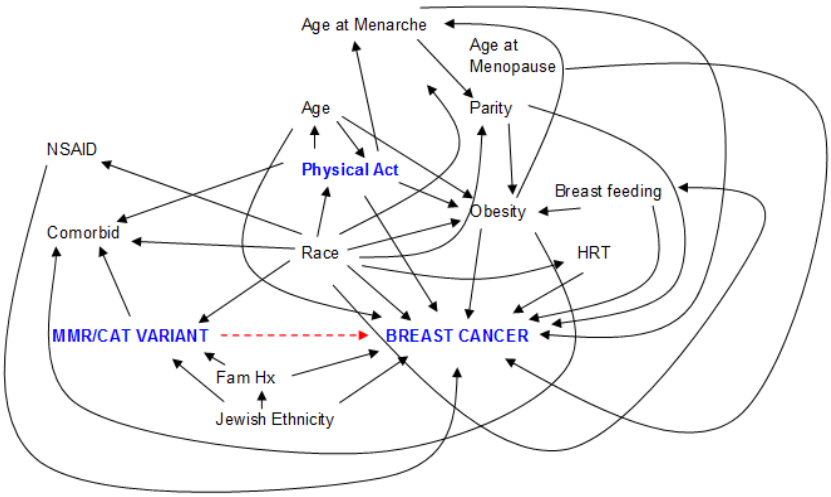

Figure 2.6 Directed acyclic graph for the association between breast cancer risk and MMR/CAT variants.

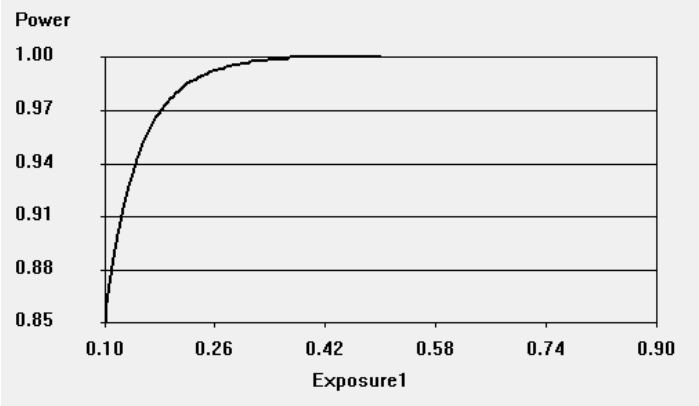

Figure 2.7 Power Curves for Main Effects of Genotype and Breast Cancer (All women Combined) – Long Island Breast Cancer Study Project, 1996-1997

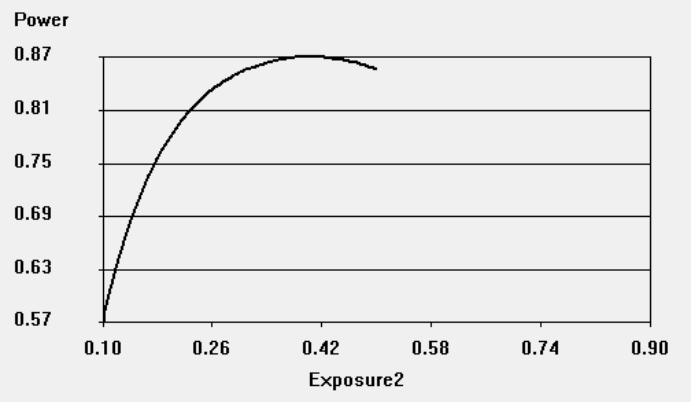

Figure 2.8 Power Curves for Multiplicative Interactions and Breast Cancer (All women Combined) – Long Island Breast Cancer Study Project, 1996-1997

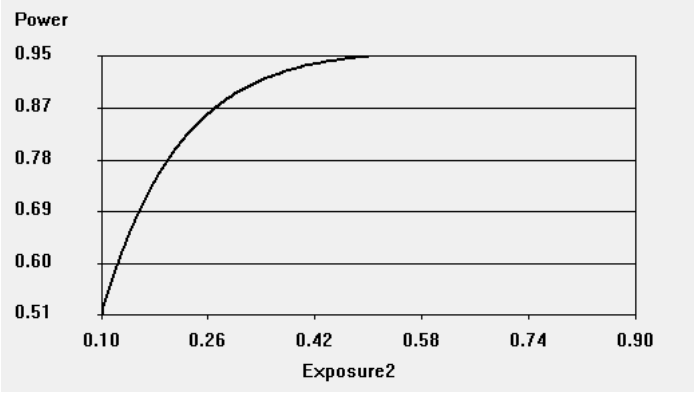

Figure 2.9 Power Curves for Additive Interactions and Breast Cancer (All women Combined) – Long Island Breast Cancer Study Project, 1996-1997