Single Nucleotide Polymorphisms and the Etiology of Basal-like and Luminal A Breast Cancer: A Pathway-Based Approach

Sarah J. Nyante

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Epidemiology.

Chapel Hill

2009

Approved by

Robert C. Millikan

Jeannette T. Bensen

Marilie D. Gammon

Jay S. Kaufman

Dan Yu Lin

Abstract

Sarah Nyante: Single Nucleotide Polymorphisms and the Etiology of Basal-like and Luminal A Breast Cancer: a Pathway-Based Approach
(Under the direction of Robert C. Millikan)

Genetic models suggest that there are breast cancer-associated genetic variants that remain uncharacterized. Heterogeneity among breast tumors may increase the difficulty of identifying these variants. The intrinsic molecular subtypes of breast cancer are associated with distinct risk factors and survival. Genetic risk factors may also differ by subtype.

312 potentially functional and tag SNPs in candidate genes related to hormone synthesis and signaling (CYP19A1, ESR1, HSD17B2, HSD3B1, PGR, SHBG) and central adiposity (ADIPOQ, LEP, LEPR, IL6, TNF) were genotyped in the Carolina Breast Cancer Study, a population-based study of African-American and white women. Genotype data was available for 1776 of 2022 controls and 1972 of 2311 cases (200 basal-like, 679 luminal A). Data from 144 ancestry informative markers was used to estimate ancestry and adjust for residual population stratification. Odds ratios (ORs) and 95% confidence intervals (CIs) for the association between genotypes and breast cancer were estimated using logistic regression. Haplotype ORs and 95% CIs were estimated using HAPSTAT.

Genotypes in LEP, LEPR, TNF, CYP19A1, ESR1, HSD3B1, HSD17B2, and PGR were associated with breast cancer overall. Genotypes in ADIPOQ, IL6, LEP, LEPR, ESR1, HSD17B2, HSD3B1, PGR, and SHBG were associated with the luminal A or basal-like subtype. Many associations were stronger when cases were stratified by subtype compared to

associations for breast cancer overall. In some cases, such as with the strongest associations in ESR1 and HSD17B2, associations were strong overall and by subtype. Haplotypes in IL6, LEP, LEPR, CYP19A1, ESR1, and PGR were associated with breast cancer overall and by subtype.

Waist-hip ratio (WHR) and combined parity and lactation were evaluated as potential effect measure modifiers. Among genotypes and haplotypes displaying evidence of multiplicative or additive interaction, genotype/haplotype associations were weaker among women with higher WHR compared to those with lower WHR. There were no clear patterns of interaction between SNPs and parity and lactation.

These results suggest that, for a subset of SNPs, SNP-breast cancer associations differ by intrinsic molecular subtype. Analyzing subtypes as distinct outcomes can increase the likelihood of identifying subtype-specific associations that may have been masked in analyses of breast cancer overall.

Acknowledgements

First, I would like to thank my dissertation committee for their help and patience with this dissertation. I would like to thank Jill Barnholtz-Sloan for her help with the ancestry calculations and discussion of the overall project. Many thanks to Yijuan Hu and Chad He for their work on HAPSCAN and HAPSTAT. I would also like to thank the UNC Mammalian Genotyping Core and the UNC Biospecimen Processing Facility for their work during the genotyping phase of this project. Thank you to the NIH Cancer Epidemiology Training Grant and the UNC Lineberger Cancer Center Cancer Control Education Program for financial support during my time in the PhD program. Finally, I would like to thank the Carolina Breast Cancer Study participants, who gave their time and more so that others could pursue this research.

Table of Contents

List of Tables

List of Figures

List of Abbreviations and Symbols

α – alpha, regression model intercept

AA – African American

AC – Afro-Caribbean

ADIPOQ – adiponectin

AIM – ancestry informative marker

ASW – individuals of African ancestry living in the southwestern US, genotyped in the International HapMap Project

ATM - ataxia telangiectasia mutated homolog

β – beta, regression model parameter estimate

BMI – body mass index

BP – base pair

BRCA1 - breast cancer 1, early onset

BRCA2 - breast cancer 2, early onset

CA - California

CBCS – Carolina Breast Cancer Study

CEU – individuals of northern and western European ancestry living in Utah, genotyped in the International HapMap Project

CHB – Han Chinese individuals, genotyped in the International HapMap Project

CEPH - Centre de'Etude du Polymorphism Humain

CI – confidence interval

CIS – carcinoma *in situ*

CLR – confidence limit ratio

CK – cytokeratin

CM - centimeter

CYP19A1 – cytochrome P450, family 19, subfamily A, polypeptide 1

DAB - 3,3'-diaminobenzidine

DAG – directed acyclic graph

DCIS – ductal carcinoma *in situ*

DF – degrees of freedom

DNA – deoxyribonucleic acid

ER – estrogen receptor alpha

ESR1 – estrogen receptor alpha

EGFR – epidermal growth factor receptor

EPIC - European Prospective Investigation into Cancer and Nutrition

FGFR2 - fibroblast growth factor receptor 2

FN – false negative

FP – false positive

GWAS – genome-wide association study

H19 - H19, imprinted maternally expressed transcript (non-protein coding)

HER2 - v-erb-b2 erythroblastic leukemia viral oncogene homolog 2

HSD3B1 – hydroxy-delta-5-steroid dehydrogenase

HSD17B2 - hydroxysteroid (17-beta) dehydrogenase 2

HWE – Hardy Weinberg Equilibrium

IGF – 1 – insulin-like growth factor 1

IGFBP3 - insulin-like growth factor binding protein 3

IHC - immunohistochemistry

IL6 – interleukin 6

JPT – Japanese individuals, genotyped in the International HapMap Project

KB - kilobase

LCIS – lobular carcinoma *in situ*

LD – linkage disequilibrium

LEP – leptin

LEPR – leptin receptor

LN – natural log

LRT – likelihood ratio test

LSP1 - lymphocyte-specific protein 1

MAF – minor allele frequency

MAP3K1 - mitogen-activated protein kinase kinase kinase 1

MLE – maximum likelihood estimation

mRNA – messenger ribonucleic acid

NAACCR – North American Association of Central Cancer Registries

NCBI - National Center for Biotechnology Information

NHANES - National Health and Nutrition Examination Survey

NT - nucleotide

OR – odds ratio

PR – progesterone receptor

PGR – progesterone receptor

R^2 – pairwise correlation coefficient

S – synergy index

SEARCH - Study of Epidemiology and Risk factors in Cancer Heredity

SEER – Surveillance, Epidemiology and End Results Program

SHBG – steroid hormone-binding globulin

SNP – single nucleotide polymorphism

SOD2 - superoxide dismutase 2, mitochondrial

TGF – transforming growth factor

TGFB1 - transforming growth factor, beta 1

TE – tris ethylenediaminetetraacetic acid

TNFA – tumor necrosis factor alpha

TNRC9 - TOX high mobility group box family member 3 (also known as TOX3)

TP53 - tumor protein p53

UNC – University of North Carolina at Chapel Hill

US – United States

UTR – untranslated region

WHR – waist to hip ratio

YRI – individuals of Yoruban ancestry living in Idaban, Nigeria, genotyped in the International HapMap Project

1. Review of the Literature

1.1 Public health burden of breast cancer

It has been several decades since the passage of the National Cancer Act (1971) and the Breast and Cervical Cancer Mortality Prevention Act (1990), but breast cancer is still a major cause of morbidity and mortality for women in the United States. As of 2006, breast cancer was the most commonly diagnosed cancer and second most common cause of cancer death for women in the US (1). Although recent data from SEER indicate that breast cancer incidence and mortality rates are decreasing, researchers estimate there will be more than 192,000 newly diagnosed cases and more than 40,000 deaths due to breast cancer in 2009 (1).

Like many cancers, the risk of breast cancer increases with age. The median age of breast cancer diagnosis in the US is 61 years old, but the incidence actually has a bi-modal distribution (2, 3). Women approximately 50 and 70 years old account for the largest proportion of breast cancer diagnoses (3). The incidence rate increases steeply with age up until about 50 years, until it starts to slow after age 50. Non-parametric models show that the incidence rate of ER-negative cancers levels off around this point, and it is mainly ER-positive cancers that increase in incidence after age 50 (4). This phenomenon was described previously by Yasui and Potter (5) in a Danish population.

Breast cancer incidence in the US differs among racial groups. During the period 2001-2005, breast cancer incidence rates were higher for white women (125.9 per 100,000) compared to African-American (111.5 per 100,000), Hispanic (91.3 per 100,000), and

Asian/Pacific Islander women (81.6 per 100,000) (6). The downward trend in the breast cancer incidence observed in data from 1996-2005 was most apparent for white women (-1.3 per 100,000); changes in breast cancer incidence rates were lower for Hispanics (-0.9 per 100,000), African Americans (-0.6 per 100,000) and Asian/Pacific Islanders (0.2 per 100,000) (6). The overall rate of breast cancer is higher in white compared to African-American women, but among those younger than 35 African Americans have higher rates, although this gap in incidence seems to have narrowed recently (7).

Breast cancer mortality rates vary by racial group as well. Breast cancer mortality is higher in African Americans (33.5 per 100,000) compared to whites (24.4 per 100,000), Asian/Pacific Islanders (12.6 per 100,000), and Hispanics (15.8 per 100,000) (6). In data collected by SEER and NAACCR registries, decreases in mortality were seen for all women during the period 1990-2002, but the largest decrease occurred in white women and among women younger than 50 years old (7). Higher mortality among African-American women may be explained in part by poorer prognostic features among African Americans. African-American women are more likely to be diagnosed with distant disease compared to whites (8). In addition, African-American women are more likely to have tumors that are larger, hormone receptor-negative, and higher grade (9-12).

Discrepancies in prognostic features and mortality between white and African-American women are unlikely to be due to differences in mammography use. The prevalence of mammography use is similar between white and African-American women (7), and the performance of mammography between the two groups has been reported to be similar (13, 14). Furthermore, controlling for stage at diagnosis, African-American women still have poorer survival compared to white women (15). Racial disparities persist when factors such

as socioeconomic status, access to healthcare, and co-morbid status are controlled for (11). An additional hypothesis is that tumor biology differs between African-American and white women. Breast tumors in African Americans are more likely to have higher mitotic activity and S-phase fraction (controlling for age and stage) (11). Younger African-American women are also more likely to have the basal-like subtype of breast cancer, which is marked by poorer prognosis and lack of targeted adjuvant treatment (16, 17).

1.2 Traditional breast cancer risk factors

Epidemiologic research has identified several characteristics that are risk factors for breast cancer. Age is one of the strongest breast cancer risk factors; the breast cancer incidence rate among women aged 85 and older is four times that of women aged 60-64 and 17 times that of women aged 40-44 (18). A family history of breast cancer is also a strong risk factor for breast cancer. One first degree relative with breast cancer results in twice the risk of breast cancer compared to no first degree relatives with breast cancer (19). The relative risk is even higher for women whose first degree relative was diagnosed at a young age (19). Increased breast cancer risk due to family history is likely due to a combination of environmental and genetic factors that are shared within the family. However, several high penetrance genetic conditions are known risk factors for breast cancer in a small proportion of the population. Germ-line mutations in BRCA1 or BRCA2 genes carry relative risk of between 10 and 30 (20). In addition, rare conditions such as Li-Fraumeni syndrome, Cowden's disease and ataxia telangiectasia greatly increase lifetime breast cancer risk (21).

Studies of perinatal and birth characteristics suggest that hormone and growth factor exposures *in utero* influence breast cancer risk later in life. Birth weight has been associated

with breast cancer in several studies, with some reporting a J-shaped association between increasing birth weight and breast cancer risk (22-24). Stratification by menopausal status shows that the association between birth weight and breast cancer exists mainly in premenopausal women and that there is little evidence for an association in postmenopausal women (25). Additionally, some studies (22, 26-29), but not all (23), have reported that maternal age, paternal age, neonatal jaundice, birth order, head size, and birth length are associated with increased breast cancer risk. Pre-eclampsia and abruptio placentae are associated with decreased breast cancer risk (22, 27). The exact mechanism for how birth characteristics interact with breast cells is unknown, but some have hypothesized that birth characteristics affect the total number of mammary gland stem cells, are markers of high intrauterine estrogen exposure levels, or are related to epigenetic alterations leading to hypermethylation or hypomethylation of key regulatory genes (25).

Studies suggest that anthropometric factors such as body weight, waist-hip ratio (WHR), height, and breast density are associated with breast cancer risk. Many studies have reported that increased body weight is associated with breast cancer risk among postmenopausal women, but not premenopausal women (30-36), though at least one study has reported a positive association between increased body weight and premenopausal breast cancer (37). Others have reported that increased body weight is inversely associated with premenopausal breast cancer (38). A non-parametric regression by van den Brandt et al. (30) showed that the breast cancer incidence rate ratio among premenopausal women increases slightly with normal and mild overweight status but then dips below the null for obese and morbidly obese women. The effect of postmenopausal HRT use on body weight and breast cancer is unclear. Studies have reported an association only among non-recent users (39),

only among current users (40), and similar associations independent of HRT status (38).

Evidence for the association between WHR and breast cancer has been less consistent than the association between weight and breast cancer. Harvie et al. (41) reviewed the literature, and reported that smaller WHR was inversely associated with breast cancer in postmenopausal women; the inverse association was seen in premenopausal women when estimates were adjusted for BMI. A meta-analysis by Connolly et al. (42) produced similar results. WHR was positively associated with breast cancer in the majority of studies of premenopausal and postmenopausal women evaluated (42). The studies closest to the null for the association between WHR and premenopausal breast cancer were ones that did not adjust for BMI (42). There were studies where WHR was associated with breast cancer without adjustment for BMI (37, 40, 43). Not all studies show a positive relationship between WHR and breast cancer. WHR was not associated with breast cancer in women 45 years old and younger, where results were adjusted for height and weight but not BMI (44). WHR was not associated with breast cancer in the EPIC study (31). Tehard et al. (38) reported an inverse association between higher WHR and breast cancer in premenopausal women and no association in postmenopausal women; BMI adjustment did not affect the results for either group.

Several studies have reported that height is positively associated with breast cancer in postmenopausal women only (34, 45, 46), but Baer et al. (47) reported that height is also associated with breast cancer in premenopausal women. The effect of height was similar for pre- and postmenopausal women in a pooled analysis by van den Brandt et al. (30). Percentage breast density is a strong risk factor for breast cancer [(48-51), reviewed by (52)], but change in breast density over time does not appear to be a risk factor for breast cancer

(53, 54).

Evidence for an association between physical activity and breast cancer is not conclusive, but much of the available evidence supports the hypothesis that physical activity is associated with a reduced risk of breast cancer. In a recent systematic review (55), approximately equal numbers of studies showed either no association or a reduced risk of breast cancer with total and leisure time physical activity. However, when the authors analyzed all studies that included a continuous measure of activity (hours per week), the pooled studies showed a 6% reduction in breast cancer risk for each hour per week of physical activity. Several of the studies reviewed also found that physical activity during adulthood was more important in reducing breast cancer risk compared to activity during childhood or adolescence (55). Variability among study results may arise from the lack of standardized methodology for assessing physical activity used in the studies (56).

Reproductive characteristics that affect endogenous estrogen exposure are associated with breast cancer. Exposure to estrogen and progesterone increases proliferation of breast epithelial cells [reviewed by (57, 58)]; the higher rate of cell division increases the chance of oncogene activation or tumor suppressor inactivation due to a replication error. Early age at menarche, later age at menopause, and a longer time interval between menarche and first full-term pregnancy are associated with increased breast cancer risk (21, 46, 59-61). Parity and lactation have a protective effect on the risk of breast cancer. The risk of breast cancer decreases with the number of children a woman has and the total number of months of breastfeeding, compared to breast cancer risk in nulliparous women (59, 62). Despite the overall long-term protective effects of pregnancy, there is also a short period of approximately 3-4 years following pregnancy where breast cancer risk is increased (63, 64).

Furthermore, what many have called the “dual effect” of parity may differ for young African-American women compared to young white women. Hall et al. (65) showed that parity is protective for white women but not for African-American women younger than 50 years old, and in Ursin et al. (66) parity had a greater protective effect for white women compared to African-American women younger than 50. In comparisons of breast cancer risk factors in white and African-American women, young black women were more likely to be parous but less likely to have breastfed (65-67). Based on the known post-birth increase in breast cancer risk and the lower rate of breastfeeding in African Americans, Pathak and colleagues (68, 69) have hypothesized that it should be expected that young black women experience higher breast cancer incidence compared to young white women.

It has not yet been shown to what extent the differences in reproductive patterns between black and white women account for differences in breast cancer incidence. Increased susceptibility to hormone-induced proliferation of undifferentiated mammary cells (70) and changes to the extracellular matrix (71) have been suggested as biological mechanisms by which pregnancy might increase the short-term risk of breast cancer. A potential explanation for the protective effect of breastfeeding is that lactation induces terminal differentiation of the mammary gland; this differentiation may protect mammary cells from elevated post-pregnancy hormone levels.

Studies suggest that characteristics of pregnancy are associated with increased breast cancer risk. Higher placental weight is associated with increased breast cancer risk (72). Early gestational age and placental abruption are associated with increased risk of breast cancer (73), though another study (72) did not find an association between gestational age and breast cancer. A multiple birth is associated with increased breast cancer rates within 5

years of the pregnancy (73, 74). Innes et al. (73) reported that pre-eclampsia was associated with reduced risk of breast cancer, especially among women age 30 and older.

Exogenous hormone use has also been linked to an increased risk of breast cancer. Several research groups have reported that oral contraceptive use is associated with a small increased risk of breast cancer in young women (75-78), though other studies have reported no association between oral contraceptive use and breast cancer in younger women (79). Subgroup analyses showed that the effects of oral contraceptive use are strongest among women with more recent use and a longer duration of use (76, 80, 81). Similarly, current and recent use of hormone replacement therapy (HRT) in postmenopausal women is associated with increased breast cancer risk (40, 82-84). Studies have found that the effect of combination (formulations containing estrogen and progesterone) hormone therapy is greater than the effect of estrogen only HRT (40, 82, 83). The pooled analysis by the Collaborative Group on Hormonal Factors in Breast Cancer (84) did not find a difference in effect for combination compared to estrogen only HRT, but it is possible that this result is due to selection bias. Information on type of HRT was available for only 39% of the women in Collaborative Group analysis.

Alcohol use is associated with increased risk of breast cancer [reviewed in (85, 86)]. The relative risk of breast cancer risk increases steadily with number of drinks consumed per day (86). In some studies (87, 88) the effect of alcohol consumption appeared to be stronger in women with BMI < 25 kg/m^2, but there was no difference in effect by BMI in a Collaborative Group meta-analysis (86).

Exposure to light at night and night shift work are associated with an increased risk of breast cancer (89-91), although at least one study reported an inverse association between

overnight shifts and breast cancer (92). Researchers hypothesize that lack of sleep during nighttime melatonin production hours inhibits melatonin production by the body (90, 93, 94). Melatonin inhibits tumor formation in rodents, and may act by regulating reproductive hormone production, cell cycle control, or p53 expression (90, 94).

Data from Japanese atomic bomb survivors and women exposed to radiation for medical treatment shows that ionizing radiation is associated with breast cancer incidence [reviewed by (95)]. The excess risk of breast cancer is proportional to the radiation dose received, though the shape of the dose-response curve differs between populations (95). Risk of radiation-associated breast cancer is modified by other breast cancer risk factors. Young age, nulliparity, low parity, and lack of breastfeeding have been associated with increased risk of breast cancer among those exposed to radiation (95).

1.3 Common genetic variation and breast cancer

The idea that breast cancer can be inherited is based on the fact that relatives of breast cancer patients are more likely to be diagnosed with breast cancer themselves. A woman's risk of breast cancer is increased if she has a first-degree relative diagnosed with breast cancer compared to women with no affected first-degree relatives; the relative risk varies inversely with the affected relative's age at diagnosis and the age of the woman at risk (19). Monozygotic twins of breast cancer cases have higher incidence of breast cancer compared to the mothers, sisters, or dizygotic twins of breast cancer cases, indicating that increased genetic similarity may be related to increased susceptibility to breast cancer (96-98).

There are two main theories that attempt to explain the genetic model underlying familial breast cancer cases. The first theory proposes that hereditary breast cancer is caused

by rare, highly penetrant alleles. In 1988 researchers identified a susceptibility locus, eventually named Breast Cancer 1 Gene (BRCA1), that explained the clustering of breast cancer cases in high risk families under an autosomal dominant model of inheritance (99). A second susceptibility allele that was strongly associated with hereditary cases not linked to BRCA1 was identified by Wooster et al. (100), and named BRCA2. The cumulative risk of breast cancer by age 70 is estimated to be 65% in BRCA1 mutation carriers and 45% in BRCA2 mutation carriers (20). Germline mutations in BRCA1 and BRCA2 are rare and only account for approximately 25% of familial breast cancer cases and less than 5% of breast cancer cases in the general population (101, 102). Identification of BRCA1 and BRCA2 demonstrates that there is a strong, heritable genetic component in some proportion of breast cancers. However, the fact that not all cases of familial patterns of inherited breast cancer can be attributed to known breast cancer genes suggests that there are susceptibility genes that have not yet been identified. Some research groups continue to search for additional high penetrance genes in non-BRCA1 or BRCA2 families (103).

The second theory is the polygenic model, which provided the best model fit (along with a recessive model) in a series of segregation analysis that compared several models of inheritance, accounting for the effects of BRCA1 and BRCA2 mutations and parity (104-106). Under the polygenic model, disease susceptibility is related to variation in multiple genes instead of a single, highly penetrant allele. Each individual disease allele is associated with a small increased risk of breast cancer, and an individual's risk of cancer increases with the number of disease alleles they carry. Each disease gene has only a small effect on overall risk, and is therefore low penetrance, but the relatively high prevalence of the disease allele in the population makes it a risk factor in a large number of cancer cases.

Research seems to support both theories. Several rare, highly penetrant alleles have been identified (mutations in BRCA1, BRCA2, TP53, and ATM); together, mutations in these genes account for as much as 50% of breast cancer associated with family history (102). On the other hand, recent genome-wide association studies have identified several low penetrance alleles associated with small increases in risk of several chronic diseases, including prostate, colorectal, lung, and breast cancers (107-119). In one of the first genome-wide multi-stage studies, Easton et al. (107) identified single nucleotide polymorphisms in FGFR2, TNRC9, MAP3K1, and LSP1 that were significantly associated with breast cancer. FGFR2 breast cancer-associated SNPs identified in a genome-wide association study (GWAS) of common genetic variants in postmenopausal white women and breast cancer were reported around the same time by Hunter et al. (108). In a third study, SNPs on 2q35 and 16q12 were reported to be positively associated with breast cancer in European populations (109). Interestingly, the SNP on 16q12 is in the same linkage block as part of the gene TNRC9, which was reported by Easton et al. (107). GWA analysis in the Shanghai Breast Cancer Study replicated associations in SNPs in LSP1, TNRC9, and FGFR2, and reported a novel association between breast cancer and SNPs located at 6q25.1 (117). Additional breast cancer-associated SNPs have been reported on 1p11.2, 3p24, 14q24.1 and 17q23.2 (118, 119).

Most breast cancer-associated GWAS SNPs have been identified in populations of European descent, and the associations have not been consistently replicated in women of African descent. The 16q12 SNP reported by Stacey et al. (109) had the opposite association in African Americans compared to Europeans. In another study, Stacey et al. (116) reported that SNPs on 5p12 were associated with ER-positive breast cancer in women of European

ancestry. When the association was tested in a Nigerian population only one of the two main SNPs was marginally associated with breast cancer, and neither SNP was associated with breast cancer in African Americans (116). Zheng et al. (120) examined 9 GWAS SNPs that were first reported in studies of European and Chinese populations, and reported that only 2 of these SNPs (rs13387042 in 2q35 and rs1219648 in FGFR2) were also associated with breast cancer in African-American women.

High risk genetic variants do not act in a deterministic manner, even for high penetrance alleles like mutations in BRCA1 and BRCA2. The distribution of breast cancer risk varies in the population (105), suggesting that there are gene-gene or gene-environment interactions that affect overall breast cancer risk. By definition, the polygenic model states that the risk conferred by multiple alleles is multiplicative. In addition, Antoniou et al. (106) found evidence for statistical interaction between the effects of unknown variants and BRCA1 and BRCA2 under the polygenic model. Many biological systems are redundant so, in some cases, interaction between multiple variants in the same biologic pathway may be required before there is an effect on breast cancer risk.

Most researchers acknowledge that the causes of cancer are not only genetic, and that non-genetic factors play a role in heritable and spontaneous breast cancer. Dizygotic twins of breast cancer cases have an elevated breast cancer risk compared to mothers and sisters of breast cancer cases, suggesting that shared environment contributes to increased breast cancer risk (97). Several studies have shown evidence of gene-environment interaction in breast cancer risk (121-125). Antoniou et al. (20) reported that the relative risk of BRCA-associated breast cancer differs by birth cohort, though it is possible that changes in screening and diagnostic patterns contributed to the observed changes in addition to any

changes in the prevalence of interacting risk factors.

1.4 Single nucleotide polymorphisms and linkage disequilibrium structure of the human genome

Single nucleotide polymorphisms (SNPs) are one type of genetic variation that may influence cancer risk in a polygenic manner. There are several ways that SNPs could affect cancer risk by disrupting the normal function of a gene and associated biological pathways. Nonsynonymous SNPs in coding regions can alter the translated amino acid, potentially impairing or destroying its function (126). Synonymous changes, or SNPs that result in the same translated amino acid, can also affect protein form and/or function by affecting folding ability, translation kinetics, or splicing (126, 127). Changes in non-coding gene sequences, such as promoters, response element binding sites, and introns, can affect gene function by changing the affinity of cis- and trans-binding sites or creating erroneous stop codons.

When details of the first draft sequence of the human genome were published in 2001, researchers mapped between 1.4 and 2.1 million SNPs across the genome (128-130), but Frazer et al. [(131)(supplementary data)] estimated that there may be as many as 9 to 10 million common SNPs. The International HapMap Project is a multi-national collaboration whose goal is to genotype common SNPs from ethnically and geographically diverse populations (131, 132). The most recent data from Phase 3 of the HapMap Project includes genotypes from African American, Chinese American, Indian American, Kenyan, Mexican American, and Italian populations in addition to the Chinese, Japanese, Yoruba, and European American populations genotyped in Phases 1 and 2. Coverage and LD characteristics of Phase 3 data have not been reported. This literature review focuses on

HapMap data from Phases 1 and 2 only. The International HapMap Project has characterized more than 3 million SNPs in individuals from four different populations – Americans of northern and western European descent (CEU), Japanese in Tokyo, Japan (JPT), Han Chinese in Beijing, China (CHB), and Nigerians of Yoruban descent (YRI population) (131). Phase 2 of the HapMap Project has genotyped approximately 1 SNP per kilobase (kb), although the SNPs are not evenly spaced (131). Phase 1 and 2 HapMap SNPs are estimated to capture most common untyped SNPs with a mean correlation coefficient (r^2) between 0.90 and 0.96, meaning that nearly all common untyped variants are in strong linkage disequilibrium (LD), or frequently co-inherited with, with a typed HapMap variant. This coverage is lower for SNPs with a low minor allele frequency (131), and therefore will be lower for nonsynonymous SNPs, the majority of which have a minor allele frequency of 0.05 or lower (133). According to Barrett and Cardon (133), nearly all common variation in the genome can be covered using a panel of 500,000 tag SNPs in European (CEU) and Asian (CHB and JPT) populations and more than 1,000,000 tags SNPs in western African (YRI) populations.

Dense mapping of SNPs across the human genome has allowed researchers to investigate genomic structure in great detail. Reich et al. (134) analyzed LD across the genome in Europeans and Nigerians. Linkage disequilibrium is the correlation between alleles at two or more loci, and is representative of the alleles originating from a common, ancestral chromosome (134). According to Reich et al. (134), LD spanned longer regions of the genome than previously predicted, but the size of LD blocks varied between gene regions and ethnic populations. For example, the average distance between SNPs in LD was 60 kb in Europeans but less than 5 kb in Nigerians (134). The authors also stated that the LD pattern

in the Nigerian population is a subset of the LD pattern seen in Europeans, and difference between the two may be due to bottlenecks or founder effects after the populations separated (134). In another analysis of LD structure in African, African-American, European, and Asian subjects, Gabriel et al. (135) reported that the distance between SNPs in LD is shorter in African and African-American populations as compared to European and Asian populations. Gabriel et al. (135) observed only between 3 and 5 common haplotypes per block in all populations and that 6 to 8 common markers could sufficiently identify these haplotypes. Similar to Reich et al. (134), the majority of the common haplotypes were shared among all populations (135). Daly and colleagues (136) conducted an in-depth analysis of haplotype structure on 5q31 in a Canadian population of European descent, and found that blocks had only a few common haplotypes that accounted for most of the chromosomal sequences. Daly et al. (136) also showed that from a given locus, LD declines with increasing distance from the locus, and that the drops in LD are abrupt, suggesting 'significant' historical recombination.

Voight et al. (137) reported evidence for recent positive selection throughout regions of the genome, and these selection signals were clustered in or near coding regions. Selection signals were more common in the YRI and CEU population data compared to that expected from simulations. The authors estimated that selection processes took place relatively recently, after the populations separated. In support of this, their data show that some but not all of the changes are shared, and that the types of genes exhibiting selection belong to different functional classes by population. Sabeti et al. (138) show evidence that positive selection occurs in genes belonging to the same biological pathway.

The genomic structure among different ethnic populations is similar, but data from the

HapMap Project highlights potential technical difficulties and sources of bias that may occur when genotyping multiple populations within a single study. LD blocks are smaller in the YRI population (134, 135). During HapMap Phase 2 genotyping, the YRI population was more likely to have untaggable SNPs, and requires substantially more tag SNPs to cover all common SNPs across the genome (131). Therefore, genotyping failures are more likely to result in a complete loss of data for a particular region in studies of African and African-American populations compared to European populations.

There have been numerous analyses of the association between SNPs and breast cancer [reviewed in (139, 140)]. The Breast Cancer Association Consortium pooled data from 12 different studies and estimated associations with borderline statistical significance for SNPs in caspase 8, TGFB1, IGFBP3, PGR, and SOD2 (141). Pharoah et al. (142) studied 710 common SNPs in 8800 subjects, and reported that SNPs in genes related to steroid hormone signaling and metabolism and cell cycle control were significantly associated with breast cancer. Several studies have also identified polymorphisms associated with breast cancer in genome-wide association studies of SNPs and breast cancer (107-119). Despite the progress made by these studies, it is likely that they have identified just some of the common genetic variants relevant to breast cancer in the US.

1.5 Tumor classification and intrinsic subtypes of breast cancer

Breast cancer subtypes are relevant to prognosis. Standard clinical practice guidelines call for measurement of ER, PR, and HER2 expression in all primary invasive breast tumors to determine treatment course (143). Hormone receptor expression is predictive of response to endocrine therapies, such as selective estrogen receptor modulators, aromatase inhibitors,

and ovarian ablation (143). HER2 overexpression or amplification is an important prognostic marker and predictive of response to several treatments, including chemotherapy, endocrine therapy, trastuzamab, and lapatinib (143). Routine use of ER, PR, and HER2 status to determine treatment is directly related to the term “triple-negative,” a tumor that does not express ER, PR or HER2 and is therefore not a candidate for endocrine or anti-HER2 therapy. Invasive breast cancer patients with the triple-negative phenotype are more likely to be younger and African American or Hispanic (12, 144). Compared to tumors expressing ER, PR or HER2, triple-negative tumors also have poorer prognostic features such as larger size, higher grade, poor differentiation, and lower survival (12, 144-146).

Perou et al. (147) undertook a more detailed classification of breast tumors by measuring the expression of more than 1700 genes using cDNA microarrays. Hierarchical clustering revealed four major clusters, or intrinsic subtypes, of breast tumors. The ER+/luminal epithelial-like tumors were characterized by expression of genes commonly expressed in luminal epithelial cells like estrogen receptor, cytokeratins 8 and 18, GATA-binding protein 3, and hepatocyte nuclear factor 3-alpha (also known as FOXA1). The basal-like group of tumors was characterized by expression of cytokeratins 5 and 17, integrin beta 4, laminin, and the absence of estrogen receptor expression. The ErbB2 group was characterized by high HER2 levels and lower levels of the estrogen receptor and ER-associated genes. The remaining tumors clustered with normal breast samples, and expressed higher levels of basal epithelial and adipose-related genes and lower levels of luminal/ER+ associated genes. When clustering was repeated using data from additional samples, what were previously described as luminal/ER+ tumors seemed to cluster further into subgroups (148, 149).

Perou et al. (147) and Sorlie et al. (148, 149) defined intrinsic molecular subtypes using tumor expression profiles of between four and six hundred genes. Based on the dominant expression patterns that defined each subtype, Nielsen et al. (150) showed that the intrinsic subtypes could be defined using immunohistochemical stains for ER, HER2, EGFR, and CK 5/6 in place of gene expression arrays. Abd El-Rehim et al. (151) and Yu et al. (152) demonstrated the identification of intrinsic molecular subtypes using gene expression signatures characterized by luminal epithelial-related proteins and hormone receptor positivity, HER2 expression, basal epithelial-related markers in independent populations. Although Abd El-Rehim et al. (151) described 6 main clusters, most of the characteristics they described parallel the intrinsic subtypes defined previously by Perou and colleagues (147, 148).

Carey et al. (16) described the prevalence of characteristics of the intrinsic molecular subtypes within a population-based study. Using primary invasive breast cancer cases from the Carolina Breast Cancer Study, Carey et al. presented a modified classification system based on previous work of Perou and colleagues (discussed above). In addition to ER, HER2, EGFR, and CK 5/6, PR was also used to define the subtypes because it is a commonly measured estrogen-related gene and is a predictor of response to hormonal therapy (16). Also, tumors expressing both hormone receptors and HER2 were defined as a separate group, based on previous gene-clustering analyses. The following definitions were used by Carey et al. (16) to define the intrinsic molecular subtypes: luminal A (ER+ and/or PR+, HER2-); luminal B (ER+ and/or PR+, HER2+); HER2+/ER- (ER-, PR-, HER2+); and basal-like (ER-, PR-, HER2-, CK5/6+ and/or EGFR+). Tumors that did not fit these definitions were called “unclassified”.

Intrinsic subtypes have been described in several populations world-wide (17, 144, 153-158). Basal-like tumors (in Fulford et al. (159) basal-like was defined based on CK14, not CK5/6) are more likely to be higher grade, solid tumors with a higher number of mitotic figures, greater necrosis, and no tubule formation (156, 157, 159-161). Patients with HER2+/ER- and luminal B tumors are most likely to have positive lymph nodes (16), those with basal-like tumors were less likely to have positive lymph nodes (156). In addition to lack of ER, PR, and HER2 expression, the basal-like phenotype is characterized by expression of smooth muscle actin and vimentin (160). Some studies have reported that basal-like tumors are more likely to be larger than other subtypes (156, 161). Expression of basal-like marker CK5/6 was more common among interval breast cancers compared to cancers detected by screening mammography (162).

Intrinsic subtypes also show different patterns of chromosomal aberrations, indicating that different genetic mechanisms may be preferred by each subtype. Basal-like tumors have the most chromosomal gains and losses, whereas luminal B tumors are more likely to have high level amplifications (163). Also, the majority of tumors arising in patients with BRCA1 mutations are basal-like tumors (164-166).

Identification of intrinsic subtypes, particularly the basal-like subtype, has also been described in breast carcinoma *in situ* (152, 167-171). Compared to luminal A DCIS, luminal B, basal-like, and HER2+/ER- DCIS are more likely to have high nuclear grade, show areas of comedo necrosis, and have a high Ki-67 index (169, 170).

Few studies have described the epidemiology of the intrinsic molecular subtypes, but those that have found that basal-like tumors are more common among younger and African-American women (12, 16, 17, 144). Millikan et al. (17) reported that increased parity,

younger age at first full term pregnancy, not breastfeeding, high waist-to-hip ratio, young age at menarche, and higher adult adiposity compared to childhood are risk factors for the basal-like subtype of breast cancer. Nulliparity and a high waist-to-hip ratio are risk factors for with luminal A breast cancer (17). In Yang et al. (172), younger age at menarche and family history were positively associated with basal-like breast cancer, but many estimates were imprecise, making it hard to interpret the results.

Among invasive tumors, intrinsic subtypes have been shown to have different breast cancer-specific and overall survival patterns. Some studies have reported that patients with luminal A tumors tend to have the best survival and patients with HER2+/ER- tumors tend to have the poorest survival (16, 156, 173). Controlling for receipt of adjuvant therapy, basal-like tumors were associated with poorer survival compared to non-basal-like tumors (174). Yamamoto et al. (175) reported that basal-like tumors were associated with significantly poorer survival compare to non-basal-like tumors. However, other studies suggest that the basal-like phenotype lacks prognostic value when compared to existing prognostic factors. In a study by Rodriguez-Pinilla et al. (161), the basal-like phenotype was associated with poor prognosis, but not independently of tumor size. In Potemski et al. (176) and Jumpannen et al. (177), the basal-like phenotype was not a significant prognostic factor among ER-negative cancers.

Even if the basal-like phenotype is not an independent prognostic factor, characterization of basal-like tumors provides an opportunity to target prevention and therapy options towards a specific type of ER-negative breast cancer. The high prevalence of basal-like tumors in African-American women may explain some of the poor clinical outcomes for this population, even after adjusting for stage at diagnosis.

If the breast cancer intrinsic molecular subtypes have different etiologies, GWAS and pooled candidate gene studies would be more likely to identify variants associated with the most common intrinsic molecular subtype (luminal A). Additionally, most of the populations in the GWAS and large pooled studies involved white and/or postmenopausal women (107-109, 116, 118, 119, 141, 142). Given that basal-like breast cancer is less prevalent among white and postmenopausal women (12, 17, 144, 172), it is unlikely that the results of these studies will be generalizable to basal-like breast cancer.

Replications of some of the GWAS SNP associations that were first reported in 2007 support the idea that they are associated with certain types of cancer. In a pooled Breast Cancer Association Consortium analysis, Garcia Closas et al. (178) reported strong differences in association for GWAS SNPs rs2981582 (FGFR2) and rs13281615 (8q24) by ER status and tumor grade; none of the associations examined differed by nodal status. Nordgard et al. (179) examined associations between TNRC9, LSP1, FGFR2, MAP3K1, and H19 gene expression, GWAS-SNP genotype and breast cancer intrinsic molecular subtype. In that study, gene expression differed by molecular subtype for all five genes, and genotype distribution differed by molecular subtype for TNRC9 (179). Kristensen and Borresen-Dale re-analyzed the association of a SNP previously reported as associated with breast cancer, and observed that the variant homozygote genotype prevalence was much higher for the basal-like subtype than other subtypes (180, 181). Further research into the etiology of the intrinsic molecular subtypes of breast cancer has the potential to increase knowledge of the biological pathways that may be active in specific tumor subtypes. In turn, knowledge of which biological pathways are active in breast tumorigenesis can provide insight as to why some groups are at increased risk for particular subtypes, and possibly influence the

development of new treatment and prevention strategies.

1.6 Estrogen and breast cancer

Estrogen is a steroid hormone that is synthesized from cholesterol through a series of conversions involving several different cytochrome P450 and hydroxysteroid dehydrogenase enzymes (182). Estrogen is active in a variety of tissues throughout the body, and is responsible for stimulating growth of reproductive organs and decreasing the physical effects of aging (183). Estrogen also inhibits osteoclasts, preserving bone density, promotes endothelial cell development, and may have neuroprotective effects in older women (183). In the normal breast, estrogen stimulates the growth and differentiation of the ductal epithelium and surrounding stroma during puberty and pregnancy (183, 184).

Estrogen production varies by stage of life and according to the menstrual cycle. Estrogen levels first rise during puberty after stimulation by gonadotropin (183). After menarche, the theca and granulosa cells in the ovary are the main source of estrogen production (183, 185). Plasma estrogen concentrations vary according to the menstrual cycle, with peak levels occurring before ovulation (183, 185). After menopause estrogen is no longer produced in the ovaries, and the estrogen is primarily produced through the aromatization of androgens in adipose tissue (183).

Estrogen is produced in 3 forms – estradiol, estrone, and estriol (183). In addition to being produced from androgen precursors, estradiol and estrone can be inter-converted, in a reaction catalyzed by 17-beta hydroxysteroid dehydrogenases 1 and 2 (186). In the normal breast epithelium, the estradiol to estrone conversion pathway has much higher activity that the estrone to estradiol pathway (187). Serum concentrations of estriol are very low

compared to estrone and estradiol in premenopausal and postmenopausal women (183). There are two receptors that bind estrogen - estrogen receptor alpha (ER-alpha) and estrogen receptor beta (ER-beta). Estradiol is considered to be the more biologically potent estrogen because it has the strongest binding affinity for ER-alpha and ER-beta; estrone's binding affinity is approximately 60% (ER-alpha) and 40% (ER-beta) compared to estradiol (183).

Despite its beneficial role in many tissues, estrogen may lead to tumor formation either through its effects via estrogen receptor signaling or through the effects of its metabolites. Once bound to the estrogen receptor, the ligand-receptor complex translocates to the nucleus where it acts as a transcription factor to specific target genes. Estrogen is metabolized via two main pathways that involve hydroxylation of the A ring (leading to catechol estrogen formation) or hydroxylation of the D ring (leading to 16-alpha hydroxyestrone formation) (182). Estrogen metabolites can bind to DNA forming DNA adducts; high rates of DNA repair after adduct removal may introduce mutations that initiate carcinogenesis (188).

Epidemiologic evidence strongly supports a role for estrogen (in addition to progesterone (57)) in many breast cancers. Serum levels of several sex steroids, including estradiol, estrone, androstenedione, testosterone, and dehydroepiandrosterone sulfate are associated with an increased risk of breast cancer, and high levels of sex hormone-binding globulin are associated with a decreased risk of breast cancer in postmenopausal women (189, 190). Lifestyle factors that increase the number of lifetime ovulatory cycles, such as early menarche, late menopause, and nulliparity, are consistently associated with increased breast cancer risk (21). Surgical removal of the ovaries in premenopausal women, and therefore removal of the main source of estrogen, can reduce the risk of breast cancer by 50%

in high risk women (191). Finally, exogenous estrogen use is associated with an increase in breast cancer risk (76, 80, 81).

Furthermore, estrogen and the estrogen receptor are present in breast tumors. In breast cancer patients, aromatase activity and estradiol concentration is highest in tumor tissue and lowest in normal tissue (192). In hormone receptor-positive breast cancer, activity in the estrone to estradiol conversion pathway is predominant over the reverse reaction (193), potentially providing an additional source of estradiol to fuel tumor growth. In a nationwide sample of infiltrating ductal cancers in African-American and white women in the US, 75% were estrogen receptor-positive (4).

Altogether, this evidence suggests that estrogen is instrumental in at least a subset of breast cancers. Therefore, factors that modify estrogen expression or estrogen activity may have an important impact on breast cancer risk.

1.7 Estrogen pathway-related candidate genes and breast cancer

In order to investigate the relationship between variation in estrogen-related genes and the intrinsic subtypes of breast cancer, this study will focus on genes involved in estrogen synthesis, estrogen chaperoning in the bloodstream, and estrogen signaling. ER-alpha is required for estrogen to exert its effects in the cell, and so SNPs in the estrogen receptor gene that increase transcription could increase estrogen-related proliferation and SNPs that reduce transcription would be expected to have a protective effect. The progesterone receptor is one of the target genes affected by estrogen signaling, and its expression is highly correlated with ER expression (194). Cytochrome P450 enzyme 19A1 (CYP19A1) converts testosterone to estradiol (186), and so polymorphisms that upregulate

CYP19A1 activity could result in higher estrogen levels and therefore greater estrogen-related proliferation. 17-beta hydroxysteroid dehydrogenase 2 converts estradiol to estrone, and is an important regulator of estradiol levels (195). 3-beta hydroxysteroid dehydrogenase 1 (HSD3B1) converts pregnolone to progesterone, and polymorphisms that affect HSD3B1 expression could increase progesterone receptor signaling (186). Finally, sex hormone binding globulin (SHBG) is a chaperone molecule that binds to sex steroids in the bloodstream. Estradiol that is bound to SHBG is not available to bind to the estrogen receptor, and so polymorphisms that reduce expression of SHBG will increase the amount of free estradiol.

1.7.1 Estrogen receptor alpha (ESR1)

Estrogen receptor alpha (ER-alpha) is a steroid hormone receptor present in the cell cytoplasm and nucleus (183). In its inactive form, ER-alpha is bound to heat shock proteins (196). After binding its ligand, estrogen, the receptor dissociates from heat shock protein, dimerizes, undergoes a conformational change, and translocates to the nucleus (183). In the nucleus, the estrogen-estrogen receptor complex modulates transcription by binding to estrogen response elements or interacting with other nuclear transcription factors like NFkB and AP-1 (183). ER-alpha is able to bind all forms of estrogen, but it has the highest affinity for 17 beta-estradiol compared to other forms of estrogen such as estrone, estriol, and 17 alpha-estradiol (197). Hundreds of genes are regulated by estrogen signaling, including up-regulation of genes associated with cell proliferation, survival (cyclin D1, replication factor C4, survivin), growth factors, and transcription factors, and down-regulation of tumor suppressors, pro-apoptotic genes (cyclin G2, IEX-1, caspase 9), and growth inhibitors (198).

Researchers have also described a cell surface-bound form of the estrogen receptor

where signaling results in almost immediate physiological effects (183). This type of membrane-bound ER has been observed in vascular, endocrine, adipose, uterine and neuronal tissues (183, 199). The estrogen receptor can be activated by growth factor signaling as well, through phosphorylation of receptor serine or tyrosine kinase residues (183). Studies have shown that the epidermal growth factor receptor, heregulin, cyclic AMP, insulin/insulin-like growth factor, and dopamine can interact with estrogen receptor in the absence of an estrogen ligand (183).

In the normal breast estrogen receptor-alpha is expressed in the nuclei of a small percentage of luminal epithelial cells that line the ducts and lobules, but not in other breast cells (200-202). In the breast, estrogen stimulates growth and differentiation of ductal epithelium (183). In ESR1 knockout mice, breast development was stunted (202), suggesting that the knockout mice lacked proliferative signaling required for further development of the mammary gland during puberty. Lack of a functional ESR1 gene has also been associated with osteoporosis and reduced fertility in female mice (183).

Biological and epidemiological evidence suggests that the estrogen receptor is a major factor in breast tumor formation and survival. Levels of ER-alpha expression are altered in neoplastic growth. Atypical ductal hyperplasia and lobular carcinoma *in situ* both express the estrogen receptor at higher levels than that of normal breast (202). Approximately 60-70% of ductal carcinoma *in situ* are ER-positive (203, 204). Also, eliminating estrogen-receptor signaling reduces tumor formation in animal models. Oncogene-driven tumor formation in ER-alpha knockout mice occurred at half the rate of wild-type mice (205).

Finally, estrogen receptor expression is a strong predictive marker of response to hormonal treatments (206). ER expression is a weak prognostic factor, and is correlated with

other prognostic factors like histologic grade, proliferation, and tumor size (4, 202, 206). Studies with long follow-up show that even though ER-positive patients have a longer time until recurrence, they eventually have recurrence rates similar to ER-negative patients (206).

There have been several association studies of ESR1 polymorphisms and breast cancer, but results have been somewhat inconsistent. The variant T allele of the ESR1 +397 C/T (*PvuII*) SNP has been associated with a small increase in breast cancer risk in two studies (207, 208), but the association was not observed in other studies (209-211). Zheng et al. reported an inverse association between the TT genotype and breast cancer (117). Interestingly, the results from one functional study suggest that the C allele and not the T allele would be associated with a greater breast cancer risk – the C allele creates a new functional *myb* binding site and is associated with increased transcriptional activity (212). For the S10S polymorphism, inverse, positive and null associations with breast cancer have been reported (211, 213, 214).

Some studies have reported an association between the P325P polymorphism and breast cancer (215, 216), and familial history of breast cancer (217). Others (218, 219) reported no association between P325P and breast cancer. It has also been reported that P325P is inversely associated with lymph node metastasis in breast cancer cases (215, 219). Other SNPs -104062 C/T and 3' UTR rs3798577 have been investigated, but no clear pattern of association has emerged.

Tag SNP studies of ESR1 have identified intronic SNPs that are potentially associated with breast cancer. In data from the SEARCH study, Mavaddat et al. (220) reported that ESR1 SNPs rs3020314, rs3020407, and rs3020401 were strongly associated with breast cancer. Dunning et al. (211) combined the ESR1 rs3020314 associations from several case-

control studies and reported a very weak but precise association with breast. Subgroup analyses revealed that the association was only observed with ER-positive breast cancer in populations of European descent (211).

ESR1 SNPs have also been associated with serum hormone levels. Sowers et al. (221) reported that several SNPs have ethnicity-specific associations with serum estradiol levels, an interesting finding given the ethnic and geographic variation in breast cancer incidence worldwide. The +397 C/T (*PvuII*) CC genotype was associated with higher serum estradiol levels in African-American women (221) and +397T - +351A haplotype (*PvuII-XbaI*) was associated with lower serum estradiol levels in postmenopausal Danish women (222), which is consistent with the higher expression associated with +397C *in vitro*. It could be hypothesized that the higher levels of circulating estradiol associated with the C allele mean that less estradiol is able to bind to the variant receptor and translocate to the nucleus, and this is why the +397T allele is associated with increased breast cancer risk. The +397T allele was associated with higher levels of androstenedione in postmenopausal women (223); it is also possible that the increased breast cancer risk is mediated through testosterone and not estrogen.

1.7.2 Progesterone receptor (PGR)

The progesterone receptor (PR) is a steroid receptor which influences DNA transcription. PR signaling is initiated by binding of the ligand, progesterone, receptor dimerization, phosphorylation, and DNA binding (224). PR is expressed in three isoforms – PR-A, PR-B, and PR-C. The PR-A and PR-B isoforms are similar in some respects. Both are capable of forming either homodimers or heterodimers, and both have the ability to bind to progesterone-response elements in DNA (224, 225). PR-A and PR-B also have similar

structures, except that PR-B contains an additional transactivation domain (224). Despite these similarities, most studies suggest that these two isoforms have unique transcriptional activities (225). Like PR-A and PR-B, the PR-C isoform contains a ligand-binding domain, but it is truncated at the N-terminus and therefore lacks the progesterone response element-binding motif and activation domains (225). PR-C has the ability to bind to PR-B and inhibits PR-B transcriptional activity, possibly by competing for progesterone or inhibiting PR-B ability to bind to DNA (225).

PR-A and PR-B are expressed in equal amounts in the luminal epithelial cells in normal breast tissue (201, 225). PR knockout mice experience anovulation, abnormal uterine morphology and histology, and impaired branching and differentiation of the breast during pregnancy (224, 226), suggesting the progesterone receptor is required for normal reproductive function.

PR expression is highly correlated with ER-alpha expression (201). In fact, PR expression is regulated by estrogen, and both estrogen and ER signaling are required for the progesterone receptor to be produced (194). PR can also induce proliferative activity through crosstalk with the estrogen receptor. In T47D cells, PR signaling stimulates proliferation via the Erk and PI3K/Akt pathways (227). Ballare et al. (227, 228) showed that progestin initiates Src/Erk signaling through a direct interaction of the PR-B N-terminus with the ER-alpha ligand-binding domain.

The progesterone receptor is expressed in 70% of invasive breast cancers (204). In breast tumors, PR expression is inversely associated with expression of HER2 and EGFR (194). The exact mechanism for loss of PR expression in breast tumors is unknown, but hypotheses include lack of functional ER, promoter methylation, loss of heterozygosity, and

signals from growth factors, such as IGF, EGF, or heregulin (194). PR-A and PR-B isoforms are expressed in equal amounts in the normal breast (225), however, in a subset of invasive tumors PR-A is present in higher amounts than PR-B, and some studies have suggested that this imbalance may be related to tamoxifen resistance (194, 225, 229). The predictive role of PR independent of ER status is unclear. A review of tamoxifen trials showed that PR status did not predict treatment benefit, but other studies have shown that among ER-positive patients, those that were PR-positive had a greater benefit from tamoxifen than PR-negative patients (194).

The V660L amino acid change in exon 4 has been studied extensively, but the relationship between V660L and breast cancer remains unclear. Groups have reported both a positive association (213, 230) and no association (141, 231, 232) between the variant allele and breast cancer. *In vitro* experiments of the functional effects of the codon 660L variant also show varying results. One group reported that the codon 660L variant produced higher transcriptional activity and was degraded more slowly within the cell (233), suggesting that the codon 660L variant could have a stronger or more prolonged effect on the transcription of PR target genes or proliferative signaling cascades. This effect would be consistent with an increased breast cancer risk for the variant allele. However, another group reported that the codon 660L variant leads to reduced receptor phosphorylation, which would lead to reduced variant allele activity and presumably would not cause increased breast cancer risk (234). It is difficult to interpret the true causal effect of the V660L SNP because it is linked to several other PGR polymorphisms. The codon 660L polymorphism is in complete linkage disequilibrium (LD) with an *Alu* insertion polymorphism that has also been shown to increase PGR transcription (234, 235). Codon 660L is also in complete LD with a

polymorphism at codon 770, and in almost complete LD with a polymorphism at codon 344 (235).

The promoter polymorphism +331 G/A has also been investigated for a possible association with breast cancer. The +331 G/A SNP is located in between the wild type transcriptional start sites for PR-B (+1) and PR-A (+751), and creates a new transcriptional start site that results in increased transcriptional activity (235). Despite this, only the Nurses' Health Study (236) has demonstrated an association between the 331A allele and breast cancer; other studies have not shown an association (213, 230, 231). After further analysis in the Nurses' Health Study, Kotsopoulos et al. (237) reported that the +331 G/A association was modified by postmenopausal hormone use, and that the increased breast cancer risk associated with the A allele is much higher in never-users compared to past or current HRT users.

1.7.3 17-beta-hydroxysteroid dehydrogenase type II (HSD17B2)

17-beta-hydroxysteroid dehydrogenase type II (HSD17B2) is a member of the short chain alcohol dehydrogenase super family of enzymes that oxidizes active sex steroids into their inactive precursor forms (195). Specifically, HSD17B2 converts estradiol into estrone, testosterone to 4-androstenedione, and 5-androstenediol into dehydroepiandrosterone (195).

HSD17B2 is expressed in a large number of normal tissues, including placenta, liver, endometrium, kidney, colon, and normal breast epithelium (187, 195). Only some studies have reported that HSD17B2 is expressed in breast tumors. Gunnarsson et al. (238) reported that HSD17B2 expression was detected in the cytoplasm of most breast tumors, but not in the surrounding stroma. In contrast, two other studies reported that HSD17B2 expression in breast cancer was very low (239) or undetectable (240). Gunnarsson et al. (238) also reported

that HSD17B2 expression is strongly correlated with expression of aromatase and cyclooxygenase 2, although Yoshimura et al. (239) did not find the same correlation.

Studies in breast cancer patients support the hypothesis that HSD17B2 may act to reduce available estradiol. In ER-positive patients, low levels of HSD17B2 are associated with distant recurrence and breast cancer-related death, but HSD17B2 levels have no effect on prognosis in ER-negative patients (238). Furthermore, a high HSD17B2:HSD17B1 ratio in ER-positive patients is associated with better prognosis (241). HSD17B1 reduces estrone to estradiol (195), and so higher amounts of HSD17B2 indicate the predominance of the estradiol to estrone oxidation pathway within the tumor.

Little has been published on the effect of HSD17B2 polymorphisms on breast cancer risk. One study examined the codon 226 M to V amino acid change and found no association with breast cancer (242). Molecular modeling did not predict any functional effects due to the codon 226 polymorphism, but *in vitro* studies were not performed to confirm this (242). Considering ER-positive cases only, the codon 226 V variant was associated with a two-fold increased risk of breast cancer for cases with two close relatives with breast cancer compared to controls carrying the codon 226 V variant, but the confidence interval was very imprecise and includes the null, and therefore should be interpreted with caution (242).

1.7.4 3-beta-hydroxysteroid dehydrogenase type I (HSD3B1)

3-beta hydroxysteroid dehydrogenase type I (HSD3B1) is a member of the short-chain oxidoreductase enzyme super family that converts pregnolone to progesterone, 17-alpha hydroxypregnenolone to 17-alpha hydroxyprogesterone, and dehydroepiandrosterone to androstenedione (186, 243). In addition to being expressed in normal breast, skin, prostate, and placenta, HSD3B1 is active in breast tumor cells, indicating that HSD3B1 activity could

act as a source of progesterone within breast tumors (186, 243, 244). Basal HSD3B1 expression is controlled by a Sp1 binding site in intron 1 (245). In normal breast cultures and breast cancer cell lines, HSD3B1 expression can be induced further by cytokines interleukin-4 and interleukin-13 (246).

Some studies have linked the HSD3B1 codon 367 N to T polymorphism with prostate cancer (247, 248), but other groups did not find an association (249). This nonsynonymous change creates a new PKC phosphorylation site in the COOH-terminal extramembrane domain, potentially causing a functional change in the HSD3B1 protein (247). The codon 367T variant is associated with increased breast density among African-American women, decreased breast density among white American women, and decreased breast density in mostly white Australian women (250, 251). Increased breast density is a risk factor for breast cancer (50, 51), so it is possible that the codon 367T variant may be associated with increased breast cancer risk, particularly in African-American women.

1.7.5 Cytochrome P450 Family 19 A1 (CYP19A1)

Cytochrome P450 Family 19 A1 (CYP19) is part of a larger family of cytochrome P450 enzymes involved in steroid hormone biosynthesis (186). The CYP19 gene product and cofactor NADPH form the aromatase enzyme complex (252). Aromatase converts androgens into estrogens, including conversion of androstenedione to estrone and testosterone to estradiol (186). The CYP19 gene has several exon I splice sites that produce different forms of exon I which are expressed in a tissue specific manner (186). CYP19 translation begins in exon II, and so each tissue has a unique 5’ region, but the enzyme produced in each tissue type is the same (252). CYP19 is expressed in the ovary, placenta, testis, adipose tissue, bone, breast, and brain (186). In normal breast tissue, CYP19 is expressed in the epithelial

cells of the terminal ductal lobular unit and stromal fibroblasts (253, 254). In the ovary, aromatase is expressed in granulosa cells and the corpus luteum (252). In adipose tissue, CYP19 is expressed in the stromal cells; expression levels are higher in subcutaneous compared to visceral fat (252).

Aromatase expression and activity is higher in breast tumors compared to normal breast tissue (255). Expression was detected in both stromal and carcinoma cells (255). In a study of aromatase mRNA levels in different quadrants of mastectomy specimens, Bulun et al. (256) reported that the tumor-bearing quadrant was significantly more likely to have the highest levels of aromatase. Aromatase transcript levels were also correlated with the number of stromal cells in each quadrant (256). In breast cancer patients, aromatase activity is highest in the tumor area and lowest in normal tissue (192). In Esteban et al. (257), aromatase activity was inversely correlated with ER-alpha but not PR expression. In Miki et al. (255), aromatase expression was positively correlated with ER-alpha.

Aromatase activity can be altered by other molecules. Treatment with aromatase inhibitors such as letrozole and exemestane inhibited proliferation in MCF-7 cells (255). When used to treat postmenopausal breast cancer patients, aromatase inhibitors provided better long-term survival compared to standard treatment regimens and lead to fewer adverse effects (258). Endogenous factors also affect aromatase activity. Keratinocyte growth factor, epidermal growth factor, transforming growth factor alpha, and leptin all stimulate aromatase activity in breast cancer cell lines (259-261).

Researchers have tested the functional effects of several CYP19 SNPs using *in vitro* studies. In the case of both the W39R and R264C amino acid changes, studies have reported that the variant allele is associated with reduced enzyme activity and that the polymorphism

has no effect on enzyme activity (262-264). Miyoshi et al. (265) reported a negative association between codon 39R and codon 264C variants and breast cancer. However, 3 other studies of R264C did not find any association with breast cancer women (264, 266-268).

Riancho et al. (269) reported that a SNP in the I.2 promoter is associated with higher levels of aromatase in adipose tissue. One would expect that higher levels of aromatase in adipose tissue would result in higher estradiol levels and higher breast cancer risk, especially in women with more adipose tissue. However, there are no published studies of the I.2 polymorphism and breast cancer. Talbott et al. (270) reported a positive association for rs108805 intron 2 A/G among premenopausal but not postmenopausal women. This SNP had been previously reported as having no association with breast cancer in a pooled study of postmenopausal women, where carriers of the G allele had increased levels of serum estradiol and estrone (267).

Kristensen reported a positive association between the CT and TT genotypes of the exon 10 3' UTR SNP rs10046, and later demonstrated that basal-like cases are more likely to have the TT genotype compared to women with other tumor subtypes (180, 181). However, other studies reported no association between rs10046 and breast cancer (271, 272). Raskin et al. (273) reported that the synonymous V80V polymorphism was positively associated with breast cancer in BRCA carriers younger than 50 years old.

1.7.6 Sex hormone-binding globulin (SHBG)

Sex hormone-binding globulin (SHGB) is an allosteric protein that binds sex steroids in the bloodstream, controlling their availability for downstream signaling processes. SHBG also binds to the membrane-bound SHBG receptor (SHBG-R); free SHBG can bind steroids

before or after it binds to its receptor, but once SHBG has bound a steroid ligand it is no longer able bind to the SHBG receptor (274).

Experimental data supports the hypothesis that SHBG is instrumental in controlling the proliferative and anti-apoptotic effects of estradiol (275, 276). These anti-proliferative effects are due in part to SHBG/SHBG-R induction of cyclic AMP and PKA in the presence of estradiol (277, 278). Additionally, SHBG suppresses estradiol up-regulation of bcl-2, c-myc, EGFR, and PR (278, 279). Estradiol down-regulation of ER-alpha is also inhibited (279). The ability of SHBG to reverse the effects of estradiol is sensitive to SHBG mutations: variants without the O-linked oligosaccharide on the threonine amino acid at codon 7 had no effect against estradiol (276).

Moore et al. (280) detected SHBG mRNA in ER-positive and ER-negative breast cancer cell lines and breast tumor samples. In breast tumors, the presence of membrane SHBG-R is positively correlated with cytosolic levels of the progesterone receptor and negatively correlated with cellular proliferation (281).

SHBG levels are associated with body size. SHBG concentration decreases with increasing BMI and waist circumference (282-285). In premenopausal African-American women, a waist-to-hip ratio of greater than 0.80 was associated with lower levels of SHBG compared to women with a waist-to-hip ratio of less than or equal to 0.75, regardless of obesity level (286).

In postmenopausal breast cancer patients, treatment with tamoxifen is associated with a significant increase in SHBG levels (287). Higher levels of SHBG were associated with a reduced risk of breast cancer in postmenopausal women in some studies (288-290), but not in others (291, 292). Several studies also reported no association between SHBG levels and

premenopausal breast cancer (291, 293-295). However, in many cases the studies of premenopausal women were limited by low statistical power; the effect estimates suggested a trend of reduced risk for the highest vs. the lowest quartiles of SHBG. SHBG was not associated with ductal carcinoma *in situ* in a small prospective study of postmenopausal women (296).

Researchers have identified two functional polymorphisms in SHBG that may be associated with breast cancer. The D356N amino acid change (also known as D327N) is associated with higher levels of SHBG (272, 297, 298). Based on the anti-proliferative effects of SHBG in cell lines, the expected effect of higher SHBG would be reduced breast cancer risk. However, in Dunning et al. (272) codon 356 NN homozygotes had an increased risk of breast cancer. The other functional SNP is a G to A nucleotide change in the 5' UTR. In Dunning et al. (272) this SNP was associated with increased SHBG levels in postmenopausal women, but no association with breast cancer was observed by Dunning et al. or in a second study (299).

1.8 Obesity, insulin resistance, and breast cancer

Obesity is associated with increased incidence of many cancers, including breast cancer (300). Data from the NHANES national survey (1999-2004) shows that 29% of adult US women are overweight and 33% are obese (301). The 2003-2004 prevalence of obesity was 31% in non-Hispanic white women and 54% in non-Hispanic African-American women, up from 15% and 31% in the 1976-1980 NHANES survey (301, 302).

In postmenopausal women, obesity is associated with an increased risk of breast cancer, (30, 32, 35, 45, 46, 303) though some studies have found no association (34). There

is some evidence that the association between body mass index (BMI) and increased breast cancer risk in postmenopausal women is modified by hormonal factors, such as HRT use or type of menopause. In the EPIC study, BMI was positively associated with breast cancer in postmenopausal women who were not HRT users, while there was no association among HRT users (31). Also, Kaaks et al. (33) reported an inverse association between BMI and breast cancer among women who experienced natural menopause and a positive association in women with surgical menopause, but both results were imprecise and did not appear statistically different from each other. In postmenopausal women, obesity is consistently associated with the incidence of ER-positive and PR-positive tumors and the relative risk increases with increasing obesity (304). Obesity is generally thought to be associated with a decreased risk of breast cancer in premenopausal women (30-32, 36), though some have reported no association (33-35) and at least one study has reported a positive (but imprecise) association (45).

Central obesity is indicative of large amounts of deep visceral fat. Waist-hip ratio (WHR), a common measure of central obesity, has been associated with increased breast cancer risk in women independently of BMI [(33, 43, 305, 306), reviewed by (41, 42)]. Unlike the association between breast cancer and BMI, studies did not demonstrate any effect modification by menopausal status.

Several mechanisms have been proposed to explain the relationship by which obesity and central obesity may increase breast cancer risk. First, obesity may increase breast cancer risk by influencing estrogen production. In postmenopausal women, estrone and estradiol levels increase with increasing BMI and WHR, and SHBG decreases with increasing BMI and WHR (285, 307, 308). Some groups contend that the effect of obesity on steroid

hormone levels explains most of the relationship between obesity and breast cancer - the effect of BMI on breast cancer risk was greatly attenuated when adjusted for serum estrogen levels (285). Others argue that this does not fully explain the extent of the association between obesity and breast cancer (309). Further evidence for the hypothesis that BMI increases breast cancer risk by estrogen-related mechanisms is the lack of association between BMI and breast cancer risk in HRT users. This suggests that among those with significantly increased hormone levels (due to exogenous hormones) the increase in hormones due to obesity has no effect, or that BMI only has an effect among women with low endogenous hormone levels (300). The lack of association between premenopausal obesity and breast cancer also fits this hypothesis. Premenopausal women have high levels of endogenous estrogens due to ovarian estrogen production. Compared to ovarian estrogens, additional estrogen produced by adipocytes and any increase in that level due to obesity is negligible. This hypothesis is also supported by the fact that estradiol levels do not vary by BMI in premenopausal women (310).

A second hypothesis is that obesity affects cancer growth by promoting insulin resistance and hyperinsulinemia. Insulin resistance occurs when muscle, liver, and adipose cells have a reduced response to insulin (311). The pancreas continues to produce insulin in order to achieve a biological response, leading to the accumulation of insulin and un-metabolized glucose in the bloodstream (311). Insulin promotes insulin-like growth factor 1 (IGF-1) production and inhibits IGF-binding protein, further increasing the amount of free IGF-1 (312). Both insulin and IGF-1 promote cell proliferation and inhibit apoptosis *in vitro* (300). Furthermore, inhibition of the IGF1-receptor inhibits growth of xenograft breast cancer cells in mice (313), suggesting that tumor growth is slowed when IGF-1 is unable to

initiate signaling in downstream pathways. Insulin receptor and the IGF-IR hybrid receptor are expressed at significantly higher levels in breast tumors compared to normal breast (314), indicating that they may play an important role in breast tumorigenesis in humans. It is also possible that insulin resistance interacts with reproductive hormones and the estrogen-related carcinogenesis pathway. High levels of insulin and IGF-1 reduce production of SHBG, leading to more bioavailable estradiol and testosterone (300).

A third hypothesis is that obesity leads to increased production of pro-inflammatory cytokines. Chronic low grade inflammation results in macrophage production of pro-inflammatory cytokines tumor necrosis factor-alpha and interleukin-1B, which then stimulate interleukin-6 production (315). This hypothesis is supported by data from the Fatless A-Zip/F1 mouse model. Fatless mice have features common to insulin resistance but lack adipose tissue or adipocytokines, which are growth factors produced by adipose tissue (315). Fatless mice grew tumors faster than the control mice, suggesting that adipocytokines were not essential for tumor formation in an insulin resistant environment, and that other pathways involved in insulin resistance may be responsible for increased cancer risk, not adipocytokines themselves (315).

Finally, a fourth hypothesis is that increased levels of adipocytokines produced by adipose tissue may promote breast tumor growth. *In vitro*, adipocytokines have effects on cell proliferation, angiogenesis, and apoptosis (reviews: (316, 317)). For example, leptin stimulates aromatase activity and increases cell proliferation (259, 318-322), and IL-6 can activate ER-alpha transcription (323) – upregulation of these genes effects could promote breast tumor growth. Adiponectin inhibits the proliferative effects of estradiol, and so obesity-induced reduction of adiponectin levels could increase breast cancer risk (324-328).

Other adipocytokines, such as vascular endothelial growth factor, hepatocyte growth factor, and heparin-binding epidermal growth factor-like growth factor, promote angiogenesis and endothelial cell migration (316).

1.9 Obesity-related candidate genes

Adipocytokines are a group of growth factors and cytokines produced primarily by white adipose tissue (329). White adipose tissue is a multi-functional organ composed of adipocytes, fibroblasts, and macrophages that acts as the body's main energy reserve through storage of fatty acids, insulation, and a source of hormones, adipocytokines, and inflammatory factors (330). Adipocytokine production levels are associated with insulin resistance syndrome and obesity (329). The circulating levels of most adipocytokines increase with increasing BMI and obesity, except for adiponectin which is inversely associated with BMI and obesity (316). Adipocytokines can have pro-carcinogenic effects and have been shown to affect aromatase activity, angiogenesis, and proliferation in cell lines (316, 317). Recent research has demonstrated that some adipocytokines and adipocytokine receptors are expressed in non-adipose tissues, including normal and malignant breast tissue. Adipocytokines produced within the breast epithelium could produce autocrine effects. On the other hand, excess adipocytokines produced in the surrounding adipose tissue in obese women could affect breast tissue or surrounding stroma in a paracrine manner (317).

Waist-hip ratio was strongly associated with basal-like breast cancer in the CBCS (17), suggesting that factors associated with central adiposity may be associated with basal-like breast cancer. Adipocytokines are one group of factors that fit into this hypothesis. In order to investigate the potential effect of common genetic variation on basal-like breast

cancer, I will focus on variation in the adipocytokines tumor necrosis factor alpha (TNF-alpha), interleukin 6 (IL-6), leptin (LEP), the leptin receptor (LEPR), and adiponectin (ADIPOQ). Each of these genes is associated with obesity or insulin resistance (331), and is also associated with cellular processes that promote tumor formation.

1.9.1 Tumor necrosis factor alpha (TNF)

Tumor necrosis factor alpha (TNF-alpha) is a cytokine with pro- and anti-tumor functions, including inhibiting tumor cell proliferation, promoting apoptosis, stimulating estrogen synthesis through aromatase activity, and promoting angiogenesis (332). TNF-alpha is secreted by adipocytes, and increased TNF-alpha levels are associated with obesity and insulin resistance (331, 333). It is unclear whether TNF-alpha levels are a cause or a consequence of insulin resistance. The addition of TNF-alpha induces insulin resistance in adipose tissue *in vitro* (334). On the other hand, weight loss reduces TNF-alpha levels in humans (331). Blocking TNF-alpha improved insulin resistance in rats, but a similar technique had no effect in humans (331).

TNF-alpha and its receptors are expressed in both normal and neoplastic breast tissue, which increases the plausibility that it acts on breast tissue. Chavey et al. (335) found that TNF-alpha is expressed at a much higher level in neoplastic compared to normal breast tissue, but other studies did not report the same result (336, 337). The tumor necrosis factor-alpha receptor 1 was present in all samples of normal, *in situ*, and invasive breast tissue, but tumor necrosis factor-alpha receptor 2 was detected increasingly in more malignant tissues (invasive tumors were the highest percentage positive, benign were lowest percentage positive) (337). Chavey et al. (335) reported that TNF-alpha was over-expressed more commonly in ER-negative and PR-negative tumors compared to ER-positive tumors.

However, in Garcia-Tunon et al. (337) TNF-alpha expression was not associated with any tumor characteristics, including ER or PR status.

There are several mechanisms by which TNF-alpha might affect breast tumor growth. TNF-alpha stimulates aromatase activity in normal and malignant breast fibroblasts cultured from breast surgical samples (318). Aromatase catalyzes the final step in the conversion of androgens to estrogens, so increased TNF-alpha could lead to higher estrogen levels and estrogen-related proliferation. Experiments by Hagemann et al. (327) showed that macrophage-produced TNF-alpha increased the invasiveness of malignant but not benign cells *in vitro* by causing an increase in matrix metalloproteinases 2, 3, 7, and 9. This action was inhibited by the addition of an anti-TNF antibody or a matrix metalloproteinase inhibitor. Some studies have shown that TNF-alpha has effects that may inhibit tumor growth. TNF-alpha inhibited epidermal growth factor-stimulated proliferation in MCF-7 cells (326), and thus TNF-alpha expression could provide protection against aberrant signaling in the epidermal growth factor pathway. TNF induced cell death in the MCF-7 breast cancer cell line, but the extent of apoptosis differed according to laboratory strain of MCF-7 (328).

Even though studies show that TNF-alpha is present and has effects in breast cells, only one study has reported that serum levels of TNF are significantly higher in women with breast cancer compared to women with a negative breast biopsy (338). Others have reported that there is no association between serum TNF-alpha levels and breast cancer (339, 340). This may be due to the fact that TNF-alpha has both pro- and anti-tumor effects and could therefore increase or decrease breast cancer risk under different circumstances. The lack of association may also be due to the fact that adipose TNF-alpha acts through autocrine and paracrine mechanisms (331), and levels of TNF-alpha within breast tissue may not be the

same as TNF-alpha levels in serum.

Multiple polymorphisms in TNF have functional effects, and may affect breast cancer risk. The T allele of the -857 C/T polymorphism has significantly higher activity after stimulation by lipopolysaccharide compared to the C allele (341, 342). The TNF-alpha -863A allele is associated with reduced transcription in reporter assays and differential binding of nuclear proteins (343, 344). Functional studies of the -308 G/A polymorphism have shown that the -308A allele drives higher expression compared to -308G in *in vitro* reporter assays (345, 346). Kroeger et al. (346) showed that the differential transcription due to the -308 G/A polymorphism only occurred in the presence of the TNF 3' UTR. However, none of the cell lines tested in these studies were breast cancer cell lines; it is possible that the functional effects differ in breast cells.

TNF polymorphisms are also associated with levels of TNF-alpha in human subjects. Skoog et al. (347) analyzed the association between TNF polymorphisms and TNF-alpha secretion in adipose tissue in non-obese individuals, and found that adipose tissue from -863A carriers secreted less than half as much TNF-alpha compared to -863 CC homozygotes. Carriers of -308A and -1031C minor alleles also demonstrated lower TNF-alpha secretion compared to -308 GG homozygotes and -1031 TT homozygotes, respectively (347). These differences in TNF-alpha levels by genotype that were apparent in the non-obese disappeared in obese subjects (347), suggesting that under an obese phenotype different factors regulate TNF-alpha expression. In another study, TNF-alpha production was significantly higher in -308A allele carriers compared to -308GG homozygotes, but this was measured in cultured blood cells, not adipocytes (348). The TNF -1031C allele was associated with lower serum TNF-alpha levels (344).

Polymorphisms in TNF have been linked to several diseases, including obesity and breast cancer. In a meta-analysis, the -308 G/A polymorphism was associated with obesity specifically among white, middle-aged subjects; -308 G/A was not associated with plasma leptin or WHR (349). In a comprehensive analysis of TNF-alpha and the neighboring gene lymphotoxin alpha (LTA), Gaudet et al. (350) reported an increased risk of breast cancer for -238 A variant allele carriers. Inferred haplotypes containing the -238 A allele were also associated with an increased risk of breast cancer (350). Women homozygous for the -857 C allele were at increased risk for breast cancer compared to TT major allele homozygotes, but this association was not confirmed in a second, Polish population (350). Gaudet et al. (350) did not find an association between the -308 G/A polymorphism and breast cancer.

In a Tunisian population, the -308 AA genotype was more common among breast cancer cases than controls (351). Other studies reported no association between –238 G/A, –308 G/A, –857 C/T, –863 C/A or –1031 T/C polymorphisms and breast cancer (350, 352-355).

Despite the reported pro- and anti-tumor effects of TNF-alpha *in vitro,* data from the studies described above (350-354) suggest that TNF-alpha polymorphisms have a positive effect or no effect on breast cancer risk. None of the results suggest that TNF-alpha polymorphisms are inversely associated with breast cancer risk.

1.9.2 Interleukin 6 (IL6)

Similar to TNF-alpha, IL6 is a cytokine associated with inflammation, immune function, and injury response (331). IL6 expression is positively associated with obesity (334). The relationship between IL6 and insulin resistance is complex. Many studies have shown that IL6 is associated with increased insulin resistance (331, 334), however at least

one study demonstrated that IL-6 increases insulin sensitivity in skeletal muscle cells and that IL6 effects on insulin signaling may be tissue specific (356).

IL6 is expressed in normal and neoplastic breast tissue, but there are conflicting reports on whether IL6 levels differ between normal and malignant tissue. In Chavey et al. (335), IL6 was expressed at a significantly higher level in neoplastic compared to normal breast tissue, but in Basolo et al. (357), IL6 was highest in cultures of normal breast epithelium and lowest in breast tumor cells. Green at al. (336) found no significant difference in IL6 levels between normal and neoplastic cell cultures. Differences may be due to the timing and laboratory methods used to measure IL6 – in Green et al. (336) and Basolo et al. (357) cells were cultured and IL6 levels were measured after the first (336) or third/fourth passages (357), whereas in Chavey et al. (335) IL6 was measured directly in fresh, non-cultured breast tissue.

In vitro studies indicate that IL6 may have pro- and anti-tumorigenic effects (reviewed by Knupfer et al. (358)). IL6 has been reported to inhibit growth in several ER-positive and ER-negative breast cancer cell lines (359-361). Danforth et al. (359) demonstrated that IL-6 and estradiol antagonize each other's growth effects in MCF-7 cells, and that IL6 causes a small decrease in ER expression and an increase in PR expression in MCF-7 cells. Also, IL-6 induced rounding, reduced cell adhesion, and decreased E-cadherin expression in the ZR-75-1-Tx, T47D, and MCF-7 cell lines (361, 362). IL-6 increased migration of T47D cells (360).

The effects of IL6 in cultured surgical breast specimens are not consistent with the reported *in vitro* effects. IL6 alone was able to activate transcription of ER-alpha in ER-positive primary breast cell cultures (323). IL6 had no effect on proliferation (measured by

DNA synthesis) in cultured breast tumor cells (357). According to Basolo et al. (357), IL6 stimulated DNA synthesis in cultures of normal breast, but Asgeirsson et al. (361) reported that IL6 inhibited growth (colony size and number) in normal breast cultures. The same study reported that IL6 did not affect adhesion or E-cadherin expression in normal mammary epithelial cultures (361), suggesting that any effects of IL6 to promote cell discohesion seen in cell lines require earlier cancer initiation and/or promotion events.

One potential mechanism by which alterations in IL6 expression or accumulation might be related to breast cancer is interaction between IL6, estrogen, and the estrogen receptor. IL6 can stimulate estrogen synthesis by inducing aromatase activity in fibroblasts cultured from normal and malignant breast tissue (318). Additionally, Chavey et al. (335) reported that IL6 is expressed more commonly in tumors that were ER-negative, PR-negative, and high grade tumors, but other studies (337, 363) found no association between ER or PR status and IL6 expression.

The most well characterized polymorphism in the IL6 gene is the -174 G/C promoter polymorphism. In luciferase reporter assays in HeLa cells, the $(AT)_{8/12}$ -174C haplotype was associated with lower baseline expression and reduced response to stimulation by lipopolysaccharide or interleukin-1. Terry et al. (364) tested not only the -174 G/C polymorphism, but other promoter polymorphisms including -572 G/C, -597 G/A, and a -373 AT repeat polymorphism in both HeLa and ECV40 cells, and showed that changes in IL6 expression level are likely due to a complex haplotype effect, not the single genotype at position -174. Some have reported that the -174G allele is associated with higher circulating IL6 levels (365), but in other studies the -174 genotype was not associated with serum IL6 (366-369). In light of the work by Terry et al. (364), these differing results may depend on

differing haplotype structures in the different populations in which the studies were conducted.

Several studies have attempted to determine whether the IL6 -174 G/C polymorphism is associated with breast cancer. In a case-control study of women of Austrian or German descent, Hefler et al. (370) the -174C allele was associated with increased breast cancer risk (370). Slattery et al. (371) reported an inverse association between -174 G/C and breast cancer. Gonzalez-Zuloeta Ladd et al. (372) reported a small, non-significant increase in the odds of breast cancer among -174C allele carriers (GC/CC vs. GG), but they also adjusted for several covariates that either do not affect nt -174 genotype or are potentially on the causal pathway between nt -174 genotype and breast cancer risk. In addition to increasing imprecision of their estimates, they may have induced confounding or attenuated their effect estimates by adjusting for factors in the causal pathway. Unadjusted results were not reported, and so it is unknown to what extent these results support those of Hefler et al. (370). Finally, Smith et al. (353) and Litovkin et al. (373) reported no association between the nt -174 genotype and breast cancer, but did not reported the corresponding odds ratio or confidence interval.

In a study of women from the southwestern US, Slattery et al. (371) reported associations for -174 G/C (rs1800795), -572 G/C (rs1800796), and intron 2 G/A (rs2069832). Haplotypes containing the minor alleles for rs1800797, rs1800795, and rs2069832 were associated with lower obesity in non-Hispanic white women, but these same haplotypes were not associated with breast cancer (374).

Studies have also reported conflicting results on the association between the -174 G/C polymorphism and breast cancer prognosis. In an Australian study, the CC genotype was

associated with poor histological grade, larger tumor size, and lower ER content (375). In the same study, the CC genotype was associated with almost twice the hazard of death compared to the GC or GG genotype, but was not an independent predictor of survival (375). In contrast, Demichele et al. (376) reported that the -174C allele is associated with improved disease-free survival and overall survival in node-positive breast cancer patients; the data also suggest that patients who are both ER-positive and -174C allele carriers have the best disease-free survival, but statistical evidence for interaction was not significant, probably due to the small number of patient years in the analysis (376).

Although serum IL6 was not associated with breast cancer incidence in a prospective study of elderly subjects (340), serum levels of IL6 were significantly higher among breast cancer patients compared to women with a negative breast biopsy (338). IL6 is produced in many different tissues in the body, including adipose tissue, and so serum IL6 levels may not reflect breast-specific levels of IL6 (331). The *in vitro* effects of IL6 combined with the demonstrated link to the estrogen-related proliferation pathway provide a plausible link as to how IL6 may be causally related to breast cancer. Since IL6 may have pro- and anti-tumor effects in breast cells, it is hard to predict what effect an IL6 polymorphism will have on breast cancer risk.

1.9.3 Leptin (LEP) and the leptin receptor (LEPR)

Leptin is a hormone produced mainly by adipocytes that is involved in regulating body weight (334). After production in adipose tissue, leptin circulates in plasma eventually reaching the central nervous system, where it binds to the leptin receptor and upregulates anorexigenic peptides and downregulates orexigenic peptides (331). As a result, leptin reduces lipid levels and improves insulin sensitivity (331). In mice, leptin deficiency causes

insulin resistance, obesity, and diabetes, and leptin treatment reverses these conditions (331, 334). Leptin treatment also reverses insulin resistance and diabetes that is due to lipodystrophy (331, 334). In non-obese humans, leptin is associated with decreased appetite and increased energy metabolism. Leptin levels are higher in the obese, suggesting that these individuals are no longer sensitive to leptin signaling, but the exact mechanism for this 'leptin resistance' has not been described (331).

Leptin is expressed in a variety of tissues, including adipose tissue, normal breast, breast cancer cell lines, and human breast tumors (377, 378). Ishikawa et al. (378) noted that there may be differences in patterns of leptin expression between malignant and benign tissues – normal breast displayed weak leptin staining whereas malignant breast cells typically displayed strong staining, similar to the levels seen in adjacent adipocytes. Caldefie-Chezet et al. (379) did not observe leptin expression in normal tissue from healthy breasts, but did observe leptin expression in phenotypically normal glands adjacent to tumor in affected breasts. Leptin receptor isoforms are present in human breast tumors and breast cancer cell lines (319, 320, 322, 378), but were not observed in normal human breast tissue (378).

There is growing evidence that leptin may play a role in the development of normal and cancerous breast tissue. Normal mammary growth (ductal branching and development) is impaired in leptin-deficient and leptin receptor–deficient mice (322, 380). Work by Cleary et al. (380) showed that transgenic TGF-α/LEP^{ob} LEP^{ob} genetically obese mice did not experience spontaneous mammary tumors, compared to 58 tumors in transgenic TGF-α/LEP^{+} LEP^{+} homozygotes and 63 tumors in TGF-α/LEP^{ob} LEP^{+} heterozygotes during the same 2 year observation period. Similar results were obtained for TGF-α/LEPR^{db} LEPR^{db} genetically

obese mice, indicating that deficiencies in the leptin ligand-receptor unit inhibit normal and neoplastic mammary growth (380).

In vitro, leptin stimulates proliferation of normal and cancerous cells. In MCF-7 and T47D breast cancer cell lines, leptin stimulates STAT3 and MAP kinase signaling (319, 320). Leptin also increases cell proliferation and DNA synthesis in MCF-7, T47D, and HBL100 cell lines (319-322), and increases anchorage-independent growth in malignant (T47D) but not normal (HBL100) breast cell lines (322). In Catalano et al. (259), leptin enhanced aromatase expression and activity in MCF-7 cells, suggesting that the effect of leptin expression on cell proliferation may be mediated by estrogen receptor signaling.

Reports of associations between serum leptin and breast cancer do not reveal a consistent trend. Studies in Chinese and Taiwanese populations reported that leptin was higher in breast cancer cases compared to controls (381, 382), but in Petridou et al. (383) leptin levels were significantly lower in premenopausal cases compared to controls and there was no difference among postmenopausal subjects. Mantzoros et al. (384) reported no association between mean leptin levels and DCIS in premenopausal women. Some of the inconsistency in the studies mentioned above may be due to the measurement of leptin after breast cancer diagnosis in cases. A Swedish nested case-control study measured leptin levels prospectively and found no association between leptin levels and breast cancer (385).

Several studies indicate that polymorphisms in the leptin and leptin receptor genes may have an effect on serum leptin levels, leptin receptor levels, and breast cancer risk. Snoussi et al. (386) reported that the leptin -2548 G/A polymorphism was associated with breast cancer in a dose-dependent manner, but in Cleveland et al. (387) only the AA genotype was associated with breast cancer convincingly. Yiannakouris et al. (388) reported

that the LEP -2548 polymorphism was associated with plasma leptin receptor levels and leptin/leptin receptor ratios in women, but not in men. *In vitro* functional assays of this SNP have not been reported so it is unclear whether these associations are directly related to the -2548 G/A SNP or other genetic polymorphisms in linkage disequilibrium with the -2548 locus.

Leptin receptor polymorphism codon 109 RR variant homozygotes had higher serum leptin levels compared to codon 109 KR heterozygotes in healthy Korean controls, but the K109R polymorphism was not associated with breast cancer (389). In van Rossum et al. (390), the 109R variant was associated with higher leptin levels among subjects who had gained weight over the course of the study, but there was no difference in leptin levels by genotype among subjects with stable weight. Some studies have reported that the LEPR codon 223R variant is associated with higher serum leptin levels postmenopausal women (391), and with breast cancer (382, 386). The LEPR codon Q223R polymorphism was not associated with breast cancer in two other studies (387, 389). Woo et al. (389) also reported no association between breast cancer and LEPR SNPs K656N and P1019P.

Some LEPR polymorphisms are also associated with obesity. Clement et al. (392) described a rare LEPR mutation (exon 16 G → A) that leads to early onset morbid obesity in homozygotes. In other studies, the LEPR amino acid change Q223R was associated with obesity among Greek men and women and British women (391, 393), but K109R and K656N were not (393). LEPR polymorphisms at codons K109R, K204R, Q223R, and K656N were not associated with obesity in Danish men (394).

1.9.4 Adiponectin (ADIPOQ)

Adiponectin is a hormone produced by mature adipocytes that plays a role in the

inflammatory response, countering the effects of insulin resistance, and inhibiting angiogenesis. Unlike other adipocytokines, plasma adiponectin and adiponectin mRNA levels are reduced in obese and insulin resistant subjects and increase with weight loss (331, 395, 396). There is evidence that adiponectin and other adipocytokines have negative effects on each other. Dietze-Schroeder et al. (397) reported that adiponectin inhibits the secretion of other adipocytokines by adipocytes and may inhibit insulin resistance through this action. In Bruun et al. (396) adiponectin expression was reduced in adipose tissue when IL6/IL6-R or TNF-alpha were added. Also, adiponectin inhibited macrophage production of TNF-alpha *in vitro* (398).

Korner et al. (399) found that adiponectin is expressed in breast tissue –the highest levels were in tumor-adjacent adipose tissue and lowest in breast tumor tissue. However, in Takahata et al. (400), adiponectin was only detected in axillary adipose tissue, and not breast tumor or normal breast tissue. Both studies detected adiponectin receptors in breast normal and tumor tissue (399, 400).

Low circulating levels of adiponectin have been associated with many cancers that are also associated with obesity, including breast cancer (381, 399, 401-403). There are conflicting reports about whether the relationship is stronger in some subgroups of women. Korner et al. (399) reported that the protective effect of adiponectin was stronger in premenopausal and obese women, Tworoger et al. (404) reported that the protective effect was observed for postmenopausal but not premenopausal women, and Miyoshi et al. (403) observed an association in both premenopausal and postmenopausal women. In Fredriksson et al. (405), serum adiponectin levels were correlated with visceral but not subcutaneous adiponectin expression levels, suggesting that central obesity, not BMI, may be more

relevant to the relationship between adiponectin and breast cancer. Kang et al. (406) reported no association between adiponectin levels and breast cancer.

Several groups have shown that adiponectin inhibits cell growth and proliferation *in vitro*. Adiponectin inhibited FGF2-stimulated endothelial cell proliferation and VEGF-stimulated cell migration, and displayed anti-angiogenic effects in mouse models, possibly by inducing tumor vessel apoptosis through the activation of caspase 8 (407). In MCF-7 cells, the addition of adiponectin reduced cyclin D1 and c-myc expression, and suppressed the proliferative effects of estradiol (324, 325). There is also some evidence that prolonged exposure to adiponectin could induce apoptosis in MCF-7 cells (324), although others found that adiponectin had no effect on apoptosis in MCF-7 or T47-D cell lines (325, 399). In Kang et al. (408), adiponectin's antiproliferative, growth arrest and apoptotic effects were seen only in ER-negative cells and not in ER-positive cells. Differences in results could be due to experimental conditions. In experiments by Pfeiler et al. (409) increases in proliferation and apoptosis were only observed when both adiponectin and 17-beta estradiol were added to the cell culture medium.

Studies have shown that two adiponectin polymorphisms have functional effects on gene expression. In luciferase reporter assays, the adiponectin -11377G variant produced lower promoter activity, and the -11391A variant construct increased promoter activity (410). Two other variants (at nucleotides +45 and +276) that are in linkage disequilibrium with alleles at -11377 and -11391 have been associated with serum adiponectin levels, obesity, and diabetes. In lean individuals the +276T carriers was associated with insulin resistance phenotype (411). According to Fredriksson et al.'s (405) study of obese subjects, changes in adiponectin expression due to +276 G/T genotype occur only in visceral fat, not

subcutaneous fat tissue. There are no functional studies of the +45 or +276 polymorphisms, and their associations with adiponectin and obesity may be due to the functional effects of -11377 C/G and -11391 G/A. Kaklamani et al. (412) reported that the +45 T/G GT and GG genotypes were inversely associated with breast cancer, and +276 G/T GG and TT genotypes were positively associated with breast cancer.

Some studies have reported interaction between adiponectin and TNF-alpha genotypes and adipocytokine serum levels. For example, serum adiponectin levels in Spanish men and women were associated with TNF -308 G/A genotype among ADIPOQ +45 G allele carriers, but adiponectin levels did not differ by TNF -308 G/A genotype among ADIPOQ +45 TT homozygotes (413).

1.10 Critique of current research

Research into the association between common genetic variation and cancer risk is advancing rapidly. Technological advances have allowed for efficient multi-marker genotyping that has resulted in a large number of studies analyzing associations between SNPs on breast cancer. Current research has many positive aspects. The low cost of genotyping has allowed for analyses to be conducted in large population-based studies (107, 109, 141, 142). Also, many studies that were implemented years ago collected blood in anticipation of the genotyping analyses that are possible today. As a result, these studies have facilitated the identification of additional low penetrance alleles associated with breast cancer, particularly through the use of GWAS.

Still, there are several areas in which different approaches might improve our ability to identify breast cancer susceptibility alleles. First, studies must recognize that breast cancer

is a heterogeneous disease, and that basing models on a single outcome of 'breast cancer' may mask effects that are not similar across all breast cancer subtypes. While an increasing number of studies are stratifying results by estrogen-receptor status, menopausal status, or clinical markers, it is rare that SNP associations are evaluated when stratified on the joint status of more than two characteristics. As Kristensen and Borresen-Dale have noted (181), refining the tumor subtype is an efficient way of detecting breast cancer-associated SNPs with modest effect than having a very large sample size.

Second, early candidate SNP studies did not analyze the effect of haplotypes on breast cancer risk, though haplotype analysis has become more common (208, 210, 230, 350, 374, 414-417). Haplotypes composed of closely-spaced SNPs are likely to represent a meaningful biological unit, and alleles in *cis* position can have synergistic effects (418). Furthermore, haplotype analysis can reduce the number of independent statistical tests performed, improving statistical power (419, 420). When multiple SNPs in the same gene region are being genotyped, use of haplotypes can reduce the number of associations evaluated because haplotypes consist of those allelic combinations that actually occur in the study population, instead of testing all possible SNP interactions. Since common genetic variants are hypothesized to have small effects, the increased power provided by haplotype analysis could aid researchers in identifying relevant allelic interactions and avoiding type II error.

Third, few of the studies examining the candidate genes in this dissertation included a sizable proportion of African-American women. This limits generalizability of their results to African-American women, with regard to the prevalence of the alleles studied and characteristics of the SNP association with breast cancer (for example, if the association is

strongest within particular stratum of tumor characteristic or prognostic factor).

This dissertation examined the association between functional and tag SNPs in estrogen synthesis and signaling-related (CYP19A1, ESR1, HSD3B1, HSD17B2, PGR, and SHBG) and adipocytokine-related (ADIPOQ, IL6, LEP, LEPR, and TNF) candidate genes and the luminal A and basal-like intrinsic molecular subtypes of breast cancer in a population of white and African-American women. The main hypothesis was that analyzing the data according to intrinsic subtype would make it easier to identify moderate effects that act in some but not all breast tumor subtypes. This approach is particularly relevant since disease-associated polymorphisms are expected to have small effects under the polygenic hypothesis.

In addition to estimation of genotype and haplotype main effects, potential gene-environment interaction was explored. I hypothesized that the effects of SNPs in genes associated with the biologic pathways associated with these subtypes may differ based on whether the risk factor is present or absent. Breast cancer is a complex disease that involves alterations in multiple cellular processes. Assessment of gene-environment interaction is one way to evaluate how genetic and non-genetic component causes might combine to affect disease risk.

1.11 References

1. Jemal A, Siegel R, Ward E, Hao Y, Xu J, Thun MJ. Cancer statistics, 2009. CA Cancer J Clin 2009;59(4):225-49.

2. Horner MJ RL, Krapcho M, Neyman N, Aminou R, Howlader N, Altekruse SF, Feuer EJ, Huang L, Mariotto A, Miller BA, Lewis DR, Eisner MP, Stinchcomb DG, Edwards BK. SEER Cancer Statistics Review, 1975-2006. Bethesda, MD: National Cancer Institute; 2009.

3. Anderson WF, Reiner AS, Matsuno RK, Pfeiffer RM. Shifting breast cancer trends in the United States. J Clin Oncol 2007;25(25):3923-9.

4. Anderson WF, Chatterjee N, Ershler WB, Brawley OW. Estrogen receptor breast cancer phenotypes in the Surveillance, Epidemiology, and End Results database. Breast Cancer Res Treat 2002;76(1):27-36.

5. Yasui Y, Potter JD. The shape of age-incidence curves of female breast cancer by hormone-receptor status. Cancer Causes Control 1999;10(5):431-7.

6. Jemal A, Thun MJ, Ries LA, Howe HL, Weir HK, Center MM, et al. Annual report to the nation on the status of cancer, 1975-2005, featuring trends in lung cancer, tobacco use, and tobacco control. J Natl Cancer Inst 2008;100(23):1672-94.

7. Smigal C, Jemal A, Ward E, Cokkinides V, Smith R, Howe HL, et al. Trends in breast cancer by race and ethnicity: update 2006. CA Cancer J Clin 2006;56(3):168-83.

8. Joslyn SA, Foote ML, Nasseri K, Coughlin SS, Howe HL. Racial and ethnic disparities in breast cancer rates by age: NAACCR Breast Cancer Project. Breast Cancer Res Treat 2005;92(2):97-105.

9. Furberg H, Millikan R, Dressler L, Newman B, Geradts J. Tumor characteristics in African American and white women. Breast Cancer Res Treat 2001;68(1):33-43.

10. Field TS, Buist DS, Doubeni C, Enger S, Fouayzi H, Hart G, et al. Disparities and survival among breast cancer patients. J Natl Cancer Inst Monogr 2005(35):88-95.

11. Amend K, Hicks D, Ambrosone CB. Breast cancer in African-American women: differences in tumor biology from European-American women. Cancer Res 2006;66(17):8327-30.

12. Morris GJ, Naidu S, Topham AK, Guiles F, Xu Y, McCue P, et al. Differences in breast carcinoma characteristics in newly diagnosed African-American and Caucasian patients: a single-institution compilation compared with the National Cancer Institute's Surveillance, Epidemiology, and End Results database. Cancer

2007;110(4):876-84.

13. Yankaskas BC, Gill KS. Diagnostic mammography performance and race: outcomes in Black and White women. Cancer 2005;104(12):2671-81.

14. Gill KS, Yankaskas BC. Screening mammography performance and cancer detection among black women and white women in community practice. Cancer 2004;100(1):139-48.

15. Howe HL, Wu X, Ries LA, Cokkinides V, Ahmed F, Jemal A, et al. Annual report to the nation on the status of cancer, 1975-2003, featuring cancer among U.S. Hispanic/Latino populations. Cancer 2006;107(8):1711-42.

16. Carey LA, Perou CM, Livasy CA, Dressler LG, Cowan D, Conway K, et al. Race, breast cancer subtypes, and survival in the Carolina Breast Cancer Study. Jama 2006;295(21):2492-502.

17. Millikan RC, Newman B, Tse CK, Moorman PG, Conway K, Dressler LG, et al. Epidemiology of basal-like breast cancer. Breast Cancer Res Treat 2008;109(1):123-39.

18. Ries L, Melbert D, Krapcho M, Mariotto A, Miller B, Feuer E, et al. SEER Cancer Statistics Review, 1975-2004. 2007 [cited; Available from: http://seer.cancer.gov/csr/1975_2004/

19. Familial breast cancer: collaborative reanalysis of individual data from 52 epidemiological studies including 58,209 women with breast cancer and 101,986 women without the disease. Lancet 2001;358(9291):1389-99.

20. Antoniou A, Pharoah PD, Narod S, Risch HA, Eyfjord JE, Hopper JL, et al. 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;72(5):1117-30.

21. Hankinson S, Hunter D. Breast Cancer. In: Adami H-O, Hunter D, Trichopoulos D, editors. Textbook of Cancer Epidemiology. New York: Oxford University Press, Inc.; 2002. p. 301-339.

22. Innes K, Byers T, Schymura M. Birth characteristics and subsequent risk for breast cancer in very young women. Am J Epidemiol 2000;152(12):1121-8.

23. Sanderson M, Williams MA, Malone KE, Stanford JL, Emanuel I, White E, et al. Perinatal factors and risk of breast cancer. Epidemiology 1996;7(1):34-7.

24. Vatten LJ, Nilsen TI, Tretli S, Trichopoulos D, Romundstad PR. Size at birth and risk of breast cancer: prospective population-based study. Int J Cancer 2005;114(3):461-4.

25. Michels KB, Xue F. Role of birthweight in the etiology of breast cancer. Int J Cancer 2006;119(9):2007-25.

26. Colditz GA, Willett WC, Stampfer MJ, Hennekens CH, Rosner B, Speizer FE. Parental age at birth and risk of breast cancer in daughters: a prospective study among US women. Cancer Causes Control 1991;2(1):31-6.

27. Ekbom A, Hsieh CC, Lipworth L, Adami HQ, Trichopoulos D. Intrauterine environment and breast cancer risk in women: a population-based study. J Natl Cancer Inst 1997;89(1):71-6.

28. Zhang Y, Cupples LA, Rosenberg L, Colton T, Kreger BE. Parental ages at birth in relation to a daughter's risk of breast cancer among female participants in the Framingham Study (United States). Cancer Causes Control 1995;6(1):23-9.

29. Hodgson ME, Newman B, Millikan RC. Birthweight, parental age, birth order and breast cancer risk in African-American and white women: a population-based case-control study. Breast Cancer Res 2004;6(6):R656-67.

30. van den Brandt PA, Spiegelman D, Yaun SS, Adami HO, Beeson L, Folsom AR, et al. Pooled analysis of prospective cohort studies on height, weight, and breast cancer risk. Am J Epidemiol 2000;152(6):514-27.

31. Lahmann PH, Hoffmann K, Allen N, van Gils CH, Khaw KT, Tehard B, et al. Body size and breast cancer risk: findings from the European Prospective Investigation into Cancer And Nutrition (EPIC). Int J Cancer 2004;111(5):762-71.

32. Ng EH, Gao F, Ji CY, Ho GH, Soo KC. Risk factors for breast carcinoma in Singaporean Chinese women: the role of central obesity. Cancer 1997;80(4):725-31.

33. Kaaks R, Van Noord PA, Den Tonkelaar I, Peeters PH, Riboli E, Grobbee DE. Breast-cancer incidence in relation to height, weight and body-fat distribution in the Dutch "DOM" cohort. Int J Cancer 1998;76(5):647-51.

34. Manjer J, Kaaks R, Riboli E, Berglund G. Risk of breast cancer in relation to anthropometry, blood pressure, blood lipids and glucose metabolism: a prospective study within the Malmo Preventive Project. Eur J Cancer Prev 2001;10(1):33-42.

35. Trentham-Dietz A, Newcomb PA, Storer BE, Longnecker MP, Baron J, Greenberg ER, et al. Body size and risk of breast cancer. Am J Epidemiol 1997;145(11):1011-9.

36. Peacock SL, White E, Daling JR, Voigt LF, Malone KE. Relation between obesity and breast cancer in young women. Am J Epidemiol 1999;149(4):339-46.

37. Shin A, Matthews CE, Shu XO, Gao YT, Lu W, Gu K, et al. Joint effects of body size, energy intake, and physical activity on breast cancer risk. Breast Cancer Res

Treat 2009;113(1):153-61.

38. Tehard B, Clavel-Chapelon F. Several anthropometric measurements and breast cancer risk: results of the E3N cohort study. Int J Obes (Lond) 2006;30(1):156-63.

39. Slattery ML, Sweeney C, Edwards S, Herrick J, Baumgartner K, Wolff R, et al. Body size, weight change, fat distribution and breast cancer risk in Hispanic and non-Hispanic white women. Breast Cancer Res Treat 2007;102(1):85-101.

40. Wu AH, Yu MC, Tseng CC, Pike MC. Body size, hormone therapy and risk of breast cancer in Asian-American women. Int J Cancer 2007;120(4):844-52.

41. Harvie M, Hooper L, Howell AH. Central obesity and breast cancer risk: a systematic review. Obes Rev 2003;4(3):157-73.

42. Connolly BS, Barnett C, Vogt KN, Li T, Stone J, Boyd NF. A meta-analysis of published literature on waist-to-hip ratio and risk of breast cancer. Nutr Cancer 2002;44(2):127-38.

43. Bruning PF, Bonfrer JM, Hart AA, van Noord PA, van der Hoeven H, Collette HJ, et al. Body measurements, estrogen availability and the risk of human breast cancer: a case-control study. Int J Cancer 1992;51(1):14-9.

44. Swanson CA, Coates RJ, Schoenberg JB, Malone KE, Gammon MD, Stanford JL, et al. Body size and breast cancer risk among women under age 45 years. Am J Epidemiol 1996;143(7):698-706.

45. Galanis DJ, Kolonel LN, Lee J, Le Marchand L. Anthropometric predictors of breast cancer incidence and survival in a multi-ethnic cohort of female residents of Hawaii, United States. Cancer Causes Control 1998;9(2):217-24.

46. 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;46(5):796-800.

47. Baer HJ, Rich-Edwards JW, Colditz GA, Hunter DJ, Willett WC, Michels KB. Adult height, age at attained height, and incidence of breast cancer in premenopausal women. Int J Cancer 2006;119(9):2231-5.

48. Brisson J, Merletti F, Sadowsky NL, Twaddle JA, Morrison AS, Cole P. Mammographic features of the breast and breast cancer risk. Am J Epidemiol 1982;115(3):428-37.

49. Threatt B, Norbeck JM, Ullman NS, Kummer R, Roselle P. Association between mammographic parenchymal pattern classification and incidence of breast cancer. Cancer 1980;45(10):2550-6.

50. Boyd NF, Lockwood GA, Byng JW, Tritchler DL, Yaffe MJ. Mammographic densities and breast cancer risk. Cancer Epidemiol Biomarkers Prev 1998;7(12):1133-44.

51. McCormack VA, dos Santos Silva I. Breast density and parenchymal patterns as markers of breast cancer risk: a meta-analysis. Cancer Epidemiol Biomarkers Prev 2006;15(6):1159-69.

52. Vachon CM, van Gils CH, Sellers TA, Ghosh K, Pruthi S, Brandt KR, et al. Mammographic density, breast cancer risk and risk prediction. Breast Cancer Res 2007;9(6):217.

53. Maskarinec G, Pagano I, Lurie G, Kolonel LN. A longitudinal investigation of mammographic density: the multiethnic cohort. Cancer Epidemiol Biomarkers Prev 2006;15(4):732-9.

54. Vachon CM, Pankratz VS, Scott CG, Maloney SD, Ghosh K, Brandt KR, et al. Longitudinal trends in mammographic percent density and breast cancer risk. Cancer Epidemiol Biomarkers Prev 2007;16(5):921-8.

55. Monninkhof EM, Elias SG, Vlems FA, van der Tweel I, Schuit AJ, Voskuil DW, et al. Physical activity and breast cancer: a systematic review. Epidemiology 2007;18(1):137-57.

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

57. Pike MC, Spicer DV, Dahmoush L, Press MF. Estrogens, progestogens, normal breast cell proliferation, and breast cancer risk. Epidemiol Rev 1993;15(1):17-35.

58. Russo J, Russo IH. The role of estrogen in the initiation of breast cancer. J Steroid Biochem Mol Biol 2006;102(1-5):89-96.

59. Bernstein L. Epidemiology of endocrine-related risk factors for breast cancer. J Mammary Gland Biol Neoplasia 2002;7(1):3-15.

60. Clavel-Chapelon F, Launoy G, Auquier A, Gairard B, Bremond A, Piana L, et al. Reproductive factors and breast cancer risk. Effect of age at diagnosis. Ann Epidemiol 1995;5(4):315-20.

61. Li CI, Malone KE, Daling JR, Potter JD, Bernstein L, Marchbanks PA, et al. Timing of Menarche and First Full-Term Birth in Relation to Breast Cancer Risk. Am J Epidemiol 2007.

62. 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;360(9328):187-95.

63. Albrektsen G, Heuch I, Kvale G. The short-term and long-term effect of a pregnancy on breast cancer risk: a prospective study of 802,457 parous Norwegian women. Br J Cancer 1995;72(2):480-4.

64. Bruzzi P, Negri E, La Vecchia C, Decarli A, Palli D, Parazzini F, et al. Short term increase in risk of breast cancer after full term pregnancy. Bmj 1988;297(6656):1096-8.

65. Hall IJ, Moorman PG, Millikan RC, Newman B. Comparative analysis of breast cancer risk factors among African-American women and White women. Am J Epidemiol 2005;161(1):40-51.

66. Ursin G, Bernstein L, Wang Y, Lord SJ, Deapen D, Liff JM, et al. Reproductive factors and risk of breast carcinoma in a study of white and African-American women. Cancer 2004;101(2):353-62.

67. Bernstein L, Teal CR, Joslyn S, Wilson J. Ethnicity-related variation in breast cancer risk factors. Cancer 2003;97(1 Suppl):222-9.

68. Pathak DR. Dual effect of first full term pregnancy on breast cancer risk: empirical evidence and postulated underlying biology. Cancer Causes Control 2002;13(4):295-8.

69. Pathak DR, Osuch JR, He J. Breast carcinoma etiology: current knowledge and new insights into the effects of reproductive and hormonal risk factors in black and white populations. Cancer 2000;88(5 Suppl):1230-8.

70. Russo J, Tay LK, Ciocca DR, Russo IH. Molecular and cellular basis of the mammary gland susceptibility to carcinogenesis. Environ Health Perspect 1983;49:185-99.

71. Polyak K. Pregnancy and breast cancer: the other side of the coin. Cancer Cell 2006;9(3):151-3.

72. Cnattingius S, Torrang A, Ekbom A, Granath F, Petersson G, Lambe M. Pregnancy characteristics and maternal risk of breast cancer. Jama 2005;294(19):2474-80.

73. Innes KE, Byers TE. First pregnancy characteristics and subsequent breast cancer risk among young women. Int J Cancer 2004;112(2):306-11.

74. Wohlfahrt J, Melbye M. Maternal risk of breast cancer and birth characteristics of offspring by time since birth. Epidemiology 1999;10(4):441-4.

75. Brinton LA, Daling JR, Liff JM, Schoenberg JB, Malone KE, Stanford JL, et al. Oral

contraceptives and breast cancer risk among younger women. J Natl Cancer Inst 1995;87(11):827-35.

76. Collaborative Group on Hormonal Factors in Breast Cancer. Breast cancer and hormonal contraceptives: collaborative reanalysis of individual data on 53 297 women with breast cancer and 100 239 women without breast cancer from 54 epidemiological studies. Lancet 1996;347(9017):1713-27.

77. Wingo PA, Lee NC, Ory HW, Beral V, Peterson HB, Rhodes P. Age-specific differences in the relationship between oral contraceptive use and breast cancer. Obstet Gynecol 1991;78(2):161-70.

78. Chu KC, Anderson WF, Fritz A, Ries LA, Brawley OW. Frequency distributions of breast cancer characteristics classified by estrogen receptor and progesterone receptor status for eight racial/ethnic groups. Cancer 2001;92(1):37-45.

79. Marchbanks PA, McDonald JA, Wilson HG, Folger SG, Mandel MG, Daling JR, et al. Oral contraceptives and the risk of breast cancer. N Engl J Med 2002;346(26):2025-32.

80. Breast cancer and combined oral contraceptives: results from a multinational study. The WHO Collaborative Study of Neoplasia and Steroid Contraceptives. Br J Cancer 1990;61(1):110-9.

81. Oral contraceptive use and breast cancer risk in young women. UK National Case-Control Study Group. Lancet 1989;1(8645):973-82.

82. Chen CL, Weiss NS, Newcomb P, Barlow W, White E. Hormone replacement therapy in relation to breast cancer. Jama 2002;287(6):734-41.

83. Beral V. Breast cancer and hormone-replacement therapy in the Million Women Study. Lancet 2003;362(9382):419-27.

84. 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;350(9084):1047-59.

85. Dumitrescu RG, Shields PG. The etiology of alcohol-induced breast cancer. Alcohol 2005;35(3):213-25.

86. Hamajima N, Hirose K, Tajima K, Rohan T, Calle EE, Heath CW, Jr., et al. Alcohol, tobacco and breast cancer--collaborative reanalysis of individual data from 53 epidemiological studies, including 58,515 women with breast cancer and 95,067 women without the disease. Br J Cancer 2002;87(11):1234-45.

87. Tjonneland A, Christensen J, Olsen A, Stripp C, Thomsen BL, Overvad K, et al. Alcohol intake and breast cancer risk: the European Prospective Investigation into Cancer and Nutrition (EPIC). Cancer Causes Control 2007;18(4):361-73.

88. Terry MB, Zhang FF, Kabat G, Britton JA, Teitelbaum SL, Neugut AI, et al. Lifetime alcohol intake and breast cancer risk. Ann Epidemiol 2006;16(3):230-40.

89. Hansen J. Increased breast cancer risk among women who work predominantly at night. Epidemiology 2001;12(1):74-7.

90. Davis S, Mirick DK, Stevens RG. Night shift work, light at night, and risk of breast cancer. J Natl Cancer Inst 2001;93(20):1557-62.

91. Schernhammer ES, Laden F, Speizer FE, Willett WC, Hunter DJ, Kawachi I, et al. Rotating night shifts and risk of breast cancer in women participating in the nurses' health study. J Natl Cancer Inst 2001;93(20):1563-8.

92. O'Leary ES, Schoenfeld ER, Stevens RG, Kabat GC, Henderson K, Grimson R, et al. Shift work, light at night, and breast cancer on Long Island, New York. Am J Epidemiol 2006;164(4):358-66.

93. Davis S, Mirick DK. Circadian disruption, shift work and the risk of cancer: a summary of the evidence and studies in Seattle. Cancer Causes Control 2006;17(4):539-45.

94. Stevens RG, Rea MS. Light in the built environment: potential role of circadian disruption in endocrine disruption and breast cancer. Cancer Causes Control 2001;12(3):279-87.

95. Ronckers CM, Erdmann CA, Land CE. Radiation and breast cancer: a review of current evidence. Breast Cancer Res 2005;7(1):21-32.

96. Lichtenstein P, Holm NV, Verkasalo PK, Iliadou A, Kaprio J, Koskenvuo M, et al. Environmental and heritable factors in the causation of cancer--analyses of cohorts of twins from Sweden, Denmark, and Finland. N Engl J Med 2000;343(2):78-85.

97. Peto J, Mack TM. High constant incidence in twins and other relatives of women with breast cancer. Nat Genet 2000;26(4):411-4.

98. Mack TM, Hamilton AS, Press MF, Diep A, Rappaport EB. Heritable breast cancer in twins. Br J Cancer 2002;87(3):294-300.

99. Newman B, Austin MA, Lee M, King MC. Inheritance of human breast cancer: evidence for autosomal dominant transmission in high-risk families. Proc Natl Acad Sci U S A 1988;85(9):3044-8.

100. Wooster R, Neuhausen SL, Mangion J, Quirk Y, Ford D, Collins N, et al. Localization of a breast cancer susceptibility gene, BRCA2, to chromosome 13q12-13. Science 1994;265(5181):2088-90.

101. Malone KE, Daling JR, Doody DR, Hsu L, Bernstein L, Coates RJ, et al. Prevalence and Predictors of BRCA1 and BRCA2 Mutations in a Population-Based Study of Breast Cancer in White and Black American Women Ages 35 to 64 Years. Cancer Res 2006;66(16):8297-308.

102. Walsh T, King MC. Ten genes for inherited breast cancer. Cancer Cell 2007;11(2):103-5.

103. Smith P, McGuffog L, Easton DF, Mann GJ, Pupo GM, Newman B, et al. A genome wide linkage search for breast cancer susceptibility genes. Genes Chromosomes Cancer 2006;45(7):646-55.

104. Antoniou AC, Pharoah PD, McMullan G, Day NE, Ponder BA, Easton D. Evidence for further breast cancer susceptibility genes in addition to BRCA1 and BRCA2 in a population-based study. Genet Epidemiol 2001;21(1):1-18.

105. Pharoah PD, Antoniou A, Bobrow M, Zimmern RL, Easton DF, Ponder BA. Polygenic susceptibility to breast cancer and implications for prevention. Nat Genet 2002;31(1):33-6.

106. Antoniou AC, Pharoah PD, McMullan G, Day NE, Stratton MR, Peto J, et al. A comprehensive model for familial breast cancer incorporating BRCA1, BRCA2 and other genes. Br J Cancer 2002;86(1):76-83.

107. Easton DF, Pooley KA, Dunning AM, Pharoah PD, Thompson D, Ballinger DG, et al. Genome-wide association study identifies novel breast cancer susceptibility loci. Nature 2007;447(7148):1087-93.

108. Hunter DJ, Kraft P, Jacobs KB, Cox DG, Yeager M, Hankinson SE, et al. A genome-wide association study identifies alleles in FGFR2 associated with risk of sporadic postmenopausal breast cancer. Nat Genet 2007;39(7):870-4.

109. Stacey SN, Manolescu A, Sulem P, Rafnar T, Gudmundsson J, Gudjonsson SA, et al. Common variants on chromosomes 2q35 and 16q12 confer susceptibility to estrogen receptor-positive breast cancer. Nat Genet 2007;39(7):865-9.

110. Nam RK, Zhang WW, Loblaw DA, Klotz LH, Trachtenberg J, Jewett MA, et al. A genome-wide association screen identifies regions on chromosomes 1q25 and 7p21 as risk loci for sporadic prostate cancer. Prostate Cancer Prostatic Dis 2007.

111. Zanke BW, Greenwood CM, Rangrej J, Kustra R, Tenesa A, Farrington SM, et al. Genome-wide association scan identifies a colorectal cancer susceptibility locus on

chromosome 8q24. Nat Genet 2007;39(8):989-94.

112. Gudmundsson J, Sulem P, Steinthorsdottir V, Bergthorsson JT, Thorleifsson G, Manolescu A, et al. Two variants on chromosome 17 confer prostate cancer risk, and the one in TCF2 protects against type 2 diabetes. Nat Genet 2007;39(8):977-83.

113. Gudmundsson J, Sulem P, Manolescu A, Amundadottir LT, Gudbjartsson D, Helgason A, et al. Genome-wide association study identifies a second prostate cancer susceptibility variant at 8q24. Nat Genet 2007;39(5):631-7.

114. Yeager M, Orr N, Hayes RB, Jacobs KB, Kraft P, Wacholder S, et al. Genome-wide association study of prostate cancer identifies a second risk locus at 8q24. Nat Genet 2007;39(5):645-9.

115. Spinola M, Leoni VP, Galvan A, Korsching E, Conti B, Pastorino U, et al. Genome-wide single nucleotide polymorphism analysis of lung cancer risk detects the KLF6 gene. Cancer Lett 2007;251(2):311-6.

116. Stacey SN, Manolescu A, Sulem P, Thorlacius S, Gudjonsson SA, Jonsson GF, et al. Common variants on chromosome 5p12 confer susceptibility to estrogen receptor-positive breast cancer. Nat Genet 2008;40(6):703-6.

117. Zheng W, Long J, Gao YT, Li C, Zheng Y, Xiang YB, et al. Genome-wide association study identifies a new breast cancer susceptibility locus at 6q25.1. Nat Genet 2009;41(3):324-8.

118. Thomas G, Jacobs KB, Kraft P, Yeager M, Wacholder S, Cox DG, et al. A multistage genome-wide association study in breast cancer identifies two new risk alleles at 1p11.2 and 14q24.1 (RAD51L1). Nat Genet 2009;41(5):579-84.

119. Ahmed S, Thomas G, Ghoussaini M, Healey CS, Humphreys MK, Platte R, et al. Newly discovered breast cancer susceptibility loci on 3p24 and 17q23.2. Nat Genet 2009;41(5):585-90.

120. Zheng W, Cai Q, Signorello LB, Long J, Hargreaves MK, Deming SL, et al. Evaluation of 11 breast cancer susceptibility Loci in african-american women. Cancer Epidemiol Biomarkers Prev 2009;18(10):2761-4.

121. Mucci LA, Wedren S, Tamimi RM, Trichopoulos D, Adami HO. The role of gene-environment interaction in the aetiology of human cancer: examples from cancers of the large bowel, lung and breast. J Intern Med 2001;249(6):477-93.

122. Low YL, Dunning AM, Dowsett M, Luben RN, Khaw KT, Wareham NJ, et al. Implications of gene-environment interaction in studies of gene variants in breast cancer: an example of dietary isoflavones and the D356N polymorphism in the sex hormone-binding globulin gene. Cancer Res 2006;66(18):8980-3.

123. Zhang Y, Wise JP, Holford TR, Xie H, Boyle P, Zahm SH, et al. Serum polychlorinated biphenyls, cytochrome P-450 1A1 polymorphisms, and risk of breast cancer in Connecticut women. Am J Epidemiol 2004;160(12):1177-83.

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

125. Sturmer T, Wang-Gohrke S, Arndt V, Boeing H, Kong X, Kreienberg R, et al. Interaction between alcohol dehydrogenase II gene, alcohol consumption, and risk for breast cancer. Br J Cancer 2002;87(5):519-23.

126. Cartegni L, Chew SL, Krainer AR. Listening to silence and understanding nonsense: exonic mutations that affect splicing. Nat Rev Genet 2002;3(4):285-98.

127. Sauna ZE, Kimchi-Sarfaty C, Ambudkar SV, Gottesman MM. Silent polymorphisms speak: how they affect pharmacogenomics and the treatment of cancer. Cancer Res 2007;67(20):9609-12.

128. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, et al. Initial sequencing and analysis of the human genome. Nature 2001;409(6822):860-921.

129. Sachidanandam R, Weissman D, Schmidt SC, Kakol JM, Stein LD, Marth G, et al. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. Nature 2001;409(6822):928-33.

130. Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, et al. The sequence of the human genome. Science 2001;291(5507):1304-51.

131. Frazer KA, Ballinger DG, Cox DR, Hinds DA, Stuve LL, Gibbs RA, et al. A second generation human haplotype map of over 3.1 million SNPs. Nature 2007;449(7164):851-61.

132. The International HapMap Project. Nature 2003;426(6968):789-96.

133. Barrett JC, Cardon LR. Evaluating coverage of genome-wide association studies. Nat Genet 2006;38(6):659-62.

134. Reich DE, Cargill M, Bolk S, Ireland J, Sabeti PC, Richter DJ, et al. Linkage disequilibrium in the human genome. Nature 2001;411(6834):199-204.

135. Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, Blumenstiel B, et al. The structure of haplotype blocks in the human genome. Science 2002;296(5576):2225-9.

136. Daly MJ, Rioux JD, Schaffner SF, Hudson TJ, Lander ES. High-resolution haplotype structure in the human genome. Nat Genet 2001;29(2):229-32.

137. Voight BF, Kudaravalli S, Wen X, Pritchard JK. A map of recent positive selection in the human genome. PLoS Biol 2006;4(3):e72.

138. Sabeti PC, Varilly P, Fry B, Lohmueller J, Hostetter E, Cotsapas C, et al. Genome-wide detection and characterization of positive selection in human populations. Nature 2007;449(7164):913-8.

139. Coughlin SS, Piper M. Genetic polymorphisms and risk of breast cancer. Cancer Epidemiol Biomarkers Prev 1999;8(11):1023-32.

140. Dunning AM, Healey CS, Pharoah PD, Teare MD, Ponder BA, Easton DF. A systematic review of genetic polymorphisms and breast cancer risk. Cancer Epidemiol Biomarkers Prev 1999;8(10):843-54.

141. Breast Cancer Association C. Commonly studied single-nucleotide polymorphisms and breast cancer: results from the Breast Cancer Association Consortium. J Natl Cancer Inst 2006;98(19):1382-96.

142. Pharoah PD, Tyrer J, Dunning AM, Easton DF, Ponder BA. Association between common variation in 120 candidate genes and breast cancer risk. PLoS Genet 2007;3(3):e42.

143. Harris L, Fritsche H, Mennel R, Norton L, Ravdin P, Taube S, et al. American Society of Clinical Oncology 2007 Update of Recommendations for the Use of Tumor Markers in Breast Cancer. J Clin Oncol 2007;25(33):5287-312.

144. Bauer KR, Brown M, Cress RD, Parise CA, Caggiano V. Descriptive analysis of estrogen receptor (ER)-negative, progesterone receptor (PR)-negative, and HER2-negative invasive breast cancer, the so-called triple-negative phenotype: a population-based study from the California cancer Registry. Cancer 2007;109(9):1721-8.

145. Rakha EA, El-Sayed ME, Green AR, Lee AH, Robertson JF, Ellis IO. Prognostic markers in triple-negative breast cancer. Cancer 2007;109(1):25-32.

146. Reis-Filho JS, Tutt AN. Triple negative tumours: a critical review. Histopathology 2008;52(1):108-18.

147. Perou CM, Sorlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, et al. Molecular portraits of human breast tumours. Nature 2000;406(6797):747-52.

148. Sorlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proc Natl Acad Sci U S A 2001;98(19):10869-74.

149. Sorlie T, Tibshirani R, Parker J, Hastie T, Marron JS, Nobel A, et al. Repeated observation of breast tumor subtypes in independent gene expression data sets. Proc

Natl Acad Sci U S A 2003;100(14):8418-23.

150. Nielsen TO, Hsu FD, Jensen K, Cheang M, Karaca G, Hu Z, et al. Immunohistochemical and clinical characterization of the basal-like subtype of invasive breast carcinoma. Clin Cancer Res 2004;10(16):5367-74.

151. Abd El-Rehim DM, Ball G, Pinder SE, Rakha E, Paish C, Robertson JF, et al. High-throughput protein expression analysis using tissue microarray technology of a large well-characterised series identifies biologically distinct classes of breast cancer confirming recent cDNA expression analyses. Int J Cancer 2005;116(3):340-50.

152. Yu K, Lee CH, Tan PH, Tan P. Conservation of breast cancer molecular subtypes and transcriptional patterns of tumor progression across distinct ethnic populations. Clin Cancer Res 2004;10(16):5508-17.

153. Calza S, Hall P, Auer G, Bjohle J, Klaar S, Kronenwett U, et al. Intrinsic molecular signature of breast cancer in a population-based cohort of 412 patients. Breast Cancer Res 2006;8(4):R34.

154. Adebamowo CA, Famooto A, Ogundiran TO, Aniagwu T, Nkwodimmah C, Akang EE. Immunohistochemical and molecular subtypes of breast cancer in Nigeria. Breast Cancer Res Treat 2007;Aug 3 - epub ahead of print.

155. Nalwoga H, Arnes JB, Wabinga H, Akslen LA. Frequency of the basal-like phenotype in African breast cancer. Apmis 2007;115(12):1391-9.

156. Kim MJ, Ro JY, Ahn SH, Kim HH, Kim SB, Gong G. Clinicopathologic significance of the basal-like subtype of breast cancer: a comparison with hormone receptor and Her2/neu-overexpressing phenotypes. Hum Pathol 2006;37(9):1217-26.

157. Ishihara A, Tsuda H, Kitagawa K, Yoneda M, Shiraishi T. Morphological characteristics of basal-like subtype of breast carcinoma with special reference to cytopathological features. Breast Cancer 2009;16(3):179-85.

158. Huo D, Ikpatt F, Khramtsov A, Dangou JM, Nanda R, Dignam J, et al. Population differences in breast cancer: survey in indigenous african women reveals over-representation of triple-negative breast cancer. J Clin Oncol 2009;27(27):4515-21.

159. Fulford LG, Easton DF, Reis-Filho JS, Sofronis A, Gillett CE, Lakhani SR, et al. Specific morphological features predictive for the basal phenotype in grade 3 invasive ductal carcinoma of breast. Histopathology 2006;49(1):22-34.

160. Livasy CA, Karaca G, Nanda R, Tretiakova MS, Olopade OI, Moore DT, et al. Phenotypic evaluation of the basal-like subtype of invasive breast carcinoma. Mod Pathol 2006;19(2):264-71.

161. Rodriguez-Pinilla SM, Sarrio D, Honrado E, Hardisson D, Calero F, Benitez J, et al. Prognostic significance of basal-like phenotype and fascin expression in node-negative invasive breast carcinomas. Clin Cancer Res 2006;12(5):1533-9.

162. Collett K, Stefansson IM, Eide J, Braaten A, Wang H, Eide GE, et al. A basal epithelial phenotype is more frequent in interval breast cancers compared with screen detected tumors. Cancer Epidemiol Biomarkers Prev 2005;14(5):1108-12.

163. Bergamaschi A, Kim YH, Wang P, Sorlie T, Hernandez-Boussard T, Lonning PE, et al. Distinct patterns of DNA copy number alteration are associated with different clinicopathological features and gene-expression subtypes of breast cancer. Genes Chromosomes Cancer 2006;45(11):1033-40.

164. Foulkes WD, Stefansson IM, Chappuis PO, Begin LR, Goffin JR, Wong N, et al. Germline BRCA1 mutations and a basal epithelial phenotype in breast cancer. J Natl Cancer Inst 2003;95(19):1482-5.

165. Lakhani SR, Reis-Filho JS, Fulford L, Penault-Llorca F, van der Vijver M, Parry S, et al. Prediction of BRCA1 status in patients with breast cancer using estrogen receptor and basal phenotype. Clin Cancer Res 2005;11(14):5175-80.

166. Turner NC, Reis-Filho JS. Basal-like breast cancer and the BRCA1 phenotype. Oncogene 2006;25(43):5846-53.

167. Dabbs DJ, Chivukula M, Carter G, Bhargava R. Basal phenotype of ductal carcinoma in situ: recognition and immunohistologic profile. Mod Pathol 2006;19(11):1506-11.

168. Hannemann J, Velds A, Halfwerk JB, Kreike B, Peterse JL, van de Vijver MJ. Classification of ductal carcinoma in situ by gene expression profiling. Breast Cancer Res 2006;8(5):R61.

169. Livasy CA, Perou CM, Karaca G, Cowan DW, Maia D, Jackson S, et al. Identification of a basal-like subtype of breast ductal carcinoma in situ. Hum Pathol 2007;38(2):197-204.

170. Paredes J, Lopes N, Milanezi F, Schmitt FC. P-cadherin and cytokeratin 5: useful adjunct markers to distinguish basal-like ductal carcinomas in situ. Virchows Arch 2007;450(1):73-80.

171. Bryan BB, Schnitt SJ, Collins LC. Ductal carcinoma in situ with basal-like phenotype: a possible precursor to invasive basal-like breast cancer. Mod Pathol 2006;19(5):617-21.

172. Yang XR, Sherman ME, Rimm DL, Lissowska J, Brinton LA, Peplonska B, et al. Differences in risk factors for breast cancer molecular subtypes in a population-based study. Cancer Epidemiol Biomarkers Prev 2007;16(3):439-43.

173. Kurebayashi J, Moriya T, Ishida T, Hirakawa H, Kurosumi M, Akiyama F, et al. The prevalence of intrinsic subtypes and prognosis in breast cancer patients of different races. Breast 2007;16 Suppl 2:S72-7.

174. Banerjee S, Reis-Filho JS, Ashley S, Steele D, Ashworth A, Lakhani SR, et al. Basal-like breast carcinomas: clinical outcome and response to chemotherapy. J Clin Pathol 2006;59(7):729-35.

175. Yamamoto Y, Ibusuki M, Nakano M, Kawasoe T, Hiki R, Iwase H. Clinical significance of basal-like subtype in triple-negative breast cancer. Breast Cancer 2009;16(4):260-7.

176. Potemski P, Kusinska R, Watala C, Pluciennik E, Bednarek AK, Kordek R. Prognostic relevance of basal cytokeratin expression in operable breast cancer. Oncology 2005;69(6):478-85.

177. Jumppanen M, Gruvberger-Saal S, Kauraniemi P, Tanner M, Bendahl PO, Lundin M, et al. Basal-like phenotype is not associated with patient survival in estrogen-receptor-negative breast cancers. Breast Cancer Res 2007;9(1):R16.

178. Garcia-Closas M, Hall P, Nevanlinna H, Pooley K, Morrison J, Richesson DA, et al. Heterogeneity of breast cancer associations with five susceptibility loci by clinical and pathological characteristics. PLoS Genet 2008;4(4):e1000054.

179. Nordgard SH, Johansen FE, Alnaes GI, Naume B, Borresen-Dale AL, Kristensen VN. Genes harbouring susceptibility SNPs are differentially expressed in the breast cancer subtypes. Breast Cancer Res 2007;9(6):113.

180. Kristensen VN, Harada N, Yoshimura N, Haraldsen E, Lonning PE, Erikstein B, et al. Genetic variants of CYP19 (aromatase) and breast cancer risk. Oncogene 2000;19(10):1329-33.

181. Kristensen VN, Borresen-Dale AL. SNPs associated with molecular subtypes of breast cancer: on the usefulness of stratified Genome-wide Association Studies (GWAS) in the identification of novel susceptibility loci. Mol Oncol 2008;2(1):12-5.

182. Mitrunen K, Hirvonen A. Molecular epidemiology of sporadic breast cancer. The role of polymorphic genes involved in oestrogen biosynthesis and metabolism. Mutat Res 2003;544(1):9-41.

183. Gruber CJ, Tschugguel W, Schneeberger C, Huber JC. Production and actions of estrogens. N Engl J Med 2002;346(5):340-52.

184. Barbieri RL. The Breast. In: Strauss JF, Barbieri RL, editors. Yen and Jaffe's Reproductive Endocrinology. 5th ed. Philadelphia: Elsevier; 2004. p. 307-326.

185. Rhen T, Cidlowski JA. Steroid Hormone Action. In: Strauss JF, Barbieri RL, editors. Yen and Jaffe's Reproductive Endocrinology. 5th ed. Philadelphia: Elsevier; 2004. p. 155-174.

186. Payne AH, Hales DB. Overview of steroidogenic enzymes in the pathway from cholesterol to active steroid hormones. Endocr Rev 2004;25(6):947-70.

187. Miettinen M, Mustonen M, Poutanen M, Isomaa V, Wickman M, Soderqvist G, et al. 17Beta-hydroxysteroid dehydrogenases in normal human mammary epithelial cells and breast tissue. Breast Cancer Res Treat 1999;57(2):175-82.

188. Cavalieri E, Frenkel K, Liehr JG, Rogan E, Roy D. Estrogens as endogenous genotoxic agents--DNA adducts and mutations. J Natl Cancer Inst Monogr 2000(27):75-93.

189. Key T, Appleby P, Barnes I, Reeves G. Endogenous sex hormones and breast cancer in postmenopausal women: reanalysis of nine prospective studies. J Natl Cancer Inst 2002;94(8):606-16.

190. Kabuto M, Akiba S, Stevens RG, Neriishi K, Land CE. A prospective study of estradiol and breast cancer in Japanese women. Cancer Epidemiol Biomarkers Prev 2000;9(6):575-9.

191. Bermejo-Perez MJ, Marquez-Calderon S, Llanos-Mendez A. Effectiveness of preventive interventions in BRCA1/2 gene mutation carriers: a systematic review. Int J Cancer 2007;121(2):225-31.

192. Chetrite GS, Cortes-Prieto J, Philippe JC, Wright F, Pasqualini JR. Comparison of estrogen concentrations, estrone sulfatase and aromatase activities in normal, and in cancerous, human breast tissues. J Steroid Biochem Mol Biol 2000;72(1-2):23-7.

193. Pasqualini JR, Chetrite G, Nguyen BL, Maloche C, Delalonde L, Talbi M, et al. Estrone sulfate-sulfatase and 17 beta-hydroxysteroid dehydrogenase activities: a hypothesis for their role in the evolution of human breast cancer from hormone-dependence to hormone-independence. J Steroid Biochem Mol Biol 1995;53(1-6):407-12.

194. Cui X, Schiff R, Arpino G, Osborne CK, Lee AV. Biology of progesterone receptor loss in breast cancer and its implications for endocrine therapy. J Clin Oncol 2005;23(30):7721-35.

195. Luu-The V. Analysis and characteristics of multiple types of human 17beta-hydroxysteroid dehydrogenase. J Steroid Biochem Mol Biol 2001;76(1-5):143-51.

196. McKenna NJ, Lanz RB, O'Malley BW. Nuclear receptor coregulators: cellular and molecular biology. Endocr Rev 1999;20(3):321-44.

197. Kuiper GG, Carlsson B, Grandien K, Enmark E, Haggblad J, Nilsson S, et al. Comparison of the ligand binding specificity and transcript tissue distribution of estrogen receptors alpha and beta. Endocrinology 1997;138(3):863-70.

198. Frasor J, Danes JM, Komm B, Chang KC, Lyttle CR, Katzenellenbogen BS. Profiling of estrogen up- and down-regulated gene expression in human breast cancer cells: insights into gene networks and pathways underlying estrogenic control of proliferation and cell phenotype. Endocrinology 2003;144(10):4562-74.

199. Osborne CK, Schiff R. Estrogen-receptor biology: continuing progress and therapeutic implications. J Clin Oncol 2005;23(8):1616-22.

200. Petersen OW, Hoyer PE, van Deurs B. Frequency and distribution of estrogen receptor-positive cells in normal, nonlactating human breast tissue. Cancer Res 1987;47(21):5748-51.

201. Anderson E. The role of oestrogen and progesterone receptors in human mammary development and tumorigenesis. Breast Cancer Res 2002;4(5):197-201.

202. Speirs V, Walker RA. New perspectives into the biological and clinical relevance of oestrogen receptors in the human breast. J Pathol 2007;211(5):499-506.

203. Leonard GD, Swain SM. Ductal carcinoma in situ, complexities and challenges. J Natl Cancer Inst 2004;96(12):906-20.

204. Anderson WF, Chu KC, Chang S, Sherman ME. Comparison of age-specific incidence rate patterns for different histopathologic types of breast carcinoma. Cancer Epidemiol Biomarkers Prev 2004;13(7):1128-35.

205. Dickson RB, Stancel GM. Estrogen receptor-mediated processes in normal and cancer cells. J Natl Cancer Inst Monogr 2000(27):135-45.

206. Osborne CK. Steroid hormone receptors in breast cancer management. Breast Cancer Res Treat 1998;51(3):227-38.

207. Kjaergaard AD, Ellervik C, Tybjaerg-Hansen A, Axelsson CK, Gronholdt ML, Grande P, et al. Estrogen receptor alpha polymorphism and risk of cardiovascular disease, cancer, and hip fracture: cross-sectional, cohort, and case-control studies and a meta-analysis. Circulation 2007;115(7):861-71.

208. Onland-Moret NC, van Gils CH, Roest M, Grobbee DE, Peeters PH. The estrogen receptor alpha gene and breast cancer risk (The Netherlands). Cancer Causes Control 2005;16(10):1195-202.

209. Sonestedt E, Ivarsson MI, Harlid S, Ericson U, Gullberg B, Carlson J, et al. The protective association of high plasma enterolactone with breast cancer is reasonably

robust in women with polymorphisms in the estrogen receptor alpha and beta genes. J Nutr 2009;139(5):993-1001.

210. Wang J, Higuchi R, Modugno F, Li J, Umblas N, Lee J, et al. Estrogen receptor alpha haplotypes and breast cancer risk in older Caucasian women. Breast Cancer Res Treat 2007;106(2):273-80.

211. Dunning AM, Healey CS, Baynes C, Maia AT, Scollen S, Vega A, et al. Association of ESR1 gene tagging SNPs with breast cancer risk. Hum Mol Genet 2009;18(6):1131-9.

212. Herrington DM, Howard TD, Brosnihan KB, McDonnell DP, Li X, Hawkins GA, et al. Common estrogen receptor polymorphism augments effects of hormone replacement therapy on E-selectin but not C-reactive protein. Circulation 2002;105(16):1879-82.

213. Fernandez LP, Milne RL, Barroso E, Cuadros M, Arias JI, Ruibal A, et al. Estrogen and progesterone receptor gene polymorphisms and sporadic breast cancer risk: a Spanish case-control study. Int J Cancer 2006;119(2):467-71.

214. Gallicchio L, Berndt SI, McSorley MA, Newschaffer CJ, Thuita LW, Argani P, et al. Polymorphisms in estrogen-metabolizing and estrogen receptor genes and the risk of developing breast cancer among a cohort of women with benign breast disease. BMC Cancer 2006;6:173.

215. Vasconcelos A, Medeiros R, Veiga I, Pereira D, Carrilho S, Palmeira C, et al. Analysis of estrogen receptor polymorphism in codon 325 by PCR-SSCP in breast cancer: association with lymph node metastasis. Breast J 2002;8(4):226-9.

216. Siddig A, Mohamed AO, Awad S, Hassan AH, Zilahi E, Al-Haj M, et al. Estrogen receptor alpha gene polymorphism and breast cancer. Ann N Y Acad Sci 2008;1138:95-107.

217. Roodi N, Bailey LR, Kao WY, Verrier CS, Yee CJ, Dupont WD, et al. Estrogen receptor gene analysis in estrogen receptor-positive and receptor-negative primary breast cancer. J Natl Cancer Inst 1995;87(6):446-51.

218. Southey MC, Batten LE, McCredie MR, Giles GG, Dite G, Hopper JL, et al. Estrogen receptor polymorphism at codon 325 and risk of breast cancer in women before age forty. J Natl Cancer Inst 1998;90(7):532-6.

219. Hsiao WC, Young KC, Lin SL, Lin PW. Estrogen receptor-alpha polymorphism in a Taiwanese clinical breast cancer population: a case-control study. Breast Cancer Res 2004;6(3):R180-6.

220. Mavaddat N, Dunning AM, Ponder BA, Easton DF, Pharoah PD. Common genetic

variation in candidate genes and susceptibility to subtypes of breast cancer. Cancer Epidemiol Biomarkers Prev 2009;18(1):255-9.

221. Sowers MR, Jannausch ML, McConnell DS, Kardia SR, Randolph JF, Jr. Endogenous estradiol and its association with estrogen receptor gene polymorphisms. Am J Med 2006;119(9 Suppl 1):S16-22.

222. Schuit SC, de Jong FH, Stolk L, Koek WN, van Meurs JB, Schoofs MW, et al. Estrogen receptor alpha gene polymorphisms are associated with estradiol levels in postmenopausal women. Eur J Endocrinol 2005;153(2):327-34.

223. Zofkova I, Zajickova K, Hill M. The estrogen receptor alpha gene determines serum androstenedione levels in postmenopausal women. Steroids 2002;67(10):815-9.

224. Conneely OM, Mulac-Jericevic B, Lydon JP. Progesterone-dependent regulation of female reproductive activity by two distinct progesterone receptor isoforms. Steroids 2003;68(10-13):771-8.

225. Mendelson CR, Hardy DB. Role of the progesterone receptor (PR) in the regulation of inflammatory response pathways and aromatase in the breast. J Steroid Biochem Mol Biol 2006;102(1-5):241-9.

226. Lydon JP, DeMayo FJ, Conneely OM, O'Malley BW. Reproductive phenotpes of the progesterone receptor null mutant mouse. J Steroid Biochem Mol Biol 1996;56(1-6 Spec No):67-77.

227. Ballare C, Vallejo G, Vicent GP, Saragueta P, Beato M. Progesterone signaling in breast and endometrium. J Steroid Biochem Mol Biol 2006;102(1-5):2-10.

228. Ballare C, Uhrig M, Bechtold T, Sancho E, Di Domenico M, Migliaccio A, et al. Two domains of the progesterone receptor interact with the estrogen receptor and are required for progesterone activation of the c-Src/Erk pathway in mammalian cells. Mol Cell Biol 2003;23(6):1994-2008.

229. Graham JD, Yeates C, Balleine RL, Harvey SS, Milliken JS, Bilous AM, et al. Progesterone receptor A and B protein expression in human breast cancer. J Steroid Biochem Mol Biol 1996;56(1-6 Spec No):93-8.

230. Pooley KA, Healey CS, Smith PL, Pharoah PD, Thompson D, Tee L, et al. Association of the progesterone receptor gene with breast cancer risk: a single-nucleotide polymorphism tagging approach. Cancer Epidemiol Biomarkers Prev 2006;15(4):675-82.

231. Johnatty SE, Spurdle AB, Beesley J, Chen X, Hopper JL, Duffy DL, et al. Progesterone receptor polymorphisms and risk of breast cancer: results from two Australian breast cancer studies. Breast Cancer Res Treat 2008;109(1):91-9.

232. De Vivo I, Hankinson SE, Colditz GA, Hunter DJ. The progesterone receptor Val660-->Leu polymorphism and breast cancer risk. Breast Cancer Res 2004;6(6):R636-9.

233. Agoulnik IU, Tong XW, Fischer DC, Korner K, Atkinson NE, Edwards DP, et al. A germline variation in the progesterone receptor gene increases transcriptional activity and may modify ovarian cancer risk. J Clin Endocrinol Metab 2004;89(12):6340-7.

234. Romano A, Delvoux B, Fischer DC, Groothuis P. The PROGINS polymorphism of the human progesterone receptor diminishes the response to progesterone. J Mol Endocrinol 2007;38(1-2):331-50.

235. De Vivo I, Huggins GS, Hankinson SE, Lescault PJ, Boezen M, Colditz GA, et al. A functional polymorphism in the promoter of the progesterone receptor gene associated with endometrial cancer risk. Proc Natl Acad Sci U S A 2002;99(19):12263-8.

236. De Vivo I, Hankinson SE, Colditz GA, Hunter DJ. A functional polymorphism in the progesterone receptor gene is associated with an increase in breast cancer risk. Cancer Res 2003;63(17):5236-8.

237. Kotsopoulos J, Tworoger SS, Devivo I, Hankinson SE, Hunter DJ, Willett WC, et al. +331G/A variant in the progesterone receptor gene, postmenopausal hormone use and risk of breast cancer. Int J Cancer 2009.

238. Gunnarsson C, Jansson A, Holmlund B, Ferraud L, Nordenskjold B, Rutqvist LE, et al. Expression of COX-2 and steroid converting enzymes in breast cancer. Oncol Rep 2006;16(2):219-24.

239. Yoshimura N, Harada N, Bukholm I, Karesen R, Borresen-Dale AL, Kristensen VN. Intratumoural mRNA expression of genes from the oestradiol metabolic pathway and clinical and histopathological parameters of breast cancer. Breast Cancer Res 2004;6(2):R46-55.

240. Suzuki T, Moriya T, Ariga N, Kaneko C, Kanazawa M, Sasano H. 17Beta-hydroxysteroid dehydrogenase type 1 and type 2 in human breast carcinoma: a correlation to clinicopathological parameters. Br J Cancer 2000;82(3):518-23.

241. Gunnarsson C, Hellqvist E, Stal O. 17beta-Hydroxysteroid dehydrogenases involved in local oestrogen synthesis have prognostic significance in breast cancer. Br J Cancer 2005;92(3):547-52.

242. Jansson A, Carlsson J, Olsson A, Storm P, Margolin S, Gunnarsson C, et al. A new polymorphism in the coding region of exon four in HSD17B2 in relation to risk of sporadic and hereditary breast cancer. Breast Cancer Res Treat 2007;106(1):57-64.

243. Duax WL, Thomas J, Pletnev V, Addlagatta A, Huether R, Habegger L, et al. Determining structure and function of steroid dehydrogenase enzymes by sequence analysis, homology modeling, and rational mutational analysis. Ann N Y Acad Sci 2005;1061:135-48.

244. Gunasegaram R, Peh KL, Loganath A, Ratnam SS. Expression of 3beta-hydroxysteroid dehydrogenase-5,4-en isomerase activity by infiltrating ductal human breast carcinoma in vitro. Breast Cancer Res Treat 1998;50(2):117-23.

245. Guerin SL, Leclerc S, Verreault H, Labrie F, Luu-The V. Overlapping cis-acting elements located in the first intron of the gene for type I 3 beta-hydroxysteroid dehydrogenase modulate its transcriptional activity. Mol Endocrinol 1995;9(11):1583-97.

246. Gingras S, Cote S, Simard J. Multiple signal transduction pathways mediate interleukin-4-induced 3beta-hydroxysteroid dehydrogenase/Delta5-Delta4 isomerase in normal and tumoral target tissues. J Steroid Biochem Mol Biol 2001;76(1-5):213-25.

247. Chang BL, Zheng SL, Hawkins GA, Isaacs SD, Wiley KE, Turner A, et al. Joint effect of HSD3B1 and HSD3B2 genes is associated with hereditary and sporadic prostate cancer susceptibility. Cancer Res 2002;62(6):1784-9.

248. Park JY, Tanner JP, Sellers TA, Huang Y, Stevens CK, Dossett N, et al. Association between polymorphisms in HSD3B1 and UGT2B17 and prostate cancer risk. Urology 2007;70(2):374-9.

249. Cunningham JM, Hebbring SJ, McDonnell SK, Cicek MS, Christensen GB, Wang L, et al. Evaluation of genetic variations in the androgen and estrogen metabolic pathways as risk factors for sporadic and familial prostate cancer. Cancer Epidemiol Biomarkers Prev 2007;16(5):969-78.

250. Haiman CA, Bernstein L, Berg D, Ingles SA, Salane M, Ursin G. Genetic determinants of mammographic density. Breast Cancer Res 2002;4(3):R5.

251. Stone J, Gurrin LC, Byrnes GB, Schroen CJ, Treloar SA, Padilla EJ, et al. Mammographic density and candidate gene variants: a twins and sisters study. Cancer Epidemiol Biomarkers Prev 2007;16(7):1479-84.

252. Kamat A, Hinshelwood MM, Murry BA, Mendelson CR. Mechanisms in tissue-specific regulation of estrogen biosynthesis in humans. Trends Endocrinol Metab 2002;13(3):122-8.

253. Hellmold H, Rylander T, Magnusson M, Reihner E, Warner M, Gustafsson JA. Characterization of cytochrome P450 enzymes in human breast tissue from reduction mammaplasties. J Clin Endocrinol Metab 1998;83(3):886-95.

254. Brodie A, Long B, Lu Q. Aromatase expression in the human breast. Breast Cancer Res Treat 1998;49 Suppl 1:S85-91; discussion S109-19.

255. Miki Y, Suzuki T, Tazawa C, Yamaguchi Y, Kitada K, Honma S, et al. Aromatase localization in human breast cancer tissues: possible interactions between intratumoral stromal and parenchymal cells. Cancer Res 2007;67(8):3945-54.

256. Bulun SE, Price TM, Aitken J, Mahendroo MS, Simpson ER. A link between breast cancer and local estrogen biosynthesis suggested by quantification of breast adipose tissue aromatase cytochrome P450 transcripts using competitive polymerase chain reaction after reverse transcription. J Clin Endocrinol Metab 1993;77(6):1622-8.

257. Esteban JM, Warsi Z, Haniu M, Hall P, Shively JE, Chen S. Detection of intratumoral aromatase in breast carcinomas. An immunohistochemical study with clinicopathologic correlation. Am J Pathol 1992;140(2):337-43.

258. Buzdar AU, Coombes RC, Goss PE, Winer EP. Summary of aromatase inhibitor clinical trials in postmenopausal women with early breast cancer. Cancer 2007;112(S3):700-709.

259. Catalano S, Marsico S, Giordano C, Mauro L, Rizza P, Panno ML, et al. Leptin enhances, via AP-1, expression of aromatase in the MCF-7 cell line. J Biol Chem 2003;278(31):28668-76.

260. Zhang Y, Kulp SK, Sugimoto Y, Farrar WB, Brueggemeier RW, Lin YC. Keratinocyte growth factor (KGF) induces aromatase activity in cultured MCF-7 human breast cancer cells. Anticancer Res 1998;18(4A):2541-6.

261. Ryde CM, Nicholls JE, Dowsett M. Steroid and growth factor modulation of aromatase activity in MCF7 and T47D breast carcinoma cell lines. Cancer Res 1992;52(6):1411-5.

262. Nativelle-Serpentini C, Lambard S, Seralini GE, Sourdaine P. Aromatase and breast cancer: W39R, an inactive protein. Eur J Endocrinol 2002;146(4):583-9.

263. Ma CX, Adjei AA, Salavaggione OE, Coronel J, Pelleymounter L, Wang L, et al. Human aromatase: gene resequencing and functional genomics. Cancer Res 2005;65(23):11071-82.

264. Watanabe J, Harada N, Suemasu K, Higashi Y, Gotoh O, Kawajiri K. Arginine-cysteine polymorphism at codon 264 of the human CYP19 gene does not affect aromatase activity. Pharmacogenetics 1997;7(5):419-24.

265. Miyoshi Y, Iwao K, Ikeda N, Egawa C, Noguchi S. Breast cancer risk associated with polymorphism in CYP19 in Japanese women. Int J Cancer 2000;89(4):325-8.

266. Lee KM, Abel J, Ko Y, Harth V, Park WY, Seo JS, et al. Genetic polymorphisms of cytochrome P450 19 and 1B1, alcohol use, and breast cancer risk in Korean women. Br J Cancer 2003;88(5):675-8.

267. Haiman CA, Dossus L, Setiawan VW, Stram DO, Dunning AM, Thomas G, et al. Genetic variation at the CYP19A1 locus predicts circulating estrogen levels but not breast cancer risk in postmenopausal women. Cancer Res 2007;67(5):1893-7.

268. Gulyaeva LF, Mikhailova ON, PustyInyak VO, Kim IVt, Gerasimov AV, Krasilnikov SE, et al. Comparative analysis of SNP in estrogen-metabolizing enzymes for ovarian, endometrial, and breast cancers in Novosibirsk, Russia. Adv Exp Med Biol 2008;617:359-66.

269. Riancho JA, Valero C, Naranjo A, Morales DJ, Sanudo C, Zarrabeitia MT. Identification of an aromatase haplotype that is associated with gene expression and postmenopausal osteoporosis. J Clin Endocrinol Metab 2007;92(2):660-5.

270. Talbott KE, Gammon MD, Kibriya MG, Chen Y, Teitelbaum SL, Long CM, et al. A CYP19 (aromatase) polymorphism is associated with increased premenopausal breast cancer risk. Breast Cancer Res Treat 2008;111(3):481-7.

271. Haiman CA, Hankinson SE, Spiegelman D, Brown M, Hunter DJ. No association between a single nucleotide polymorphism in CYP19 and breast cancer risk. Cancer Epidemiol Biomarkers Prev 2002;11(2):215-6.

272. Dunning AM, Dowsett M, Healey CS, Tee L, Luben RN, Folkerd E, et al. Polymorphisms associated with circulating sex hormone levels in postmenopausal women. J Natl Cancer Inst 2004;96(12):936-45.

273. Raskin L, Lejbkowicz F, Barnett-Griness O, Dishon S, Almog R, Rennert G. BRCA1 breast cancer risk is modified by CYP19 polymorphisms in Ashkenazi Jews. Cancer Epidemiol Biomarkers Prev 2009;18(5):1617-23.

274. Hryb DJ, Khan MS, Romas NA, Rosner W. The control of the interaction of sex hormone-binding globulin with its receptor by steroid hormones. J Biol Chem 1990;265(11):6048-54.

275. Fortunati N, Fissore F, Fazzari A, Becchis M, Comba A, Catalano MG, et al. Sex steroid binding protein exerts a negative control on estradiol action in MCF-7 cells (human breast cancer) through cyclic adenosine 3',5'-monophosphate and protein kinase A. Endocrinology 1996;137(2):686-92.

276. Catalano MG, Frairia R, Boccuzzi G, Fortunati N. Sex hormone-binding globulin antagonizes the anti-apoptotic effect of estradiol in breast cancer cells. Mol Cell Endocrinol 2005;230(1-2):31-7.

277. Fissore F, Fortunati N, Comba A, Fazzari A, Gaidano G, Berta L, et al. The receptor-mediated action of sex steroid binding protein (SBP, SHBG): accumulation of cAMP in MCF-7 cells under SBP and estradiol treatment. Steroids 1994;59(11):661-7.

278. Fazzari A, Catalano MG, Comba A, Becchis M, Raineri M, Frairia R, et al. The control of progesterone receptor expression in MCF-7 breast cancer cells: effects of estradiol and sex hormone-binding globulin (SHBG). Mol Cell Endocrinol 2001;172(1-2):31-6.

279. Catalano MG, Costantino L, Frairia R, Boccuzzi G, Fortunati N. Sex hormone-binding globulin selectively modulates estradiol-regulated genes in MCF-7 cells. Horm Metab Res 2007;39(4):288-94.

280. Moore KH, Bertram KA, Gomez RR, Styner MJ, Matej LA. Sex hormone binding globulin mRNA in human breast cancer: detection in cell lines and tumor samples. J Steroid Biochem Mol Biol 1996;59(3-4):297-304.

281. Catalano MG, Comba A, Fazzari A, Benedusi-Pagliano E, Sberveglieri M, Revelli A, et al. Sex steroid binding protein receptor (SBP-R) is related to a reduced proliferation rate in human breast cancer. Breast Cancer Res Treat 1997;42(3):227-34.

282. Baer HJ, Colditz GA, Willett WC, Dorgan JF. Adiposity and sex hormones in girls. Cancer Epidemiol Biomarkers Prev 2007;16(9):1880-8.

283. Ingram D, Nottage E, Ng S, Sparrow L, Roberts A, Willcox D. Obesity and breast disease. The role of the female sex hormones. Cancer 1989;64(5):1049-53.

284. Dorgan JF, Reichman ME, Judd JT, Brown C, Longcope C, Schatzkin A, et al. The relation of body size to plasma levels of estrogens and androgens in premenopausal women (Maryland, United States). Cancer Causes Control 1995;6(1):3-8.

285. Key TJ, Appleby PN, Reeves GK, Roddam A, Dorgan JF, Longcope C, et al. Body mass index, serum sex hormones, and breast cancer risk in postmenopausal women. J Natl Cancer Inst 2003;95(16):1218-26.

286. Barnett JB, Woods MN, Rosner B, McCormack C, Longcope C, Houser RF, Jr., et al. Sex hormone levels in premenopausal African-American women with upper and lower body fat phenotypes. Nutr Cancer 2001;41(1-2):47-56.

287. Lonning PE, Johannessen DC, Lien EA, Ekse D, Fotsis T, Adlercreutz H. Influence of tamoxifen on sex hormones, gonadotrophins and sex hormone binding globulin in postmenopausal breast cancer patients. J Steroid Biochem Mol Biol 1995;52(5):491-6.

288. Adly L, Hill D, Sherman ME, Sturgeon SR, Fears T, Mies C, et al. Serum

concentrations of estrogens, sex hormone-binding globulin, and androgens and risk of breast cancer in postmenopausal women. Int J Cancer 2006;119(10):2402-7.

289. Schairer C, Hill D, Sturgeon SR, Fears T, Mies C, Ziegler RG, et al. Serum concentrations of estrogens, sex hormone binding globulin, and androgens and risk of breast hyperplasia in postmenopausal women. Cancer Epidemiol Biomarkers Prev 2005;14(7):1660-5.

290. Zeleniuch-Jacquotte A, Shore RE, Koenig KL, Akhmedkhanov A, Afanasyeva Y, Kato I, et al. Postmenopausal levels of oestrogen, androgen, and SHBG and breast cancer: long-term results of a prospective study. Br J Cancer 2004;90(1):153-9.

291. Yu H, Shu XO, Shi R, Dai Q, Jin F, Gao YT, et al. Plasma sex steroid hormones and breast cancer risk in Chinese women. Int J Cancer 2003;105(1):92-7.

292. Manjer J, Johansson R, Berglund G, Janzon L, Kaaks R, Agren A, et al. Postmenopausal breast cancer risk in relation to sex steroid hormones, prolactin and SHBG (Sweden). Cancer Causes Control 2003;14(7):599-607.

293. Meyer F, Brown JB, Morrison AS, MacMahon B. Endogenous sex hormones, prolactin, and breast cancer in premenopausal women. J Natl Cancer Inst 1986;77(3):613-6.

294. Micheli A, Muti P, Secreto G, Krogh V, Meneghini E, Venturelli E, et al. Endogenous sex hormones and subsequent breast cancer in premenopausal women. Int J Cancer 2004;112(2):312-8.

295. Eliassen AH, Missmer SA, Tworoger SS, Spiegelman D, Barbieri RL, Dowsett M, et al. Endogenous steroid hormone concentrations and risk of breast cancer among premenopausal women. J Natl Cancer Inst 2006;98(19):1406-15.

296. Zeleniuch-Jacquotte A, Gu Y, Shore RE, Koenig KL, Arslan AA, Kato I, et al. Postmenopausal levels of sex hormones and risk of breast carcinoma in situ: results of a prospective study. Int J Cancer 2005;114(2):323-7.

297. Cousin P, Dechaud H, Grenot C, Lejeune H, Pugeat M. Human variant sex hormone-binding globulin (SHBG) with an additional carbohydrate chain has a reduced clearance rate in rabbit. J Clin Endocrinol Metab 1998;83(1):235-40.

298. Cousin P, Calemard-Michel L, Lejeune H, Raverot G, Yessaad N, Emptoz-Bonneton A, et al. Influence of SHBG gene pentanucleotide TAAAA repeat and D327N polymorphism on serum sex hormone-binding globulin concentration in hirsute women. J Clin Endocrinol Metab 2004;89(2):917-24.

299. Thompson DJ, Healey CS, Baynes C, Kalmyrzaev B, Ahmed S, Dowsett M, et al. Identification of common variants in the SHBG gene affecting sex hormone-binding

globulin levels and breast cancer risk in postmenopausal women. Cancer Epidemiol Biomarkers Prev 2008;17(12):3490-8.

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

301. Ogden CL, Carroll MD, Curtin LR, McDowell MA, Tabak CJ, Flegal KM. Prevalence of overweight and obesity in the United States, 1999-2004. Jama 2006;295(13):1549-55.

302. Baskin ML, Ard J, Franklin F, Allison DB. Prevalence of obesity in the United States. Obes Rev 2005;6(1):5-7.

303. Adebamowo CA, Ogundiran TO, Adenipekun AA, Oyesegun RA, Campbell OB, Akang EU, et al. Obesity and height in urban Nigerian women with breast cancer. Ann Epidemiol 2003;13(6):455-61.

304. Althuis MD, Fergenbaum JH, Garcia-Closas M, Brinton LA, Madigan MP, Sherman ME. Etiology of hormone receptor-defined breast cancer: a systematic review of the literature. Cancer Epidemiol Biomarkers Prev 2004;13(10):1558-68.

305. Ballard-Barbash R, Schatzkin A, Carter CL, Kannel WB, Kreger BE, D'Agostino RB, et al. Body fat distribution and breast cancer in the Framingham Study. J Natl Cancer Inst 1990;82(4):286-90.

306. Huang Z, Willett WC, Colditz GA, Hunter DJ, Manson JE, Rosner B, et al. Waist circumference, waist:hip ratio, and risk of breast cancer in the Nurses' Health Study. Am J Epidemiol 1999;150(12):1316-24.

307. McTiernan A, Wu L, Chen C, Chlebowski R, Mossavar-Rahmani Y, Modugno F, et al. Relation of BMI and physical activity to sex hormones in postmenopausal women. Obesity (Silver Spring) 2006;14(9):1662-77.

308. Mahabir S, Baer DJ, Johnson LL, Hartman TJ, Dorgan JF, Campbell WS, et al. Usefulness of body mass index as a sufficient adiposity measurement for sex hormone concentration associations in postmenopausal women. Cancer Epidemiol Biomarkers Prev 2006;15(12):2502-7.

309. Berstein LM. Increased risk of breast cancer in women with central obesity: additional considerations. J Natl Cancer Inst 1990;82(24):1943-4.

310. Potischman N, Swanson CA, Siiteri P, Hoover RN. Reversal of relation between body mass and endogenous estrogen concentrations with menopausal status. J Natl Cancer Inst 1996;88(11):756-8.

311. National Institute of Diabetes and Digestive and Kidney Diseases, NIH. Insulin

Resistance and Pre-Diabetes. [cited 2007 December 18]; Available from: http://www.diabetes.niddk.nih.gov/dm/pubs/insulinresistance/

312. Ceschi M, Gutzwiller F, Moch H, Eichholzer M, Probst-Hensch NM. Epidemiology and pathophysiology of obesity as cause of cancer. Swiss Med Wkly 2007;137(3-4):50-6.

313. Sachdev D, Yee D. Inhibitors of insulin-like growth factor signaling: a therapeutic approach for breast cancer. J Mammary Gland Biol Neoplasia 2006;11(1):27-39.

314. Frasca F, Pandini G, Vigneri R, Goldfine ID. Insulin and hybrid insulin/IGF receptors are major regulators of breast cancer cells. Breast Dis 2003;17:73-89.

315. Hursting SD, Nunez NP, Varticovski L, Vinson C. The obesity-cancer link: lessons learned from a fatless mouse. Cancer Res 2007;67(6):2391-3.

316. Rose DP, Komninou D, Stephenson GD. Obesity, adipocytokines, and insulin resistance in breast cancer. Obes Rev 2004;5(3):153-65.

317. Schaffler A, Scholmerich J, Buechler C. Mechanisms of disease: adipokines and breast cancer - endocrine and paracrine mechanisms that connect adiposity and breast cancer. Nat Clin Pract Endocrinol Metab 2007;3(4):345-54.

318. Macdiarmid F, Wang D, Duncan LJ, Purohit A, Ghilchick MW, Reed MJ. Stimulation of aromatase activity in breast fibroblasts by tumor necrosis factor alpha. Mol Cell Endocrinol 1994;106(1-2):17-21.

319. Dieudonne MN, Machinal-Quelin F, Serazin-Leroy V, Leneveu MC, Pecquery R, Giudicelli Y. Leptin mediates a proliferative response in human MCF7 breast cancer cells. Biochem Biophys Res Commun 2002;293(1):622-8.

320. Laud K, Gourdou I, Pessemesse L, Peyrat JP, Djiane J. Identification of leptin receptors in human breast cancer: functional activity in the T47-D breast cancer cell line. Mol Cell Endocrinol 2002;188(1-2):219-26.

321. Okumura M, Yamamoto M, Sakuma H, Kojima T, Maruyama T, Jamali M, et al. Leptin and high glucose stimulate cell proliferation in MCF-7 human breast cancer cells: reciprocal involvement of PKC-alpha and PPAR expression. Biochim Biophys Acta 2002;1592(2):107-16.

322. Hu X, Juneja SC, Maihle NJ, Cleary MP. Leptin--a growth factor in normal and malignant breast cells and for normal mammary gland development. J Natl Cancer Inst 2002;94(22):1704-11.

323. Speirs V, Kerin MJ, Walton DS, Newton CJ, Desai SB, Atkin SL. Direct activation of oestrogen receptor-alpha by interleukin-6 in primary cultures of breast cancer

epithelial cells. Br J Cancer 2000;82(7):1312-6.

324. Dieudonne MN, Bussiere M, Dos Santos E, Leneveu MC, Giudicelli Y, Pecquery R. Adiponectin mediates antiproliferative and apoptotic responses in human MCF7 breast cancer cells. Biochem Biophys Res Commun 2006;345(1):271-9.

325. Arditi JD, Venihaki M, Karalis KP, Chrousos GP. Antiproliferative effect of adiponectin on MCF7 breast cancer cells: a potential hormonal link between obesity and cancer. Horm Metab Res 2007;39(1):9-13.

326. Flury N, Eppenberger U, Mueller H. Tumor-necrosis factor-alpha modulates mitogen-activated protein kinase activity of epidermal-growth-factor-stimulated MCF-7 breast cancer cells. Eur J Biochem 1997;249(2):421-6.

327. Hagemann T, Robinson SC, Schulz M, Trumper L, Balkwill FR, Binder C. Enhanced invasiveness of breast cancer cell lines upon co-cultivation with macrophages is due to TNF-alpha dependent up-regulation of matrix metalloproteases. Carcinogenesis 2004;25(8):1543-9.

328. Burow ME, Weldon CB, Tang Y, Navar GL, Krajewski S, Reed JC, et al. Differences in susceptibility to tumor necrosis factor alpha-induced apoptosis among MCF-7 breast cancer cell variants. Cancer Res 1998;58(21):4940-6.

329. Matsuzawa Y, Funahashi T, Nakamura T. Molecular mechanism of metabolic syndrome X: contribution of adipocytokines adipocyte-derived bioactive substances. Ann N Y Acad Sci 1999;892:146-54.

330. Trayhurn P, Beattie JH. Physiological role of adipose tissue: white adipose tissue as an endocrine and secretory organ. Proc Nutr Soc 2001;60(3):329-39.

331. Ronti T, Lupattelli G, Mannarino E. The endocrine function of adipose tissue: an update. Clin Endocrinol (Oxf) 2006;64(4):355-65.

332. Balkwill F. Tumor necrosis factor or tumor promoting factor? Cytokine Growth Factor Rev 2002;13(2):135-41.

333. Zinman B, Hanley AJ, Harris SB, Kwan J, Fantus IG. Circulating tumor necrosis factor-alpha concentrations in a native Canadian population with high rates of type 2 diabetes mellitus. J Clin Endocrinol Metab 1999;84(1):272-8.

334. Sell H, Dietze-Schroeder D, Eckel J. The adipocyte-myocyte axis in insulin resistance. Trends Endocrinol Metab 2006;17(10):416-22.

335. Chavey C, Bibeau F, Gourgou-Bourgade S, Burlinchon S, Boissiere F, Laune D, et al. Oestrogen receptor negative breast cancers exhibit high cytokine content. Breast Cancer Res 2007;9(1):R15.

336. Green AR, Green VL, White MC, Speirs V. Expression of cytokine messenger RNA in normal and neoplastic human breast tissue: identification of interleukin-8 as a potential regulatory factor in breast tumours. Int J Cancer 1997;72(6):937-41.

337. Garcia-Tunon I, Ricote M, Ruiz A, Fraile B, Paniagua R, Royuela M. Role of tumor necrosis factor-alpha and its receptors in human benign breast lesions and tumors (in situ and infiltrative). Cancer Sci 2006;97(10):1044-9.

338. Lyon DE, McCain NL, Walter J, Schubert C. Cytokine comparisons between women with breast cancer and women with a negative breast biopsy. Nurs Res 2008;57(1):51-8.

339. Krajcik RA, Massardo S, Orentreich N. No association between serum levels of tumor necrosis factor-alpha (TNF-alpha) or the soluble receptors sTNFR1 and sTNFR2 and breast cancer risk. Cancer Epidemiol Biomarkers Prev 2003;12(9):945-6.

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

341. Lv K, Chen R, Cai Q, Fang M, Sun S. Effects of a single nucleotide polymorphism on the expression of human tumor necrosis factor-alpha. Scand J Immunol 2006;64(2):164-9.

342. Soga Y, Nishimura F, Ohyama H, Maeda H, Takashiba S, Murayama Y. Tumor necrosis factor-alpha gene (TNF-alpha) -1031/-863, -857 single-nucleotide polymorphisms (SNPs) are associated with severe adult periodontitis in Japanese. J Clin Periodontol 2003;30(6):524-31.

343. Skoog T, van't Hooft FM, Kallin B, Jovinge S, Boquist S, Nilsson J, et al. A common functional polymorphism (C-->A substitution at position -863) in the promoter region of the tumour necrosis factor-alpha (TNF-alpha) gene associated with reduced circulating levels of TNF-alpha. Hum Mol Genet 1999;8(8):1443-9.

344. Sharma S, Sharma A, Kumar S, Sharma SK, Ghosh B. Association of TNF haplotypes with asthma, serum IgE levels, and correlation with serum TNF-alpha levels. Am J Respir Cell Mol Biol 2006;35(4):488-95.

345. Wilson AG, Symons JA, McDowell TL, McDevitt HO, Duff GW. Effects of a polymorphism in the human tumor necrosis factor alpha promoter on transcriptional activation. Proc Natl Acad Sci U S A 1997;94(7):3195-9.

346. Kroeger KM, Carville KS, Abraham LJ. The -308 tumor necrosis factor-alpha promoter polymorphism effects transcription. Mol Immunol 1997;34(5):391-9.

347. Skoog T, Eriksson P, Hoffstedt J, Ryden M, Hamsten A, Armer P. Tumour necrosis factor-alpha (TNF-alpha) polymorphisms-857C/A and -863C/A are associated with TNF-alpha secretion from human adipose tissue. Diabetologia 2001;44(5):654-5.

348. Louis E, Franchimont D, Piron A, Gevaert Y, Schaaf-Lafontaine N, Roland S, et al. Tumour necrosis factor (TNF) gene polymorphism influences TNF-alpha production in lipopolysaccharide (LPS)-stimulated whole blood cell culture in healthy humans. Clin Exp Immunol 1998;113(3):401-6.

349. Sookoian SC, Gonzalez C, Pirola CJ. Meta-analysis on the G-308A tumor necrosis factor alpha gene variant and phenotypes associated with the metabolic syndrome. Obes Res 2005;13(12):2122-31.

350. Gaudet MM, Egan KM, Lissowska J, Newcomb PA, Brinton LA, Titus-Ernstoff L, et al. Genetic variation in tumor necrosis factor and lymphotoxin-alpha (TNF-LTA) and breast cancer risk. Hum Genet 2007;121(3-4):483-90.

351. Mestiri S, Bouaouina N, Ahmed SB, Khedhaier A, Jrad BB, Remadi S, et al. Genetic variation in the tumor necrosis factor-alpha promoter region and in the stress protein hsp70-2: susceptibility and prognostic implications in breast carcinoma. Cancer 2001;91(4):672-8.

352. Giordani L, Bruzzi P, Lasalandra C, Quaranta M, Schittulli F, Della Ragione F, et al. Association of breast cancer and polymorphisms of interleukin-10 and tumor necrosis factor-alpha genes. Clin Chem 2003;49(10):1664-7.

353. Smith KC, Bateman AC, Fussell HM, Howell WM. Cytokine gene polymorphisms and breast cancer susceptibility and prognosis. Eur J Immunogenet 2004;31(4):167-73.

354. Park KS, Mok JW, Ko HE, Tokunaga K, Lee MH. Polymorphisms of tumour necrosis factors A and B in breast cancer. Eur J Immunogenet 2002;29(1):7-10.

355. Kohaar I, Tiwari P, Kumar R, Nasare V, Thakur N, Das BC, et al. Association of single nucleotide polymorphisms (SNPs) in TNF-LTA locus with breast cancer risk in Indian population. Breast Cancer Res Treat 2009;114(2):347-55.

356. Weigert C, Hennige AM, Brodbeck K, Haring HU, Schleicher ED. Interleukin-6 acts as insulin sensitizer on glycogen synthesis in human skeletal muscle cells by phosphorylation of Ser473 of Akt. Am J Physiol Endocrinol Metab 2005;289(2):E251-7.

357. Basolo F, Fiore L, Fontanini G, Conaldi PG, Calvo S, Falcone V, et al. Expression of and response to interleukin 6 (IL6) in human mammary tumors. Cancer Res 1996;56(13):3118-22.

358. Knupfer H, Preiss R. Significance of interleukin-6 (IL-6) in breast cancer (review). Breast Cancer Res Treat 2007;102(2):129-35.

359. Danforth DN, Jr., Sgagias MK. Interleukin-1 alpha and interleukin-6 act additively to inhibit growth of MCF-7 breast cancer cells in vitro. Cancer Res 1993;53(7):1538-45.

360. Badache A, Hynes NE. Interleukin 6 inhibits proliferation and, in cooperation with an epidermal growth factor receptor autocrine loop, increases migration of T47D breast cancer cells. Cancer Res 2001;61(1):383-91.

361. Asgeirsson KS, Olafsdottir K, Jonasson JG, Ogmundsdottir HM. The effects of IL-6 on cell adhesion and e-cadherin expression in breast cancer. Cytokine 1998;10(9):720-8.

362. Tamm I, Kikuchi T, Cardinale I, Krueger JG. Cell-adhesion-disrupting action of interleukin 6 in human ductal breast carcinoma cells. Proc Natl Acad Sci U S A 1994;91(8):3329-33.

363. Robinson EK, Sneige N, Grimm EA. Correlation of interleukin 6 with interleukin 1alpha in human mammary tumours, but not with oestrogen receptor expression. Cytokine 1998;10(12):970-6.

364. Terry CF, Loukaci V, Green FR. Cooperative influence of genetic polymorphisms on interleukin 6 transcriptional regulation. J Biol Chem 2000;275(24):18138-44.

365. Fishman D, Faulds G, Jeffery R, Mohamed-Ali V, Yudkin JS, Humphries S, et al. The effect of novel polymorphisms in the interleukin-6 (IL-6) gene on IL-6 transcription and plasma IL-6 levels, and an association with systemic-onset juvenile chronic arthritis. J Clin Invest 1998;102(7):1369-76.

366. Hulkkonen J, Pertovaara M, Antonen J, Pasternack A, Hurme M. Elevated interleukin-6 plasma levels are regulated by the promoter region polymorphism of the IL6 gene in primary Sjogren's syndrome and correlate with the clinical manifestations of the disease. Rheumatology (Oxford) 2001;40(6):656-61.

367. Veres A, Prohaszka Z, Kilpinen S, Singh M, Fust G, Hurme M. The promoter polymorphism of the IL-6 gene is associated with levels of antibodies to 60-kDa heat-shock proteins. Immunogenetics 2002;53(10-11):851-6.

368. Sie MP, Sayed-Tabatabaei FA, Oei HH, Uitterlinden AG, Pols HA, Hofman A, et al. Interleukin 6 -174 g/c promoter polymorphism and risk of coronary heart disease: results from the rotterdam study and a meta-analysis. Arterioscler Thromb Vasc Biol 2006;26(1):212-7.

369. Hegedus CM, Skibola CF, Bracci P, Holly EA, Smith MT. Screening the human serum proteome for genotype-phenotype associations: an analysis of the IL6 -

174G>C polymorphism. Proteomics 2007;7(4):548-57.

370. Hefler LA, Grimm C, Lantzsch T, Lampe D, Leodolter S, Koelbl H, et al. Interleukin-1 and interleukin-6 gene polymorphisms and the risk of breast cancer in caucasian women. Clin Cancer Res 2005;11(16):5718-21.

371. Slattery ML, Curtin K, Baumgartner R, Sweeney C, Byers T, Giuliano AR, et al. IL6, aspirin, nonsteroidal anti-inflammatory drugs, and breast cancer risk in women living in the southwestern United States. Cancer Epidemiol Biomarkers Prev 2007;16(4):747-55.

372. Gonzalez-Zuloeta Ladd AM, Arias Vasquez A, Witteman J, Uitterlinden AG, Coebergh JW, Hofman A, et al. Interleukin 6 G-174 C polymorphism and breast cancer risk. Eur J Epidemiol 2006;21(5):373-6.

373. Litovkin KV, Domenyuk VP, Bubnov VV, Zaporozhan VN. Interleukin-6 -174G/C polymorphism in breast cancer and uterine leiomyoma patients: a population-based case control study. Exp Oncol 2007;29(4):295-8.

374. Slattery ML, Curtin K, Sweeney C, Wolff RK, Baumgartner RN, Baumgartner KB, et al. Modifying effects of IL-6 polymorphisms on body size-associated breast cancer risk. Obesity (Silver Spring) 2008;16(2):339-47.

375. Iacopetta B, Grieu F, Joseph D. The -174 G/C gene polymorphism in interleukin-6 is associated with an aggressive breast cancer phenotype. Br J Cancer 2004;90(2):419-22.

376. DeMichele A, Martin AM, Mick R, Gor P, Wray L, Klein-Cabral M, et al. Interleukin-6 -174G-->C polymorphism is associated with improved outcome in high-risk breast cancer. Cancer Res 2003;63(22):8051-6.

377. O'Brien S N, Welter BH, Price TM. Presence of leptin in breast cell lines and breast tumors. Biochem Biophys Res Commun 1999;259(3):695-8.

378. Ishikawa M, Kitayama J, Nagawa H. Enhanced expression of leptin and leptin receptor (OB-R) in human breast cancer. Clin Cancer Res 2004;10(13):4325-31.

379. Caldefie-Chezet F, Damez M, de Latour M, Konska G, Mishellani F, Fusillier C, et al. Leptin: a proliferative factor for breast cancer? Study on human ductal carcinoma. Biochem Biophys Res Commun 2005;334(3):737-41.

380. Cleary MP, Phillips FC, Getzin SC, Jacobson TL, Jacobson MK, Christensen TA, et al. Genetically obese MMTV-TGF-alpha/Lep(ob)Lep(ob) female mice do not develop mammary tumors. Breast Cancer Res Treat 2003;77(3):205-15.

381. Chen DC, Chung YF, Yeh YT, Chaung HC, Kuo FC, Fu OY, et al. Serum

adiponectin and leptin levels in Taiwanese breast cancer patients. Cancer Lett 2006;237(1):109-14.

382. Han CZ, Du LL, Jing JX, Zhao XW, Tian FG, Shi J, et al. Associations among lipids, leptin, and leptin receptor gene Gin223Arg polymorphisms and breast cancer in China. Biol Trace Elem Res 2008;126(1-3):38-48.

383. Petridou E, Papadiamantis Y, Markopoulos C, Spanos E, Dessypris N, Trichopoulos D. Leptin and insulin growth factor I in relation to breast cancer (Greece). Cancer Causes Control 2000;11(5):383-8.

384. Mantzoros CS, Bolhke K, Moschos S, Cramer DW. Leptin in relation to carcinoma in situ of the breast: a study of pre-menopausal cases and controls. Int J Cancer 1999;80(4):523-6.

385. Stattin P, Soderberg S, Biessy C, Lenner P, Hallmans G, Kaaks R, et al. Plasma leptin and breast cancer risk: a prospective study in northern Sweden. Breast Cancer Res Treat 2004;86(3):191-6.

386. Snoussi K, Strosberg AD, Bouaouina N, Ben Ahmed S, Helal AN, Chouchane L. Leptin and leptin receptor polymorphisms are associated with increased risk and poor prognosis of breast carcinoma. BMC Cancer 2006;6:38.

387. Cleveland RJ, Gammon MD, Long CM, Gaudet MM, Eng SM, Teitelbaum SL, et al. Common genetic variations in the LEP and LEPR genes, obesity and breast cancer incidence and survival. Breast Cancer Res Treat 2009.

388. Yiannakouris N, Melistas L, Yannakoulia M, Mungal K, Mantzoros CS. The-2548G/A polymorphism in the human leptin gene promoter region is associated with plasma free leptin levels; interaction with adiposity and gender in healthy subjects. Hormones (Athens) 2003;2(4):229-36.

389. Woo HY, Park H, Ki CS, Park YL, Bae WG. Relationships among serum leptin, leptin receptor gene polymorphisms, and breast cancer in Korea. Cancer Lett 2006;237(1):137-42.

390. van Rossum CT, Hoebee B, van Baak MA, Mars M, Saris WH, Seidell JC. Genetic variation in the leptin receptor gene, leptin, and weight gain in young Dutch adults. Obes Res 2003;11(3):377-86.

391. Quinton ND, Lee AJ, Ross RJ, Eastell R, Blakemore AI. A single nucleotide polymorphism (SNP) in the leptin receptor is associated with BMI, fat mass and leptin levels in postmenopausal Caucasian women. Hum Genet 2001;108(3):233-6.

392. Clement K, Vaisse C, Lahlou N, Cabrol S, Pelloux V, Cassuto D, et al. A mutation in the human leptin receptor gene causes obesity and pituitary dysfunction. Nature

1998;392(6674):398-401.

393. Yiannakouris N, Yannakoulia M, Melistas L, Chan JL, Klimis-Zacas D, Mantzoros CS. The Q223R polymorphism of the leptin receptor gene is significantly associated with obesity and predicts a small percentage of body weight and body composition variability. J Clin Endocrinol Metab 2001;86(9):4434-9.

394. Echwald SM, Sorensen TD, Sorensen TI, Tybjaerg-Hansen A, Andersen T, Chung WK, et al. Amino acid variants in the human leptin receptor: lack of association to juvenile onset obesity. Biochem Biophys Res Commun 1997;233(1):248-52.

395. Lihn AS, Pedersen SB, Richelsen B. Adiponectin: action, regulation and association to insulin sensitivity. Obes Rev 2005;6(1):13-21.

396. Bruun JM, Lihn AS, Verdich C, Pedersen SB, Toubro S, Astrup A, et al. Regulation of adiponectin by adipose tissue-derived cytokines: in vivo and in vitro investigations in humans. Am J Physiol Endocrinol Metab 2003;285(3):E527-33.

397. Dietze-Schroeder D, Sell H, Uhlig M, Koenen M, Eckel J. Autocrine action of adiponectin on human fat cells prevents the release of insulin resistance-inducing factors. Diabetes 2005;54(7):2003-11.

398. Yokota T, Oritani K, Takahashi I, Ishikawa J, Matsuyama A, Ouchi N, et al. Adiponectin, a new member of the family of soluble defense collagens, negatively regulates the growth of myelomonocytic progenitors and the functions of macrophages. Blood 2000;96(5):1723-32.

399. Korner A, Pazaitou-Panayiotou K, Kelesidis T, Kelesidis I, Williams CJ, Kaprara A, et al. Total and high-molecular-weight adiponectin in breast cancer: in vitro and in vivo studies. J Clin Endocrinol Metab 2007;92(3):1041-8.

400. Takahata C, Miyoshi Y, Irahara N, Taguchi T, Tamaki Y, Noguchi S. Demonstration of adiponectin receptors 1 and 2 mRNA expression in human breast cancer cells. Cancer Lett 2007;250(2):229-36.

401. Kelesidis I, Kelesidis T, Mantzoros CS. Adiponectin and cancer: a systematic review. Br J Cancer 2006;94(9):1221-5.

402. Mantzoros C, Petridou E, Dessypris N, Chavelas C, Dalamaga M, Alexe DM, et al. Adiponectin and breast cancer risk. J Clin Endocrinol Metab 2004;89(3):1102-7.

403. Miyoshi Y, Funahashi T, Kihara S, Taguchi T, Tamaki Y, Matsuzawa Y, et al. Association of serum adiponectin levels with breast cancer risk. Clin Cancer Res 2003;9(15):5699-704.

404. Tworoger SS, Eliassen AH, Kelesidis T, Colditz GA, Willett WC, Mantzoros CS, et

al. Plasma adiponectin concentrations and risk of incident breast cancer. J Clin Endocrinol Metab 2007;92(4):1510-6.

405. Fredriksson J, Carlsson E, Orho-Melander M, Groop L, Ridderstrale M. A polymorphism in the adiponectin gene influences adiponectin expression levels in visceral fat in obese subjects. Int J Obes (Lond) 2006;30(2):226-32.

406. Kang JH, Yu BY, Youn DS. Relationship of serum adiponectin and resistin levels with breast cancer risk. J Korean Med Sci 2007;22(1):117-21.

407. Brakenhielm E, Veitonmaki N, Cao R, Kihara S, Matsuzawa Y, Zhivotovsky B, et al. Adiponectin-induced antiangiogenesis and antitumor activity involve caspase-mediated endothelial cell apoptosis. Proc Natl Acad Sci U S A 2004;101(8):2476-81.

408. Kang JH, Lee YY, Yu BY, Yang BS, Cho KH, Yoon DK, et al. Adiponectin induces growth arrest and apoptosis of MDA-MB-231 breast cancer cell. Arch Pharm Res 2005;28(11):1263-9.

409. Pfeiler GH, Buechler C, Neumeier M, Schaffler A, Schmitz G, Ortmann O, et al. Adiponectin effects on human breast cancer cells are dependent on 17-beta estradiol. Oncol Rep 2008;19(3):787-93.

410. Bouatia-Naji N, Meyre D, Lobbens S, Seron K, Fumeron F, Balkau B, et al. ACDC/adiponectin polymorphisms are associated with severe childhood and adult obesity. Diabetes 2006;55(2):545-50.

411. Filippi E, Sentinelli F, Trischitta V, Romeo S, Arca M, Leonetti F, et al. Association of the human adiponectin gene and insulin resistance. Eur J Hum Genet 2004;12(3):199-205.

412. Kaklamani VG, Sadim M, Hsi A, Offit K, Oddoux C, Ostrer H, et al. Variants of the adiponectin and adiponectin receptor 1 genes and breast cancer risk. Cancer Res 2008;68(9):3178-84.

413. Gonzalez-Sanchez JL, Martinez-Calatrava MJ, Martinez-Larrad MT, Zabena C, Fernandez-Perez C, Laakso M, et al. Interaction of the -308G/A promoter polymorphism of the tumor necrosis factor-alpha gene with single-nucleotide polymorphism 45 of the adiponectin gene: effect on serum adiponectin concentrations in a Spanish population. Clin Chem 2006;52(1):97-103.

414. Gold B, Kalush F, Bergeron J, Scott K, Mitra N, Wilson K, et al. Estrogen receptor genotypes and haplotypes associated with breast cancer risk. Cancer Res 2004;64(24):8891-900.

415. Feigelson HS, Cox DG, Cann HM, Wacholder S, Kaaks R, Henderson BE, et al. Haplotype analysis of the HSD17B1 gene and risk of breast cancer: a comprehensive

approach to multicenter analyses of prospective cohort studies. Cancer Res 2006;66(4):2468-75.

416. Haiman CA, Stram DO, Pike MC, Kolonel LN, Burtt NP, Altshuler D, et al. A comprehensive haplotype analysis of CYP19 and breast cancer risk: the Multiethnic Cohort. Hum Mol Genet 2003;12(20):2679-92.

417. Einarsdottir K, Darabi H, Li Y, Low YL, Li YQ, Bonnard C, et al. ESR1 and EGF genetic variation in relation to breast cancer risk and survival. Breast Cancer Res 2008;10(1):R15.

418. Schaid DJ. Evaluating associations of haplotypes with traits. Genet Epidemiol 2004;27(4):348-64.

419. Akey J, Jin L, Xiong M. Haplotypes vs single marker linkage disequilibrium tests: what do we gain? Eur J Hum Genet 2001;9(4):291-300.

420. Clark AG. The role of haplotypes in candidate gene studies. Genet Epidemiol 2004;27(4):321-33.

2. Research Design and Methods

2.1. Specific Aims

The strong association between family history of breast cancer and disease incidence is evidence that heritable traits can play a strong role in breast carcinogenesis (1). Positive family history of breast cancer is estimated to account for approximately 5%-10% of all breast cancer cases (2). However, alterations in genes that are known to confer a high risk of breast cancer, such as BRCA1, BRCA2, ATM, and others, account for approximately 50% of familial breast cancer cases and only about 5% of breast cancer cases in the general population (3, 4). Family and twin studies suggest that other genetic traits that have not yet been identified are involved in the majority of the other familial breast cancer cases (2).

The polygenic model of breast cancer arises from complex segregation analyses suggesting that the remaining familial cases are the result of several common allelic variants, controlling for some environmental risk factors (5). Under this model each variant confers a small increase in cancer risk, and the effects of the alleles are multiplicative resulting in a higher level of risk with each additional susceptibility variant. Single nucleotide polymorphisms (SNPs) are specific locations in the genetic code whose sequence varies in a relatively large proportion (>1%) of the population, and are one type of common genetic variation that could increase cancer susceptibility under the polygenic model. Although these "low-penetrance" genetic variants are thought to increase cancer risk by a small amount, because the variants are common their total impact could involve a large number of cancer cases. Also, common genetic variants could affect biologic processes already known to play

a role in sporadic breast cancer cases; environmental risk factors may interact with common genetic variants to increase breast cancer risk. Therefore, low penetrance genetic variants are likely to affect susceptibility to sporadic as well as familial breast cancer.

A central part of determining the etiology of breast cancer is the recognition that breast cancer is composed of several different subtypes of disease that could have different underlying etiologies. Microarray-based gene expression profiling has led to the characterization of five different intrinsic subtypes of breast cancer – luminal A, luminal B, HER2+/ER-, basal-like, and an unclassified type that clustered with normal breast tissue (6-8). Studies suggest that the intrinsic breast cancer subtypes have unique risk factors (9-12), indicating that the cellular pathways involved in carcinogenesis may differ by subtype. The luminal A and basal-like subtypes are of particular interest. Luminal A is the most common subtype, whereas basal-like tumors have a poor prognosis and do not express any tumor markers for which there are targeted therapies (13).

Analysis of the epidemiology of intrinsic breast cancer subtypes in the Carolina Breast Cancer Study (CBCS) indicated that parity without breastfeeding (vs. nulliparity), lactation suppressant use, younger age at menarche, and high waist-to-hip ratio were risk factors for the basal-like subtype (9). Additionally, earlier first full term pregnancy (vs. nulliparity) was associated more strongly with basal-like breast cancer compared to later first full-term pregnancy (9). In contrast, nulliparity, and high WHR were risk factors for luminal A breast cancer (9). Older age at first full term pregnancy (vs. nulliparity) was not a risk factor for luminal A breast cancer; younger age at first full-term pregnancy (vs. nulliparity) appeared to have a protective effect (9). Reproductive history is a known risk factor for breast cancer, and is thought to influence risk through exposure to estrogen and progesterone

(14-16). Adiposity may be linked to breast cancer through regulation of pro-inflammatory adipocytokines and stimulation of sex steroid production (17-19). Additionally, the basal-like subtype may be more strongly related to family history of breast cancer compared to the other intrinsic subtypes (10). If the polygenic model of disease holds for basal-like breast cancer, association studies of common genetic variants should be able to identify loci that contain or are in linkage disequilibrium with causal disease alleles in hormone metabolism or central adiposity-related pathways.

I hypothesized that SNPs in genes associated with estrogen synthesis and signaling activity and central adiposity are associated with the incidence of the luminal A and basal-like subtypes of breast cancer, and that the effects of these SNPs may be modified by SNPs or environmental factors. To investigate this hypothesis, functional and tag SNPs were genotyped in cases and controls from the Carolina Breast Cancer Study. The Carolina Breast Cancer Study (CBCS) is a population-based case control study of African-American and white women in North Carolina. Subjects were sampled so that African American women and women younger than 50 would be over-represented compared to the general population. Basal-like breast tumors are most common in premenopausal African-American women (9, 20, 21), and so the CBCS sampling design provides increased statistical power for analyses of basal-like breast cancer.

In this study, the following research questions were addressed:

1. Based on the current literature, which candidate genes and candidate SNPs are most likely to be associated with basal-like and luminal A breast cancers?
2. What is the association between candidate SNPs and basal-like and luminal A

breast cancer?

3. For each candidate gene, are there haplotypes that are risk factors for basal-like and luminal A breast cancers?

4. Are the effects of genotypes and/or haplotypes modified by environmental factors?

These research questions were answered by completing the following specific aims:

Specific aim 1: Identify pathway-specific genes and relevant SNPs that may be important to the etiology of the basal-like and luminal A breast cancer subtypes

a. Conduct a literature review to identify candidate genes associated with risk factors for basal-like and luminal A breast cancers.
b. Evaluate the LD structure for each candidate gene using data from the International HapMap Project and previously published studies.
c. Select and genotype functional and tag SNPs in CBCS cases and controls.
d. Perform quality control and data cleaning steps, including tests of Hardy Weinberg equilibrium and evaluation of missing data.

Specific aim 2: Determine the association between genotypes and the basal-like and luminal A subtypes of breast cancer.

a. Calculate allele and genotype frequencies for each SNP (stratified by self-identified race).
b. Estimate ORs and 95% CIs for the association between each genotype and the basal-like and luminal A subtypes of breast cancer.

c. Adjust estimates from b. for population stratification.

Specific aim 3: Determine the association between specific haplotypes and the basal-like and luminal A subtypes of breast cancer.

a. Identify SNP regions for haplotype analyses
b. Estimate ORs and 95% CIs for the association between haplotypes and the basal-like and luminal A subtypes of breast cancer.
c. Adjust estimates from c. for population stratification.

Specific Aim 4: Determine whether genotype or haplotype effects are modified by interaction with environmental factors.

a. Evaluate multiplicative interaction between genotypes and haplotypes in ADIPOQ, IL6, LEP, LEPR, and TNF and WHR using the likelihood ratio test.
b. Evaluate multiplicative interaction between genotypes and haplotypes in CYP19A1, ESR1, HSD3B1, HSD17B2, PGR, and SHBG and parity and lactation using the likelihood ratio test.
c. Evaluate additive interaction between genotypes and haplotypes in ADIPOQ, IL6, LEP, LEPR, and TNF and WHR using the synergy index.
d. Evaluate additive interaction between genotypes and haplotypes in CYP19A1, ESR1, HSD3B1, HSD17B2, PGR, and SHBG and parity and lactation using the synergy index.

2.2 Overview of methods

The aim of this dissertation was to estimate the association between selected potentially functional and tag SNPs and the risk of the luminal A and basal-like breast cancer subtypes. This analysis was carried out using data from the Carolina Breast Cancer Study, a population-based case-control study of breast cancer in North Carolina (22, 23).

SNPs were genotyped from blood samples collected from cases and controls at the time of the in-home study interview. A panel of SNPs was genotyped for each gene of interest: (1) SNPs previously identified to have a functional effect on gene expression; function, or the circulating gene product; (2) SNPs evaluated previously for an association with breast cancer; and (3) tag SNPs, or the smallest set of SNPs that represents the majority of known genetic variation in the gene, as determined by linkage disequilibrium with other SNPs in the gene. Likelihood-based estimation methods were used to estimate the relative odds of basal-like and luminal A breast cancers due to SNP genotypes and haplotypes among African-American and white women in North Carolina. Odds ratios estimating genotype and haplotype associations in all breast cancer cases compared to controls were calculated as a representation of the overall association in the study population. Potential for additive and/or multiplicative interaction was evaluated for SNPs in the two candidate gene sets. Interaction between adipocytokine-related SNPs and WHR, and interaction between estrogen and progesterone-related SNPs and parity and lactation was explored for genotypes and haplotypes that were associated with luminal A or basal-like breast cancer.

2.3 Carolina Breast Cancer Study (CBCS)

2.3.1 CBCS study design

The Carolina Breast Cancer Study (CBCS) is a population-based case control study of

the genetic and environmental causes of breast cancer in African-American and white women in North Carolina. Study methods have been published previously by Newman et al. (22) and Millikan et al. (23). Study participants were recruited from 24 counties in central and eastern North Carolina (Figure 2.1). Counties were selected so that there would be a large proportion of African-American and rural residents.

Women were recruited to the study in two phases – between 1993 and 1996 (Phase 1) and between 1996 and 2001 (Phase 2). Cases were defined as any woman between the ages of 20 and 74, who lived in one of the 24 study counties and was diagnosed with primary invasive breast cancer during the study accrual period. Rapid case ascertainment systems implemented through the North Carolina Central Cancer Registry were used to identify all eligible cases. During Phase 2, women diagnosed with breast carcinoma *in situ* (CIS) were also enrolled in the CBCS.

Among eligible invasive breast cancer cases (eligibility based on age, gender, residence, diagnosis confirmation), a randomized recruitment case sampling strategy was used to insure adequate numbers of African-American and younger cases (24, 25). The study design sampled 100% of African-American women aged 20-49, 75% of African-American women aged 50-74, 67% of white women aged 20-49, and 20% of white women aged 50-74. No sampling was used for CIS cases recruited in Phase 2; all eligible CIS cases were included in the study.

Women residing within the study geographic area and without a history of breast cancer were eligible to be controls in the study. Potential controls younger than age 65 were identified through Department of Motor Vehicles records, while potential controls 65-74 years old were identified through Health Care Financing Administration records. Potential

controls were identified from these two sources using modified randomized recruitment (24) and were frequency matched to cases by race and 5-year age group distributions.

For cases, a letter was first sent to the patient's physician requesting permission to contact the patient about the study. Potential cases and controls selected by sampling were contacted by telephone. If a woman agreed to participate, an in-person interview was conducted by a trained registered nurse interviewer. During the interview, participants answered detailed questions about social and demographic characteristics, family history of cancer, reproductive history, menstrual history, exogenous hormone use, alcohol use, and occupational history. Anthropometric features, including height, weight, waist circumference and hip circumference, were measured by the interviewer. Participants were also asked to provide a 30 ml blood sample. DNA was extracted from blood and re-suspended in Tris-EDTA buffer for storage.

2.3.2 Case review and immunohistochemistry

Cases were asked to provide written consent for study access to their medical records and archival tumor tissue blocks. A centralized pathology review was performed to confirm each breast cancer diagnosis. CBCS immunohistochemistry (IHC) procedures are described fully in previous CBCS publications (13, 23, 26)

For invasive breast cancers, estrogen receptor (ER) and progesterone receptor (PR) status was abstracted from the patient's medical record (80% of invasive cases). If ER or PR status was not recorded in the medical record but archival tissue was available, the assay was performed at the UNC Immunohistochemistry Core Laboratory (20% of invasive cases). ER and PR IHC was repeated at the UNC Core Laboratory for a random sample of 10% of ER-positive and 10% of ER-negative invasive tumors where ER status

was initially recorded from the medical record (13). A comparison of the two assays resulted in a kappa statistic of 0.62 and an overall concordance of 81%, indicating agreement between the two data sources.

All IHC staining for epidermal growth factor receptor (EGFR), v-erb-b2 erythroblastic leukemia viral oncogene homolog 2 (HER2), and cytokeratin 5/6 (CK 5/6) was performed at the UNC Immunohistochemistry Core Laboratory. ER- and PR-positivity was defined as at least 5 percent of cells showing nuclei-specific staining (26). HER2 status was measured using the CB11 monoclonal antibody, as described previously by Millikan et al. (23). Tumors were scored as HER2-positive if they showed unambiguous membrane staining with weak, moderate, or strong intensity in at least 10% of tumor cells. Immunohistochemistry staining for EGFR and CK 5/6 was performed using criteria adapted from Nielsen et al. (27). Tumors were scored as positive for EGFR and CK 5/6 if they showed any cytoplasmic or membranous staining.

All immunohistochemistry for CIS tumors was performed by the UNC Immunohistochemistry Core Laboratory, and was described by Livasy et al. (28). ER-positivity was defined as an Allred score of 2 with nuclear staining. CIS cases were considered HER2-positive if they displayed greater than 10% membranous staining in CIS cells with an intensity equivalent to 3+ by DAB chromogen or 2+/3+ by SG chromogen. CIS cases were positive for EGFR if they showed any membranous staining and positive for CK 5/6 if they showed any cytoplasmic staining. PR expression was not used in classifying CIS cases due to the high correlation between ER and PR expression, and the need to preserve tissue (13).

2.3.3 IHC definition of intrinsic breast cancer subtypes

Breast cancer intrinsic molecular subtype was determined based on the joint expression of ER, PR, HER2, EGFR, and CK 5/6. Tumors were classified into one of five intrinsic subtypes using the hierarchy described in Nielsen et al. (27) and Carey et al. (13)(Figure 2.2). For invasive cases, the luminal A subtype included tumors that were ER+ and/or PR+ and HER2-. Luminal B tumors were ER+ and/or PR+ and HER2+. HER2+/ER- tumors were HER2+, ER- and PR-. Basal-like tumors were EGFR+ and/or CK 5/6+ and ER-, PR-, and HER2-. Tumors that did not express any of the IHC markers were categorized as 'unclassified'. CIS tumors were classified using a similar scheme, except PR status was not utilized.

2.3.4 CBCS participation rates

Overall response rates (proportion of women who completed the interview out of eligible, selected women) for invasive cases and controls were 76% and 55%, respectively. Among cases, rates were highest among non-African Americans younger than 50 years old (81%) and lowest among African Americans age 50 and older (70%). Among controls, rates were highest among non-African Americans age 50 and older (65%) and lowest among African Americans younger than 50 years (47%).

Overall response rates for CIS cases and matched controls were 83% and 65%, respectively. Among CIS cases, the highest response rate was from non-African Americans younger than 50 (86%) and the lowest response was from African Americans age 50 and older (76%). Among CIS controls the highest response was from non-African Americans age 50 and older (69%), the lowest response was from African Americans age 50 and older (51%). A total of 2311 cases (894 African American/1417 non-African American) and 2022

controls (788 African American/1234 non-African American) enrolled in the study.

Tumor tissue was available for 1845 of 2311 (80%) cases [1446 of 1808 (80%) invasive cases; 399 of 503 (80%) CIS cases]. IHC assays were completed successfully for 1424 of 2311 (62%) cases [1149 of 1808 (64%) invasive cases; 275 of 503 (55%) CIS cases]. Cases with subtype data were more likely to be African American and to have a later stage at diagnosis (9). African American cases in the CBCS had larger tumors compared to non-African American women, and larger tumors were more likely to have sufficient tissue for IHC assays (9, 13).

2.4 Genotyping analyses

2.4.1 SNP selection

Millikan et al. (9) reported that increased parity, younger age at first full-term pregnancy, not breastfeeding, high waist-to-hip ratio, and self-reported higher adult adiposity compared to childhood are risk factors for the basal-like subtype of breast cancer in the CBCS. These risk factors can be divided into two broad categories – pregnancy/hormonal-related factors and central obesity-related factors. In order to explore whether polymorphisms in genes related to these two broad categories of risk factors are associated with basal-like or luminal A breast cancer, a literature review was performed and several candidate genes hypothesized to play a role in breast carcinogenesis were selected for genotyping. The candidate genes in the pregnancy/hormonal factors-related pathway are CYP19A1, ESR1, HSD3B1, HSD17B2, PGR, and SHBG. Candidate genes in the central obesity-related factors are ADIPOQ, IL6, LEP, LEPR, and TNF.

For each candidate gene, a combination of potentially functional SNPs and tags SNPs

were selected for genotyping. A potentially functional SNP was defined as a single base pair change that either has been shown to affect gene transcription, transcription factor binding, serum protein levels, or breast cancer risk in previously published studies, or is suspected of having such a function. Genotyping of functional SNPs is essential because functional changes can indicate a potential biological mechanism through which a particular genotype is causally related to breast cancer. Due to the methods used for genotyping, only biallelic SNPs were included in this study. An allele is defined as the state, or base pair identity, in which either copy of a gene can exist (29); a biallelic SNP has two possible alleles.

Potentially functional SNPs were identified through a literature search in the National Library of Medicine (NLM) PubMed database (30). For each candidate gene, search terms combined all gene names and abbreviations (identified through the NLM OMIM and MeSH databases) with several MeSH terms and keywords that identified single nucleotide polymorphisms. The abstracts and articles from each gene search were reviewed for evidence of functional SNPs to be included in this analysis. Works cited in articles identified through this search were also reviewed for additional functional SNPs. Only potentially functional SNPs with a minor allele frequency of at least 0.05 in populations of European and African or African-American descent were genotyped.

Example: search terms for candidate gene interleukin-6

("Interleukin-6"[Mesh] OR "IL6 protein, human "[Substance Name] OR IL6 OR IL-6 OR Interleukin-6 OR “interleukin 6” OR IFNB2 OR “interferon beta 2” OR “interferon beta-2”) AND (single nucleotide polymorphism OR genetic polymorphism OR SNP OR polymorphism[TITLE] OR “germline mutation” OR “germ-line mutation”)

Tag SNPs, or SNPs that are highly correlated with other SNPs, were also selected for

genotyping (29, 31). Once a tag SNP is genotyped, the identity of untyped SNPs can be inferred, based on high linkage disequilibrium between the tag and untyped SNPs. As a result, genotyping tag SNPs identifies a large portion of the known genetic variation in a genomic region without genotyping every polymorphism.

Tag SNPs were selected using data from the International HapMap Project database (32). The CBCS is composed of white and African-American participants. African Americans are of both European and African ancestry and have different allelic frequencies and LD structures compared to European and African populations (33, 34). The International HapMap Project database currently has SNP data for multiple populations from around the world, including a population of African Americans from the southwestern US. However, this data was not available at the time that tag SNPs were selected. Therefore, in order to select tag SNPs that defined genetic variation in both white and African-American CBCS participants, tag SNPs were selected from HapMap populations representing both European and African ancestry. The HapMap genotype data for populations of European ancestry is based on 90 individuals from Utah, US who are of northern and western European ancestry (CEU population) (35). Genotypes for West African populations are based on data from 90 individuals of Yoruban descent in Ibadan, Nigeria (YRI population) (35).

The gene region to be tagged was identified based on the most 3' and most 5' SNP mapped to the candidate gene in the NCBI dbSNP database (36). Files containing the genotype data for this region in the CEU and YRI populations were downloaded from the HapMap website.

Tag SNPs were selected using the Tagger SNP selection program in Haploview version 3.32 (37, 38). Tagger constructs lists of tag SNPs using methods described by

Carlson et al. (39) and the correlation coefficient r^2. Pairwise r^2 is one measure of how well the identity of one allele at a polymorphic locus predicts the identity of the allele at another polymorphic locus (40) . The single SNP that is correlated with the greatest number of other SNPs at a pre-specified r^2 threshold is identified and grouped with its correlated SNPs into a bin. R^2 is re-calculated for SNPs in the bin to identify the best tag SNP. This process is repeated using the non-binned SNPs until only SNPs that are not in high LD with other SNPs are left. These are assigned to single-SNP bins. In this study, a minimum r^2 of 0.80 was required for tagging (39). Only SNPs with a minor allele frequency of 0.10 or greater were included in the tagging procedure, to increase the power to detect an association between the tag SNP and breast cancer and also to reduce the total number of tag SNPs per gene. Any potentially functional SNPs identified from the literature that was also genotyped in the HapMap database were included in the tagging procedure using the "force include" option in Tagger.

SNPs from other investigators were genotyped in the same GoldenGate assay, and so the number of SNPs that could be genotyped for this dissertation was limited. Aggressive tagging was used for ESR1, LEPR, HSD17B2, and PGR in order to reduce the total number of tag SNPs for each gene. The aggressive tagging procedure attempts to replace tag SNPs with 2-marker haplotypes composed of singleton SNPs (37). If the 2-marker haplotype can replace the tag SNP at the required r^2 threshold, then only 2 SNPs will be genotyped to determine the same amount of information that would require genotyping 3 SNPs using simple pairwise tagging. In this study, aggressive tagging reduced the number of required SNPs by an average of 3.5 SNPs when selecting tags for the CEU population and 11.5 SNPs for the YRI population.

Analyses of Phase 1 HapMap data indicate that coverage of common SNPs (minor allele frequency > 0.05) is high, though it varies throughout the genome (41). In white populations, use of CEU tags captured 86% to 100% of common SNPs with $r^2 > 80\%$ (41, 42). In African Americans, YRI tags captured approximately 80% of common SNPs with $r^2 > 80\%$ (41, 42). Coverage for African-American populations when SNPs were chosen from both YRI and CEU panels was not determined. In simulated case-control studies, the use of tag SNPs from both CEU and YRI populations increased the power to detect a causal allele in African Americans compared to when tags were selected from the YRI genotype panel only, particularly when the causal allele was untyped (90% vs. 81%) (42). In the same study, power for detecting untyped causal alleles in white American populations was between 91% and 95% using CEU tags (42). Therefore, the final tag SNP list for each candidate gene was a combination of the CEU and YRI tag SNPs (Table 2.1). All SNPs were genotyped in all CBCS study participants regardless of self-reported race.

Tagging procedures described above were used to select SNPs for ADIPOQ, IL6, LEP, LEPR, TNF, ESR1, PGR, HSD3B1, HSD17B2, and SHBG. For the CYP19A1, a set of 19 haplotype tag SNPs identified by Haiman et al. (43, 44) was submitted for genotyping, along with additional SNPs identified in the literature review. Using data from the Multi-Ethnic Cohort study, Haiman et al. (43, 44) used a panel of high density markers to map regions of LD in white, Hispanic, Japanese, Hawaiian, and African-American women. They identified four blocks of LD that could be defined in all five ethnic groups in their study population using 19 haplotype tag SNPs.

2.4.2 Ancestry informative markers (AIMs)

AIMs are a series of unlinked markers with large differences in allele frequencies that can

distinguish between ancestral populations (45, 46). Fisher's information content is the inverse of the maximum likelihood estimation of the ancestral proportion, and can be used in addition to allele frequency differences to increase the efficiency of AIM selection (46). AIMs were selected from a larger panel of AIMs that were used to estimate ancestry in African Americans (45). 200 AIMs that maximized Fisher's information content and the difference in allele frequencies for the CEU and YRI populations were submitted for genotyping (45).

2.4.3 Genotyping laboratory methods

SNPs were genotyped by the UNC Mammalian Genotyping Core using the Custom GoldenGate Genotyping assay from Illumina (Illumina, Inc., San Diego, CA). The GoldenGate assay allows multiple SNPs to be genotyped in a single assay using 2μg of study subject DNA. First, the SNP identity and position was validated by Illumina based on the SNP rs number, a unique identifier assigned by the NLM dbSNP database (36). An assay designability score was assigned to each SNP, indicating the likelihood for a successful genotyping in the multiplex assay. Possible reasons for SNP validation failure included a short flanking sequence, formatting error, more than two possible alleles, close proximity to another SNP being assayed, or the presence of degenerate nucleotides in the assay design region (47).

The GoldenGate assay has been described in detail by Shen et al. (48). Genotypes were determined using a solid-phase bead array with 1536 bead stations; 1536 SNPs were genotyped per study subject in each assay. DNA was activated and combined with hybridization buffer, paramagnetic particles, and oligonucleotides specific to each of the possible SNP alleles plus a unique oligonucleotide specific to a locus just downstream of the

SNP. In the first step of the assay, allele-specific and locus-specific oligonucleotides hybridized to the DNA sample. Then, allele-specific oligonucleotides were extended and joined with the locus-specific oligonucleotide, creating a double stranded DNA fragment. This fragment was amplified in a PCR reaction using primers labeled with Cy3 and Cy5 fluorescent dyes, where the dye color corresponds to a specific SNP allele. DNA was then hybridized to a bead array where the DNA strand localized to a specific bead station based on the locus-specific oligonucleotide sequence. The ratio of Cy3:Cy5 fluorescence was measured at each bead station on the array, where each bead corresponded to a unique SNP locus.

A list of all SNPs to be genotyped was submitted to Illumina, Inc. (Illumina, Inc., San Diego, CA) for validation. Of the 366 candidate gene SNPs submitted, 47 failed the Illumina validation assay. Failed SNPs were replaced by another SNP in the same bin with high correlation ($r^2 \geq 0.8$), if one existed. When multiple replacement SNPs were available to tag a given bin, preference was given to SNP with the highest correlation with the failed SNP. After repeated assay validation tests, 335 SNPs were genotyped in the GoldenGate assay. 158 of 200 AIMs passed Illumina validation and were genotyped in the GoldenGate assay.

2.4.4 Genotyping quality control

Several quality control measures were used to reduce the chances of genotype misclassification or other bias. Assay intensity data and genotype cluster images for all SNPs were reviewed individually. SNPs were excluded from the dataset due to low signal intensity or inability to clearly distinguish between genotype clusters. 163 of 1536 (11 %) SNPs were excluded from the entire dataset based on cluster analysis.

Blind duplicates of 169 study samples were assayed order to verify the reproducibility

of genotype calls from the same sample. 7 SNPs had 1 genotype miscall and 2 SNPs had 2 genotype miscalls. Lab controls (Coriell CEPH trios) were also genotyped in each 96-well plate; each control was repeated between 11 and 14 times over the course of the entire assay. Out of 184 lab control samples, there were 2 instances of genotype disagreement with duplicate samples. These error rates were within our pre-specified range of acceptable values, and no SNPs were excluded from the analysis on the basis of these results.

4155 CBCS DNA samples were genotyped, representing 3857 unique study subjects. 204 samples representing 135 unique subjects had genotype calls for less than 95% of SNPs and were excluded from the dataset. 103 subjects were excluded from the dataset because of low call rates (< 95%), 5 subjects were excluded because of an apparent gender mismatch, and 1 subject was excluded because of discordance between 2 non-blind DNA samples. Subjects excluded due to low call rates did not differ from the overall group submitted for genotyping with regard to case status, self-identified race, AJCC stage (cases only), or intrinsic molecular subtype (cases only).

A total of 312 of 335 (93 %) candidate gene SNPs were genotyped successfully in 3748 CBCS samples (ADIPOQ - 16, IL6 – 11, LEP - 14, LEPR - 74, TNF – 2, CYP19A1 - 24, ESR1 - 97, HSD3B1 - 7, HSD17B2 - 40, PGR - 26, SHBG – 1). All 23 SNPs with unacceptable genotype calls were excluded during the initial quality control review performed by the UNC Mammalian Genotyping Core, due to low signal intensity and/or poor genotype cluster definition.

2.4.5 Hardy Weinberg Equilibrium

Tests of Hardy-Weinberg equilibrium (HWE) were conducted for candidate gene SNPs to determine whether alleles were inherited independently in the CBCS source

population. Deviations from Hardy Weinberg equilibrium in controls can indicate genotyping error, selection bias, population stratification, new mutations, or a violation of the HWE population assumptions (49, 50). In cases, deviation from HWE can signal an association between a particular allele and disease. Deviations from HWE may also occur due to chance. Tests of HWE were conducted for each SNP in controls stratified by self-identified race using the method of Wigginton et al. (51) in Plink v1.05. The exact test was used because some SNPs had a low minor allele frequency. The asymptotic χ^2 test has high type I error rates at low minor allele frequencies (51, 52). The exact test was used for all SNPs for consistency. Genotyping cluster images were re-reviewed for all SNPs with HWE test P-values less than 0.01 to ensure data quality (Table 2.2). Data quality for these SNPs was confirmed, and no SNPs were excluded during this process.

2.4.6 Genotyping study participation rates

2045 of 2311 (88%) enrolled cases and 1818 of 2022 (90%) enrolled controls provided a blood sample at the time of interview. Non-African American women were more likely to provide DNA than African American women. Blood donation did not differ by other breast cancer risk factors or by stage at diagnosis among cases. 6 cases had insufficient amounts of DNA available for genotyping and were excluded from further analyses. An additional 67 cases and 42 controls were excluded during quality control analyses. A total of 1776 of 2022 (88%) controls and 1972 of 2311 (85%) cases were successfully genotyped. Overall, subjects without genotyping data were more likely to be cases, from the Phase 2 invasive study, and African American.

Among cases, 978 of 1808 (54%) invasive cases had both genotyping and tumor subtype data and 242 of 503 (48%) CIS cases had genotyping and subtype data, including

200 basal-like cases (182 invasive, 18 CIS) and 679 luminal A cases (528 invasive, 151 CIS). Stage at diagnosis, lymph node status, or the distribution of intrinsic molecular subtypes did not differ between enrolled cases with and without genotyping data.

A flow chart of participants' inclusion in the genotyping study is shown in Figure 2.3. Characteristics of CBCS participants included in the genotyping study are shown in Table 2.3. It is important to note that the denominator of the participation rate for the intrinsic molecular subtypes is the proportion of subtype cases successfully identified by IHC, not the total number of subtype cases enrolled in the CBCS.

2.5 Population stratification

Population stratification is a potential source of bias in any genetic association study conducted in an admixed population, like the CBCS. Population stratification occurs when a population is composed of multiple ancestral groups, and the ancestral groups have different allelic frequencies for the genetic marker of interest (46, 53). If the outcome is more common in one ancestral group, then members of that subpopulation are more likely to be among the case group, and any genotype present at a higher frequency in that subpopulation will be associated with the outcome, regardless of whether it is in linkage disequilibrium with the true causal allele (54).

Although self-identified race is expected to be correlated with ancestry, there is still potential for residual bias due to population stratification because of admixture of African and European ancestry in African Americans (46). There is also the possibility of cryptic stratification among white CBCS participants who are descended from multiple European populations with distinct genetic backgrounds (54). Studies have reported that African

Americans have approximately 5% to 20% European ancestry (33, 45, 55-59). Others have demonstrated that there is identifiable population substructure in Americans of European descent (60). Population stratification can also occur when the amount of admixture in the population varies among individuals (53).

2.5.1 Methods of assessing population stratification

There are several methods of assessing population stratification in genetic association studies. One method is to estimate individual ancestry and adjust parameter estimates for ancestry. Individual estimates of ancestry can be calculated using maximum likelihood estimation (MLE), and was described previously Barnholtz-Sloan et al. (56). This method requires that the study population is genotyped for a set of AIMs specifically selected to maximize the differences between ancestral populations, and knowledge of the allele frequencies in the ancestral populations (46). The MLE maximizes a log-likelihood equation that is a function of the observed allele frequencies in the admixed population, the contributions from the ancestral populations, and the difference in allele frequency between the ancestral populations (56). The likelihood is maximized using the Newton-Raphson grid search method, yielding proportions of ancestry for each ancestral population that sum to 1(56).

Structured association methods can also be used to estimate individual ancestry. Structured association uses genetic marker information to infer subpopulation membership, based on the pre-specified number of subpopulations [reviewed by (46)]. Structured association can use markers pre-selected to differ between ancestral populations (AIMs) or random genetic markers (46). Structure, a commonly used structured association program, uses Bayesian Markov Chain Monte Carlo estimation to simultaneously estimate allele

frequencies in subpopulations and individual ancestry proportions (54, 58, 61, 62). If the number of subpopulations is unknown ancestry can be inferred based on the posterior probability for a range of subpopulations, though the number of subpopulations may not have a valid interpretation in the context of the data (58). Studies comparing maximum likelihood and structured association ancestry estimates report that the two methods are highly correlated (55, 59, 63).

Genomic control was proposed by Devlin and Roeder to eliminate the problems of type I error that may occur due to population stratification or cryptic relatedness in case control studies (64). Association test statistics can be inflated when there is population substructure (64). The genomic control method of adjusting for population stratification involves calculating a variance inflation factor for a set of random, unlinked SNPs across the genome, and adjusting all SNP association tests for the extra variance due to population substructure (64, 65). Genomic control assumes that the variance inflation is roughly constant across all loci being tested, but this is not always true leading to possible over- or under-adjustment of variance for different loci. Case-control simulations by Devlin and Roeder showed that despite controlling for type I error, genomic control also resulted in reduced power to detect risk alleles (64). Marchini et al. (66) demonstrated that using too few markers for genomic control can lead to false positives, and the degree of bias increases with increasing sample size. While using more markers can reduce the chance of type I error it can also lead to loss of power (66).

Principal components analysis is a method of transforming correlated variables into new, uncorrelated variables based on the linear relationships that can be defined within the data (67). Applied to population genetics, the goal is to identify the principal components that

describe the variation in a set of genetic markers (68). Price et al. (60) described the use of principal components analysis to control for population stratification by identifying the axes of genetic variation, continuously adjusting genotypes and phenotypes for ancestry along each axis, and calculating associations using the adjusted genotypes and phenotypes. One of the major advantages of principal components compared to structured association is that it is much more efficient at determining population structure using a very large number of markers. This is not an advantage for ancestry estimation in the CBCS because of the prior decision to genotype using the 1536 marker custom GoldenGate platform instead of a much larger GWAS platform. Furthermore, the principal components method has a higher rate of type I error when the number of markers used to identify genetic variation is low (60). This method is more suited to large GWAS datasets rather than candidate gene studies with a limited number of SNPs, such as the CBCS genotyping panel.

MLE, structured association, and principal components all require some prior knowledge of the number and types of subpopulations present in the data, either for AIM selection or interpretation of the number of populations inferred from the data. However, unlike MLE and structured association methods, genomic control and principal components analysis do not explicitly model individual ancestry estimates. As such, MLE and structured association provide more information that allows for characterization of the distribution of ancestry and admixture in the study population as opposed to providing methods mainly intended to adjust for population stratification.

2.5.2 African and European ancestry in the CBCS

144 ancestry informative markers (AIMs) were genotyped in the same Illumina GoldenGate assay as the candidate gene SNPs, using methods described above. Individual

proportions of African and European ancestry were estimated using maximum likelihood and structured association. Under the maximum likelihood method, the ancestry proportions were estimated for each study subject by solving likelihood equations described by Barnholtz-Sloan et al. (56) using the Newton-Raphson method. AIMs were selected to describe ancestry with regard to African and European populations only, and the maximum likelihood equation is restricted such that the ancestry proportions add up to 1 (56). The proportion of African ancestry in CBCS subjects is described and utilized in the remainder of this dissertation; the proportion of European ancestry is equal to one minus the proportion of African ancestry.

The structured association estimates were generated using Structure v.2.0, which uses Bayesian estimation to determine the proportion of a subject's genome that belongs to each ancestral population cluster, K (58, 61, 62). Preliminary estimates were calculated for K=1 through K=5 and the most likely number of populations was determined to be 2, based on a plateau reach in the estimated log probability of the data (69). Cluster membership and ancestry estimates were re-calculated for K=2 using the admixture and correlated allele frequencies models (58).

Results from the maximum likelihood ancestry estimation are shown in Figure 2.4. The median proportion of African ancestry was 81% in African Americans and 6% in whites. Subjects who reported that they were white, Hispanic, and Asian/Pacific Islander had less than 50% African ancestry. Subjects who described themselves as American Indian or Eskimo had varying amounts of African ancestry, ranging from 2% to 89%. The majority of African Americans had between 50% and 96% African ancestry. There were some self-identified white subjects with relatively high proportions of African ancestry, and some self-identified African Americans with relatively low proportions of African ancestry (Figure

2.4). Subjects with seemingly discordant race and ancestry results are not concentrated in any particular genotyping plate, column, or row, so it is unlikely that these results are due to some systematic error during plating and/or genotyping (data not shown).

Ancestry estimates obtained from Structure were similar to MLE estimates, but were skewed towards the ends of the distributions (Figure 2.5). The median Structure-determined African ancestry was 92% in African Americans and 1% in whites. Ancestry estimates from the two methods were highly correlated (all subjects $r^2 = 0.97$, $P < 0.0001$; African American $r^2 = 0.99$, $P < 0.0001$; non-African American $r^2 = 0.87$, $P < 0.0001$) (Figure 2.6).

Recent studies using MLE and Structure have estimated that African Americans have between 77% and 87% African ancestry (45, 55, 56, 58, 59). Parra et al. (33) reviewed early studies of ancestry estimation and found that estimated African ancestry in African Americans from South Carolina and Georgia varied from approximately 85% to 96%. Others have reported that MLE and Structure ancestry estimates are highly correlated in admixed populations (55, 59, 63). Shriver et al. (59) also reported that the correlation between Structure and MLE individual ancestry estimates was higher in African Americans from Washington, DC compared to white Americans from State College, PA, which is what was observed in the CBCS.

2.6 Genotype associations

2.6.1 Determination of genotype and allele frequencies

Genotype and allele frequencies were calculated stratified by case status and self-identified race. Genotype and allele frequencies were adjusted for the sampling probabilities used during study recruitment so that frequencies reflect the prevalence in the general

population.

2.6.2 Genotype exposure definitions

There are several potential model forms that can be used to model genotype effects. A genotype refers to the allelic identity at a particular locus on both copies of a chromosome. All SNPs included in this study were biallelic, so there are three possible genotypes for each SNP locus – homozygous for the major (common) allele, heterozygous, or homozygous for the minor (rare) allele. It is possible that a SNP may have only two genotypes if the third genotype is rare or does not occur in the study population. Under the dominant genetic model, one variant allele is sufficient to affect disease risk and the effects of the heterozygote and minor allele homozygote genotypes are the same. Under the recessive genetic model, two variant alleles are needed to affect disease risk, and the risk of disease is the same among major allele homozygotes and heterozygotes. Under the additive genetic model, the effects of genotype are linear, and the change in disease risk is proportional to the number of variant alleles in the genotype. In a logistic regression model, this means that the log odds ratio for the minor allele homozygote genotype is twice that of the heterozygote genotype. Under the general model, no relationship is assumed between heterozygote and minor allele homozygote effects.

All SNP associations were estimated initially using the general model with 2 degrees of freedom. In simulations of a linear regression model, Lettre et al. (70) showed that the general model has only slightly less power than the true, correctly specified mode of inheritance; this is expected because the general model requires two degrees of freedom whereas the additive, recessive, and dominant models all have one degree of freedom, assuming no covariates. However, when the true model form is not known the general model

has greater power compared to if the wrong model form is chosen (for example, if the effects are modeled as dominant but the true underlying mode of inheritance is recessive) (70). The difference in power between the wrong model and the true and general models was greatest at low minor allele frequencies, whereas the difference between the true and general models was constant across the range of minor allele frequencies (70). Because the mode of inheritance for most SNPs in this study is unknown, not imposing a particular mode of inheritance by using a general model enables a more accurate characterization of genotype effects.

For each locus, 2 indicator variables were created to model the effects of the heterozygote and variant allele homozygote genotypes compared to the major allele homozygote genotype. Estimated ORs contrasted the effect of minor allele homozygote vs. major allele homozygote and heterozygote vs. major allele homozygote, where the major allele homozygote served as the reference group.

When the general model ORs suggested that a particular locus had an additive, dominant, or recessive mode of inheritance, a second model was created using the more appropriate genetic model. When there were too few minor allele homozygotes in the study population, then minor allele homozygotes and heterozygotes were combined into a single category and compared to major allele homozygotes (equivalent to the dominant genetic model). Homozygote count cutoffs of less than 5 cases, or less than 5 controls, or less than 10 cases and less than 10 controls were used to define when there were too few homozygotes for odds ratio estimation.

2.6.3 Maximum likelihood logistic regression

Logistic regression was used to estimate the association between candidate SNP

genotypes and breast cancer. First, unconditional binary maximum likelihood regression models were used to estimate the effect of a given genotype on case or control status. In this model, the outcome was 'any breast cancer case,' and represents the association that is comparable to other studies that are unable to stratify breast tumors by intrinsic molecular subtype. Results from the binary logistic regression models were also used to compare genotype associations in the CBCS to previously reported associations from other studies. The binary logistic model function is:

$$\text{Logit}[D=1|X=x] = \alpha + \beta_1 X_1 + \beta_2 X_2$$

where α = regression model intercept

β_1 = regression coefficient corresponding to heterozygote genotype

β_2 = regression coefficient corresponding to minor allele homozygote genotype

X_1 = presence or absence of heterozygote genotype

X_2 = presence or absence of minor allele homozygote genotype

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

Unconditional polytomous logistic regression models were used to estimate the association between SNPs and the intrinsic breast cancer subtypes. The polytomous logistic regression model simultaneously estimates regression parameters for n-1 comparisons when there is an n-level outcome variable. Here, the models estimated the log odds of luminal A, basal-like, HER2+/ER-, luminal B, and unclassified breast cancer compared to controls, yielding 5 intercept parameters and 5 effect estimates for each independent variable in the model (71). Parameter estimates were determined using maximum likelihood estimation, similar to when the outcome is binary. The logit functions for the polytomous model are:

$$\text{Logit}[D=\text{basal-like}|X=x] = \alpha_{basal} + \beta_{basal(1)}X_1 + \beta_{basal(2)}X_2$$

$$\text{Logit}[D=\text{luminal A}|X=x] = \alpha_{lum\ A} + \beta_{lum\ A(1)}X_1 + \beta_{lum\ A(2)}X_2$$

$$\text{Logit}[D=\text{HER2+/ER-}|X=x] = \alpha_{HER2} + \beta_{HER2(1)}X_1 + \beta_{HER2(2)}X_2$$

$$\text{Logit}[D=\text{luminal B}|X=x] = \alpha_{lum\ B} + \beta_{lum\ B(1)}X_1 + \beta_{lum\ B(2)}X_2$$

$$\text{Logit}[D=\text{unclassified}|X=x] = \alpha_{unclass} + \beta_{unclass(1)}X_1 + \beta_{unclass(2)}X_2$$

According to Hosmer and Lemeshow (71), parameter estimates from individual binary logistic regression models where the cases are restricted to a particular subtype should be close to the estimates from the polytomous model. However, it is recommended that final parameter estimates and standard errors come from the polytomous model (71).

Odds ratios (ORs) and 95% confidence intervals (95% CIs) for the association between genotypes in the candidate genes and the basal-like and luminal A subtypes of breast cancer will be estimated using both binary and polytomous logistic regression models. An offset term will be included in all regression models to account for the sampling structure in the study design (24, 25). A statistic testing the equality of parameter estimates for basal-like and luminal A subtypes was calculated based on the asymptotic chi-square distribution of the Wald statistic. The null hypothesis for this test was H_0: $\beta_{basal(i)} = \beta_{lum\ A(i)}$.

2.7 Haplotype associations

A haplotype is a sequence of alleles on the same DNA strand, and represents the biologically relevant unit of DNA sequence. Haplotype analyses can have greater power compared to multiple single SNP analyses when the causal allele is unknown, though power may be reduced when there are many haplotypes (72). Haplotype analyses may also have greater power than single SNP analyses if multiple causal alleles have greater joint effects

when they are inherited together on the same DNA strand (73, 74).

In case-control studies like the CBCS, family genotype data is unavailable and haplotypes must be inferred based on the observed genotypes in the study population. Several methods have been described for inferring haplotypes in case-control data, such as Clark's algorithm, the expectation-maximization (EM) algorithm, coalescence-based algorithms, and the partition-ligation method (75). However, when inferred haplotypes are used directly in the estimation of regression parameters without consideration of the probability of haplotype assignment for each individual, regression parameters can be biased (72).

2.7.1 HAPSTAT and HAPSCAN

Haplotype frequencies, ORs, and standard errors were estimated using a modified version of the HAPSTAT program, which estimates the probability of a given haplotype using maximum likelihood estimation and the EM algorithm (76, 77). The modified HAPSTAT program relaxes the assumption of independence between genetic and non-genetic variables, allowing adjustment for self-identified race, African ancestry, and age. HAPSTAT was also modified to incorporate the offset term in parameter estimation. HAPSTAT uses a probability distribution to estimate the haplotype-associated ORs and standard errors, yielding unbiased parameters estimates with appropriate variance (76). Additionally, HAPSTAT allows the user to model gene-environment interactions and can accommodate missing genotype data and deviations from Hardy Weinberg equilibrium (76). HAPSCAN uses HAPSTAT algorithms to estimate a global test of haplotype association for a given set of SNPs, and can be programmed to test multiple groups of SNPs.

SNPs were selected for haplotype analysis using two methods. First, haplotype effects were estimated across each candidate gene using HAPSCAN. Overlapping sliding 3-SNP and

5-SNP windows were used. HAPSCAN was run several times, defining the outcome as all breast cancer cases, luminal A cases only, and basal-like cases only. Regions with the highest $-\log_{10}$ P-values were similar for 3-SNP and 5-SNP windows, and analysis continued with 3-SNP windows. Sliding windows with global test $-\log_{10}$ P-values greater than 2 were chosen for haplotype OR estimation. If the global test $-\log_{10}$ P-values indicated the haplotype association extended over multiple windows, then SNPs from consecutive windows were combined for haplotype frequency and OR estimation. Next, genotype ORs were reviewed for consecutive SNPs with non-null ORs. There were 4 SNP regions that were identified by both HAPSCAN and review of single SNP ORs.

For each region identified as described above, preliminary estimates of all possible haplotype ORs and 95% CIs were estimated using SAS/Genetics v9.1.3. ORs were estimated for breast cancer overall and the luminal A and basal-like subtypes, regardless of which outcome was associated with a potential haplotype effect in the review of single SNP ORs or the HAPSCAN analysis. Haplotypes with an OR of 1.5 or greater or less than 0.67 were re-calculated using HAPSTAT. A comparison of ORs and 95% CIs showed that estimates from SAS/Genetics were systematically further from the null compared to HAPSTAT estimates, so it is unlikely that any associations were missed by not calculating all possible estimates in HAPSTAT (data not shown). All haplotype associations were estimated using the general model (2-d.f.).

2.8 Interaction

The biologic mechanism that leads to the formation of breast cancer is likely to have multiple component causes that interact. In addition to estimating main effects, statistical

interaction with non-genetic breast cancer risk factors was evaluated in order to estimate the extent to which those factors influence genotype and haplotype associations. WHR and combined parity and lactation were chosen for interaction analyses because parity without lactation and WHR were two strong risk factors for basal-like breast cancer in the CBCS, the association for those factors and luminal A breast cancer was either weaker or in the opposite direction, and the candidate genes were chosen based on biology associated with these two risk factors (9). I hypothesized that candidate gene associations would be modified in the presence of the environmental risk factor.

Evidence for a biological interaction between genotypes or haplotypes was inferred from measures of statistical interaction calculated from regression models. Statistical interaction, or effect measure modification, occurs when the joint effects of two exposures are not additive for difference measures or not multiplicative for ratio measures. Some have argued that independent steps of a multistage process, such as cancer, have multiplicative effects (78). On the other hand, additive interaction on the risk scale may be more reflective of biological interaction in simple systems (78). As Greenland and colleagues note (78), if both exposures have marginal effects, the presence of exact additivity on one scale implies departure from additivity on another scale.

To limit the number of comparisons, potential effect measure modifiers were selected based on the plausibility of biologic interaction. Two-way interactions between the effect measure modifier and SNPs with a main effect OR greater than 1.5 or less than 0.67 were evaluated to determine whether the joint effect is associated with a departure from additivity or multiplicativity. Although there is a chance that synergistic interactions were missed [SNPs have no effect individually, but have a causal effect when both the effect modifier is

present (78)] by limiting this analysis to SNPs that showed a main effect, it was necessary to limit the number of interaction terms that were evaluated.

Effect measure modification was evaluated through the introduction of a multiplicative interaction term to the regression model. Departures from the multiplicative null were evaluated using the likelihood ratio test (LRT), which compares the -2 log-likelihood of nested models under the null hypothesis that the addition of the interaction term does not improve model fit. Likelihood ratio test P-values less than 0.10 were considered evidence of multiplicative interaction, and stratified ORs were calculated for those SNPs.

Departures from the additive null were evaluated by calculating the synergy index (S) and the corresponding 90% CI (79). S compares the excess risk of joint exposure allowing for interaction to the excess risk of joint exposure under the additive null. The null value for S is 1; values less than 1 indicate less than additive interaction and values greater than 1 indicate greater than additive interaction. Although the interaction contrast ratio (ICR) is commonly used to assess additive interaction and may have a more accessible interpretation, the ICR is not valid when covariates are included in the logistic regression model (80). All models in this analysis must be adjusted for age and self-identified race due to frequency matching. This problem may be avoided if the ICR is estimated using a linear odds ratio model instead of a log-linear model (80, 81), however procedures available to fit linear odds ratio models in SAS did not allow for the incorporation of an offset term.

A basic logistic regression model allowing for interaction is:

$$\text{logit}[D=1|X=x] = \alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 (X_1)(X_2)$$

where α = regression model intercept

X1 = exposure

X2 = effect modifier

β_1 = effect of exposure on outcome

β_2 = effect of effect modifier on outcome

β_3 = excess effect due to joint exposure

The synergy index is calculated as:

$$S= [e^{(\beta 1 + \beta 2 + \beta 3)} – 1]/[(e^{\beta 1} -1) + (e^{\beta 2} -1)]$$

2.8.1 Parity and lactation variable definition

Parity was measured by self-report during the study interview. Women were asked how many times they had been pregnant in their lifetime, including the current pregnancy if they were pregnant at the time of the interview. Women were then asked the duration of each pregnancy and the outcome. Parity was defined as the total number of full-term births reported by the study subject. Lactation was measured by self-report during the study interview. For each live birth reported, women were asked whether they breastfed the baby and for how long (in months). Only 5 subjects with genotyping data were missing information on lactation history. This constitutes less than 1% of the population and is unlikely to bias the results. There was no missing data for parity.

The association between parity and basal-like breast cancer did not differ for CBCS participants with 1-2 children compared to 3 or more children (where nulliparous women were the referent group) for women with the same lactation status (9). Likewise, the association between parity and lactation and luminal A breast cancer was the same for women with 1-2 children compared to women with 3 or more children for women with the same lactation status (9). Therefore, parity and lactation was defined as a single 3-category variable: nulliparous (controls N=201; all cases N=301; luminal A N=111; basal-like N=24),

parous/never breastfed (controls N=878; all cases N=983; luminal A N=317; basal-like N=124), and parous/ever breastfed (controls N=694; all cases N=686; luminal A N=251; basal-like N=52).

2.8.2 Waist-hip ratio (WHR) variable definition

WHR was calculated as the ratio of waist circumference to hip circumference. Waist and hip circumference were measured using a tape measure by a trained nurse-interviewer during the study interview and were recorded to the nearest half centimeter. The waist measurement was taken at the natural indentation of the waist. Hip circumference was measured at the greatest protrusion of the buttocks. Both measurements were taken twice and averaged. If measurements differed by more than 1 cm, a third measurement was taken and the two closest measurements were averaged.

WHR was categorized based on the tertile distribution CBCS controls. Associations between WHR and basal-like breast cancer and WHR and luminal A breast cancer were similar for tertile 2 and tertile 3 (vs. tertile 1) in the CBCS (9), and so the top two tertiles were combined. Data on WHR was missing for 21 controls and 29 cases (5 basal-like cases, 7 luminal A cases). The proportion of missing WHR data was similar for cases compared to controls as well as basal-like cases compared to controls and luminal A cases compared to controls.

2.8.3 Body mass index (BMI) variable definition

Studies suggest that body mass index (BMI, weight in kg/height in m^2) is a confounder of the association between WHR and breast cancer in premenopausal women [reviewed by (82, 83)]. Preliminary analyses showed that BMI was a confounder of the association between WHR and breast cancer in the CBCS; ORs estimated from models not

adjusting for BMI were closer to the null (data not shown). Therefore, all models estimating WHR parameters were adjusted for BMI. BMI was calculated from height and weight measured during the study visit. Weight was the average of two measurements taken using a standardized scale and recorded to the nearest half kilogram. Height was the average of two measurements made to the nearest half centimeter.

BMI data was missing for 78 subjects with genotyping data (37 controls, 41 cases, 5 basal-like cases and 15 luminal A cases), 10 of whom were classified as WHR < 0.77, 30 of whom were classified as WHR ≥ 0.77, and 38 of whom were missing WHR data.

A total of 90 of 3748 (2%) genotyped subjects (40 controls, 50 cases, 6 basal-like cases, 17 luminal A cases) were missing data for either WHR or BMI and were excluded from the WHR effect measure modification analysis. Proportions of subjects missing data for either WHR or BMI did not differ by case status. The low proportion of missing data combined with the fact that missingness was unrelated to case status strongly indicate that missing WHR or BMI data was not a source of bias in the WHR interaction analyses.

2.9 Methodological issues

2.9.1 Inclusion of breast carcinoma *in situ* (CIS)

There is evidence for and against the theory that CIS is an intermediate step in the progression of cells from hyperplastic to malignant. CIS and invasive lesions from the same woman often share genetic changes, grade, and gene expression patterns, and regions of the breast with invasive tumors are more likely to also have CIS [reviewed by (84-88)]. However, early studies of the natural history of DCIS show that not all women with CIS progress to invasive breast cancer [reviewed in (89)]. Morrow et al. (90) details the

ambivalence of some towards the grouping of CIS with invasive breast cancer because of the lack of predictors to reliably determine which CIS will progress to invasive disease.

CIS cases were included in this dissertation for several reasons. First, many risk factors for DCIS are similar to risk factors for invasive breast cancer including family history, which is likely to have a strong genetic component [reviewed in (89)]. Two studies that compared multiple risk factors for invasive breast cancer and DCIS in the same screening population reported that associations were similar for family history of breast cancer, previous breast surgery, postmenopausal hormone use, and hysterectomy (91, 92). Associations differed between invasive breast cancer and DCIS for age, early age at menarche (among women 50 and older), and presence of a palpable mass (91). There were conflicting results between the two studies for age at first birth and BMI in older or postmenopausal women (91, 92).

Another reason CIS were included in this dissertation is that CIS and invasive tumors often share similar tumor characteristics. Most importantly for this study, intrinsic molecular subtypes have been observed in pure DCIS lesions and in DCIS observed alongside invasive breast tumors (28, 93-95), indicating that changes associated with the basal-like phenotype occur early. Like invasive tumors, most basal-like DCIS showed strong expression of CK 5/6, vimentin, EGFR, and Ki-67; expression of p63 and smooth muscle actin was rare (28, 93).

It has not been reported whether intrinsic molecular subtypes are associated with recurrence. However, the similarity between risk factors and molecular features suggest that there are a common set of factors that lead to both types of lesions. For these reasons, subjects recruited for the CIS study from this analysis were not excluded from the study

population. A sensitivity analysis comparing genotype ORs for all cases (invasive and CIS) and controls to ORs for invasive cases and controls sampled for the invasive study was performed to evaluate the effect of retaining CIS cases and matched controls.

2.9.2 Selection bias and missing data

Selection bias is defined as the distortion of an estimated effect due to procedures used to select subjects for inclusion in a study or from factors that affect participation (96). Selection bias can be thought of as a type of missing data problem, where in a complete case analysis non-participants are not included in the analysis. If missing data is unrelated to case, exposure, or covariate status it is said to be missing completely at random and will not bias effect estimates. Under this scenario, a complete case analysis will yield an unbiased odds ratio. For selection bias, this is equivalent to the situation where non-participants are a random sample of the eligible source population. Most of the analyses performed in this study involve a 3-level exposure, and the severity of bias due to selection or missing data for a specific SNP or haplotype depends on the true distribution of cases and controls across those three levels.

Selection bias and missing data have the potential to affect the results of this study at each step where data was obtained from participants. Initial study recruitment was related to case and race status, and inclusion in the genotyping phase of the study was related to race. If factors related to selection bias are measured, they can be adjusted for in statistical analyses (97). Race was a frequency matching factor and was automatically included in all models. This should minimize the possibility of selection bias, assuming that genotype distributions are similar within self-identified race groups. All analyses also adjust for ancestry. So even if there are residual differences in allelic distributions within race groups due to admixture,

adjustment for African ancestry should control for this potential bias. Several hundred enrolled participants were not included in the genotyping study. Similar to the overall study, inclusion in the genotyping study was related to case status and race. Adjustment for race and ancestry in statistical models should control for any potential bias from this step.

Potential selection bias at a third step of the study only affects the analyses where the outcome is basal-like or luminal A breast cancer. Not all cases had IHC data available in order to determine the breast tumor intrinsic molecular subtype. The presence of subtype data was related to race and tumor stage. There is the potential for selection bias if the molecular subtype distribution for cases without IHC data differs from the distribution of molecular subtypes in the case source population.

It was difficult to assess how likely bias is to occur in this situation. It has been observed that the prevalence of the basal-like and luminal A subtypes is related to race, but whether this is true depends on the subtype distribution in the uncharacterized tumors. The actual distribution of these mostly smaller tumors will probably never be measured accurately. A central question is what is the distribution of intrinsic molecular subtypes among early stage tumors? Data from external populations will be subject to the same problem of smaller tumors not having sufficient tissue for research purposes after diagnostic obligations are met, and thus will have the same potentially skewed distributions as in the CBCS.

To summarize, there were several parts of the study where selection bias may influence the results. Most of the potential selection bias in the analysis of breast cancer overall is due to lower inclusion rates among African Americans and cases. Adjusting for race and ancestry in models should control for this potential bias, assuming that the sampling

procedures selecting cases and controls were unbiased. Inclusion in molecular subtype-specific analyses was related to stage in addition to race. One way to limit possible bias would be to conduct subtype analyses with cases stratified by stage. However, it was not feasible to stratify cases by stage because of sample size limitations.

2.9.3 Confounding

Confounding is defined as the mixing of effects of a third covariate with that of the exposure on disease, resulting in a biased effect estimate. In order for a covariate to be a confounder, it must be causally related to the exposure of interest and the outcome of interest (98). In this study, the main exposure in all analyses was a genotype or a haplotype. Therefore, any confounders would have to causally affect a genotype or haplotype to meet the confounding criteria described by Rothman and Greenland. The effect of potential confounders was also evaluated using statistical models. If the addition of a covariate changed the |ln(OR)| of the exposure of interest by more than 0.10, then that covariate was considered a confounder.

While some associations between environmental variables and genotype may be observed due to the random error, no environmental variables were expected to be causally associated with candidate gene genotypes (meaning that on a directed acyclic graph (DAG), genotype was not the descendent of an environmental variable). Therefore even if the covariate was associated with genotype and outcome, it still did not meet the definition of a confounder. Even more importantly, if that covariate was on the causal pathway between genotype and the outcome, adjusting for that covariate could bias the association between exposure and outcome.

The only breast cancer risk factor that was likely to be causally associated with

candidate genotypes was ancestry. The proportion of African ancestry was estimated for each subject and included in all models, as described above. Adjustment for ancestry had little effect on ORs estimated for breast cancer overall and basal-like breast cancer. Ancestry adjustment did affect ORs for a small number of genotypes and luminal A breast cancer, and was adjusted for in all models to control for bias in these associations.

Effect measure modifiers in analyses of genotype-environment interactions can be susceptible to confounding by other environmental variables. However, this was problematic because adjusting for a confounder of the effect modifier has the potential to bias the association of the genotype. Even more importantly, if the potential confounder was on the causal pathway between genotype and the outcome, adjusting for that covariate could drive the estimated genotype effect towards the null.

Relationships between WHR and basal-like and parity, lactation, and luminal A breast cancer were explored using DAGs. In the WHR-breast cancer DAG menopausal status and parity/lactation status were identified as potential confounders. Potential confounders from DAG analysis were evaluated for a statistical effect on the parameter estimate of interest. Adjustment for these two risk factors did not alter the parameter estimates for WHR, and they were not included in further WHR interaction analyses. Reviews of the literature suggest that BMI is a confounder of the association between WHR and breast cancer in premenopausal women, and failure to adjust for BMI biases associations towards the null (82, 83). In CBCS data, BMI adjustment biased the association for between WHR and breast cancer overall by more than ln(OR) = 0.10. The bias for basal-like and luminal A associations was lower than this threshold, but closer to 0.10 than to 0. Based on the effect of BMI adjustment in CBCS data and the acknowledgement of BMI as a confounder in the

literature, BMI was kept as a confounder in WHR interaction models.

In the DAG examination of relationships between parity, lactation, and breast cancer, menopausal status and age at menarche were identified as potential confounders. Neither of these risk factors affected parameter estimates for the association between the combined parity and lactation variable and basal-like or luminal A breast cancer. Neither factor was included in further analyses as a confounder.

2.9.4 WHR misclassification

Waist and hip circumference were measured at the time of interview. Cases were interviewed a median 3.9 months (range, 0.8 – 42.5 months) after diagnosis, meaning that waist circumference and hip circumference may have been measured after the start of adjuvant therapy for some cases. If case WHR at the time of interview was systematically different from pre-diagnosis WHR, there is the potential for misclassification. There was no systematic event that would have led to WHR change in controls, so misclassification would be non-differential.

Weight change is a commonly documented side effect of breast cancer-related therapy [reviewed by (99, 100)]. In most patient series, patients gained approximately 2 to 20 pounds, and the amount of weight gained varied by study cohort and treatment (99, 100). Most studies reported that weight gain began shortly after breast cancer diagnosis, and the amount of weight gained increased over time (100-104). In some studies, patients experienced weight loss during the year following diagnosis (101, 105). Freedman et al. (106) reported that a group of healthy controls gained more weight on average than breast cancer patients receiving adjuvant therapy, but that the breast cancer patients had a greater fluctuation in weight during the time period shortly before the initiation of chemotherapy

until 6 months post-chemotherapy completion.

Ingram et al. (107) reported that post-diagnosis weight change was related to the type of adjuvant therapy, but other studies found no difference by chemotherapy type or regimen (103, 108, 109). Studies have also reported that weight change in breast cancer patients is associated with being premenopausal (99, 101, 106, 110). Two studies reported that lower pre-diagnosis BMI was associated with weight change, but another study did not find an association (101, 103, 109). There is also evidence that African-American breast cancer patients experienced greater weight gain compared to white patients, especially following adjuvant chemotherapy (101, 109).

In addition to changes in weight, studies have reported that breast cancer patients experienced increases in body fat percentage, fat mass, waist size, and hip size (102, 106, 111-113). Goodwin et al. (114) reported that although waist size and hip size increased 1 year after diagnosis, WHR was unchanged over the same time period. However women may have already started chemotherapy at the time of baseline WHR measurement, biasing the association toward the null.

A sensitivity analysis was conducted in order to estimate the potential effect that WHR misclassification due to chemotherapy might have on the association between WHR and basal-like breast cancer. The sensitivity analysis was conducted using a publicly available probabilistic bias analysis program <https://sites.google.com/site/biasanalysis/> (115), which calculates a simulated data table of "true" classification based on the observed data table and estimated sensitivity and specificity of the classification. The CBCS lacks data on whether cases had started chemotherapy by the time of interview. Sensitivity and specificity ranges were estimated based on the stage and race distribution in CBCS basal-like

breast cancer cases and the prevalence of chemotherapy treatment by stage in North Carolina Central Cancer Registry data (116).

2.9.5 Outcome misclassification

Probabilistic sensitivity analyses were also conducted to evaluate the potential effects of molecular subtype misclassification. There has been some discussion in the literature as to the true definition of 'basal-like' breast cancer (117, 118). Not all studies use the same set of markers to define 'basal-like', and in studies that have used markers similar to those used by CBCS, there was not 100% agreement between tumors defined as basal-like using microarray expression profiles and immunohistochemistry definitions (27, 119, 120). Simulations of genotype and basal-like vs. luminal A associations were conducted, assuming non-differential misclassification of case status. Sensitivity and specificity ranges were based on previously published data (27, 119, 120). Sensitivity analyses were conducted using a publicly available program (115). All analyses were run for 5000 simulations.

2.10 Data interpretation

The results from this analysis were interpreted based on effect size, precision, and any trends or patterns in the data. The precision of the effect estimates were measured by calculating the confidence interval ratio (CLR), which is equal to the upper 95% confidence limit divided by lower 95% confidence limit. A single CLR has relatively little meaning, but it can be useful for comparing several effect estimates to each other. A lower CLR indicates a more precise estimate. Null hypothesis testing was not used to draw conclusions about SNP or haplotype main effects.

Only associations for basal-like breast cancer, luminal A breast cancer, and breast

cancer overall (all cases) were reported here. The basal-like subtype is of interest because it is largely uncharacterized, and is a unique type of hormone-receptor negative breast cancer. The luminal A subtype is of interest because luminal A is the most common subtype and therefore a logical point of reference. Also, the candidate genes under study and potential effect measure modifiers were selected based on risk factors for these two subtypes. The luminal A and basal-like subtypes were also the two subtypes with the largest sample size. Even though parameters were estimated for luminal B, HER2+/ER-, and unclassified subtypes in the polytomous model, the associations were not reported due to limited sample size and imprecise OR estimates.

2.11 References

1. Familial breast cancer: collaborative reanalysis of individual data from 52 epidemiological studies including 58,209 women with breast cancer and 101,986 women without the disease. Lancet 2001;358(9291):1389-99.

2. Balmain A, Gray J, Ponder B. The genetics and genomics of cancer. Nat Genet 2003;33 Suppl:238-44.

3. Walsh T, King MC. Ten genes for inherited breast cancer. Cancer Cell 2007;11(2):103-5.

4. Malone KE, Daling JR, Doody DR, Hsu L, Bernstein L, Coates RJ, et al. Prevalence and Predictors of BRCA1 and BRCA2 Mutations in a Population-Based Study of Breast Cancer in White and Black American Women Ages 35 to 64 Years. Cancer Res 2006;66(16):8297-308.

5. Pharoah PD, Antoniou A, Bobrow M, Zimmern RL, Easton DF, Ponder BA. Polygenic susceptibility to breast cancer and implications for prevention. Nat Genet 2002;31(1):33-6.

6. Perou CM, Sorlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, et al. Molecular portraits of human breast tumours. Nature 2000;406(6797):747-52.

7. Sorlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proc Natl Acad Sci U S A 2001;98(19):10869-74.

8. Sorlie T, Tibshirani R, Parker J, Hastie T, Marron JS, Nobel A, et al. Repeated observation of breast tumor subtypes in independent gene expression data sets. Proc Natl Acad Sci U S A 2003;100(14):8418-23.

9. Millikan RC, Newman B, Tse CK, Moorman PG, Conway K, Dressler LG, et al. Epidemiology of basal-like breast cancer. Breast Cancer Res Treat 2008;109(1):123-39.

10. Yang XR, Sherman ME, Rimm DL, Lissowska J, Brinton LA, Peplonska B, et al. Differences in risk factors for breast cancer molecular subtypes in a population-based study. Cancer Epidemiol Biomarkers Prev 2007;16(3):439-43.

11. Nordgard SH, Johansen FE, Alnaes GI, Naume B, Borresen-Dale AL, Kristensen VN. Genes harbouring susceptibility SNPs are differentially expressed in the breast cancer subtypes. Breast Cancer Res 2007;9(6):113.

12. Kristensen VN, Borresen-Dale AL. SNPs associated with molecular subtypes of breast cancer: on the usefulness of stratified Genome-wide Association Studies

(GWAS) in the identification of novel susceptibility loci. Mol Oncol 2008;2(1):12-5.

13. Carey LA, Perou CM, Livasy CA, Dressler LG, Cowan D, Conway K, et al. Race, breast cancer subtypes, and survival in the Carolina Breast Cancer Study. Jama 2006;295(21):2492-502.

14. Hankinson S, Hunter D. Breast Cancer. In: Adami H-O, Hunter D, Trichopoulos D, editors. Textbook of Cancer Epidemiology. New York: Oxford University Press, Inc.; 2002. p. 301-339.

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

16. Pike MC, Spicer DV, Dahmoush L, Press MF. Estrogens, progestogens, normal breast cell proliferation, and breast cancer risk. Epidemiol Rev 1993;15(1):17-35.

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

18. Ceschi M, Gutzwiller F, Moch H, Eichholzer M, Probst-Hensch NM. Epidemiology and pathophysiology of obesity as cause of cancer. Swiss Med Wkly 2007;137(3-4):50-6.

19. Schaffler A, Scholmerich J, Buechler C. Mechanisms of disease: adipokines and breast cancer - endocrine and paracrine mechanisms that connect adiposity and breast cancer. Nat Clin Pract Endocrinol Metab 2007;3(4):345-54.

20. Bauer KR, Brown M, Cress RD, Parise CA, Caggiano V. Descriptive analysis of estrogen receptor (ER)-negative, progesterone receptor (PR)-negative, and HER2-negative invasive breast cancer, the so-called triple-negative phenotype: a population-based study from the California cancer Registry. Cancer 2007;109(9):1721-8.

21. Morris GJ, Naidu S, Topham AK, Guiles F, Xu Y, McCue P, et al. Differences in breast carcinoma characteristics in newly diagnosed African-American and Caucasian patients: a single-institution compilation compared with the National Cancer Institute's Surveillance, Epidemiology, and End Results database. Cancer 2007;110(4):876-84.

22. Newman B, Moorman PG, Millikan R, Qaqish BF, Geradts J, Aldrich TE, et al. The Carolina Breast Cancer Study: integrating population-based epidemiology and molecular biology. Breast Cancer Res Treat 1995;35(1):51-60.

23. Millikan R, Eaton A, Worley K, Biscocho L, Hodgson E, Huang WY, et al. HER2 codon 655 polymorphism and risk of breast cancer in African Americans and whites. Breast Cancer Res Treat 2003;79(3):355-64.

24. Weinberg CR, Sandler DP. Randomized recruitment in case-control studies. Am J Epidemiol 1991;134(4):421-32.

25. Weinberg CR, Wacholder S. The design and analysis of case-control studies with biased sampling. Biometrics 1990;46(4):963-75.

26. Huang WY, Newman B, Millikan RC, Schell MJ, Hulka BS, Moorman PG. Hormone-related factors and risk of breast cancer in relation to estrogen receptor and progesterone receptor status. Am J Epidemiol 2000;151(7):703-14.

27. Nielsen TO, Hsu FD, Jensen K, Cheang M, Karaca G, Hu Z, et al. Immunohistochemical and clinical characterization of the basal-like subtype of invasive breast carcinoma. Clin Cancer Res 2004;10(16):5367-74.

28. Livasy CA, Perou CM, Karaca G, Cowan DW, Maia D, Jackson S, et al. Identification of a basal-like subtype of breast ductal carcinoma in situ. Hum Pathol 2007;38(2):197-204.

29. Thomas DC. Statistical Methods in Genetic Epidemiology. New York: Oxford University Press; 2004.

30. PubMed. [Database] [cited 2007; Available from: www.pubmed.gov

31. Johnson GC, Esposito L, Barratt BJ, Smith AN, Heward J, Di Genova G, et al. Haplotype tagging for the identification of common disease genes. Nat Genet 2001;29(2):233-7.

32. Frazer KA, Ballinger DG, Cox DR, Hinds DA, Stuve LL, Gibbs RA, et al. A second generation human haplotype map of over 3.1 million SNPs. Nature 2007;449(7164):851-61.

33. Parra EJ, Marcini A, Akey J, Martinson J, Batzer MA, Cooper R, et al. Estimating African American admixture proportions by use of population-specific alleles. Am J Hum Genet 1998;63(6):1839-51.

34. Xu S, Huang W, Wang H, He Y, Wang Y, Wang Y, et al. Dissecting linkage disequilibrium in African-American genomes: roles of markers and individuals. Mol Biol Evol 2007;24(9):2049-58.

35. The International HapMap Project. Nature 2003;426(6968):789-96.

36. Sherry ST, Ward MH, Kholodov M, Baker J, Phan L, Smigielski EM, et al. dbSNP: the NCBI database of genetic variation. Nucleic Acids Res 2001;29(1):308-11.

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

38. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics 2005;21(2):263-5.

39. Carlson CS, Eberle MA, Rieder MJ, Yi Q, Kruglyak L, Nickerson DA. Selecting a maximally informative set of single-nucleotide polymorphisms for association analyses using linkage disequilibrium. Am J Hum Genet 2004;74(1):106-20.

40. Devlin B, Risch N. A comparison of linkage disequilibrium measures for fine-scale mapping. Genomics 1995;29(2):311-22.

41. Gu S, Pakstis AJ, Li H, Speed WC, Kidd JR, Kidd KK. Significant variation in haplotype block structure but conservation in tagSNP patterns among global populations. Eur J Hum Genet 2007;15(3):302-12.

42. de Bakker PI, Burtt NP, Graham RR, Guiducci C, Yelensky R, Drake JA, et al. Transferability of tag SNPs in genetic association studies in multiple populations. Nat Genet 2006;38(11):1298-303.

43. Haiman CA, Dossus L, Setiawan VW, Stram DO, Dunning AM, Thomas G, et al. Genetic variation at the CYP19A1 locus predicts circulating estrogen levels but not breast cancer risk in postmenopausal women. Cancer Res 2007;67(5):1893-7.

44. Haiman CA, Stram DO, Pike MC, Kolonel LN, Burtt NP, Altshuler D, et al. A comprehensive haplotype analysis of CYP19 and breast cancer risk: the Multiethnic Cohort. Hum Mol Genet 2003;12(20):2679-92.

45. Tian C, Hinds DA, Shigeta R, Kittles R, Ballinger DG, Seldin MF. A genomewide single-nucleotide-polymorphism panel with high ancestry information for African American admixture mapping. Am J Hum Genet 2006;79(4):640-9.

46. Barnholtz-Sloan JS, McEvoy B, Shriver MD, Rebbeck TR. Ancestry estimation and correction for population stratification in molecular epidemiologic association studies. Cancer Epidemiol Biomarkers Prev 2008;17(3):471-7.

47. Illumina, Inc. [cited; Available from: http://www.illumina.com/

48. Shen R, Fan JB, Campbell D, Chang W, Chen J, Doucet D, et al. High-throughput SNP genotyping on universal bead arrays. Mutat Res 2005;573(1-2):70-82.

49. Balding DJ. A tutorial on statistical methods for population association studies. Nat Rev Genet 2006;7(10):781-91.

50. Hosking L, Lumsden S, Lewis K, Yeo A, McCarthy L, Bansal A, et al. Detection of genotyping errors by Hardy-Weinberg equilibrium testing. Eur J Hum Genet 2004;12(5):395-9.

51. Wigginton JE, Cutler DJ, Abecasis GR. A note on exact tests of Hardy-Weinberg equilibrium. Am J Hum Genet 2005;76(5):887-93.

52. Guo SW, Thompson EA. Performing the exact test of Hardy-Weinberg proportion for multiple alleles. Biometrics 1992;48(2):361-72.

53. Hoggart CJ, Parra EJ, Shriver MD, Bonilla C, Kittles RA, Clayton DG, et al. Control of confounding of genetic associations in stratified populations. Am J Hum Genet 2003;72(6):1492-1504.

54. Pritchard JK, Donnelly P. Case-control studies of association in structured or admixed populations. Theor Popul Biol 2001;60(3):227-37.

55. Aldrich MC, Selvin S, Hansen HM, Barcellos LF, Wrensch MR, Sison JD, et al. Comparison of statistical methods for estimating genetic admixture in a lung cancer study of African Americans and Latinos. Am J Epidemiol 2008;168(9):1035-46.

56. Barnholtz-Sloan JS, Chakraborty R, Sellers TA, Schwartz AG. Examining population stratification via individual ancestry estimates versus self-reported race. Cancer Epidemiol Biomarkers Prev 2005;14(6):1545-51.

57. Chakraborty R, Kamboh MI, Nwankwo M, Ferrell RE. Caucasian genes in American blacks: new data. Am J Hum Genet 1992;50(1):145-55.

58. Falush D, Stephens M, Pritchard JK. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. Genetics 2003;164(4):1567-87.

59. Shriver MD, Parra EJ, Dios S, Bonilla C, Norton H, Jovel C, et al. Skin pigmentation, biogeographical ancestry and admixture mapping. Hum Genet 2003;112(4):387-99.

60. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet 2006;38(8):904-9.

61. Pritchard JK, Rosenberg NA. Use of unlinked genetic markers to detect population stratification in association studies. Am J Hum Genet 1999;65(1):220-8.

62. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. Genetics 2000;155(2):945-59.

63. Tsai HJ, Choudhry S, Naqvi M, Rodriguez-Cintron W, Burchard EG, Ziv E. Comparison of three methods to estimate genetic ancestry and control for stratification in genetic association studies among admixed populations. Hum Genet 2005;118(3-4):424-33.

64. Devlin B, Roeder K. Genomic control for association studies. Biometrics 1999;55(4):997-1004.

65. Ziv E, Burchard EG. Human population structure and genetic association studies. Pharmacogenomics 2003;4(4):431-41.

66. Marchini J, Cardon LR, Phillips MS, Donnelly P. The effects of human population structure on large genetic association studies. Nat Genet 2004;36(5):512-7.

67. Jackson J. A User's Guide to Principal Components. New York: John Wiley & Sons; 1991.

68. Foulkes AS. Applied Statistical Genetics with R For Population-based Association Studies. New York: Springer; 2009.

69. Pritchard JK, Wen X, Falush D. Documentation for structure software: Version 2.2. In. http://pritch.bsd.uchicago.edu/structure.html; 2007.

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

71. Hosmer DW, Lemeshow S. Special Topics: Polytomous Logistic Regression. In: Applied Logistic Regression. New York: John Wiley & Sons; 1989.

72. Schaid DJ. Evaluating associations of haplotypes with traits. Genet Epidemiol 2004;27(4):348-64.

73. Akey J, Jin L, Xiong M. Haplotypes vs single marker linkage disequilibrium tests: what do we gain? Eur J Hum Genet 2001;9(4):291-300.

74. Morris RW, Kaplan NL. On the advantage of haplotype analysis in the presence of multiple disease susceptibility alleles. Genet Epidemiol 2002;23(3):221-33.

75. Niu T. Algorithms for inferring haplotypes. Genet Epidemiol 2004;27(4):334-47.

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

77. Lin DY, Zeng D. Likelihood-Based Inference on Haplotype Effects in Genetic Association Studies. Journal of the American Statistical Association 2006;101(473):89-118.

78. Greenland S, Lash TL, Rothman KJ. Concepts of Interaction. In: Rothman KJ, Greenland S, Lash TL, editors. Modern Epidemiology. Third ed. Philadelphia:

Lippincott Williams & Wilkins; 2008. p. 71-86.

79. Zou GY. On the estimation of additive interaction by use of the four-by-two table and beyond. Am J Epidemiol 2008;168(2):212-24.

80. Skrondal A. Interaction as departure from additivity in case-control studies: a cautionary note. Am J Epidemiol 2003;158(3):251-8.

81. Richardson DB, Kaufman JS. Estimation of the relative excess risk due to interaction and associated confidence bounds. Am J Epidemiol 2009;169(6):756-60.

82. Connolly BS, Barnett C, Vogt KN, Li T, Stone J, Boyd NF. A meta-analysis of published literature on waist-to-hip ratio and risk of breast cancer. Nutr Cancer 2002;44(2):127-38.

83. Harvie M, Hooper L, Howell AH. Central obesity and breast cancer risk: a systematic review. Obes Rev 2003;4(3):157-73.

84. Arpino G, Laucirica R, Elledge RM. Premalignant and in situ breast disease: biology and clinical implications. Ann Intern Med 2005;143(6):446-57.

85. Jones JL. Overdiagnosis and overtreatment of breast cancer: progression of ductal carcinoma in situ: the pathological perspective. Breast Cancer Res 2006;8(2):204.

86. Lakhani SR. The transition from hyperplasia to invasive carcinoma of the breast. J Pathol 1999;187(3):272-8.

87. Leonard GD, Swain SM. Ductal carcinoma in situ, complexities and challenges. J Natl Cancer Inst 2004;96(12):906-20.

88. Wiechmann L, Kuerer HM. The molecular journey from ductal carcinoma in situ to invasive breast cancer. Cancer 2008;112(10):2130-42.

89. Erbas B, Provenzano E, Armes J, Gertig D. The natural history of ductal carcinoma in situ of the breast: a review. Breast Cancer Res Treat 2006;97(2):135-44.

90. Morrow M, O'Sullivan MJ. The dilemma of DCIS. Breast 2007;16 Suppl 2:S59-62.

91. Kerlikowske K, Barclay J, Grady D, Sickles EA, Ernster V. Comparison of risk factors for ductal carcinoma in situ and invasive breast cancer. J Natl Cancer Inst 1997;89(1):76-82.

92. Reinier KS, Vacek PM, Geller BM. Risk factors for breast carcinoma in situ versus invasive breast cancer in a prospective study of pre- and post-menopausal women. Breast Cancer Res Treat 2007;103(3):343-8.

93. Dabbs DJ, Chivukula M, Carter G, Bhargava R. Basal phenotype of ductal carcinoma in situ: recognition and immunohistologic profile. Mod Pathol 2006;19(11):1506-11.

94. Bryan BB, Schnitt SJ, Collins LC. Ductal carcinoma in situ with basal-like phenotype: a possible precursor to invasive basal-like breast cancer. Mod Pathol 2006;19(5):617-21.

95. Hannemann J, Velds A, Halfwerk JB, Kreike B, Peterse JL, van de Vijver MJ. Classification of ductal carcinoma in situ by gene expression profiling. Breast Cancer Res 2006;8(5):R61.

96. Rothman KJ, Greenland S. Precision and validity of studies. In: Rothman KJ, Greenland S, editors. Modern Epidemiology. 2nd ed. Philadelphia: Lippincott Williams & Wilkins; 1998.

97. Savitz DA. Interpreting Epidemiologic Evidence. New York: Oxford University Press; 2003.

98. Rothman KJ, Greenland S. Modern Epidemiology. In. 2nd ed. Philadelphia: Lippincott-Raven; 1998. p. 123-125.

99. Demark-Wahnefried W, Winer EP, Rimer BK. Why women gain weight with adjuvant chemotherapy for breast cancer. J Clin Oncol 1993;11(7):1418-29.

100. McInnes JA, Knobf MT. Weight gain and quality of life in women treated with adjuvant chemotherapy for early-stage breast cancer. Oncol Nurs Forum 2001;28(4):675-84.

101. Rock CL, Flatt SW, Newman V, Caan BJ, Haan MN, Stefanick ML, et al. Factors associated with weight gain in women after diagnosis of breast cancer. Women's Healthy Eating and Living Study Group. J Am Diet Assoc 1999;99(10):1212-21.

102. Harvie MN, Campbell IT, Baildam A, Howell A. Energy balance in early breast cancer patients receiving adjuvant chemotherapy. Breast Cancer Res Treat 2004;83(3):201-10.

103. Costa LJ, Varella PC, del Giglio A. Weight changes during chemotherapy for breast cancer. Sao Paulo Med J 2002;120(4):113-7.

104. Kumar N, Allen KA, Riccardi D, Bercu BB, Cantor A, Minton S, et al. Fatigue, weight gain, lethargy and amenorrhea in breast cancer patients on chemotherapy: is subclinical hypothyroidism the culprit? Breast Cancer Res Treat 2004;83(2):149-59.

105. Han HS, Lee KW, Kim JH, Kim SW, Kim IA, Oh DY, et al. Weight changes after adjuvant treatment in Korean women with early breast cancer. Breast Cancer Res Treat 2008.

106. Freedman RJ, Aziz N, Albanes D, Hartman T, Danforth D, Hill S, et al. Weight and body composition changes during and after adjuvant chemotherapy in women with breast cancer. J Clin Endocrinol Metab 2004;89(5):2248-53.

107. Ingram C, Brown JK. Patterns of weight and body composition change in premenopausal women with early stage breast cancer: has weight gain been overestimated? Cancer Nurs 2004;27(6):483-90.

108. Makari-Judson G, Judson CH, Mertens WC. Longitudinal patterns of weight gain after breast cancer diagnosis: observations beyond the first year. Breast J 2007;13(3):258-65.

109. Saquib N, Flatt SW, Natarajan L, Thomson CA, Bardwell WA, Caan B, et al. Weight gain and recovery of pre-cancer weight after breast cancer treatments: evidence from the women's healthy eating and living (WHEL) study. Breast Cancer Res Treat 2007;105(2):177-86.

110. Camoriano JK, Loprinzi CL, Ingle JN, Therneau TM, Krook JE, Veeder MH. Weight change in women treated with adjuvant therapy or observed following mastectomy for node-positive breast cancer. J Clin Oncol 1990;8(8):1327-34.

111. Campbell KL, Lane K, Martin AD, Gelmon KA, McKenzie DC. Resting energy expenditure and body mass changes in women during adjuvant chemotherapy for breast cancer. Cancer Nurs 2007;30(2):95-100.

112. Demark-Wahnefried W, Peterson BL, Winer EP, Marks L, Aziz N, Marcom PK, et al. Changes in weight, body composition, and factors influencing energy balance among premenopausal breast cancer patients receiving adjuvant chemotherapy. J Clin Oncol 2001;19(9):2381-9.

113. Del Rio G, Zironi S, Valeriani L, Menozzi R, Bondi M, Bertolini M, et al. Weight gain in women with breast cancer treated with adjuvant cyclophosphomide, methotrexate and 5-fluorouracil. Analysis of resting energy expenditure and body composition. Breast Cancer Res Treat 2002;73(3):267-73.

114. Goodwin PJ, Ennis M, Pritchard KI, McCready D, Koo J, Sidlofsky S, et al. Adjuvant treatment and onset of menopause predict weight gain after breast cancer diagnosis. J Clin Oncol 1999;17(1):120-9.

115. Fox MP, Lash TL, Greenland S. A method to automate probabilistic sensitivity analyses of misclassified binary variables. Int J Epidemiol 2005;34(6):1370-6.

116. Ali S. Female Breast Cancer Incidence, Stage at Diagnosis, Treatment, and Mortality in North Carolina. In: State Center for Health Statistics, North Carolina Department of Health and Human Services; 2006. p. 1-6.

117. Gusterson B. Do 'basal-like' breast cancers really exist? Nat Rev Cancer 2009;9(2):128-34.

118. Moinfar F. Is 'basal-like' carcinoma of the breast a distinct clinicopathological entity? A critical review with cautionary notes. Pathobiology 2008;75(2):119-31.

119. Calza S, Hall P, Auer G, Bjohle J, Klaar S, Kronenwett U, et al. Intrinsic molecular signature of breast cancer in a population-based cohort of 412 patients. Breast Cancer Res 2006;8(4):R34.

120. Livasy CA, Karaca G, Nanda R, Tretiakova MS, Olopade OI, Moore DT, et al. Phenotypic evaluation of the basal-like subtype of invasive breast carcinoma. Mod Pathol 2006;19(2):264-71.

2.12 Tables

Table 2.1 SNPs selected for genotyping[1]

	Number of tag SNPs			
	CEU	YRI	CEU + YRI	Shared by CEU and YRI
ADIPOQ	11	14	19	6
IL6	6	9	13	2
LEP	5	12	16	1
LEPR	23	67	82	8
TNFA	3	3	3	3
ESR1	38	83	105	16
HSD3B1	3	5	8	0
HSD17B2	7	43	46	4
PGR	9	23	32	0
SHBG	1	0	1	0

1 - aggressive tagging was used for LEPR, ESR1, HSD17B2, and PGR; pairwise tagging was used for all other genes

Table 2.2 Candidate gene SNPs with extreme Hardy Weinberg P-values

	Gene	SNP	Exact HWE P-value
White controls	ADIPOQ	rs9877202	< 0.0001
	HSD17B2	rs16956274	0.0046
	HSD17B2	rs8049423	0.0046
	HSD17B2	rs8045494	0.0046
	ESR1	rs6914211	0.0068
African-American controls	LEPR	rs9436740	< 0.0001
	HSD17B2	rs2955162	< 0.0001
	CYP19A1	rs2470144	0.0021
	LEPR	rs9436748	0.0041
	LEPR	rs11808888	0.0064

Table 2.3 Characteristics of CBCS participants with genotyping data

	Controls	All cases	Luminal A	Basal-like
N	1776	1972	679	200
Median age (range)	51 (21 – 74)	50 (23 – 74)	52 (23-74)	46 (25-74)
Self-identified race				
African American	658	742	233	108
Non-African American	1117	1229	446	92
Unknown	1	1		
Menopausal status				
Premenopausal	761	879	277	111
Postmenopausal	1015	1093	402	89
Stage				
CIS		838	151	18
1		615	224	43
2		635	237	108
3		146	41	19
4		43	8	6
Missing[1]		1471	18	6

1- invasive breast cancer cases

2.13 Figures

Figure 2.1 Carolina Breast Cancer Study geographic area

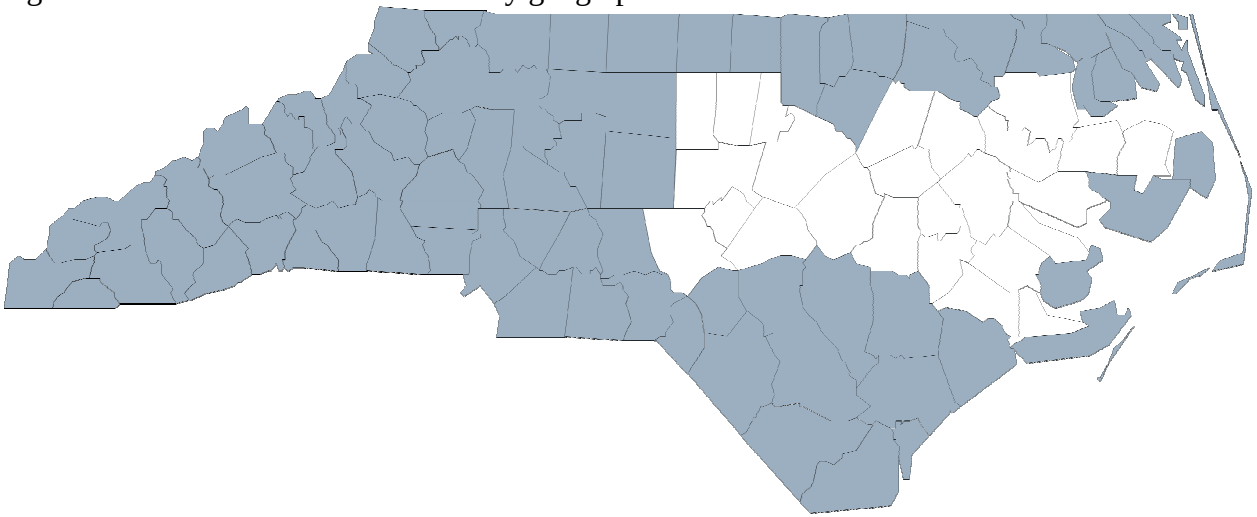


Figure 2.2 Definition of intrinsic molecular subtypes by immunohistochemical staining in the CBCS

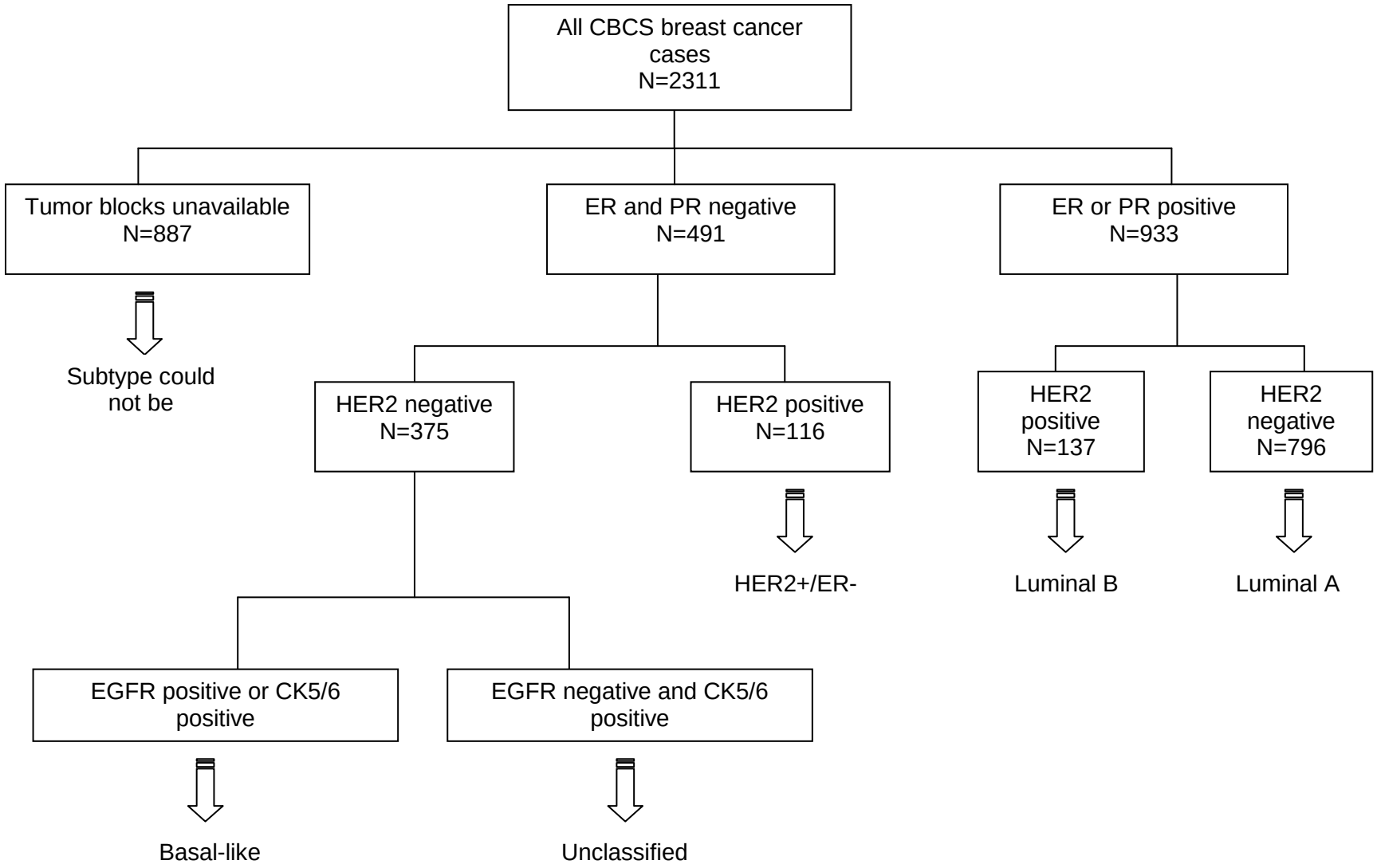


Figure 2.3 CBCS participants flow chart

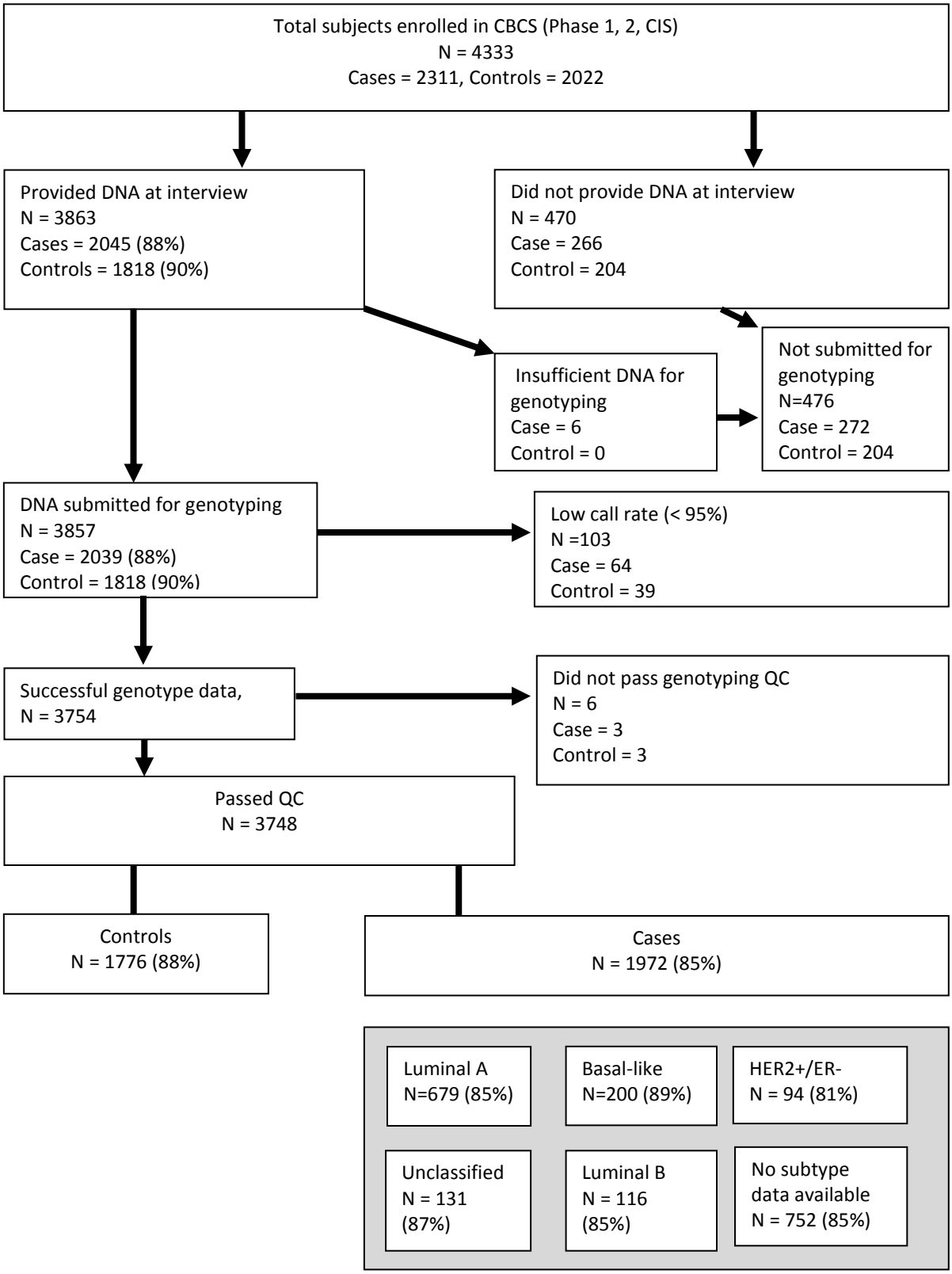


Figure 2.4 Maximum likelihood African ancestry stratified by self-identified race, with median individual ancestry estimates

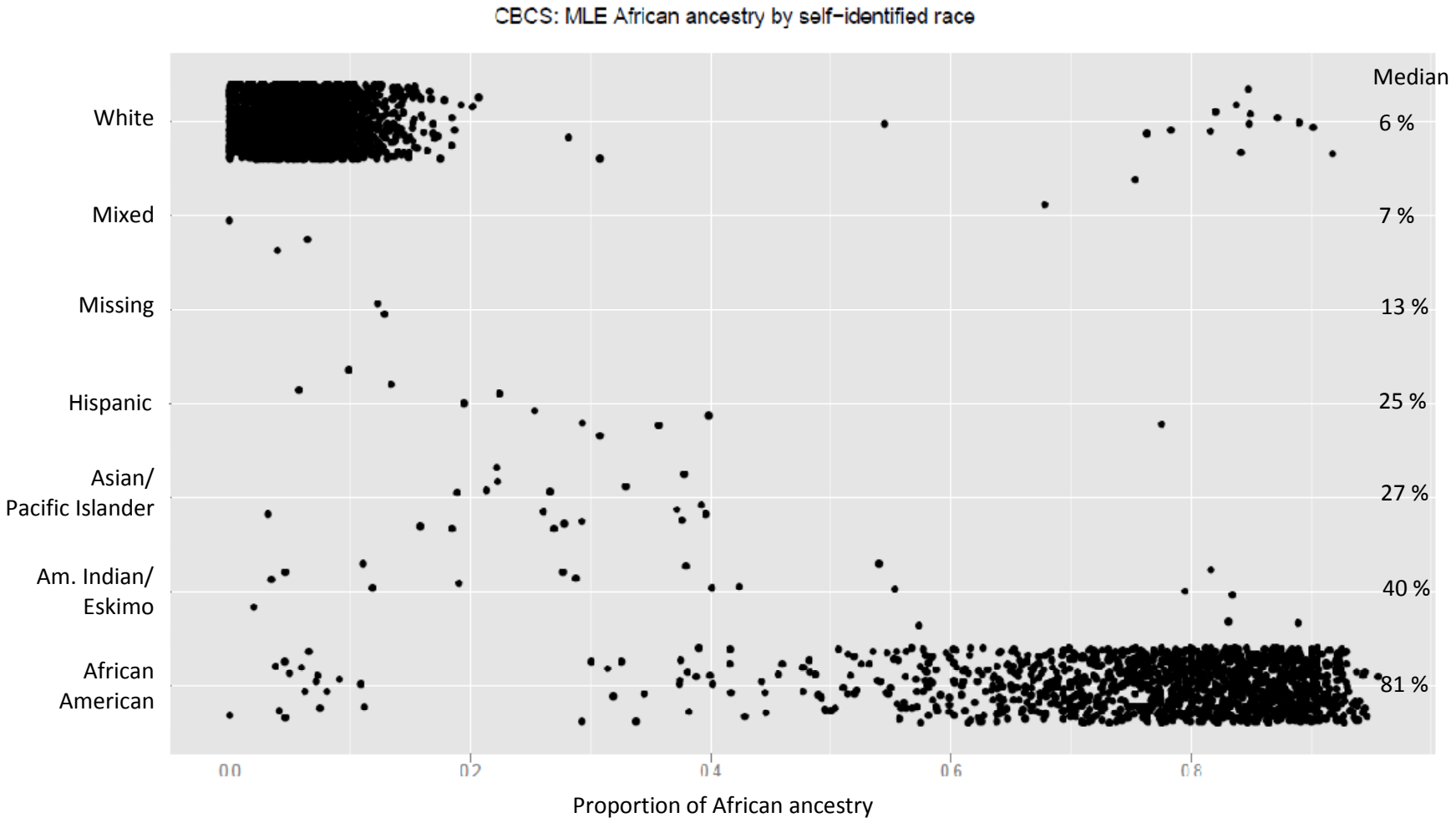


Figure 2.5 Structure African ancestry stratified by self-identified race, with median individual ancestry estimates

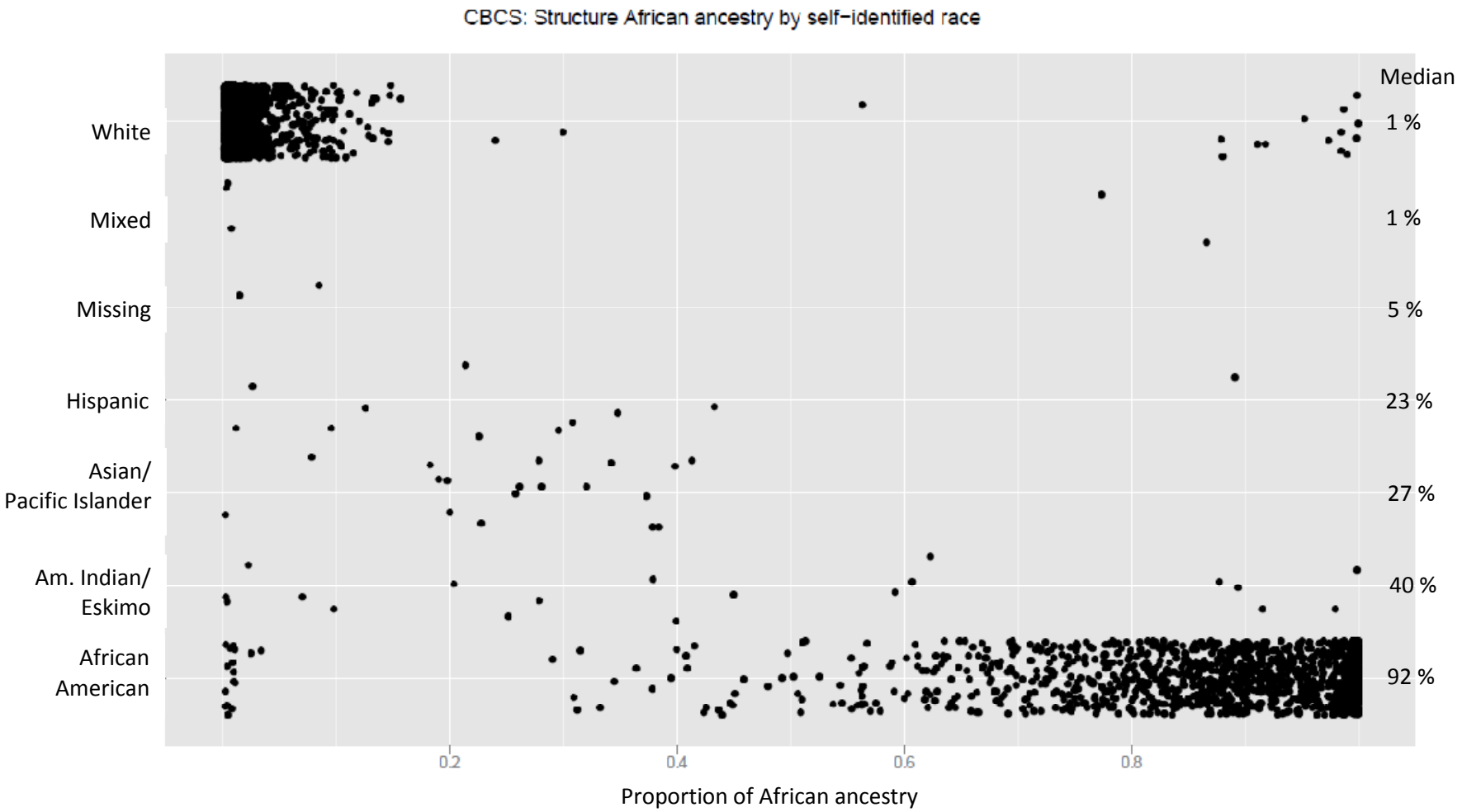
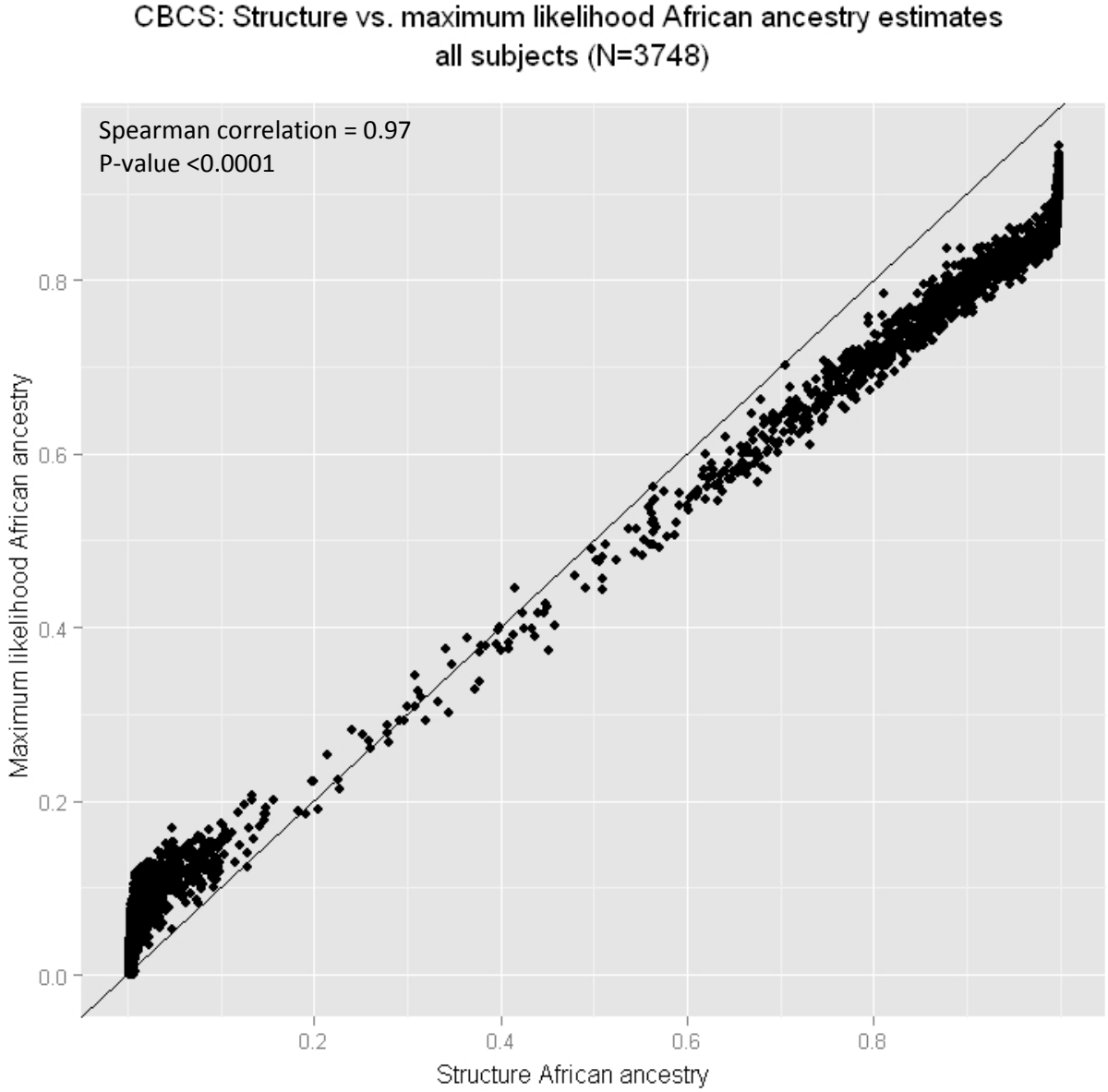


Figure 2.6 Correlation between maximum likelihood and Structure estimates of African ancestry



3. Results Paper 1: Association between genetic variation in adipocytokines and basal-like and luminal A breast cancer

3.1 Abstract

Introduction: ADIPOQ, IL6, LEP, LEPR, and TNF are associated with central obesity and influence tumorigenic activities in cell culture. We investigated whether single nucleotide polymorphisms (SNPs) in these genes are associated with breast cancer. We previously reported that the association between waist-hip ratio (WHR) and basal-like breast tumor subtype was stronger than the association between WHR and the luminal A subtype. Therefore, we analyzed SNP associations focusing on the basal-like and luminal A subtypes.

Materials and Methods: Eligible cases were women aged 20-74, diagnosed with primary breast cancer between 1993 and 2001, and living in North Carolina. ER. PR, HER2, CK5/6, and EGFR immunohistochemistry was used to determine the intrinsic molecular subtype of case tumors. Controls were cancer-free women living in the same geographic area, and were frequency matched to cases by age and race. 143 tag and functional SNPs were genotyped in cases and controls using the Illumina GoldenGate assay. Genotype data was available for 1776 of 2022 controls and 1972 of 2311 cases. There were 200 basal-like and 679 luminal A cases. Odds ratios and 95% confidence intervals estimating the associations between SNPs and all breast cancer cases, basal-like breast cancer, and luminal A breast cancer were estimated using logistic regression. Haplotype frequencies and odds ratios were estimated

using HAPSTAT. Additive interaction of genotype effects and waist-hip ratio (WHR) was estimated by calculating the synergy index and multiplicative interaction was evaluated using the likelihood ratio test.

Results: Genotypes in ADIPOQ, IL6, LEP, LEPR, and TNF had moderate to weak associations with breast cancer overall. Stronger associations were apparent when cases were stratified by subtype. Genotypes in IL6 were associated with luminal A but not basal-like breast cancer. Genotypes in ADIPOQ, LEP, and LEPR were associated with both subtypes. Functional variants were associated with luminal A breast cancer, but no functional variants were associated with the basal-like subtype. Haplotypes in IL6, LEP, and LEPR were associated with breast cancer overall and by subtype. There was evidence of interaction on additive and multiplicative scales between WHR and SNPs in ADIPOQ (luminal A cases), IL6 (basal-like cases), and LEPR (all cases, luminal A cases). There was evidence of interaction on the multiplicative scale between WHR and haplotypes in IL6 and LEPR, where the outcome was all breast cancer cases.

Conclusions: SNPs in ADIPOQ, LEP, IL6, and LEPR were associated with basal-like and/or luminal A breast cancer subtypes. Some of these associations were not apparent when all breast tumors were analyzed as a single outcome. These results are supportive of a role of adipocytokine SNPs in the etiology of different types of breast tumors, including those that do and do not express hormone receptors.

Keywords: adipocytokines, adiponectin, interleukin 6, leptin, leptin receptor, tumor necrosis factor-alpha, breast cancer, single nucleotide polymorphism, basal-like, luminal A

3.2 Introduction

Gene expression-based characterization of breast tumors has revived interest in the 'basal-like' subtype of breast cancer (1, 2, 3). The basal-like subtype is characterized by expression of cytokeratins 5 and 17, EGFR, smooth muscle actin, and vimentin, a lack of ER, PR, or HER2 expression, tumor characteristics associated with poor prognosis, and high prevalence among BRCA1-associated breast cancers (4-10). The basal-like subtype is also associated with poorer survival compared to the other breast cancer subtypes (6, 11, 12). Comparative studies of breast cancer risk factors suggest that associations between some traditional breast cancer risk factors and the basal-like subtype differ from the associations with other breast tumors (13, 14).This study investigated whether common genetic variants are also uniquely associated with the basal-like subtype. Identification of genetic risk factors specific to the basal-like subtype could help explain underlying biological mechanisms involved in basal-like breast cancer risk.

In the Carolina Breast Cancer Study (CBCS), we previously reported that the basal-like subtype is strongly associated with elevated waist-hip ratio (WHR) in addition to some other breast cancer risk factors (13). One potential mechanism that could explain the association between central obesity and breast cancer is the induction of pro-tumor pathways by adipocytokines (15, 16). Adipocytokines are hormones produced in visceral adipose tissue, and circulating levels are correlated with obesity (15, 16). Adipocytokines are also expressed in breast tissue, and have the ability to increase proliferation, interact with the estrogen receptor, and promote cell migration and invasion [reviewed in (15-17)]. In this study, we investigated single nucleotide polymorphisms (SNPs) in adiponectin (ADIPOQ), interleukin 6 (IL6), leptin (LEP), leptin receptor (LEPR), and tumor necrosis factor-alpha

(TNF) and their association with breast cancer risk.

Reported associations between adipocytokine functional SNPs, such as IL6 -174 G/C, LEPR Q223R, and TNF 863 C > A, and breast cancer have been inconsistent (18-27). It is possible that some of the variation in reported associations has to do with tumor heterogeneity, and that stratification by tumor type will allow for the identification of unique associations that cannot be detected reliably in a pooled tumor population. In the CBCS, elevated WHR was more strongly associated with the basal-like subtype compared to the more common luminal A subtype, suggesting that genetic risk factors associated with central obesity may also be more strongly associated with the basal-like subtype (13). To evaluate this, we estimated SNP associations for breast cancer subtypes in addition to estimating associations for all breast cancer cases.

Consideration of breast cancer subtypes is essential for identifying risk factors for and improving characterization of the basal-like subtype. Basal-like breast cancer is a poor prognosis breast cancer that does not express molecular targets of breast cancer treatment such as the HER2 receptor or estrogen receptor (2, 3, 11). Focusing on the basal-like subtype may provide some insight into the molecular mechanisms involved in tumor development and provide information for treatments. The luminal A subtype is included in this analysis as a point of reference. Luminal A breast cancer was the most common molecular subtype in the CBCS and has a relatively good prognosis (11).

3.3 Methods

3.3.1 Study population

The CBCS is a population-based case control study of breast cancer in North

Carolina. Details of the CBCS study design have been described in detail by Newman et al. (28) and Millikan et al.(29). Eligible cases included all women ages 20-74 who were diagnosed with primary invasive breast cancer from 1993 to 2001, and who lived within the 24-county study area at the time of diagnosis. Cases were identified through the North Carolina Central Cancer Registry using rapid case ascertainment. Randomized recruitment was used to oversample African American cases and cases younger than 50 years old (30). Women diagnosed with breast carcinoma *in situ* (CIS) were also enrolled in the study from 1996-2001. All eligible CIS cases were asked to participate in the study.

Eligible controls were defined as women ages 20 to 74 years, residing within the study area, and who did not have a history of breast cancer. Controls younger than age 65 were identified through Department of Motor Vehicles records, while controls 65-74 years old were identified through Health Care Financing Administration records. Controls were frequency-matched to cases by race and 5-year age groups.

Cases and controls were contacted by mail followed by a telephone call. Women who agreed to participate in the study provided informed consent and completed an in-home interview conducted by a trained nurse. During the interview, women were asked about known and suspected breast cancer risk factors, including social and demographic characteristics, family history of cancer, reproductive history, menstrual history, exogenous hormone use, alcohol use, and occupational history. Height, weight, waist circumference and hip circumference were measured by the nurse. Women were also asked to provide a 30 ml blood sample. DNA was extracted from the blood sample and stored at -80^{o}C in TE buffer.

Overall response rates for invasive cases and controls were 76% and 55%, respectively. Among cases, rates were highest among non-African Americans younger than

50 years old (81%) and lowest among African Americans age 50 and older (70%). Among controls, rates were highest among non-African Americans age 50 and older (65%) and lowest among African Americans younger than 50 years (47%). Overall response rates for CIS cases and matched controls were 83% and 65%, respectively. Among CIS cases, the highest response rate was from non-African Americans younger than 50 (86%) and the lowest response was from African Americans age 50 and older (76%). Among CIS controls the highest response was from non-African Americans age 50 and older (69%), the lowest response was from African Americans age 50 and older (51%). A total of 2311 cases (894 African American/1417 non-African American) and 2022 controls (788 African American/1234 non-African American) were enrolled in the study.

Among cases, tumor subtype was determined by immunohistochemistry (IHC) analysis of archival tumor tissue. IHC procedures for invasive breast cancers have been described by Carey et al. (11). Cases were asked to provide written consent for access to their medical records and formalin-fixed paraffin-embedded tumor tissue blocks. A centralized pathology review was performed to confirm each breast cancer diagnosis. For invasive breast cancers, estrogen receptor (ER) and progesterone receptor (PR) status was abstracted from the patient's medical record (80% of invasive cases). If ER or PR status was not recorded in the medical record but archival tissue was available, the assay was performed at the UNC Immunohistochemistry Core Laboratory (20% of invasive cases). All IHC staining for epidermal growth factor receptor (EGFR), v-erb-b2 erythroblastic leukemia viral oncogene homolog 2 (HER2), and cytokeratin 5/6 (CK 5/6) was performed at the UNC Immunohistochemistry Core Laboratory. Scoring for ER, PR, HER2, EGFR, and CK 5/6 has been described previously (13, 31, 32).

All immunohistochemistry for CIS cases was performed by the UNC Immunohistochemistry Core Laboratory, and was described by Livasy et al. (33). ER-positivity was defined as an Allred score of 2 with nuclear staining. CIS cases were considered HER2-positive if they displayed greater than 10% membranous staining in CIS cells with an intensity equivalent to 3+ by DAB chromogen or 2+/3+ by SG chromogen. Finally, CIS cases were positive for EGFR if they showed any membranous staining and positive for CK 5/6 if they showed any cytoplasmic staining. PR expression was not used in classifying CIS cases due to the high correlation between ER and PR expression, and the need to preserve tissue (11).

Tumor tissue was available for 1845 of 2311 (80%) cases [1446 of 1808 (80%) invasive cases; 399 of 503 (80%) CIS cases]. IHC assays were completed successfully for 1424 of 2311 (62%) cases [1149 of 1808 (64%) invasive cases; 275 of 503 (55%) CIS cases]. 225 cases were classified as basal-like (ER -, PR -, HER2 -, CK5/6 + and/or EGFR +) and 796 cases were classified as luminal A (ER + and/or PR +, HER2 -). The remaining subtypes consisted of 137 luminal B (ER + and/or PR +, HER2 +), 116 HER2+/ER- (ER -, PR -, HER2 +), and 150 unclassified (ER -, PR - , HER2 -, CK 5/6 -, and EGFR -) tumors. Tumor subtype could not be determined for 887 cases (38%). Cases with missing subtype data were more likely to be non-African American and to have an earlier stage at diagnosis (13). As described above, this analysis focuses on basal-like and luminal A breast cancer, and results for HER2+/ER-, luminal B, and unclassified tumors are not shown.

3.3.2 Genotyping

3.3.2.1 SNP selection

A combination of tag SNPs and potentially functional SNPs was selected for

genotyping. Potentially functional SNPs were defined as SNPs with a minor allele frequency of 0.05 or greater that were reported to have a functional effect *in vitro*, or had been investigated previously for an association with breast cancer (ADIPOQ rs1501299, rs2241766, rs822396, IL6 rs1800795, rs1800796, rs2069832, LEPR rs1137100, rs1137101, rs8179183, and TNF rs1800630; no potentially functional SNPs were genotyped for LEP).

Tag SNPs for each gene were selected using genotype frequency information from the International HapMap Project (34). At the time of SNP selection only Phase 1 and 2 genotypes were available from HapMap, so tag SNPs intended to represent genetic variation from European ancestral populations was selected using CEU data and tag SNPs intended to represent genetic variation from African ancestral populations were selected using YRI data. Tag SNPs with a minor allele frequency of 0.10 or greater were selected using Tagger in Haploview (35, 36). Pairwise tagging with a minimum r^2 of 0.80 was used to select tags (37). CEU and YRI tag SNPs were combined into a single list and genotyped in all subjects. Pairwise tagging for LEPR required more than 90 tag SNPs, which stretched the limits of the space available on the genotyping chip. Tagging for LEPR was repeated using aggressive tagging with 2-marker haplotypes, which reduced the number of SNPs required by 15. To increase the efficiency of tag SNP selection, SNPs selected from the literature were also used as tag SNPs, using the "force include" option in Tagger.

3.3.2.2 Genotyping results and quality control

The SNPs in this study were genotyped as part of a larger panel of 1536 SNPs. In addition to candidate gene SNPs, 158 ancestry informative markers (AIMs) were genotyped in order to adjust for population stratification. Genotyping was performed by the UNC Mammalian Genotyping Core using the Illumina GoldenGate assay (Illumina, Inc., San

Diego, CA). Assay intensity data and genotype cluster images for all SNPs were reviewed individually. SNPs were excluded from the dataset due to low signal intensity or inability to clearly distinguish between genotype clusters. 163 of 1536 (11 %) SNPs were excluded from the entire dataset based on cluster analysis.

Additionally, blind duplicates of 169 study samples were assayed in order to verify the reproducibility of genotype calls from the same sample. 7 SNPs had 1 genotype miscall and 2 SNPs had 2 genotype miscalls. Lab controls (Coriell CEPH trios) were also genotyped in each 96-well plate – each control was repeated between 11 and 14 times over the course of the entire assay. Out of 184 lab control samples, there were 2 instances of genotype disagreement with duplicate samples. These error rates were within our pre-specified range of acceptable values, and no SNPs were excluded from the analysis on the basis of these results. 1373 of 1536 (89%) SNPs in the panel had data that was acceptable for analysis. Among SNPs in ADIPOQ, IL6, LEP, LEPR, and TNF, 117 of 143 (82%) SNPs provided DNA data of acceptable quality and are included in this analysis (ADIPOQ - 16, IL6 – 11, LEP - 14, LEPR - 74, TNF - 2). 144 of 158 (91%) AIMs passed quality control and were used to estimate ancestry.

Exact tests for deviation from Hardy Weinberg equilibrium (HWE) were conducted in controls stratified by self-identified race to determine whether genotype frequencies were distributed as expected given the allele frequencies. Deviations from HWE in controls can indicate several things, including genotyping error, selection bias, new mutations, or failure of the source population to fulfill HWE assumptions (38, 39). Deviations from HWE can also occur by chance. HWE test statistics and P-values were calculated in Plink v1.05 using methods described by Wigginton et al. (40). In order to confirm that HWE deviations were

not due to erroneous genotype calls, genotyping cluster images were re-reviewed for all SNPs with HWE test P-values less than 0.01. All SNPs reviewed during this process were judged to have acceptable signal intensity and genotype cluster definition, and none were excluded.

Of the 2311 cases and 2022 controls enrolled in the CBCS, 2045 (88%) cases and 1818 (90%) controls provided a blood sample at the time of interview. 2039 (88%) cases and 1818 (90%) controls had sufficient amounts of DNA available for genotyping. Of these, 64 cases and 39 controls had genotype calls for less than 95% of SNPs in the Illumina panel and were excluded from the analysis. An additional 2 cases and 3 controls were excluded due to an apparent gender mismatch. One case was excluded because of suspected contamination identified though the analysis of non-blind duplicate samples.

Ultimately, 1776 of 2022 (88%) controls and 1972 of 2311 (85%) cases were successfully genotyped. Subjects without genotype data were more likely to be cases, recruited during phase 2 of the study, and African American. The presence or absence of genotype data did not differ by any breast cancer risk factors other than African American race. Among cases the presence of genotype data was not associated with stage at diagnosis, lymph node status, or molecular subtype.

Among cases, 978 of 1808 (54%) invasive cases had both genotyping and tumor subtype data and 242 of 503 (48%) CIS cases had genotyping and subtype data, including 200 basal-like cases (182 invasive, 18 CIS) and 679 luminal A cases (528 invasive, 151 CIS). The distribution of intrinsic molecular subtypes did not differ between enrolled cases with and without genotyping data.

3.3.3 Variable definitions and statistical methods

3.3.3.1 Variables

Age was defined as age in years at breast cancer diagnosis for cases, and age in years at the time of sampling for study participation for controls. Self-identified race was reported during the study interview. Of the 3748 CBCS subjects with genotyping data 2293 identified themselves as white and 1400 identified themselves as African American. Less than 2% of CBCS participants reported that they were Native American/Eskimo (N=19), Asian or Pacific Islander (N=18), Hispanic (N=11), or mixed race (N=5). For regression analyses, Native American, Eskimo, Asian/Pacific Islander, Hispanic, and mixed race women and self-described white women were grouped together as non-African American. Self-identified race information was missing for 2 study subjects, who were excluded from analyses that adjusted for or stratified by self-identified race.

Individual estimates of African and European ancestry were estimated from 144 AIMs. Two methods were used to estimate ancestry in study subjects – maximum likelihood estimation and structured association. Under the maximum likelihood method, the proportion of African ancestry was estimated for each study subject by solving likelihood equations described by Barnholtz-Sloan et al. (41). The structured association estimates were generated using Structure v.2.0, which uses Bayesian estimation to determine the proportion of a subject's genome that belongs to each ancestral population cluster (K) (42-44). Preliminary estimates were calculated for K=1 through K=5 and the most likely number of populations was determined to be 2. Estimates were re-calculated for K=2 using the admixture and correlated allele frequencies models.

By maximum likelihood, the median proportion of African ancestry was 81% among subjects whose self-reported race was African American and 6% among those whose self-

reported race was white. Using Structure, African ancestry was estimated to be 92% among self-reported African Americans and 1% among whites. The correlation between STRUCTURE and maximum likelihood ancestry estimates was very high (Spearman correlation, all subjects $r = 0.97$, $P < 0.0001$; African Americans $r = 0.98645$, $P < 0.0001$, non-African Americans $r = 0.86767$, $P < 0.0001$). Data analysis continued using the maximum likelihood estimates only. The distribution of African ancestry estimated by maximum likelihood is shown in Figure 3.1.

Waist-hip ratio (WHR) is the ratio of waist circumference to hip circumference. Waist and hip circumference were measured using a tape measure by a trained nurse during the study visit and were recorded to the nearest half centimeter. The waist measurement was taken at the natural indentation of the waist. Hip circumference was measured at the greatest protrusion of the buttocks. Measurements were taken twice and averaged. If the first two measurements differed by more than 1 cm, a third measurement was taken and the two closest measurements were averaged. The WHR variable used in this analysis is based on the tertile distribution of WHR in CBCS controls. The associations between basal-like breast cancer and WHR and luminal A breast cancer and WHR were similar for tertile 2 and tertile 3 (vs. tertile 1) (13), and so those two categories were combined and WHR was categorized as < 0.77 and ≥ 0.77 for this analysis. Body mass index (BMI, weight in kg/height in m^2) was calculated from height and weight measured during the study visit and was included in regression models as a continuous variable. Weight was the average of two measurements taken using a standardized scale and recorded to the nearest half kilogram. Height was the average of two measurements made to the nearest half centimeter.

A total of 90 of 3748 (2%) genotyped subjects (40 controls, 50 cases, 6 basal-like

cases, 17 luminal A cases) were missing data for either WHR or BMI and were excluded from the effect measure modification analysis. Proportions of subjects missing data for either WHR or BMI did not differ by case status. The low proportion of missing data combined with the fact that missingness was unrelated to case status strongly indicate that missing WHR data was not a source of bias in this analysis.

3.3.3.2 Genotype associations

Genotype frequencies for each SNP were calculated stratified by self-identified race. Genotype proportions were adjusted for the sampling probabilities used to select eligible participants. Linkage disequilibrium measures were calculated using Haploview (36).

Odds ratios (ORs) and 95% confidence intervals (CIs) for the association between genotypes and breast cancer overall were estimated using unconditional binary logistic regression models. ORs and 95% CIs for basal-like and luminal A breast cancer were estimated using unconditional polytomous logistic regression models that simultaneously estimated parameters for all breast cancer subtypes. Statistics testing the equality of parameter estimates for basal-like and luminal A subtypes were calculated based on the asymptotic chi-square distribution of the Wald statistic. The polytomous regression model did not converge for rs9436748, so parameters were estimated using separate binary logistic regression models for basal-like cases and controls, and luminal A cases and controls. Parameter estimates from individual binary logistic regression models where the cases are restricted to a particular subtype should be similar to the parameter estimates from the polytomous model (45). All single SNP regression models were run using SAS v9.1.3 (SAS, Cary, NC).

Genotype effects were modeled using the general model form with 2 degrees of

freedom, unless the rare homozygote cell counts were too small. In that case, the model compared the rare homozygote and heterozygote genotypes to the common homozygote genotype. If the results using the general model indicated that the underlying genetic model form may be recessive, dominant, or additive, additional analyses specific to the likely genetic model were conducted. All models were adjusted for the frequency matching factors age at recruitment into the study (continuous) and self-identified race (African American, non-African American). An offset term was included in all models to account for randomized recruitment sampling (30). Confidence limit ratios (CLR, upper 95% confidence limit divided by lower 95% confidence limit) were calculated as a measure of relative precision.

Self-identified race was included in all models due to the study design. The inclusion of ancestry information was based on whether adjusting for ancestry improved model fit and confounding control in the presence of self-identified race. Comparing parameter estimates for models adjusted for and not adjusted for ancestry, parameter estimates changed by more than 0.10 for some SNPs when estimating associations for luminal A cases and controls. Based on this, African ancestry was included in all models as a continuous variable in order to provide additional control for residual confounding due to population stratification. SNPs with a relatively strong odds ratio (≥ 1.5 or ≤ 0.67) or a P-value less than 0.05, and a precise confidence interval (CLR ≤ 5) were considered to be the best candidates for association with breast cancer and are described in the results section.

3.3.3.3 Haplotype associations

Haplotype frequencies and ORs were estimated using a modified version of the HAPSTAT program (46, 47). HAPSTAT estimates the probability of a given haplotype using maximum likelihood estimation and the EM algorithm, and incorporates that probability into

the estimation of the haplotype effects. This yields unbiased parameter estimates with appropriate variance given that the true haplotype phase of CBCS subjects is unknown. Modifications to the original HAPSTAT program allowed for inclusion of the offset term and a relaxation of the assumption of independence between genotypes and environmental variables, allowing for adjustment of age, self-identified race, and AIM-estimated African ancestry in all haplotype analyses. SNPS were selected for haplotype analysis using a 3-SNP sliding window scanning method (HAPSCAN) and by reviewing single SNP ORs for a consecutive string of associated SNPs. All haplotype ORs were estimated using the general model, and were adjusted for age, self-identified race, and African ancestry.

3.3.3.4 Effect measure modification

Associations between WHR and breast cancer in the CBCS have been reported previously (13, 48, 49). Multiplicative genotype-WHR and haplotype-WHR interaction was evaluated using the likelihood ratio test (LRT). Systematic reviews of the literature have shown that the association between WHR and premenopausal breast cancer was biased towards the null in studies that did not adjust for BMI (50, 51). In the CBCS, not adjusting for BMI pushed the WHR OR towards the null (data not shown). Therefore, BMI was included in models evaluating WHR-genotype interaction. SNPs and haplotypes yielding LRT P-values less than 0.10 were considered to demonstrate evidence of interaction. ORs and 95% CIs were calculated stratified by WHR and genotype for these SNPs. Additive interaction for genotypes and WHR was explored by calculating the synergy index (S) and 90% confidence intervals (52). S estimates above 1 indicate greater than additive interaction, and S estimates below 1 indicate less than additive interaction. Additive interaction was not evaluated for haplotypes.

3.3.3.5 Interpretation

In the analysis of genotype and haplotype associations, P-values were used in conjunction with point estimates and confidence limit ratios to evaluate the combined strength and precision of estimated associations. Strict hypothesis testing was not performed, and so P-values were not adjusted for multiple comparisons. Decisions to display stratified ORs for the evaluation of multiplicative and additive interaction were based on P-values. The intent was to display stratified ORs for the SNPs and haplotypes that showed the strongest evidence of heterogeneity. Interaction P-values were not adjusted for multiple comparisons.

3.3.3.6 Sensitivity analysis

Probabilistic sensitivity analyses were performed to evaluate the potential effects of WHR misclassification. WHR was measured at the time of interview, which was a median of 3.9 months (range, 0.8 – 42.5 months) after diagnosis for cases and a median of 3.8 months (range, 0 – 45.2 months) after sampling for controls. Weight change is a commonly documented side effect of breast cancer-related therapy, and so it is possible that WHR may have also changed among cases that started treatment before the study interview (53-64).

Studies that have reported on waist and hip measures after breast cancer diagnosis have found that though waist size and hip size increased following breast cancer-related chemotherapy, WHR remained the same (56, 61, 65). In Goodwin et al. (65), some women may have already started chemotherapy at the time of baseline WHR measurement, biasing any association between chemotherapy and WHR change. None of these three studies reported on differences in waist and hip measures by race, so it is unknown whether chemotherapy-related waist and hip change affects African-American cases differently from non-African Americans.

Complete information is not available on when CBCS cases started treatment, so a range of sensitivity and specificity values were calculated by estimating the likely false negative (FN) and false positive (FP) rates. FN and FP are based on the expected number of CBCS women in each category of WHR who likely received chemotherapy, based on the prevalence of chemotherapy treatment by stage in North Carolina Central Cancer Registry data (66). Simulations estimated a bias-corrected OR for the association between WHR and basal-like breast cancer.

Probabilistic sensitivity analyses were also conducted to evaluate the potential effects of molecular subtype misclassification. Not all studies used the same set of markers to define 'basal-like' breast cancer, and in studies that have used markers similar to those used by CBCS, there has not been 100% agreement between tumors defined as basal-like using microarray expression profiles and immunohistochemistry definitions (5, 32, 67). Simulations of genotype and basal-like vs. luminal A associations were conducted, assuming non-differential misclassification of case status. Sensitivity and specificity ranges were based on previously published data (5, 32, 67). Sensitivity analyses were conducted using a publicly available program (68). All analyses were run for 5000 simulations.

A simple sensitivity analysis was also performed to evaluate the effect of including CIS cases in the analysis. There is evidence to support that CIS is an intermediate step in the progression of cells from hyperplastic to malignant [reviewed by (69, 70)]. However, some argue against grouping CIS with breast cancer because of the lack of predictors to reliably determine which CIS will progress to invasive disease (71).

CIS were included in this study for the following reasons. First, most risk factors for DCIS are similar to risk factors for invasive breast cancer [reviewed in (72)]. Studies that

compared risk factors for invasive breast cancer and DCIS in the same screening population reported that were similar for family history of breast cancer, previous breast surgery, postmenopausal hormone use, and hysterectomy (73, 74). Associations differed between invasive and DCIS for age, early age at menarche (among women 50 and older), and presence of a palpable mass (73). There were conflicting results between the two studies for age at first birth and BMI in older or postmenopausal women (73, 74).

Second, intrinsic molecular subtypes have been observed in pure DCIS and in DCIS observed alongside invasive breast tumors (33, 75-77). Like invasive tumors, most basal-like DCIS showed strong expression of CK 5/6, vimentin, EGFR, and Ki-67; expression of p63 smooth muscle actin was rare (33, 75). Others did not report that cases were consecutive case series or systematically sampled from a defined population (77).

It has not yet been reported whether intrinsic molecular subtypes are associated with recurrence or survival in CIS. However the similarity between risk factors and molecular features suggest that there are a common set of factors that lead to both types of lesions. For these reasons, we chose not to exclude subjects recruited for the CIS study from this analysis. In a simple sensitivity analysis, the exclusion of CIS cases and matched controls did not systematically change the estimated genotype ORs (data not shown).

3.4 Results

3.4.1 Genotype associations

Genotype frequencies adjusted for sampling probabilities are shown in Table 3.1. None of the SNPs were monomorphic in African Americans or non-African Americans. Several SNPs did have very low minor allele frequencies, most commonly among non-

African American subjects. Odds ratios for SNPs associated with breast cancer overall are shown in Table 3.2 (the subtype-specific associations for these SNPs are included in Table 3.3 for comparison to overall effects). SNPs in LEP, LEPR, and TNF were associated with breast cancer overall. Few ORs were strong; most ranged from approximately 0.7 to 1.4, with many effect estimates near the null. The strongest association was for LEPR rs1409802 [all cases, adjusted for age, self-identified race, African ancestry, and offset term, AA vs. AG+GG OR, 1.50; 95% CI, 1.10 - 2.06)].

SNPs associated with basal-like and luminal A breast cancer are shown in Table 3.3 (associations with breast cancer overall are included in Table 3.2 for comparison). SNPs in IL6 (rs1800796, rs2069824, rs2069827) were associated with luminal A but not basal-like breast cancer. SNPs in ADIPOQ were inversely associated with both luminal A and basal-like subtypes, while SNPs in LEP were positively associated with basal-like and luminal A breast cancer.

SNPs in LEPR were associated with both basal-like and luminal A breast cancer. Nonsynonymous SNPs K109R (rs1137100) and Q223R (rs1137101) were associated with the luminal A subtype but not the basal-like subtype (Table 3.3). A cluster of 3 SNPs in LEPR intron 2 (rs17412175, rs9436746, rs9436748) were inversely associated with basal-like breast cancer, with odds ratios ranging from 0.48 to 0.57 (Table 3.3, Figure 3.2). Two additional SNPs (rs17097182 and rs970467) were also inversely associated with basal-like breast cancer but the estimated ORs had very wide confidence intervals and therefore are not presented. SNPs in TNF were not strongly associated with basal-like or luminal A breast cancer. In sensitivity analyses of the difference between basal-like and luminal A subtype associations bias-corrected ORs were similar to observed ORs, suggesting that any potential

subtype misclassification likely had minimal effects on the results (data not shown).

Effect measure modification of SNP ORs by WHR was explored on the additive and multiplicative scales for SNPs with a marginal association with breast cancer. For breast cancer overall, the LRT suggested evidence of multiplicative interaction for LEPR rs1137100 and rs12042877 (Table 3.4a). For basal-like breast cancer, only IL6 rs2069824 showed evidence of genotype-WHR interaction (Table 3.4b). For luminal A breast cancer, ADIPOQ rs16861194 and LEPR rs12042877, rs6588147, rs6704167, and rs9436746 showed evidence of WHR-genotype interaction (Table 3.4c). For breast cancer overall and by subtype, the joint effect of the index genotype and high WHR was usually less than multiplicative. Effects were greater than multiplicative only for rs6704167 (TT) and rs9436746 and luminal A breast cancer. Confidence intervals for estimates of additive interaction were imprecise, but for most SNPs S statistics were less than 1, indicating that interaction was also less than additive (Tables 3.4a-3.4c). Exceptions were rs1137100 (AG) in breast cancer overall where interaction was greater than additive, and rs6704167 (TT) in luminal A breast cancer for which an S statistic could not be calculated. There was little difference between bias corrected ORs and ORs estimated from the observed data in sensitivity analyses of potential WHR misclassification due to the effects of breast cancer treatment (data not shown).

3.4.2 Haplotype associations

Haplotype effects were estimated for regions identified by HAPSCAN and regions where several consecutive SNP main effects were associated with basal-like or luminal A breast cancer, and are shown in Table 3.5. Overall, haplotype associations tended to be stronger than single SNP associations. The strongest haplotype association was that of haplotype 5 with luminal A breast cancer (OR = 3.21), whereas the strongest single SNP

association was an OR of 0.48 (LEPR rs9436748, Table 3.3). There were several additional strong haplotype associations with ORs of 2 or greater, including those in IL6 (haplotype 1a), LEP (haplotypes 3, 4a, 4b, 5) and LEPR (haplotype 8b), where two copies of the target haplotype were strongly associated with breast cancer overall and/or luminal A breast cancer (Table 3.5). Only LEP haplotype 4a had a similarly strong association with the basal-like subtype, but the estimate was very imprecise (Table 3.5). Haplotype associations for luminal A breast cancer were similar to associations for breast cancer overall, except for haplotype 4a, where the association for breast cancer overall was greater than for the luminal A, and haplotype 5 where the association for breast cancer overall was weaker than the association for the luminal A subtype. Haplotype associations for basal-like breast cancer were less similar to haplotype associations for breast cancer overall, and there were several instances where the estimated haplotype association for the basal-like subtype (haplotypes 5, 6a, 6b, 9b) was not also observed in breast cancer overall.

Based on the LRT, there was evidence of multiplicative haplotype-WHR interaction for haplotypes 1a and 8b in breast cancer overall. No haplotypes showed evidence of multiplicative interaction for the luminal A or basal-like subtypes. Associations for haplotypes 1a and 8b with breast cancer overall stratified by WHR are shown in Table 6. Although the ORs for 2 copies vs. 0 copies of the at risk haplotype were imprecise, the association for 2 copies was greater among women with lower WHR.

3.5 Discussion

SNPs in ADIPOQ, IL6, LEP, LEPR, and TNF were genotyped in order to determine whether polymorphisms in these genes were associated with breast cancer. ORs were

estimated for basal-like and luminal A breast cancer subtypes in addition to estimating effects for all breast cancer cases combined under the hypothesis that the breast cancer molecular subtypes have unique risk factors, and that SNPs may be associated specifically with the basal-like or luminal A subtype. Effects unique to one tumor subtype may not be apparent if all breast cancer cases are analyzed together.

SNPs associated with breast cancer overall had weak to moderate effects, including the potentially functional variant LEPR rs1137100 (K109R). Work by Cleary et al. (78) demonstrated that mice with two functionally silent copies of LEPR did not develop spontaneous mammary tumors while mice with one or more functional copies of LEPR were susceptible to mammary tumors. This suggests that a polymorphism that disables leptin receptor signaling would not increase breast cancer risk. An alternative mechanism may be that impaired leptin receptor signaling leads to an accumulation of leptin, and excess leptin leads to increased breast cancer risk. A breakdown in the leptin/leptin receptor negative feedback loop leading to an accumulation of leptin in the bloodstream has been described as a feature of obesity (79). There are several other pathways through which leptin signaling could proceed. Leptin has been shown to enhance aromatase activity and activate the estrogen receptor in an estrogen receptor-positive cell line (80, 81). Leptin has also been shown to stimulate signaling of the HER2 and IGF-1 receptors in some breast cancer cell lines (82, 83).

Reports vary on the biological effect of LEPR K109R in women. Woo et al. (21) and van Rossum et al. (84) reported that healthy women with codon 109 RR variants have higher serum leptin levels compared to women with 109 QR variants, though in van Rossum et al. the difference in leptin levels was limited to women who had gained weight after a mean of

6.7 years of follow-up. This is consistent with the hypothesis that leptin accumulation is involved in the biological mechanism, if K109R is causally associated with breast cancer. In contrast, when Wauters et al. (85) examined leptin levels in relation to K109R genotype in overweight and obese women, leptin levels were lower among postmenopausal women with at least one copy of the 109R variant. There was no difference in leptin by K109R genotype in premenopausal women (85).

For many SNPs, the genotype-breast cancer association was stronger when the cases were stratified by molecular subtype. SNPs in LEP and LEPR were associated with basal-like and/or luminal A breast cancer; SNPs in ADIPOQ and IL6 were also associated with basal-like and/or luminal A breast cancer, but were not associated with all cases. Most of the associated SNPs were tag SNPs with unknown function, so it is unclear what biological link exists between the presence of one or more polymorphic alleles and an increased risk of breast cancer. There are numerous genomic variants that were not measured in this study that could be responsible for the observed associations, including insertion-deletion polymorphisms, repeat polymorphisms and untyped SNPs. Therefore, future studies should focus on variants in LD with the SNPs identified in this study in addition to the associated tag SNPs themselves.

SNP effects were stronger when analyzed in combination as haplotypes, and in many cases the haplotype OR magnitude exceeded that of any single SNP within the haplotype. Generally, haplotype effect estimates were the most stable for all cases and luminal A cases compared to controls. Corresponding estimates for the basal-like subtype were less precise and are therefore somewhat inconclusive. Estimates for several haplotypes and basal-like breast cancer could not be calculated due to the small number of basal-like cases. An

exception to this was LEPR haplotype 6b, which was positively associated with basal-like breast cancer with a precise confidence interval. Haplotype analysis using the sliding window technique also enabled the identification of a breast cancer-associated region (LEP haplotype 3) where the individual SNPs showed no association. Thus, in this study employing haplotype analysis in addition to single SNP analysis enhanced our ability to identify breast cancer-associated SNPs.

There were some very general themes in the observed associations. For ADIPOQ and LEP, the associations were consistently in one direction for both luminal A and basal-like subtypes, which suggests that alterations to the sequence or structure of the genes have a similar effect on both breast cancer subtypes. For example, the strongest associations among LEP SNPs were consistently positive. LEP -2548 G/A was not genotyped in this study, but has been reported to be positively associated with breast cancer in Tunisian and American women (20, 25). In ADIPOQ, associated SNPs were inversely associated with both subtypes. A similar pattern was reported by Kaklamani et al. (86), although their results were not always presented with the most common genotype as the referent group. In the case of ADIPOQ, IL6, and LEP, no single SNP was associated with both subtypes. This is very consistent with the idea that molecular subtypes have unique sets of risk factors that are only evident in a stratified analysis. In contrast, several SNPs in LEPR were associated with both subtypes.

The 3 associated SNPs in IL6 were associated with luminal A but not basal-like breast cancer. This is consistent with several lines of evidence that IL6 is active in ER-positive breast cancer cells. The addition of IL6 to ER-positive cell lines induces tumorigenic effects such as cell rounding, reduced cell adhesion, decreased E-cadherin expression, and

increased cell migration (87-89). In cell cultures, IL6 is able to activate transcription of ER-alpha in ER-positive breast cells and stimulate estrogen synthesis by inducing aromatase activity (90, 91). IL6 polymorphisms -174 G/C, -597 G/A, and -373A_8T_{12} were associated with disease-free survival among patients with ER-positive but not ER-negative tumors (92). However, the presence of IL6 in breast tumors does not appear to be related to ER status. Several studies reported that IL6 was expressed more commonly in ER-negative breast tumors or that there was no association between IL6 expression and ER status (93-95).

Table 7 shows CBCS ORs for SNPs that have been studied previously alongside the associations reported by others. The only SNP whose effect was consistent with previous studies was IL6 -572 G/C (rs1800796), which was weakly associated with breast cancer overall in the CBCS. The association was stronger comparing luminal A cases to controls, and was similar to results reported by Slattery et al. (19) for premenopausal women and postmenopausal women not on hormone therapy (Table 7). Other previously studied SNPs were associated with luminal A but not basal-like breast cancer. LEPR Q223R (rs1137101) was associated with luminal A breast cancer, but the association was not as strong as that observed by Snoussi et al. (20)(Table 7). Some studies reported that the LEPR codon 223R variant was associated with higher serum leptin levels in postmenopausal women (96), and with breast cancer risk (20, 24). The LEPR Q223R polymorphism was not associated with breast cancer in two other studies (21, 25). LEPR variant K109R (rs1137100) was also associated with the luminal A subtype and not the basal-like. Woo et al. (21) reported that LEPR codon 109 RR homozygotes have higher serum leptin levels compared to codon 109 KR heterozygotes in healthy Korean controls, but the K109R polymorphism was not associated with breast cancer.

In many cases, ORs estimated in the CBCS were not in agreement with previous studies (Table 7). Kaklamani et al. (86) reported associations for +45 T/G (rs2241766) and +276 G/T (rs1501299) but no association was seen for either SNP in the CBCS. IL6 -174 G/C (rs1800795) and intron 2 G/A (rs2069832) were not associated with all cases, luminal A breast cancer, or basal-like breast cancer although other investigators reported associations (18, 19). Terry et al. (97) tested not only the -174 G/C polymorphism, but other promoter polymorphisms including -572 G/C, -597 G/A, and a -373 A_8T_{12} repeat polymorphism in both HeLa and ECV40 cells, and showed that changes in IL6 expression level are likely due to a complex haplotype effect, not the single genotype at position -174. In this study the -572/-174 C-G/C-G diplotype is positively associated with all breast cancer cases and one copy of the C-G haplotype is associated with luminal A breast cancer (Table 5). The full haplotype described by Terry et al. was not analyzed because polymorphisms at nt -597 and -373 were not genotyped in the CBCS. Other studies also reported no association between -174 G/C and breast cancer (26, 27). In TNF -863 C/A (rs1800630), the AA vs. CC genotype was inversely associated with all breast cancer cases in the CBCS, whereas Gaudet et al. (22) reported no association. *In vitro*, the A allele reduced TNF transcription levels and serum TNF levels (98, 99).

None of the previously studied SNPs in Table 7 were associated with the basal-like subtype. This may be due to greater imprecision of basal-like estimates because of the small number of basal-like cases. Another explanation could be that SNPs of interest from previous studies have been defined based on the etiology of “all cases”, the majority of which are luminal A breast cancers. It is possible that different functional variants are associated with basal-like tumors. Even though the effects of potentially “functional” SNPs in the CBCS did

not agree with the results of other studies, the minor allele frequencies of these SNPs in CBCS controls are comparable to those found in other white and African American populations (Table 8).

Previous analyses in the CBCS found that higher WHR was a strong risk factor for basal-like breast cancer (13), so effect measure modification of genetic effects by WHR was evaluated on the multiplicative and additive scales. Interaction was less than multiplicative and less than additive for most joint SNP-WHR effects. These results should be interpreted with caution. In many cases the confidence intervals of the effect estimates were similar across strata even though the odds ratios differed. Joint effects could not be calculated for haplotypes, but stratified analysis of haplotypes with low LRT P-values showed that haplotype associations with breast cancer overall were greater in magnitude among women with lower WHR compared to women with higher WHR. Cleveland et al. (25) reported that there was no evidence of interaction between BMI and LEPR Q223R. We also found no evidence of interaction between this SNP and WHR in the CBCS.

Data from Slattery et al. (100) showed evidence of multiplicative interaction between IL6 rs1800795 and rs1800796 and WHR, but genotype-WHR interaction was not observed for these SNPs in the CBCS. Differences in WHR categorization limit direct comparison of results. The IL6 SNP that showed evidence of WHR interaction in the CBCS was not in LD with rs1800795 or rs1800796 in CBCS white or African-American cases or controls. Also, joint genotype-WHR effects reported by Slattery et al. (100) women showed patterns of being both less than additive and greater than additive, which is not what was observed in the CBCS. In the CBCS rs1800796 was part of a haplotype (1a) that showed evidence of interaction with WHR; 2 copies of the haplotype with the C allele was strongly associated

with breast cancer among women with low WHR. The data presented by Slattery et al. are not consistent with this trend.

The results presented here offer new insight into the etiology of breast cancer. Many of these tag SNPs have not been analyzed previously for their association with breast cancer. Additionally, this is one of few reports where SNP associations have been presented stratified by molecular subtype. Although estimates for the basal-like subtype were less precise compared to the luminal A estimates due to smaller sample size, the results presented were chosen based on both the strength of the OR and precision in order to avoid bias in reporting results by subtype. P-values were not adjusted for multiple comparisons because hypothesis testing was not used as strict criteria for evaluating ORs. Instead, the P-value was used as a proxy measure for the OR magnitude and standard error. Nevertheless, random error and/or systematic bias may have influenced the results, and thus approach used here and the results need to be replicated by others. Probabilistic sensitivity analyses were conducted to determine the effect that WHR misclassification or molecular subtype classification may have had on the results. Results of these analyses show that the ORs estimated from the data are close in magnitude to the range of corrected estimates, given the estimated sensitivity and specificity ranges.

The results of this study are promising, but we must keep in mind the potential effects that unequal study participation may have had on the study population. Study response rates were higher for cases compared to controls. Response also differed by self-identified race and age group – among invasive cases, CIS cases, and CIS study controls the lowest response rates were for African Americans age 50 and older. Adjustment for self-identified race and age should control for possible selection bias in an analysis of all cases compared to controls,

assuming that genotype distributions within case, race, and age strata are a representative sample of genotypes in those strata of the source population (101).

Selection bias may also have occurred during genotyping. This analysis includes genotyping data for 86% of enrolled CBCS subjects, which is 57% of all women invited to participate in the study. Enrolled subjects without genotyping data were more likely to be cases, from the Phase 2 invasive study, and African American compared to enrolled subjects with genotyping data. Enrolled subjects were missing genotype data because they were unwilling or unable to provide a blood sample at the time of interview, a mechanism of missingness that is likely a combination of self-selection and severity of illness (among cases). As with overall study participation, adjustment for factors associated with participation should control for the possible selection bias. However, if a particular genotype is associated with severity of disease, cases with genotyping data would most likely not be a random sample of all eligible cases and the possibility of selection bias would exist even after adjustment for other factors associated with study participation.

The third stage of the study where selection bias could occur was in determining the breast tumor intrinsic molecular subtype in cases. Tumor subtype was determined for 62% of enrolled cases, or 48% of eligible cases invited to participate in the study. Subtype and genotype was available for 1220 cases (53% of enrolled cases, 41% of eligible cases). Study results could be biased if the genotypes of cases with sufficient tumor tissue for subtyping were systematically different from the genotypes of women without subtype information. Genotype distributions between cases with and without subtype information were comparable, differing for only a few SNPs. In African American cases, distributions differed for ADIPOQ rs822391, LEP rs10954173 and rs11760956 (which were in perfect LD,

r^2=1.0), and LEPR rs10889563, rs11208654, rs12042877, rs6413506. In non-African American cases, distributions differed for LEPR rs10889569, rs11585329, rs6678033, rs6700896, rs8179183 (rs10889569, rs6678033, and rs6700896 were in strong LD, r^2 > 0.94). The molecular subtype distribution was similar between cases with and without genotyping data.

In conclusion, SNPs in ADIPOQ, IL6, LEP, LEPR, and TNFA were associated with breast cancer; some of these associations appeared to be subtype-specific. The use of haplotypes to estimate the effect of several alleles in combination produced even stronger associations. Furthermore, these effects may be modified by WHR. Further research into adipocytokine-WHR interaction could play a role in identifying the causal relationship between adipocytokines, central obesity, and breast cancer.

The identification of a group of LEPR SNPs associated with basal-like breast cancer is target for future research. In-depth analysis of other variants in the region, including non-SNP variants, will increase the chance of locating the causal variant. Fine-mapping and analysis of conserved DNA sequences in the region could also help determine the location of the causal variant (s). Contingent on replication of the results in other studies, these results provide evidence that genetic polymorphisms in adipocytokines are associated with breast cancer risk.

3.6 References

1. Perou CM, Sorlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, et al. Molecular portraits of human breast tumours. Nature 2000;406(6797):747-52.

2. Sorlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proc Natl Acad Sci U S A 2001;98(19):10869-74.

3. Sorlie T, Tibshirani R, Parker J, Hastie T, Marron JS, Nobel A, et al. Repeated observation of breast tumor subtypes in independent gene expression data sets. Proc Natl Acad Sci U S A 2003;100(14):8418-23.

4. Fulford LG, Easton DF, Reis-Filho JS, Sofronis A, Gillett CE, Lakhani SR, et al. Specific morphological features predictive for the basal phenotype in grade 3 invasive ductal carcinoma of breast. Histopathology 2006;49(1):22-34.

5. Livasy CA, Karaca G, Nanda R, Tretiakova MS, Olopade OI, Moore DT, et al. Phenotypic evaluation of the basal-like subtype of invasive breast carcinoma. Mod Pathol 2006;19(2):264-71.

6. Kim MJ, Ro JY, Ahn SH, Kim HH, Kim SB, Gong G. Clinicopathologic significance of the basal-like subtype of breast cancer: a comparison with hormone receptor and Her2/neu-overexpressing phenotypes. Hum Pathol 2006;37(9):1217-26.

7. Rodriguez-Pinilla SM, Sarrio D, Honrado E, Hardisson D, Calero F, Benitez J, et al. Prognostic significance of basal-like phenotype and fascin expression in node-negative invasive breast carcinomas. Clin Cancer Res 2006;12(5):1533-9.

8. Foulkes WD, Stefansson IM, Chappuis PO, Begin LR, Goffin JR, Wong N, et al. Germline BRCA1 mutations and a basal epithelial phenotype in breast cancer. J Natl Cancer Inst 2003;95(19):1482-5.

9. Lakhani SR, Reis-Filho JS, Fulford L, Penault-Llorca F, van der Vijver M, Parry S, et al. Prediction of BRCA1 status in patients with breast cancer using estrogen receptor and basal phenotype. Clin Cancer Res 2005;11(14):5175-80.

10. Turner NC, Reis-Filho JS. Basal-like breast cancer and the BRCA1 phenotype. Oncogene 2006;25(43):5846-53.

11. Carey LA, Perou CM, Livasy CA, Dressler LG, Cowan D, Conway K, et al. Race, breast cancer subtypes, and survival in the Carolina Breast Cancer Study. Jama 2006;295(21):2492-502.

12. Kurebayashi J, Moriya T, Ishida T, Hirakawa H, Kurosumi M, Akiyama F, et al. The prevalence of intrinsic subtypes and prognosis in breast cancer patients of different

races. Breast 2007;16 Suppl 2:S72-7.

13. Millikan RC, Newman B, Tse CK, Moorman PG, Conway K, Dressler LG, et al. Epidemiology of basal-like breast cancer. Breast Cancer Res Treat 2008;109(1):123-39.

14. Yang XR, Sherman ME, Rimm DL, Lissowska J, Brinton LA, Peplonska B, et al. Differences in risk factors for breast cancer molecular subtypes in a population-based study. Cancer Epidemiol Biomarkers Prev 2007;16(3):439-43.

15. Schaffler A, Scholmerich J, Buechler C. Mechanisms of disease: adipokines and breast cancer - endocrine and paracrine mechanisms that connect adiposity and breast cancer. Nat Clin Pract Endocrinol Metab 2007;3(4):345-54.

16. Rose DP, Komninou D, Stephenson GD. Obesity, adipocytokines, and insulin resistance in breast cancer. Obes Rev 2004;5(3):153-65.

17. Vona-Davis L, Howard-McNatt M, Rose DP. Adiposity, type 2 diabetes and the metabolic syndrome in breast cancer. Obes Rev 2007;8(5):395-408.

18. Hefler LA, Grimm C, Lantzsch T, Lampe D, Leodolter S, Koelbl H, et al. Interleukin-1 and interleukin-6 gene polymorphisms and the risk of breast cancer in caucasian women. Clin Cancer Res 2005;11(16):5718-21.

19. Slattery ML, Curtin K, Baumgartner R, Sweeney C, Byers T, Giuliano AR, et al. IL6, aspirin, nonsteroidal anti-inflammatory drugs, and breast cancer risk in women living in the southwestern United States. Cancer Epidemiol Biomarkers Prev 2007;16(4):747-55.

20. Snoussi K, Strosberg AD, Bouaouina N, Ben Ahmed S, Helal AN, Chouchane L. Leptin and leptin receptor polymorphisms are associated with increased risk and poor prognosis of breast carcinoma. BMC Cancer 2006;6:38.

21. Woo HY, Park H, Ki CS, Park YL, Bae WG. Relationships among serum leptin, leptin receptor gene polymorphisms, and breast cancer in Korea. Cancer Lett 2006;237(1):137-42.

22. Gaudet MM, Egan KM, Lissowska J, Newcomb PA, Brinton LA, Titus-Ernstoff L, et al. Genetic variation in tumor necrosis factor and lymphotoxin-alpha (TNF-LTA) and breast cancer risk. Hum Genet 2007;121(3-4):483-90.

23. Kohaar I, Tiwari P, Kumar R, Nasare V, Thakur N, Das BC, et al. Association of single nucleotide polymorphisms (SNPs) in TNF-LTA locus with breast cancer risk in Indian population. Breast Cancer Res Treat 2009;114(2):347-55.

24. Han CZ, Du LL, Jing JX, Zhao XW, Tian FG, Shi J, et al. Associations among lipids,

leptin, and leptin receptor gene Gin223Arg polymorphisms and breast cancer in China. Biol Trace Elem Res 2008;126(1-3):38-48.

25. Cleveland RJ, Gammon MD, Long CM, Gaudet MM, Eng SM, Teitelbaum SL, et al. Common genetic variations in the LEP and LEPR genes, obesity and breast cancer incidence and survival. Breast Cancer Res Treat 2009.

26. Smith KC, Bateman AC, Fussell HM, Howell WM. Cytokine gene polymorphisms and breast cancer susceptibility and prognosis. Eur J Immunogenet 2004;31(4):167-73.

27. Litovkin KV, Domenyuk VP, Bubnov VV, Zaporozhan VN. Interleukin-6 -174G/C polymorphism in breast cancer and uterine leiomyoma patients: a population-based case control study. Exp Oncol 2007;29(4):295-8.

28. Newman B, Moorman PG, Millikan R, Qaqish BF, Geradts J, Aldrich TE, et al. The Carolina Breast Cancer Study: integrating population-based epidemiology and molecular biology. Breast Cancer Res Treat 1995;35(1):51-60.

29. Millikan R, Eaton A, Worley K, Biscocho L, Hodgson E, Huang WY, et al. HER2 codon 655 polymorphism and risk of breast cancer in African Americans and whites. Breast Cancer Res Treat 2003;79(3):355-64.

30. Weinberg CR, Sandler DP. Randomized recruitment in case-control studies. Am J Epidemiol 1991;134(4):421-32.

31. Huang WY, Newman B, Millikan RC, Schell MJ, Hulka BS, Moorman PG. Hormone-related factors and risk of breast cancer in relation to estrogen receptor and progesterone receptor status. Am J Epidemiol 2000;151(7):703-14.

32. Nielsen TO, Hsu FD, Jensen K, Cheang M, Karaca G, Hu Z, et al. Immunohistochemical and clinical characterization of the basal-like subtype of invasive breast carcinoma. Clin Cancer Res 2004;10(16):5367-74.

33. Livasy CA, Perou CM, Karaca G, Cowan DW, Maia D, Jackson S, et al. Identification of a basal-like subtype of breast ductal carcinoma in situ. Hum Pathol 2007;38(2):197-204.

34. Frazer KA, Ballinger DG, Cox DR, Hinds DA, Stuve LL, Gibbs RA, et al. A second generation human haplotype map of over 3.1 million SNPs. Nature 2007;449(7164):851-61.

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

36. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD

and haplotype maps. Bioinformatics 2005;21(2):263-5.

37. Carlson CS, Eberle MA, Rieder MJ, Yi Q, Kruglyak L, Nickerson DA. Selecting a maximally informative set of single-nucleotide polymorphisms for association analyses using linkage disequilibrium. Am J Hum Genet 2004;74(1):106-20.

38. Balding DJ. A tutorial on statistical methods for population association studies. Nat Rev Genet 2006;7(10):781-91.

39. Hosking L, Lumsden S, Lewis K, Yeo A, McCarthy L, Bansal A, et al. Detection of genotyping errors by Hardy-Weinberg equilibrium testing. Eur J Hum Genet 2004;12(5):395-9.

40. Wigginton JE, Cutler DJ, Abecasis GR. A note on exact tests of Hardy-Weinberg equilibrium. Am J Hum Genet 2005;76(5):887-93.

41. Barnholtz-Sloan JS, Chakraborty R, Sellers TA, Schwartz AG. Examining population stratification via individual ancestry estimates versus self-reported race. Cancer Epidemiol Biomarkers Prev 2005;14(6):1545-51.

42. Pritchard JK, Rosenberg NA. Use of unlinked genetic markers to detect population stratification in association studies. Am J Hum Genet 1999;65(1):220-8.

43. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. Genetics 2000;155(2):945-59.

44. Falush D, Stephens M, Pritchard JK. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. Genetics 2003;164(4):1567-87.

45. Hosmer DW, Lemeshow S. Special Topics: Polytomous Logistic Regression. In: Applied Logistic Regression. New York: John Wiley & Sons; 1989.

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

47. Lin DY, Zeng D. Likelihood-Based Inference on Haplotype Effects in Genetic Association Studies. Journal of the American Statistical Association 2006;101(473):89-118.

48. Hall IJ, Moorman PG, Millikan RC, Newman B. Comparative analysis of breast cancer risk factors among African-American women and White women. Am J Epidemiol 2005;161(1):40-51.

49. Hall IJ, Newman B, Millikan RC, Moorman PG. Body size and breast cancer risk in

black women and white women: the Carolina Breast Cancer Study. Am J Epidemiol 2000;151(8):754-64.

50. Harvie M, Hooper L, Howell AH. Central obesity and breast cancer risk: a systematic review. Obes Rev 2003;4(3):157-73.

51. Connolly BS, Barnett C, Vogt KN, Li T, Stone J, Boyd NF. A meta-analysis of published literature on waist-to-hip ratio and risk of breast cancer. Nutr Cancer 2002;44(2):127-38.

52. Zou GY. On the estimation of additive interaction by use of the four-by-two table and beyond. Am J Epidemiol 2008;168(2):212-24.

53. Demark-Wahnefried W, Winer EP, Rimer BK. Why women gain weight with adjuvant chemotherapy for breast cancer. J Clin Oncol 1993;11(7):1418-29.

54. McInnes JA, Knobf MT. Weight gain and quality of life in women treated with adjuvant chemotherapy for early-stage breast cancer. Oncol Nurs Forum 2001;28(4):675-84.

55. Rock CL, Flatt SW, Newman V, Caan BJ, Haan MN, Stefanick ML, et al. Factors associated with weight gain in women after diagnosis of breast cancer. Women's Healthy Eating and Living Study Group. J Am Diet Assoc 1999;99(10):1212-21.

56. Harvie MN, Campbell IT, Baildam A, Howell A. Energy balance in early breast cancer patients receiving adjuvant chemotherapy. Breast Cancer Res Treat 2004;83(3):201-10.

57. Costa LJ, Varella PC, del Giglio A. Weight changes during chemotherapy for breast cancer. Sao Paulo Med J 2002;120(4):113-7.

58. Kumar N, Allen KA, Riccardi D, Bercu BB, Cantor A, Minton S, et al. Fatigue, weight gain, lethargy and amenorrhea in breast cancer patients on chemotherapy: is subclinical hypothyroidism the culprit? Breast Cancer Res Treat 2004;83(2):149-59.

59. Goodwin PJ, Panzarella T, Boyd NF. Weight gain in women with localized breast cancer--a descriptive study. Breast Cancer Res Treat 1988;11(1):59-66.

60. Han HS, Lee KW, Kim JH, Kim SW, Kim IA, Oh DY, et al. Weight changes after adjuvant treatment in Korean women with early breast cancer. Breast Cancer Res Treat 2008.

61. Freedman RJ, Aziz N, Albanes D, Hartman T, Danforth D, Hill S, et al. Weight and body composition changes during and after adjuvant chemotherapy in women with breast cancer. J Clin Endocrinol Metab 2004;89(5):2248-53.

62. Ingram C, Brown JK. Patterns of weight and body composition change in premenopausal women with early stage breast cancer: has weight gain been overestimated? Cancer Nurs 2004;27(6):483-90.

63. Camoriano JK, Loprinzi CL, Ingle JN, Therneau TM, Krook JE, Veeder MH. Weight change in women treated with adjuvant therapy or observed following mastectomy for node-positive breast cancer. J Clin Oncol 1990;8(8):1327-34.

64. Saquib N, Flatt SW, Natarajan L, Thomson CA, Bardwell WA, Caan B, et al. Weight gain and recovery of pre-cancer weight after breast cancer treatments: evidence from the women's healthy eating and living (WHEL) study. Breast Cancer Res Treat 2007;105(2):177-86.

65. Goodwin PJ, Ennis M, Pritchard KI, McCready D, Koo J, Sidlofsky S, et al. Adjuvant treatment and onset of menopause predict weight gain after breast cancer diagnosis. J Clin Oncol 1999;17(1):120-9.

66. Ali S. Female Breast Cancer Incidence, Stage at Diagnosis, Treatment, and Mortality in North Carolina. In: State Center for Health Statistics, North Carolina Department of Health and Human Services; 2006. p. 1-6.

67. Calza S, Hall P, Auer G, Bjohle J, Klaar S, Kronenwett U, et al. Intrinsic molecular signature of breast cancer in a population-based cohort of 412 patients. Breast Cancer Res 2006;8(4):R34.

68. Fox MP, Lash TL, Greenland S. A method to automate probabilistic sensitivity analyses of misclassified binary variables. Int J Epidemiol 2005;34(6):1370-6.

69. Lakhani SR. The transition from hyperplasia to invasive carcinoma of the breast. J Pathol 1999;187(3):272-8.

70. Leonard GD, Swain SM. Ductal carcinoma in situ, complexities and challenges. J Natl Cancer Inst 2004;96(12):906-20.

71. Morrow M, O'Sullivan MJ. The dilemma of DCIS. Breast 2007;16 Suppl 2:S59-62.

72. Erbas B, Provenzano E, Armes J, Gertig D. The natural history of ductal carcinoma in situ of the breast: a review. Breast Cancer Res Treat 2006;97(2):135-44.

73. Kerlikowske K, Barclay J, Grady D, Sickles EA, Ernster V. Comparison of risk factors for ductal carcinoma in situ and invasive breast cancer. J Natl Cancer Inst 1997;89(1):76-82.

74. Reinier KS, Vacek PM, Geller BM. Risk factors for breast carcinoma in situ versus invasive breast cancer in a prospective study of pre- and post-menopausal women. Breast Cancer Res Treat 2007;103(3):343-8.

75. Dabbs DJ, Chivukula M, Carter G, Bhargava R. Basal phenotype of ductal carcinoma in situ: recognition and immunohistologic profile. Mod Pathol 2006;19(11):1506-11.

76. Bryan BB, Schnitt SJ, Collins LC. Ductal carcinoma in situ with basal-like phenotype: a possible precursor to invasive basal-like breast cancer. Mod Pathol 2006;19(5):617-21.

77. Hannemann J, Velds A, Halfwerk JB, Kreike B, Peterse JL, van de Vijver MJ. Classification of ductal carcinoma in situ by gene expression profiling. Breast Cancer Res 2006;8(5):R61.

78. Cleary MP, Phillips FC, Getzin SC, Jacobson TL, Jacobson MK, Christensen TA, et al. Genetically obese MMTV-TGF-alpha/Lep(ob)Lep(ob) female mice do not develop mammary tumors. Breast Cancer Res Treat 2003;77(3):205-15.

79. Ronti T, Lupattelli G, Mannarino E. The endocrine function of adipose tissue: an update. Clin Endocrinol (Oxf) 2006;64(4):355-65.

80. Catalano S, Marsico S, Giordano C, Mauro L, Rizza P, Panno ML, et al. Leptin enhances, via AP-1, expression of aromatase in the MCF-7 cell line. J Biol Chem 2003;278(31):28668-76.

81. Catalano S, Mauro L, Marsico S, Giordano C, Rizza P, Rago V, et al. Leptin induces, via ERK1/ERK2 signal, functional activation of estrogen receptor alpha in MCF-7 cells. J Biol Chem 2004;279(19):19908-15.

82. Cirillo D, Rachiglio AM, la Montagna R, Giordano A, Normanno N. Leptin signaling in breast cancer: an overview. J Cell Biochem 2008;105(4):956-64.

83. Soma D, Kitayama J, Yamashita H, Miyato H, Ishikawa M, Nagawa H. Leptin augments proliferation of breast cancer cells via transactivation of HER2. J Surg Res 2008;149(1):9-14.

84. van Rossum CT, Hoebee B, van Baak MA, Mars M, Saris WH, Seidell JC. Genetic variation in the leptin receptor gene, leptin, and weight gain in young Dutch adults. Obes Res 2003;11(3):377-86.

85. Wauters M, Mertens I, Chagnon M, Rankinen T, Considine RV, Chagnon YC, et al. Polymorphisms in the leptin receptor gene, body composition and fat distribution in overweight and obese women. Int J Obes Relat Metab Disord 2001;25(5):714-20.

86. Kaklamani VG, Sadim M, Hsi A, Offit K, Oddoux C, Ostrer H, et al. Variants of the adiponectin and adiponectin receptor 1 genes and breast cancer risk. Cancer Res 2008;68(9):3178-84.

87. Asgeirsson KS, Olafsdottir K, Jonasson JG, Ogmundsdottir HM. The effects of IL-6

on cell adhesion and e-cadherin expression in breast cancer. Cytokine 1998;10(9):720-8.

88. Tamm I, Kikuchi T, Cardinale I, Krueger JG. Cell-adhesion-disrupting action of interleukin 6 in human ductal breast carcinoma cells. Proc Natl Acad Sci U S A 1994;91(8):3329-33.

89. Badache A, Hynes NE. Interleukin 6 inhibits proliferation and, in cooperation with an epidermal growth factor receptor autocrine loop, increases migration of T47D breast cancer cells. Cancer Res 2001;61(1):383-91.

90. Macdiarmid F, Wang D, Duncan LJ, Purohit A, Ghilchick MW, Reed MJ. Stimulation of aromatase activity in breast fibroblasts by tumor necrosis factor alpha. Mol Cell Endocrinol 1994;106(1-2):17-21.

91. Speirs V, Kerin MJ, Walton DS, Newton CJ, Desai SB, Atkin SL. Direct activation of oestrogen receptor-alpha by interleukin-6 in primary cultures of breast cancer epithelial cells. Br J Cancer 2000;82(7):1312-6.

92. DeMichele A, Gray R, Horn M, Chen J, Aplenc R, Vaughan WP, et al. Host genetic variants in the interleukin-6 promoter predict poor outcome in patients with estrogen receptor-positive, node-positive breast cancer. Cancer Res 2009;69(10):4184-91.

93. Chavey C, Bibeau F, Gourgou-Bourgade S, Burlinchon S, Boissiere F, Laune D, et al. Oestrogen receptor negative breast cancers exhibit high cytokine content. Breast Cancer Res 2007;9(1):R15.

94. Garcia-Tunon I, Ricote M, Ruiz A, Fraile B, Paniagua R, Royuela M. Role of tumor necrosis factor-alpha and its receptors in human benign breast lesions and tumors (in situ and infiltrative). Cancer Sci 2006;97(10):1044-9.

95. Robinson EK, Sneige N, Grimm EA. Correlation of interleukin 6 with interleukin 1alpha in human mammary tumours, but not with oestrogen receptor expression. Cytokine 1998;10(12):970-6.

96. Quinton ND, Lee AJ, Ross RJ, Eastell R, Blakemore AI. A single nucleotide polymorphism (SNP) in the leptin receptor is associated with BMI, fat mass and leptin levels in postmenopausal Caucasian women. Hum Genet 2001;108(3):233-6.

97. Terry CF, Loukaci V, Green FR. Cooperative influence of genetic polymorphisms on interleukin 6 transcriptional regulation. J Biol Chem 2000;275(24):18138-44.

98. Skoog T, van't Hooft FM, Kallin B, Jovinge S, Boquist S, Nilsson J, et al. A common functional polymorphism (C-->A substitution at position -863) in the promoter region of the tumour necrosis factor-alpha (TNF-alpha) gene associated with reduced circulating levels of TNF-alpha. Hum Mol Genet 1999;8(8):1443-9.

99. Sharma S, Sharma A, Kumar S, Sharma SK, Ghosh B. Association of TNF haplotypes with asthma, serum IgE levels, and correlation with serum TNF-alpha levels. Am J Respir Cell Mol Biol 2006;35(4):488-95.

100. Slattery ML, Curtin K, Sweeney C, Wolff RK, Baumgartner RN, Baumgartner KB, et al. Modifying effects of IL-6 polymorphisms on body size-associated breast cancer risk. Obesity (Silver Spring) 2008;16(2):339-47.

101. Savitz DA. Interpreting Epidemiologic Evidence. New York: Oxford University Press; 2003.

102. Bouatia-Naji N, Meyre D, Lobbens S, Seron K, Fumeron F, Balkau B, et al. ACDC/adiponectin polymorphisms are associated with severe childhood and adult obesity. Diabetes 2006;55(2):545-50.

103. Fishman D, Faulds G, Jeffery R, Mohamed-Ali V, Yudkin JS, Humphries S, et al. The effect of novel polymorphisms in the interleukin-6 (IL-6) gene on IL-6 transcription and plasma IL-6 levels, and an association with systemic-onset juvenile chronic arthritis. J Clin Invest 1998;102(7):1369-76.

104. Balasubramanian SP, Azmy IA, Higham SE, Wilson AG, Cross SS, Cox A, et al. Interleukin gene polymorphisms and breast cancer: a case control study and systematic literature review. BMC Cancer 2006;6:188.

105. Chiu KC, Chu A, Chuang LM, Saad MF. Association of leptin receptor polymorphism with insulin resistance. Eur J Endocrinol 2004;150(5):725-9.

106. Pechlivanis S, Bermejo JL, Pardini B, Naccarati A, Vodickova L, Novotny J, et al. Genetic variation in adipokine genes and risk of colorectal cancer. Eur J Endocrinol 2009;160(6):933-40.

3.7 Tables

Table 3.1 Genotype frequencies in adipocytokines, adjusted for sampling probabilities[1]

		Controls				Cases			
		African American		Non-African American		African American		Non-African American	
		N	%	N	%	N	%	N	%
ADIPOQ									
rs16861194	AA	387	62	945	86	438	59	1038	86
	AG	220	31	159	13	256	34	186	14
	GG	51	7	12	2	48	6	5	0
	Missing			1					
rs16861205	AA	29	3	12	2	40	6	3	0
	AG	203	30	150	12	223	30	178	13
	GG	426	67	955	87	479	64	1048	87
rs822391	CC	3	0	44	3	4	1	58	4
	CT	64	8	359	40	57	8	430	36
	TT	591	91	713	57	681	92	741	60
	Missing			1					
rs16861210	AA	14	2	11	1	20	3	10	1
	AG	196	30	184	15	197	26	214	19
	GG	448	68	922	84	525	71	1005	81
rs822396	AA	407	60	730	60	467	63	767	62
	AG	222	36	346	38	238	32	405	34
	GG	29	4	41	3	37	5	57	4
rs12495941	GG	282	43	476	39	302	41	545	45
	GT	304	47	502	50	350	47	554	44
	TT	72	10	139	11	90	12	130	11
rs7649121	AA	493	73	763	71	564	76	811	67
	AT	147	25	320	25	168	23	384	30
	TT	18	2	34	4	9	1	34	3
	Missing					1			
rs9877202	AA	481	72	1109	100	519	70	1224	100
	AG	162	26	6	0	207	28	5	0
	GG	15	2	2	0	16	2		

		Controls				Cases			
		African American		Non-African American		African American		Non-African American	
		N	%	N	%	N	%	N	%
rs2241766	GG			26	2	5	1	25	2
	GT	46	7	231	18	58	8	250	20
	TT	612	93	860	80	679	91	954	78
rs1501299	AA	95	14	74	5	89	12	78	7
	AC	304	45	434	40	327	44	486	40
	CC	259	41	609	55	325	44	664	53
	Missing					1		1	
rs3821799	CC	134	19	348	34	157	21	391	31
	CT	312	51	542	48	336	45	605	49
	TT	212	30	227	18	249	34	233	20
rs6444174	CC	11	2			23	3		
	CT	167	27	15	1	201	27	6	0
	TT	480	71	1102	99	518	70	1223	100
rs6773957	AA	208	30	166	12	239	33	171	15
	AG	312	50	529	49	339	45	592	49
	GG	138	20	422	40	164	22	466	37
rs1063537	CC	626	96	870	81	696	94	963	78
	CT	32	4	222	18	43	6	244	20
	TT			25	2	3	0	22	2
rs9842733	AA	539	82	1114	100	593	80	1228	100
	AT	116	18	3	0	142	19	1	0
	TT	3	0			7	1		
rs1403697	CC	5	1			18	2		
	CT	155	24	3	0	181	25	2	0
	TT	498	75	1114	100	543	73	1227	100
IL6									
rs2069824	CC	11	2	6	0	9	1	12	1
	CT	135	19	161	12	127	17	162	12
	TT	507	79	950	88	606	82	1055	87

		Controls				Cases			
		African American		Non-African American		African American		Non-African American	
		N	%	N	%	N	%	N	%
	Missing	5							
rs2069827	GG	632	97	918	84	720	97	1000	81
	GT	26	3	188	15	22	3	219	19
	TT			11	1			10	1
rs1800796	CC	11	1	6	0	8	1	6	1
	CG	96	15	81	8	118	16	115	10
	GG	551	83	1030	92	616	83	1108	90
rs1800795	CC	5	0	204	17	4	1	230	19
	CG	95	13	543	56	104	14	589	50
	GG	556	87	367	27	630	85	408	31
	Missing	2		3		4		2	
rs2069832	AA	5	0	202	16	4	1	224	18
	AG	94	13	542	56	105	14	592	50
	GG	559	87	371	28	633	85	411	32
	Missing			2				2	
rs2069835	CC	4	1	2	0	11	2	4	0
	CT	110	15	125	10	145	20	107	8
	TT	544	84	990	90	585	79	1118	91
	Missing					1			
rs2069840	CC	470	67	476	40	526	71	540	43
	CG	171	30	489	49	194	26	545	46
	GG	16	3	151	11	22	3	144	11
	Missing	1		1					
rs2069842	AA	4	1			8	1	1	0
	AG	87	16	6	0	99	13	2	0
	GG	567	83	1111	100	635	86	1226	100
rs1548216	CC	23	3	1	0	26	3	1	0
	CG	199	31	41	4	226	30	58	5
	GG	435	66	1075	96	490	66	1170	95

		Controls				Cases			
		African American		Non-African American		African American		Non-African American	
		N	%	N	%	N	%	N	%
	Missing	1							
rs2069843	AA	7	1	1	0	7	1	1	0
	AG	162	22	39	4	183	24	56	4
	GG	489	77	1077	96	552	75	1172	95
rs2069845	AA	281	43	342	26	312	42	370	28
	AG	299	46	548	55	344	46	604	51
	GG	77	10	226	19	86	12	252	20
	Missing	1		1				3	
LEP									
rs6976701	AA	7	1	1	0	5	1		
	AG	125	19	12	1	154	21	25	2
	GG	526	80	1103	99	581	78	1204	98
	Missing			1		2			
rs4236625	AA	460	68	973	87	510	69	1063	86
	AT	182	30	135	13	202	27	153	13
	TT	15	2	7	1	27	4	5	1
	Missing	1		2		3		8	
rs12706832	AA	403	61	248	23	453	61	251	22
	AG	225	33	539	41	246	33	619	50
	GG	29	7	329	35	43	6	358	28
	Missing	1		1				1	
rs10244329	AA	154	25	272	33	169	23	304	24
	AT	331	49	543	43	369	50	626	51
	TT	173	25	301	24	204	27	299	24
	Missing			1					
rs11763517	CC	30	4	290	24	24	3	287	24
	CT	225	33	537	43	237	32	617	50
	TT	402	63	288	34	481	65	325	26
	Missing	1		2					
rs7795794	AA	6	1	6	0	15	2	6	1

		Controls				Cases			
		African American		Non-African American		African American		Non-African American	
		N	%	N	%	N	%	N	%
	AG	157	26	129	13	167	23	149	13
	GG	495	73	982	86	560	75	1073	86
	Missing							1	
rs11760956	AA	24	3	176	16	19	3	165	15
	AG	201	29	504	39	222	30	580	47
	GG	433	68	437	45	501	68	483	39
	Missing							1	
rs10954173	AA	24	3	176	16	19	3	165	15
	AG	201	29	504	39	220	30	580	47
	GG	433	68	437	45	503	68	483	39
	Missing							1	
rs3793162	AA	19	3			19	3		
	AG	192	28	10	1	163	22	15	1
	GG	447	69	1107	99	560	76	1214	99
rs3828942	AA	20	6	210	27	36	5	231	19
	AG	198	28	520	41	245	33	603	48
	GG	438	66	385	32	460	62	394	34
	Missing	2		2		1		1	
rs17151919	AA	4	1			8	1		
	AG	100	15	5	0	131	18	1	0
	GG	554	85	1112	100	603	81	1228	100
rs17151922	GG	375	57	1099	99	383	52	1203	98
	GT	242	37	17	1	308	41	26	2
	TT	41	6	1	0	51	7		
rs10954174	AA	8	1			6	1		
	AG	90	11	8	0	116	16	6	0
	GG	560	88	1109	100	620	83	1223	100
rs11761556	AA	20	2	332	35	30	4	354	28
	AC	188	27	528	41	211	29	614	49

		Controls				Cases			
		African American		Non-African American		African American		Non-African American	
		N	%	N	%	N	%	N	%
	CC	448	71	255	23	499	67	259	23
	Missing	2		2		2		2	
LEPR									
rs3806318	AA	573	84	586	53	664	90	629	51
	AG	81	16	439	40	72	10	521	42
	GG	4	1	90	7	6	1	77	7
	Missing			2				2	
rs1327118	CC	115	19	234	19	127	17	231	20
	CG	308	51	546	45	341	46	661	53
	GG	235	30	336	35	273	37	337	27
	Missing			1		1			
rs12145690	AA	234	30	334	35	272	37	337	27
	AC	310	51	547	45	340	46	661	53
	CC	114	19	234	19	129	17	231	20
	Missing			2		1			
rs9436738	AA	9	1	24	2	9	1	29	2
	AG	159	22	257	19	177	24	290	24
	GG	490	77	835	79	556	75	909	74
	Missing			1				1	
rs9436297	CC	9	1	25	2	13	2	30	2
	CT	176	24	258	19	189	25	290	24
	TT	473	74	833	79	540	73	909	74
	Missing			1					
rs9436740	AA	96	15	562	47	124	17	635	53
	AT	418	65	473	44	459	62	496	39
	TT	143	20	82	9	157	21	97	8
	Missing	1				2		1	
rs9436299	AA	438	69	531	45	485	65	580	47
	AC	191	28	490	48	230	31	503	41
	CC	29	3	96	8	27	4	146	11

		Controls				Cases			
		African American		Non-African American		African American		Non-African American	
		N	%	N	%	N	%	N	%
rs17127608	CC	472	73	1113	100	559	75	1220	99
	CT	172	24	4	0	165	22	8	1
	TT	14	3			18	2	1	0
rs3790433	AA	335	49	74	6	373	50	95	8
	AG	252	43	403	31	306	41	438	35
	GG	71	9	640	63	63	9	696	57
rs17127618	CC	427	64	799	74	464	63	878	72
	CG	196	31	292	24	249	33	317	25
	GG	34	5	26	2	26	3	33	2
	Missing	1				3		1	
rs7534511	AA	37	4	84	7	36	5	127	10
	AG	206	35	478	46	241	33	501	41
	GG	414	62	553	48	465	62	600	49
	Missing	1		2				1	
rs9436301	CC	93	18	73	6	114	15	89	8
	CT	309	45	430	35	370	50	443	34
	TT	256	37	614	59	258	35	697	58
rs1887285	CC	9	3	13	1	15	2	14	1
	CT	122	22	183	15	158	21	196	16
	TT	526	75	921	85	569	77	1017	83
	Missing	1						2	
rs17097182	AA	337	55	1041	94	395	53	1127	91
	AT	251	36	73	6	288	39	99	9
	TT	70	9	3	0	59	8	3	0
rs17412175	AA	7	1	257	23	8	1	245	22
	AT	104	18	553	55	121	16	628	50
	TT	547	81	307	23	613	83	356	28
rs970467	AA	39	7	14	1	37	5	25	2
	AG	229	33	232	18	270	36	249	19

		Controls				Cases			
		African American		Non-African American		African American		Non-African American	
		N	%	N	%	N	%	N	%
	GG	390	60	871	81	435	59	955	79
rs9436746	AA	259	38	153	12	294	40	190	15
	AC	285	44	532	49	338	46	587	48
	CC	112	18	430	39	106	14	449	37
	Missing	2		2		4		3	
rs9436748	GG	449	68	331	25	504	68	382	30
	GT	175	27	555	55	214	29	627	50
	TT	34	4	228	21	24	3	220	19
	Missing			3					
rs6657868	AA	175	25	133	11	184	25	180	14
	AG	308	45	532	50	366	50	553	45
	GG	174	30	452	39	192	25	496	41
	Missing	1							
rs17127655	CC	417	66	1106	99	508	69	1220	100
	CT	206	29	8	1	198	27	7	0
	TT	33	5	1	0	30	4	1	0
	Missing	2		2		6		1	
rs6588147	AA	473	74	511	43	511	68	560	46
	AG	160	23	503	48	211	29	524	43
	GG	25	3	103	8	20	3	145	11
rs7531110	GG	221	29	146	12	227	31	195	15
	GT	319	51	562	52	384	52	578	47
	TT	118	21	408	36	131	17	456	38
	Missing			1					
rs7555955	AA	25	3	103	8	23	3	145	11
	AG	177	25	503	48	216	30	524	43
	GG	455	72	511	43	503	67	560	46
	Missing	1							
rs6704167	AA	397	57	347	26	450	61	394	31

		Controls				Cases			
		African American		Non-African American		African American		Non-African American	
		N	%	N	%	N	%	N	%
	AT	229	39	544	55	256	34	620	51
	TT	32	4	224	19	34	5	215	18
	Missing			2		2			
rs7529650	AA	85	15	400	34	99	13	450	37
	AG	316	51	561	53	354	48	576	47
	GG	256	34	156	13	287	39	203	16
	Missing	1				2			
rs2025804	CC	27	3	104	8	44	6	141	11
	CT	210	33	494	48	240	32	515	42
	TT	421	64	519	44	458	61	573	47
rs7518849	CC	26	4	4	0	25	3	6	0
	CT	172	23	127	10	197	27	132	11
	TT	460	73	986	90	520	70	1091	89
rs10158579	CC	92	16	16	1	87	12	28	3
	CT	274	44	260	21	335	45	277	22
	TT	292	40	841	78	320	43	922	76
	Missing							2	
rs11808888	AA	176	28	18	1	178	24	29	3
	AG	294	47	268	22	366	50	281	22
	GG	188	26	831	77	198	27	919	75
rs17127677	GG	415	59	843	78	435	59	928	76
	GT	213	35	259	21	275	37	274	21
	TT	30	7	15	1	32	4	27	3
rs17127686	AA	450	70	1112	100	528	72	1224	100
	AG	185	28	5	0	191	26	4	0
	GG	22	2			14	2	1	0
	Missing	1				9			
rs6694528	CC	247	34	839	78	268	36	923	76
	CT	295	48	262	21	353	48	278	22

		Controls				Cases			
		African American		Non-African American		African American		Non-African American	
		N	%	N	%	N	%	N	%
	TT	116	18	16	1	121	16	28	3
rs11208654	CC	28	4	104	8	39	6	141	11
	CT	211	32	495	48	253	34	519	43
	TT	415	64	516	44	447	60	567	47
	Missing	4		2		3		2	
rs10889556	AA	229	30	619	52	264	36	674	55
	AG	326	56	432	43	364	49	451	37
	GG	103	13	65	5	114	15	104	8
	Missing			1					
rs7526141	CC	545	82	297	22	625	84	333	26
	CT	108	18	556	55	108	14	638	52
	TT	5	0	262	23	9	1	257	21
	Missing			2				1	
rs1171275	AA	49	7	35	2	51	7	44	4
	AG	257	40	325	28	323	44	345	27
	GG	351	53	756	70	368	49	840	69
	Missing	1		1					
rs1475397	CC	92	13	613	59	113	15	669	55
	CT	297	51	418	35	340	46	470	37
	TT	269	36	86	6	289	39	90	7
rs1627238	CC	215	32	751	70	220	29	833	68
	CT	299	49	328	28	389	53	350	28
	TT	143	19	37	2	132	18	46	4
	Missing	1		1		1			
rs11208662	CC	26	3	10	1	23	3	12	1
	CG	187	27	168	13	212	29	185	15
	GG	445	69	939	87	507	68	1032	84
rs1171279	CC	92	13	613	59	112	15	669	55
	CT	297	51	418	35	340	46	469	37

		Controls				Cases			
		African American		Non-African American		African American		Non-African American	
		N	%	N	%	N	%	N	%
	TT	269	36	86	6	290	39	91	8
rs1751492	CC	52	10	79	6	63	9	116	9
	CT	270	42	451	44	315	42	495	40
	TT	336	48	587	50	364	49	618	51
rs6697315	CC	92	16	120	9	115	15	161	13
	CT	324	50	501	47	366	49	542	45
	TT	241	34	495	44	260	35	526	43
	Missing	1		1		1			
rs1171267	GG	291	40	496	44	313	43	529	43
	GT	299	46	502	47	336	45	539	45
	TT	68	13	118	9	92	12	156	12
	Missing			1		1		5	
rs1782763	CC	43	9	116	9	66	9	160	12
	CT	263	39	475	45	289	39	489	41
	TT	352	51	524	46	387	52	580	47
	Missing			2					
rs1409802	AA	22	4	54	4	27	4	93	7
	AG	201	30	417	41	241	32	444	36
	GG	435	66	646	55	474	64	692	56
rs10157610	CC	496	76	1114	100	590	79	1224	100
	CT	150	23	3	0	142	19	4	0
	TT	12	1			10	1	1	0
rs3790431	CC	40	4	57	4	47	7	58	5
	CT	287	46	385	33	303	41	409	34
	TT	331	50	674	63	392	53	760	60
	Missing			1				2	
rs1137100	AA	449	66	633	54	490	66	684	56
	AG	189	31	425	41	225	30	448	37
	GG	20	3	59	4	25	3	97	7

		Controls				Cases			
		African American		Non-African American		African American		Non-African American	
		N	%	N	%	N	%	N	%
	Missing					2			
rs3790429	AA	528	78	777	63	622	84	875	72
	AT	128	22	305	35	115	16	334	27
	TT	2	0	35	2	5	1	20	2
rs3790426	GG	374	58	679	63	440	60	766	61
	GT	232	35	383	33	242	33	405	34
	TT	52	6	53	4	56	8	56	5
	Missing			2		4		2	
rs1343982	AA	55	10	68	5	75	10	108	8
	AG	268	44	444	42	314	42	471	39
	GG	335	46	605	53	353	48	650	53
rs10493380	AA	547	81	741	59	644	87	823	68
	AC	110	18	337	39	94	13	375	30
	CC	1	0	39	2	4	1	31	2
rs1938489	AA	361	51	677	63	431	58	767	61
	AG	265	45	385	33	274	37	408	34
	GG	32	4	55	4	37	5	54	5
rs10889563	AA	137	20	297	26	166	23	323	26
	AG	358	55	567	54	366	49	605	48
	GG	163	26	253	19	210	28	300	25
	Missing							1	
rs12042877	CC	341	50	606	53	361	49	648	53
	CT	268	43	444	42	314	42	473	39
	TT	49	7	67	5	67	9	107	8
	Missing							1	
rs10749754	AA	176	26	211	16	220	30	252	21
	AG	341	52	548	52	357	48	588	47
	GG	141	22	358	32	165	22	389	32

		Controls				Cases			
		African American		Non-African American		African American		Non-African American	
		N	%	N	%	N	%	N	%
rs12564626	AA	121	20	207	16	137	19	249	21
	AG	324	46	549	52	363	49	591	47
	GG	213	34	361	32	242	33	388	32
	Missing							1	
rs1137101	AA	127	20	358	32	138	19	387	32
	AG	326	49	547	52	367	49	585	47
	GG	204	31	212	17	237	32	257	22
	Missing	1							
rs4655537	AA	89	12	160	16	111	15	169	15
	AG	324	50	514	42	356	48	557	44
	GG	245	37	443	43	275	37	503	41
rs3828034	CC	1	0	38	2	1	0	43	3
	CT	58	10	332	36	40	5	342	27
	TT	599	90	747	61	701	95	844	70
rs12405556	GG	461	71	650	56	515	69	696	57
	GT	174	26	406	40	203	27	434	36
	TT	23	3	61	5	24	3	99	7
rs3762274	AA	125	20	420	37	139	19	483	39
	AG	335	49	544	51	363	49	558	44
	GG	198	30	153	12	238	32	188	17
	Missing					2			
rs11585329	GG	608	92	789	74	691	93	865	71
	GT	48	8	300	25	51	7	331	26
	TT	2	0	28	2			33	3
rs11801408	CC	226	36	715	59	281	38	812	68
	CT	311	46	360	39	352	48	363	29
	TT	121	18	42	3	109	15	54	4
rs8179183	CC	35	6	34	2	20	3	40	3
	CG	215	30	326	36	228	31	340	27

		Controls				Cases			
		African American		Non-African American		African American		Non-African American	
		N	%	N	%	N	%	N	%
	GG	408	64	757	62	494	67	849	70
rs6678033	AA	201	30	154	19	210	28	172	13
	AG	336	52	522	43	373	50	564	45
	GG	119	18	441	38	158	21	493	42
	Missing	2				1			
rs4655555	AA	20	5	32	2	20	3	46	3
	AT	165	24	322	33	204	28	340	28
	TT	473	71	763	65	518	70	842	68
	Missing							1	
rs10889569	AA	110	15	441	38	148	20	493	42
	AT	331	54	518	43	371	50	563	45
	TT	211	31	155	19	219	30	171	13
	Missing	6		3		4		2	
rs6693573	CC	1	0			2	0		
	CG	90	11	6	0	101	14	3	0
	GG	567	89	1111	100	639	86	1226	100
rs17127807	AA	481	76	1071	96	527	71	1177	96
	AG	162	22	45	3	204	28	51	4
	GG	14	2	1	0	11	1		
	Missing	1						1	
rs6700896	CC	191	28	427	37	211	29	476	40
	CT	321	46	530	44	383	51	576	46
	TT	146	26	160	19	148	20	177	14
rs17127826	AA	386	59	1065	96	416	56	1177	96
	AG	232	34	51	4	292	39	51	4
	GG	40	7	1	0	34	5	1	0
rs17127828	AA	465	72	1068	96	536	72	1182	97
	AG	183	28	48	4	191	26	47	3
	GG	9	1	1	0	15	2		

		Controls				Cases			
		African American		Non-African American		African American		Non-African American	
		N	%	N	%	N	%	N	%
	Missing	1							
rs6413506	AA	600	90	1117	100	662	89	1224	99
	AG	57	9			79	11	5	1
	GG	1	0			1	0		
TNFA									
rs1799964	CC	19	3	56	4	12	2	50	4
	CT	181	28	346	31	211	28	445	36
	TT	458	69	715	65	519	70	734	59
rs1800630	AA	11	2	35	2	5	1	30	3
	AC	126	19	285	26	148	20	334	27
	CC	521	79	797	71	589	80	865	71

1 - counts (N) reflect raw data. Percentages are adjusted for study sampling probabilities.

Table 3.2 Association between adipocytokine SNPs and breast cancer

	Cases	Controls	OR (95% CI)[1]	CLR[2]	P-value
ADIPOQ					
rs16861194					
GG	53	63	0.70 (0.47 , 1.04)	2.2	0.0794
AG	443	380	1.00 (0.85 , 1.19)	1.4	0.9709
AA	1476	1332	Referent		
rs16861205					
AA+AG	445	395	0.98 (0.83 , 1.16)	1.4	0.8182
GG	1527	1381	Referent		
rs3821799					
CC	548	482	1.02 (0.85 , 1.24)	1.5	0.8197
CT	942	854	0.99 (0.84 , 1.17)	1.4	0.9232
TT	482	440	Referent		
IL6					
rs2069824					
CC+CT	311	313	0.84 (0.70 , 1.01)	1.4	0.0604
TT	1661	1458	Referent		
rs2069827					
GT+TT	251	225	1.03 (0.84 , 1.27)	1.5	0.7805
GG	1721	1551	Referent		
rs1800796					
CC+CG	247	194	1.18 (0.96 , 1.46)	1.5	0.1253
GG	1725	1582	Referent		
LEP					
rs6976701					
AA+AG	184	145	1.16 (0.91 , 1.50)	1.7	0.2366
GG	1786	1630	Referent		
rs3793162					
AA+AG	197	221	0.75 (0.59 , 0.95)	1.6	0.0168
GG	1775	1555	Referent		
rs17151922					
TT	51	42	1.17 (0.75 , 1.83)	2.5	0.4915
GT	334	259	1.30 (1.05 , 1.62)	1.5	0.0166
GG	1587	1475	Referent		
rs10954174					
AA+AG	128	106	1.13 (0.84 , 1.50)	1.8	0.4205
GG	1844	1670	Referent		
LEPR					
rs9436299					
CC	173	126	1.30 (1.00 , 1.69)	1.7	0.05

	Cases	Controls	OR (95% CI)[1]	CLR[2]	P-value
AC	733	681	0.97 (0.84 , 1.12)	1.3	0.6815
AA	1066	969	Referent		
rs17412175					
AA	253	264	0.87 (0.70 , 1.06)	1.5	0.1677
AT+TT	1719	1512	Referent		
rs9436746					
CC	555	542	0.88 (0.75 , 1.02)	1.4	0.0951
AA+AC	1410	1230	Referent		
rs9436748					
TT	244	262	0.81 (0.66 , 0.99)	1.5	0.0429
GG+GT	1728	1511	Referent		
rs6657868					
AA	364	309	1.07 (0.88 , 1.31)	1.5	0.4803
AG	919	840	0.99 (0.85 , 1.15)	1.4	0.8541
GG	689	626	Referent		
rs17127655					
TT+CT	236	248	0.78 (0.62 , 0.98)	1.6	0.0301
CC	1729	1524	Referent		
rs6588147					
GG	165	129	1.21 (0.93 , 1.58)	1.7	0.1541
AG	735	663	1.02 (0.88 , 1.18)	1.3	0.838
AA	1072	984	Referent		
rs6704167					
TT	249	256	0.86 (0.69 , 1.08)	1.6	0.1929
AT	876	773	1.02 (0.88 , 1.19)	1.3	0.7866
AA	845	745	Referent		
rs7529650					
AA	550	485	0.97 (0.79 , 1.18)	1.5	0.7401
AG	930	877	0.89 (0.75 , 1.06)	1.4	0.1841
GG	490	413	Referent		
rs2025804					
CC	185	132	1.31 (1.02 , 1.69)	1.7	0.0369
CT	755	704	0.96 (0.83 , 1.11)	1.3	0.5751
TT	1032	940	Referent		
rs11808888					
AA	208	194	0.99 (0.77 , 1.27)	1.7	0.9159
AG	647	562	1.05 (0.89 , 1.23)	1.4	0.5842
GG	1117	1020	Referent		
rs11208654					
CC	180	133	1.29 (0.99 , 1.66)	1.7	0.0547

	Cases	Controls	OR (95% CI)[1]	CLR[2]	P-value
CT	772	706	1.00 (0.86 , 1.15)	1.3	0.9493
TT	1015	931	Referent		
rs10889556					
GG	218	168	1.19 (0.94 , 1.51)	1.6	0.1397
AG	815	759	0.95 (0.82 , 1.10)	1.3	0.5011
AA	939	848	Referent		
rs7526141					
TT	266	267	0.83 (0.66 , 1.05)	1.6	0.1292
CT	746	664	0.96 (0.81 , 1.14)	1.4	0.6383
CC	959	843	Referent		
rs1751492					
CC	179	131	1.30 (1.01 , 1.67)	1.7	0.0431
CT	810	722	1.06 (0.92 , 1.22)	1.3	0.4436
TT	983	923	Referent		
rs1171267					
TT	248	186	1.25 (1.01 , 1.54)	1.5	0.0402
GG+GT	1718	1589	Referent		
rs1782763					
CC	226	159	1.36 (1.09 , 1.71)	1.6	0.0069
TT+CT	1746	1615	Referent		
rs1409802					
AA	120	76	1.50 (1.10 , 2.06)	1.9	0.0113
AG	685	619	1.02 (0.88 , 1.17)	1.3	0.8225
GG	1167	1081	Referent		
rs1137100					
GG	122	79	1.45 (1.06 , 1.97)	1.9	0.0202
AG	673	615	1.00 (0.87 , 1.16)	1.3	0.9629
AA	1175	1082	Referent		
rs1343982					
AA	183	123	1.43 (1.11 , 1.83)	1.6	0.0048
GG+AG	1789	1653	Referent		
rs10889563					
AA	490	434	0.91 (0.75 , 1.10)	1.5	0.3435
AG	971	926	0.80 (0.68 , 0.95)	1.4	0.0091
GG	510	416	Referent		
rs12042877					
TT	174	116	1.43 (1.11 , 1.84)	1.7	0.0063
CC+CT	1797	1660	Referent		
rs10749754					
AA	472	387	1.11 (0.91, 1.34)	1.5	0.2992

	Cases	Controls	OR (95% CI)[1]	CLR[2]	P-value
AG	945	890	0.92 (0.79, 1.08)	1.4	0.3183
GG	555	499	Referent		
rs1137101					
GG	526	485	0.91 (0.75 , 1.10)	1.5	0.3114
AG	952	874	0.89 (0.75 , 1.06)	1.4	0.1868
AA	494	416	Referent		
rs4655537					
AA	281	249	0.99 (0.81 , 1.23)	1.5	0.9611
AG	913	839	0.96 (0.83 , 1.11)	1.3	0.5877
GG	778	688	Referent		
rs3828034					
CC	44	39	0.96 (0.60 , 1.54)	2.5	0.8742
CT	382	390	0.83 (0.69 , 0.98)	1.4	0.0324
TT	1546	1347	Referent		
rs12405556					
TT	123	84	1.35 (1.00 , 1.83)	1.8	0.0529
GT	637	581	1.00 (0.86 , 1.15)	1.3	0.9583
GG	1212	1111	Referent		
rs3762274					
GG	426	351	1.09 (0.89 , 1.33)	1.5	0.3948
AG	921	880	0.90 (0.77 , 1.05)	1.4	0.1912
AA	623	545	Referent		
rs11801408					
TT	163	163	0.84 (0.66 , 1.08)	1.6	0.1676
CC+CT	1809	1613	Referent		
rs17127826					
GG	35	41	0.70 (0.43 , 1.14)	2.6	0.1495
AA+AG	1937	1735	Referent		
TNF					
rs1799964					
CC	62	75	0.71 (0.49 , 1.03)	2.1	0.068
CT	657	527	1.16 (1.01 , 1.35)	1.3	0.0419
TT	1253	1174	Referent		

1 - odds ratio, 95% confidence interval, adjusted for age, self-identified race, African ancestry, offset term
2 - confidence limit ratio, upper 95% confidence limit divided by lower 95% confidence limit

Table 3.3 Odds ratios for SNPs associated with luminal A and basal-like breast cancer

	Luminal A					Basal-like					Luminal A vs. basal-like[3]
	Case	Control	OR (95% CI)[1]	CLR[2]	P-value	Case	Control	OR (95% CI)[1]	CLR[2]	P-value	P-value
ADIPOQ											
rs16861194											
GG	14	63	0.56 (0.30 , 1.03)	3.5	0.0634	8	63	0.74 (0.34 , 1.64)	4.9	0.4574	0.5382
AG	144	380	0.98 (0.77 , 1.23)	1.6	0.8330	42	380	0.78 (0.53 , 1.14)	2.1	0.1947	0.2746
AA	521	1332	Referent			150	1332	Referent			
rs16861205											
AA+AG	150	395	1.02 (0.81 , 1.29)	1.6	0.8559	39	395	0.66 (0.45 , 0.97)	2.2	0.0345	0.0369
GG	529	1381	Referent			161	1381	Referent			
rs3821799											
CC	191	482	0.99 (0.76 , 1.28)	1.7	0.9200	58	482	1.01 (0.68 , 1.50)	2.2	0.9695	0.9235
CT	323	854	0.97 (0.77 , 1.22)	1.6	0.7792	80	854	0.70 (0.48 , 1.00)	2.1	0.0499	0.0969
TT	165	440	Referent			62	440	Referent			
IL6											
rs2069824											
CC+CT	89	313	0.70 (0.54 , 0.92)	1.7	0.0090	37	313	1.01 (0.68 , 1.48)	2.2	0.9781	0.1021
TT	590	1458	Referent			163	1458	Referent			
rs2069827											
GT+TT	107	225	1.31 (1.01 , 1.72)	1.7	0.0455	21	225	1.05 (0.64 , 1.73)	2.7	0.8491	0.3942
GG	572	1551	Referent			179	1551	Referent			
rs1800796											
CC+CG	97	194	1.44 (1.09 , 1.90)	1.7	0.0095	22	194	0.86 (0.53 , 1.40)	2.6	0.5468	0.0461
GG	582	1582	Referent			178	1582	Referent			
LEP											

	Luminal A					Basal-like					Luminal A vs. basal-like[3]
	Case	Control	OR (95% CI)[1]	CLR[2]	P-value	Case	Control	OR (95% CI)[1]	CLR[2]	P-value	P-value
rs6976701											
AA+AG	69	145	1.42 (1.01 , 1.99)	2.0	0.0415	23	145	0.99 (0.60 , 1.64)	2.7	0.9748	0.1922
GG	610	1630	Referent			177	1630	Referent			
rs3793162											
AA+AG	64	221	0.82 (0.58 , 1.14)	2.0	0.2393	29	221	0.79 (0.50 , 1.26)	2.5	0.3174	0.8953
GG	615	1555	Referent			171	1555	Referent			
rs17151922											
TT	12	42	0.89 (0.44 , 1.77)	4.0	0.7315	7	42	0.98 (0.41 , 2.34)	5.7	0.9699	0.8374
GT	111	259	1.41 (1.04 , 1.92)	1.8	0.0248	40	259	1.01 (0.65 , 1.55)	2.4	0.9809	0.1572
GG	556	1475	Referent			153	1475	Referent			
rs10954174											
AA+AG	35	106	1.01 (0.66 , 1.54)	2.3	0.9581	23	106	1.66 (0.99 , 2.78)	2.8	0.0524	0.0976
GG	644	1670	Referent			177	1670	Referent			
LEPR											
rs9436299											
CC	69	126	1.46 (1.04 , 2.05)	2.0	0.0273	10	126	0.91 (0.46 , 1.82)	4.0	0.7876	0.1913
AC	251	681	0.95 (0.78 , 1.16)	1.5	0.6108	84	681	1.25 (0.91 , 1.71)	1.9	0.1671	0.1124
AA	359	969	Referent			106	969	Referent			
rs17412175											
AA	92	264	0.85 (0.64 , 1.12)	1.7	0.2578	14	264	0.56 (0.31 , 1.01)	3.3	0.0528	0.1765
AT+TT	587	1512	Referent			186	1512	Referent			
rs9436746											
CC	197	542	0.87 (0.70 , 1.07)	1.5	0.1781	37	542	0.57 (0.39 , 0.85)	2.2	0.0050	0.0468
AA+AC	481	1230	Referent			161	1230	Referent			
rs9436748											

	Luminal A					Basal-like					Luminal A vs. basal-like[3]
	Case	Control	OR (95% CI)[1]	CLR[2]	P-value	Case	Control	OR (95% CI)[1]	CLR[2]	P-value	P-value
TT	88	262	0.82 (0.62, 1.08)	1.7	0.1526	14	262	0.48 (0.27, 0.87)	3.2	0.0148	0.1305
GG+GT	591	1511	Referent			186	1511	Referent			
rs6657868											
AA	132	309	1.18 (0.91 , 1.55)	1.7	0.2178	36	309	1.16 (0.74 , 1.84)	2.5	0.5177	0.9452
AG	310	840	0.99 (0.80 , 1.21)	1.5	0.9003	110	840	1.45 (1.02 , 2.06)	2.0	0.0383	0.0434
GG	237	626	Referent			54	626	Referent			
rs17127655											
CT+TT	70	248	0.76 (0.54 , 1.06)	1.9	0.1047	29	248	0.65 (0.41 , 1.04)	2.5	0.0729	0.5662
CC	606	1524	Referent			170	1524	Referent			
rs6588147											
GG	66	129	1.36 (0.96 , 1.91)	2.0	0.0805	8	129	0.74 (0.35 , 1.59)	4.6	0.4412	0.1295
AG	260	663	1.04 (0.85 , 1.27)	1.5	0.7223	86	663	1.39 (1.01 , 1.91)	1.9	0.0417	0.0905
AA	353	984	Referent			106	984	Referent			
rs6704167											
TT	91	256	0.87 (0.65 , 1.17)	1.8	0.3663	16	256	0.67 (0.37 , 1.19)	3.2	0.1742	0.3920
AT	297	773	0.97 (0.79 , 1.19)	1.5	0.7989	94	773	1.18 (0.85 , 1.63)	1.9	0.3289	0.2881
AA	291	745	Referent			89	745	Referent			
rs7529650											
AA	196	485	0.90 (0.69 , 1.17)	1.7	0.4211	39	485	0.89 (0.55 , 1.43)	2.6	0.6285	0.9784
AG	307	877	0.78 (0.62 , 0.98)	1.6	0.0362	112	877	1.21 (0.84 , 1.76)	2.1	0.3071	0.0287
GG	176	413	Referent			49	413	Referent			
rs2025804											
CC	73	132	1.51 (1.08 , 2.09)	1.9	0.0148	13	132	1.11 (0.59 , 2.06)	3.5	0.7512	0.3463
CT	269	704	1.02 (0.84 , 1.24)	1.5	0.8366	85	704	1.20 (0.88 , 1.65)	1.9	0.2526	0.3431
TT	337	940	Referent			102	940	Referent			

	Luminal A					Basal-like					Luminal A vs. basal-like[3]
	Case	Control	OR (95% CI)[1]	CLR[2]	P-value	Case	Control	OR (95% CI)[1]	CLR[2]	P-value	P-value
rs11808888											
AA	67	194	0.97 (0.68 , 1.39)	2.0	0.8775	20	194	0.60 (0.34 , 1.06)	3.1	0.0792	0.1243
AG	206	562	0.95 (0.76 , 1.18)	1.6	0.6424	76	562	0.98 (0.69 , 1.39)	2.0	0.9089	0.8694
GG	406	1020	Referent			104	1020	Referent			
rs11208654											
CC	69	133	1.43 (1.03 , 2.00)	2.0	0.0346	12	133	1.03 (0.54 , 1.96)	3.6	0.9199	0.3336
CT	277	706	1.08 (0.89 , 1.31)	1.5	0.4479	88	706	1.25 (0.92 , 1.71)	1.9	0.1572	0.3807
TT	331	931	Referent			100	931	Referent			
rs10889556											
GG	78	168	1.40 (1.02 , 1.92)	1.9	0.0365	23	168	1.24 (0.74 , 2.07)	2.8	0.4155	0.6560
AG	287	759	1.03 (0.85 , 1.26)	1.5	0.7384	98	759	1.21 (0.87 , 1.67)	1.9	0.2531	0.3787
AA	314	848	Referent			79	848	Referent			
rs7526141											
TT	103	267	0.95 (0.69 , 1.29)	1.9	0.7377	17	267	0.66 (0.36 , 1.21)	3.4	0.1766	0.2580
CT	270	664	1.04 (0.83 , 1.32)	1.6	0.7124	68	664	0.98 (0.67 , 1.44)	2.2	0.9086	0.7499
CC	306	843	Referent			115	843	Referent			
rs1751492											
CC	59	131	1.33 (0.94 , 1.87)	2.0	0.1059	17	131	1.19 (0.68 , 2.09)	3.1	0.5329	0.7514
CT	290	722	1.13 (0.93 , 1.37)	1.5	0.2165	85	722	1.09 (0.80 , 1.49)	1.9	0.5878	0.8609
TT	330	923	Referent			98	923	Referent			
rs1171267											
TT	85	186	1.29 (0.97 , 1.71)	1.8	0.0796	25	186	1.21 (0.77 , 1.92)	2.5	0.4078	0.9120
GG+GT	593	1589	Referent			174	1589	Referent			
rs1782763											
CC	78	159	1.38 (1.03 , 1.87)	1.8	0.0332	21	159	1.29 (0.79 , 2.12)	2.7	0.3094	0.7955

	Luminal A					Basal-like					Luminal A vs. basal-like[3]
	Case	Control	OR (95% CI)[1]	CLR[2]	P-value	Case	Control	OR (95% CI)[1]	CLR[2]	P-value	P-value
CT+TT	601	1615	Referent			179	1615	Referent			
rs1409802											
AA	42	76	1.64 (1.09 , 2.48)	2.3	0.0176	8	76	1.08 (0.50 , 2.33)	4.7	0.8499	0.2962
AG	254	619	1.13 (0.93 , 1.37)	1.5	0.2307	72	619	1.05 (0.76 , 1.44)	1.9	0.7690	0.6762
GG	383	1081	Referent			120	1081	Referent			
rs1137100											
GG	45	79	1.64 (1.10 , 2.45)	2.2	0.0155	8	79	1.02 (0.48 , 2.21)	4.6	0.9505	0.2417
AG	248	615	1.10 (0.91 , 1.34)	1.5	0.3280	70	615	1.04 (0.75 , 1.42)	1.9	0.8319	0.7139
AA	384	1082	Referent			122	1082	Referent			
rs1343982											
AA	62	123	1.47 (1.05 , 2.04)	1.9	0.0228	19	123	1.37 (0.81 , 2.30)	2.8	0.2380	0.7985
AG+GG	617	1653	Referent			181	1653	Referent			
rs10889563											
AA	157	434	0.75 (0.58 , 0.97)	1.7	0.0301	42	434	0.81 (0.52 , 1.26)	2.4	0.3515	0.7346
AG	331	926	0.71 (0.57 , 0.89)	1.6	0.0028	105	926	0.84 (0.59 , 1.21)	2.1	0.3470	0.3983
GG	191	416	Referent			52	416	Referent			
rs12042877											
TT	60	116	1.49 (1.06 , 2.09)	2.0	0.0200	20	116	1.55 (0.93 , 2.58)	2.8	0.0931	0.8886
CC+CT	619	1660	Referent			180	1660	Referent			
rs10749754											
AA	177	387	1.35 (1.04 , 1.74)	1.7	0.0238	48	387	1.21 (0.78 , 1.87)	2.4	0.4015	0.6436
AG	319	890	0.97 (0.78 , 1.21)	1.6	0.7719	106	890	1.15 (0.79 , 1.67)	2.1	0.4577	0.3902
GG	183	499	Referent			46	499	Referent			
rs1137101											
GG	182	416	1.35 (1.04, 1.75)	1.7	0.0252	53	416	1.10 (0.72, 1.69)	2.4	0.6648	0.3831

	Luminal A					Basal-like					Luminal A vs. basal-like[3]
	Case	Control	OR (95% CI)[1]	CLR[2]	P-value	Case	Control	OR (95% CI)[1]	CLR[2]	P-value	P-value
AG	327	874	1.07 (0.86, 1.35)	1.6	0.5379	99	874	1.01 (0.70, 1.47)	2.1	0.9443	0.7760
AA	170	485	Referent			48	485	Referent			
rs4655537											
AA	79	249	0.72 (0.54 , 0.98)	1.8	0.0341	27	249	0.92 (0.57 , 1.47)	2.6	0.7145	0.3623
AG	312	839	0.87 (0.71 , 1.06)	1.5	0.1683	93	839	0.92 (0.67 , 1.28)	1.9	0.6337	0.7322
GG	288	688	Referent			80	688	Referent			
rs3828034											
CC	14	39	0.92 (0.48 , 1.76)	3.7	0.8060	5	39	1.52 (0.57 , 4.08)	7.2	0.4037	0.3162
CT	152	390	0.96 (0.76 , 1.21)	1.6	0.7496	28	390	0.70 (0.45 , 1.09)	2.4	0.1127	0.1816
TT	513	1347	Referent			167	1347	Referent			
rs12405556											
TT	47	84	1.58 (1.06 , 2.33)	2.2	0.0233	9	84	1.02 (0.49 , 2.12)	4.3	0.9587	0.2573
GT	231	581	1.06 (0.87 , 1.30)	1.5	0.5419	62	581	0.97 (0.70 , 1.35)	1.9	0.8762	0.6209
GG	401	1111	Referent			129	1111	Referent			
rs3762274											
GG	161	351	1.40 (1.07 , 1.82)	1.7	0.0132	49	351	1.08 (0.71 , 1.67)	2.4	0.7102	0.2753
AG	311	880	0.96 (0.77 , 1.20)	1.5	0.7325	93	880	0.87 (0.61 , 1.25)	2.0	0.4552	0.6132
AA	205	545	Referent			58	545	Referent			
rs11801408											
TT	48	163	0.77 (0.54 , 1.09)	2.0	0.1420	16	163	0.65 (0.37 , 1.14)	3.0	0.1353	0.6105
CC+CT	631	1613	Referent			184	1613	Referent			
rs17127826											
GG	8	41	0.50 (0.23 , 1.10)	4.9	0.0856	6	41	0.78 (0.32 , 1.93)	6.0	0.5965	0.4168
AA+AG	671	1735	Referent			194	1735	Referent			
TNF											

	Luminal A					Basal-like					Luminal A vs. basal-like[3]
	Case	Control	OR (95% CI)[1]	CLR[2]	P-value	Case	Control	OR (95% CI)[1]	CLR[2]	P-value	P-value
rs1799964											
CC	24	75	0.79 (0.48 , 1.30)	2.7	0.3506	5	75	0.60 (0.23 , 1.53)	6.6	0.2852	0.6064
CT	223	527	1.14 (0.93 , 1.39)	1.5	0.1963	52	527	0.83 (0.59 , 1.17)	2.0	0.2884	0.0942
TT	432	1174	Referent			143	1174	Referent			

1 - odds ratio, 95% confidence interval, adjusted for age, self-identified race, African ancestry, offset term

2 - confidence limit ratio, upper 95% confidence limit divided by lower 95% confidence limit

3 - H_0: β (luminal A) = β(basal-like)

Table 3.4a Joint effects for SNPs with multiplicative interaction between genotype and WHR[1], comparing all cases and controls

		WHR < 0.77	WHR ≥ 0.77		
		Case/control OR (95% CI)[2]	Case/control OR (95% CI)[2]	LRT[3] P-value	S (90% CI)[4]
		<0.77	>=0.77		
LEPR					
rs1137100	GG	50/24	70/53	0.06055	
		2.32 (1.36, 3.95)	1.37 (0.91, 2.07)		0.33 (0.12, 0.96)
	AG	204/220	456/390		
		0.92 (0.71, 1.18)	1.27 (1.02, 1.57)		1.32 (0.56, 3.12)
	AA	356/352	805/716		
		Referent	1.23 (1.01, 1.49)		
rs12042877	TT	57/31	115/81	0.08941	
		2.07 (1.29, 3.32)	1.61 (1.16, 2.22)		0.55 (0.25, 1.18)
	CC+CT	552/565	1218/1078		
		Referent	1.27 (1.09, 1.49)		

1 - waist-hip ratio
2 - odds ratio, 95% confidence interval adjusted for BMI, age, self-identified race, African ancestry, and offset term
3 - likelihood ratio test
4 - synergy index, 90% confidence interval

Table 3.4b Joint effects for SNPs with multiplicative interaction between genotype and WHR[1], comparing basal-like cases and controls

		WHR < 0.77	WHR ≥ 0.77		
		Case/control OR (95% CI)[2]	Case/control OR (95% CI)[2]	LRT[3] P-value	S (90% CI)[4]
		<0.77	>=0.77		
IL6					
rs2069824	TT	32/492	127/948	0.0756	
		Referent	2.25 (1.45, 3.48)		0.45 (0.19, 1.06)
	CC+CT	11/103	25/207		
		1.96 (0.92, 4.15)	1.92 (1.07, 3.47)		

1 - waist-hip ratio
2 - odds ratio, 95% confidence interval adjusted for BMI, age, self-identified race, African ancestry, and offset term
3 - likelihood ratio test
4 - synergy index, 90% confidence interval

Table 3.4c Joint effects for SNPs with multiplicative interaction between genotype and WHR[1], comparing luminal A cases and controls

		WHR < 0.77	WHR ≥ 0.77		
		Case/control OR (95% CI)[2]	Case/control OR (95% CI)[2]	LRT[3] P-value	S (90% CI)[4]
ADIPOQ					
rs16861194	AA	155/481	363/834	0.0454	
		Referent	1.28 (1.00, 1.63)		
	AG	38/95	103/281		
		1.08 (0.69, 1.68)	1.19 (0.86, 1.65)		0.64 (0.42, 0.97)
	GG	1/20	12/43		
		---[5]	0.92 (0.45, 1.89)		---[5]
LEPR					
rs9436746	AA+AC	134/371	340/839	0.0895	
		Referent	1.16 (0.89 , 1.50)		
	CC	60/223	137/318		
		0.71 (0.49, 1.02)	1.13 (0.84, 1.53)		---[5]
rs6588147	AA	88/315	265/657	0.0508	
		Referent	1.54 (1.14, 2.08)		
	AG	78/234	176/423		
		1.18 (0.81, 1.70)	1.46 (1.07, 2.00)		0.67 (0.38, 1.16)
	GG	28/47	37/79		
		2.44 (1.40, 4.24)	1.55 (0.95, 2.52)		0.37 (0.14, 0.93)
rs6704167	AA	74/224	212/509	0.0561	
		Referent	1.30 (0.93, 1.81)		
	AT	95/254	200/511		
		1.09 (0.75, 1.58)	1.18 (0.85, 1.63)		0.56 (0.24, 1.31)
	TT	25/118	66/137		
		0.63 (0.37, 1.06)	1.38 (0.91, 2.10)		---[5]
rs12042877	CC+CT	175/565	438/1078	0.0762	
		Referent	1.35 (1.08, 1.69)		
	TT	19/31	40/81		
		2.42 (1.30, 4.53)	1.73 (1.11, 2.69)		0.43 (0.16, 1.16)

1 - waist-hip ratio
2 - odds ratio, 95% confidence interval adjusted for BMI, age, self-identified race, African ancestry, and offset term
3 - likelihood ratio test of multiplicative interaction term
4 - synergy index, 90% confidence interval
5 - parameters could not be estimated

Table 3.5 Association between adipocytokine haplotypes and breast cancer

Haplotype		All Cases			Luminal A			Basal-like		
	No. Copies	OR (95% CI)[1]	P-Value	CLR[2]	OR (95% CI) [1]	P-Value	CLR[2]	OR (95% CI) [1]	P-Value	CLR[2]
IL6										
1: rs2069827, rs1800796										
1a: G-C	0	Referent			Referent			Referent		
	1	1.07 (0.88, 1.30)	0.4930	1.5	1.32 (1.02, 1.72)	0.0382	1.7	0.86 (0.53, 1.38)	0.5239	2.6
	2	2.05 (1.13, 3.72)	0.0189	3.3	1.88 (0.67, 5.24)	0.2284	7.8	---[3]		
1b: G-G	0	Referent			Referent			Referent		
	1	0.85 (0.60, 1.20)	0.3599	2.0	0.81 (0.48, 1.37)	0.4399	2.8	---[3]		
	2	0.82 (0.56, 1.20)	0.3004	2.1	0.64 (0.37, 1.09)	0.1021	2.9	---[3]		
2: rs1800796, rs1800795										
C-G	0	Referent			Referent			Referent		
	1	1.06 (0.87, 1.29)	0.5672	1.5	1.30 (1.00, 1.70)	0.0465	1.7	0.86 (0.53, 1.39)	0.5304	2.6
	2	1.96 (1.08, 3.57)	0.0275	3.3	1.81 (0.65, 5.05)	0.2576	7.8	---[3]		
LEP										
3: rs12706832, rs10244329, rs11763517, rs7795794										
A-T-T-G	0	Referent			Referent			Referent		
	1	0.82 (0.71, 0.95)	0.0088	1.3	1.31 (1.00, 1.73)	0.0514	1.7	0.89 (0.64, 1.26)	0.5176	2.0
	2	2.14 (1.57, 2.93)	< 0.0001	1.9	2.27 (1.32, 3.89)	0.0029	2.9	1.14 (0.43, 2.99)	0.7910	6.9
4: rs11760956, rs10954173, rs3793162, rs3828942, rs17151922										
4a: G-G-A-G-G	0	Referent			Referent			Referent		
	1	0.79 (0.64, 0.96)	0.0202	1.5	0.79 (0.57, 1.09)	0.1501	1.9	0.65 (0.40, 1.07)	0.0909	2.7
	2	2.69 (1.62, 4.46)	0.0001	2.8	1.78 (0.75, 4.25)	0.1945	5.7	2.51 (0.78, 8.06)	0.1226	10.3
4b: G-G-G-G-T	0	Referent			Referent			Referent		
	1	0.90 (0.77, 1.05)	0.1665	1.4	1.27 (0.95, 1.68)	0.1012	1.8	0.94 (0.66, 1.33)	0.7220	2.0
	2	2.32 (1.62, 3.33)	< 0.0001	2.1	2.09 (1.09, 4.02)	0.0265	3.7	1.11 (0.34, 3.58)	0.8670	10.5
5: rs17151922, rs10954174, rs11761556										
T-G-C	0	Referent			Referent			Referent		
	1	0.92 (0.77, 1.11)	0.3876	1.4	1.22 (0.92, 1.63)	0.1694	1.8	0.63 (0.39, 1.00)	0.0504	2.5
	2	2.00 (1.26, 3.17)	0.0031	2.5	3.21 (1.49, 6.90)	0.0029	4.6	---[3]		

Haplotype		All Cases			Luminal A			Basal-like		
	No. Copies	OR (95% CI)[1]	P-Value	CLR[2]	OR (95% CI) [1]	P-Value	CLR[2]	OR (95% CI) [1]	P-Value	CLR[2]
LEPR										
6: rs17412175, rs9436746, rs9436748										
6a: A-C-T	0	Referent			Referent			Referent		
	1	0.79 (0.69, 0.91)	0.0010	1.3	0.79 (0.65, 0.97)	0.0261	1.5	0.83 (0.58, 1.20)	0.3180	2.1
	2	0.97 (0.78, 1.22)	0.8214	1.6	0.93 (0.68, 1.27)	0.6460	1.9	0.61 (0.32, 1.19)	0.1507	3.8
6b: T-A-G	0	Referent			Referent			Referent		
	1	0.99 (0.87, 1.12)	0.8254	1.3	0.95 (0.78, 1.15)	0.6107	1.5	1.52 (1.03, 2.23)	0.0336	2.2
	2	1.15 (0.95, 1.39)	0.1388	1.5	1.24 (0.96, 1.61)	0.0948	1.7	1.63 (1.03, 2.58)	0.0352	2.5
7: rs10749754, rs12042877, rs12564626										
7a: A-T-A	0									
	1	1.03 (0.90, 1.17)	0.6941	1.3	1.11 (0.92, 1.34)	0.2758	1.5	1.04 (0.75, 1.43)	0.8250	1.9
	2	1.37 (1.06, 1.76)	0.0158	1.7	1.52 (1.07, 2.16)	0.0196	2.0	1.15 (0.57, 2.33)	0.6969	4.1
7b: G-C-G	0									
	1	0.87 (0.77, 0.99)	0.0342	1.3	0.82 (0.67, 1.00)	0.0458	1.5	0.86 (0.61, 1.20)	0.3707	2.0
	2	0.88 (0.73, 1.06)	0.1787	1.4	0.74 (0.57, 0.95)	0.0204	1.7	0.87 (0.57, 1.33)	0.5230	2.4
8: rs12405556, rs3762274										
8a: G-A		Referent								
		0.84 (0.74, 0.96)	0.0127	1.3	0.74 (0.61, 0.91)	0.0048	1.5	0.92 (0.65, 1.31)	0.6440	2.0
		0.88 (0.73, 1.06)	0.1648	1.5	0.71 (0.55, 0.92)	0.0092	1.7	0.87 (0.56, 1.35)	0.5355	2.4
8b: T-A	0	Referent			Referent			Referent		
	1	1.24 (0.96, 1.59)	0.0938	1.6	1.19 (0.85, 1.66)	0.3092	2.0	---[3]		
	2	2.59 (1.31, 5.11)	0.0060	3.9	2.39 (0.83, 6.85)	0.1053	8.2	---[3]		
8c: T-G	0	Referent			Referent			Referent		
	1	0.96 (0.83, 1.11)	0.5535	1.3	1.05 (0.86, 1.28)	0.6479	1.5	0.86 (0.60, 1.23)	0.4141	2.1
	2	1.32 (0.96, 1.80)	0.0832	1.9	1.60 (1.05, 2.46)	0.0302	2.4	0.96 (0.38, 2.41)	0.9336	6.3
9: rs2025804, rs7518849, rs10158579										
9a: T-C-T	0	Referent			Referent			Referent		

Haplotype		All Cases			Luminal A			Basal-like		
	No. Copies	OR (95% CI)[1]	P-Value	CLR[2]	OR (95% CI) [1]	P-Value	CLR[2]	OR (95% CI) [1]	P-Value	CLR[2]
	1	0.90 (0.76, 1.06)	0.2163	1.4	0.74 (0.58, 0.95)	0.0172	1.6	1.25 (0.88, 1.78)	0.2152	2.0
	2	1.43 (0.83, 2.20)	0.1011	2.4	1.57 (0.83, 3.00)	0.1680	3.6	1.99 (0.78, 5.04)	0.1485	6.4
9b: T-T-T	0	Referent			Referent			Referent		
	1	0.91 (0.81, 1.03)	0.1331	1.3	0.82 (0.68, 0.99)	0.0349	1.4	0.79 (0.58, 1.08)	0.1435	1.9
	2	0.91 (0.75, 1.11)	0.3628	1.5	1.00 (0.77, 1.29)	0.9748	1.7	0.65 (0.40, 1.07)	0.0927	2.7

1 - odds ratio, 95% confidence interval, adjusted for age, self-identified race, African ancestry, BMI, and offset term

2 - confidence limit ratio, upper 95% confidence limit divided by lower 95% confidence limit

3 - parameters not estimated due to small sample size

Table 3.6 Association between adipocytokine haplotypes and breast cancer overall, stratified by WHR

Haplotype		WHR < 0.77			WHR ≥ 0.77			
	No. copies	OR (95% CI)[1]	P-Value	CLR[2]	OR (95% CI)[1]	P-Value	CLR[2]	LRT[3] P-value
IL6 1a: G-C[4]	0	Referent			Referent			0.0822
	1	1.65 (1.11, 2.43)	0.0124	2.2	0.91 (0.73, 1.15)	0.4477	1.6	
	2	3.80 (1.05, 13.75)	0.0415	13.1	1.67 (0.85, 3.28)	0.1403	3.9	
LEPR 8b: T-A[5]	0	Referent			Referent			0.0997
	1	1.13 (0.75, 1.71)	0.5440	2.3	1.29 (0.94, 1.77)	0.1160	1.9	
	2	4.32 (1.72, 10.83)	0.0018	6.3	1.45 (0.48, 4.32)	0.5092	8.9	

1 - odds ratio and 95% confidence interval, adjusted for age, self-identified race, African ancestry, BMI, and offset term
2 - confidence limit ratio, upper 95% confidence limit divided by lower 95% confidence limit
3 - likelihood ratio test
4 - rs2069827and rs1800796
5 - rs12405556 and rs3762274

Table 3.7 Adipocytokine functional SNP odds ratios and 95% confidence intervals for CBCS and previously published studies

	CBCS all cases[1]	CBCS luminal A[1]	CBCS basal-like[1]	Kaklamani et al. (86)			
ADIPOQ							
rs2241766							
+45 T/G							
GG+GT	1.00 (0.84, 1.21)	0.99 (0.77, 1.27)	1.18 (0.78, 1.77)	0.64 (0.49, 0.83)			
TT	Referent	Referent	Referent	Referent			
rs1501299							
+276 G/T							
GG	1.13 (0.89, 1.45)	1.22 (0.86, 1.73)	0.84 (0.52, 1.37)	1.80 (1.14, 2.85)			
GT	1.10 (0.86, 1.41)	1.14 (0.81, 1.63)	0.72 (0.44, 1.19)	1.59 (1.03, 2.48)			
TT	Referent	Referent	Referent	Referent			
	CBCS all cases[1]	CBCS luminal A[1]	CBCS basal-like[1]	Slattery et al.(19), premenopausal[2]	Slattery et al.(19), postmenopausal/ no HRT[2]	Slattery et al.(19), postmenopausal/ HRT[2]	Hefler et al.(18)
IL6							
rs1800795							
-174 G/C							
CC	0.99 (0.78, 1.26)	1.06 (0.77, 1.46)	1.23 (0.70, 2.17)				2.0 (1.1, 3.6)
CG	1.00 (0.85, 1.18)	1.00 (0.80, 1.26)	1.03 (0.71, 1.53)				1.5 (1.0, 2.3)
CC+CG	1.00 (0.85, 1.17)	1.02 (0.82, 1.26)	1.08 (0.75, 1.56)	0.76 (0.56, 1.02)	0.68 (0.47, 0.99)	0.78 (0.61, 1.00)	
GG	Referent	Referent	Referent	Referent	Referent	Referent	Referent
rs1800796							
-572 G/C							
CC+CG	1.18 (0.96, 1.46)	1.44 (1.09, 1.90)	0.86 (0.53, 1.40)	1.53 (0.99, 2.37)	1.33 (0.78, 2.26)	1.12 (0.77, 1.64)	
GG	Referent	Referent	Referent	Referent	Referent	Referent	
rs2069832							
Intron 2 G/A							
AA+AG	1.01 (0.86, 1.18)	1.03 (0.83, 1.27)	1.07 (0.74, 1.55)	0.76 (0.56, 1.03)	0.70 (0.48, 1.02)	0.77 (0.60, 1.00)	
GG	Referent	Referent	Referent	Referent	Referent	Referent	

	CBCS all cases[1]	CBCS luminal A[1]	CBCS basal-like[1]	Snoussi et al. (20)	Woo et al. (21)	Han et al. (24)	Cleveland et al. (25)
LEPR							
rs1137101							
Q223R							
GG	1.10 (0.91, 1.34)	1.35 (1.04, 1.75)	1.10 (0.72, 1.69)	2.26 (1.31, 3.90)	0.59 (0.19, 1.81)	Referent	1.04 (0.81, 1.34)
AG	0.99 (0.84, 1.16)	1.07 (0.86, 1.35)	1.01 (0.70, 1.47)	1.68 (1.12, 2.50)	Referent[3]	1.30 (1.03, 2.70)	1.00 (0.78, 1.27)
AA	Referent	Referent	Referent	Referent		7.14 (1.92, 25.60)	Referent
rs1137100							
K109R							
GG	1.45 (1.06, 1.97)	1.64 (1.10, 2.45)	1.02 (0.48, 2.21)		1.08 (0.40, 2.93)		
AG	1.00 (0.87, 1.16)	1.10 (0.91, 1.34)	1.04 (0.75, 1.42)		Referent[3]		
AA	Referent	Referent	Referent				
rs8179183							
K656N							
CC	0.70 (0.48, 1.03)	0.70 (0.41, 1.20)	0.78 (0.34, 1.77)				
CG	0.90 (0.77, 1.04)	1.05 (0.86, 1.28)	0.84 (0.60, 1.18)		0.63 (0.14, 2.81)		
GG	Referent	Referent	Referent		Referent		
	CBCS all cases[1]	CBCS luminal A[1]	CBCS basal-like[1]	Gaudet et al.(22)	Kohaar et al. (23)		
TNFA							
rs1800630							
-863 C/A							
AA	0.65 (0.40, 1.05)	0.76 (0.41, 1.44)	0.67 (0.20, 2.21)	0.95 (0.64, 1.42)			
AC	1.07 (0.91, 1.26)	1.12 (0.90, 1.38)	0.84 (0.58, 1.22)	0.89 (0.76, 1.05)			
AA+AC	1.03 (0.88, 1.20)	1.08 (0.88, 1.33)	0.82 (0.57, 1.19)		0.86 (0.41, 1.80)		
CC	Referent	Referent	Referent	Referent	Referent		

1 - odds ratio, 95% confidence interval adjusted for age, self-identified race, African ancestry, offset term

2 - white, non-Hispanic women only

3 - no subjects with AA genotype in study

Table 3.8 Minor allele frequencies in adipocytokine functional SNPs

	Study	Subjects			
			rs822396 (-3971 A>G) minor allele: G	rs2241766 (+45 T>G) minor allele: G	rs1501299 (+276 G>T) minor allele: T
ADIPOQ	CBCS	AA and white female population-based controls from North Carolina	AA: 0.22 non-AA: 0. 22	AA: 0.40 non-AA: 0.11	AA: 0.38 non-AA: 0.26
	International HapMap Project		CEU: 0.16 ASW: 0.15 YRI: 0.24	CEU: -- ASW: -- YRI: --	CEU: 0.28 ASW: 0.32 YRI: 0.38
	Bouatia-Naji et al. (102)	Unrelated controls from Lille and Paris, France		0.13	0.26
			rs1800796 (-572 G>C) minor allele: C	rs1800795 (-174 G>C) minor allele: C	rs2069832 (Intron 2 G>A) minor allele: A
IL6	CBCS	AA and white female population-based controls from North Carolina	AA: 0.09 non-AA: 0.0 4	AA: 0.07 non-AA: 0.45	AA: 0.07 non-AA: 0.44
	International HapMap Project		CEU: 0.04 ASW: -- YRI: 0.09	CEU: 0.54 ASW: 0.10 YRI: 0	CEU: 0.54 ASW: 0.10 YRI: --
	Seattle SNPs Program for Genomic Applications	DNA samples from the Coriell Cell Repository (PGA-UW-FHCRC)	AA: 0.05 white: 0	AA: 0 white: 0.50	AA: 0.02 white: 0.50
	Fishman et al. (103)	White controls: healthy men and women recruited from general practice in north London Afro-Caribbean controls: A random sample drawn from Family Practitioner Committee population registers in northwest London		AC: 0.05 white: 0.40	
	Terry et al. (97)	182 unrelated		0.41	

		individuals from a study of hypertension			
	Hefler et al. (18)	Study controls - women of Austrian or German descent with no history of breast cancer who attended study hospital outpatient departments in Halle-Wittenberg, Germany and Vienna, Austria		0.37	
	Balasubramanian et al. (104)	Study controls: recruited from Sheffield Breast Screening Service. White women only		0.42	
	Slattery et al. (19)	Controls from 4-Corners Study: Population-based female controls living in Arizona, Colorado, New Mexico, Utah.	Non-Hispanic white: 0.06	Non-Hispanic white: 0.44	Non-Hispanic white: 0.44
			rs1137101 (Q223R) minor allele: G(R)	rs1137100 (K109R) minor allele: G (R)	
LEPR	CBCS	AA and white female population-based controls from North Carolina	AA: 0.56 non-AA: 0.43	AA: 0.19 non-AA: 0.25	
	International HapMap Project		CEU: 0.47 ASW: 0.62 YRI: 0.60	CEU: 0.29 ASW: 0.21 YRI: 0.17	
	Seattle SNPs Program for Genomic Applications	Individuals of African American and European descent from the Coriell Cell Repository (AFD_AFR, AFD_EUR)	AA: 0.54 white: 0.54		
	Chiu et al. (105)	36 female and 31 male	0.47		

		healthy white controls from west Los Angeles who were normotensive and glucose tolerant			
	Snoussi et al. (20)	Blood donor study controls with no family history of breast cancer from Sousse Hospital, Tunisia	0.34		
	Cleveland et al. (25)	Controls from a case-control study in Long Island, New York	0.58		
	Pechlivanis et al. (106)	Male and female controls from a colorectal cancer case-control study in the Czech Republic	0.46		
	Wauters et al. (85)	Overweight or obese white women, without diabetes, visiting an obesity clinic in Antwerp, Belgium		0.28	
			rs1800630 (-863 C>A) minor allele: A		
TNFA	CBCS	AA and white female population-based controls from North Carolina	AA: 0.12 non-AA: 0.15		
	International HapMap Project		CEU: 0.15 ASW: 0.10 YRI: 0.10		
	Seattle SNPs Program for Genomic Applications	DNA samples from the Coriell Cell Repository (PGA-UW-FHCRC)	AA: 0.08 European: 0.21		
	Skoog et al. (98)	254 healthy men of Swedish origin, randomly recruited using	0.17		

		a population registry			
	Gaudet et al. (22)	White female controls living in Wisconsin, Massachusetts, or New Hampshire, ages 20-74, selected through drivers license or Medicare lists	0.17		

Abbreviations: AA-African American, AC – Afro-Caribbean.

3.8 Figures

Figure 3.1 Maximum likelihood African ancestry stratified by self-identified race, with median individual ancestry estimates

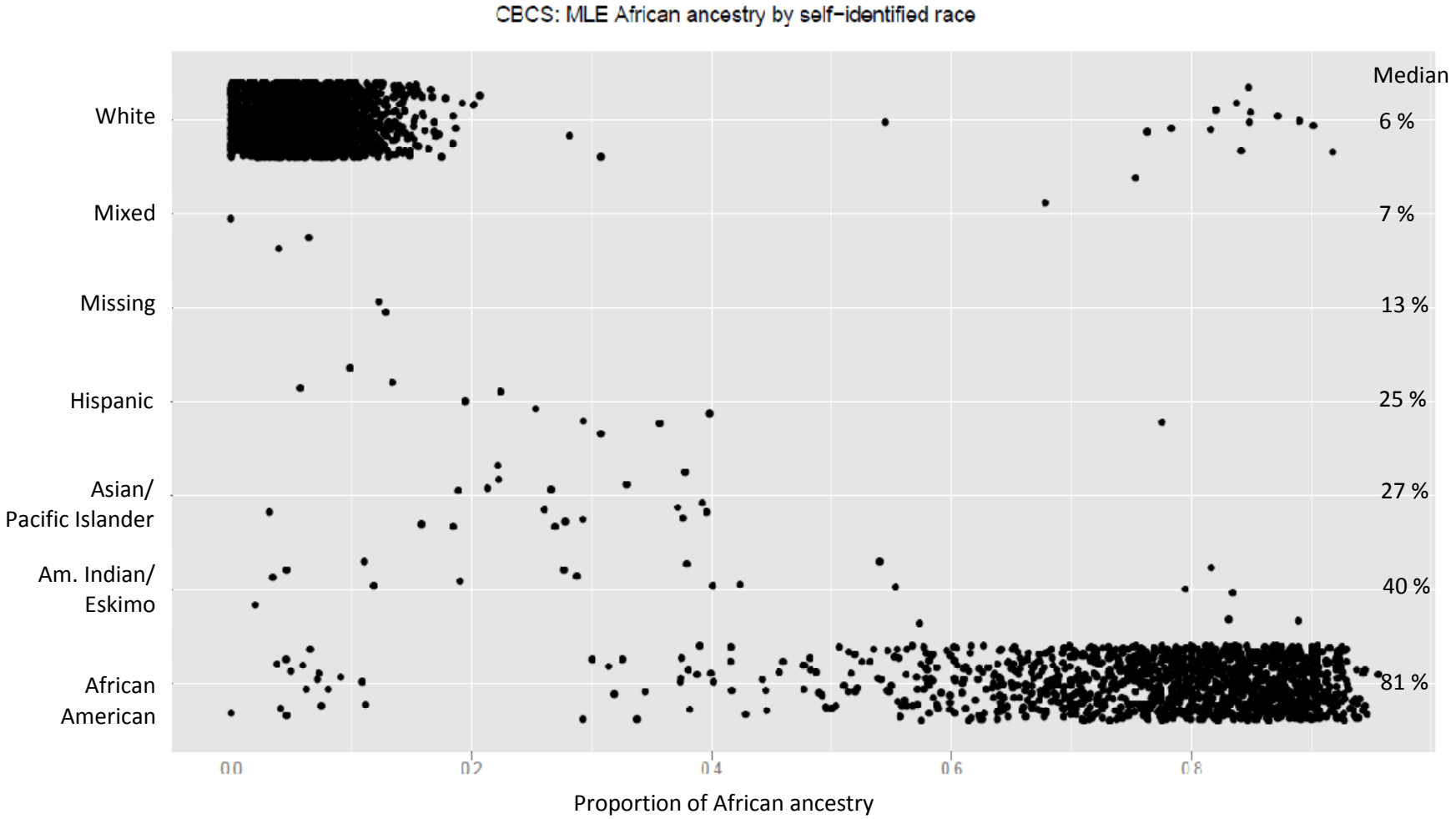
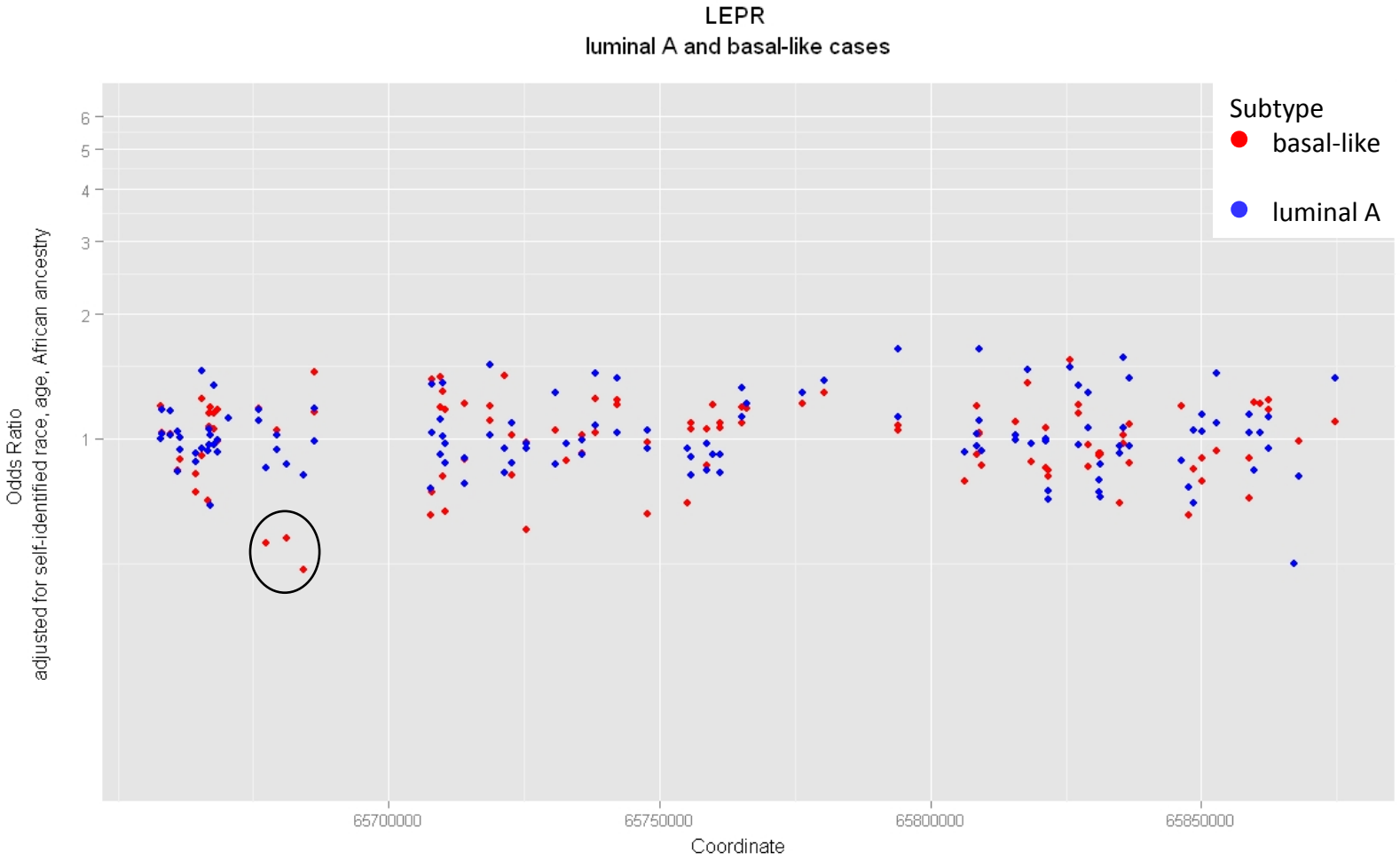


Figure 3.2 LEPR associations by molecular subtype
a.



b. Pairwise r^2 in African American controls

	rs17097182	rs17412175	rs970467	rs9436746	rs9436748
rs17097182	1				
rs17412175	0.04	1			
rs970467	0.25	0.15	1		
rs9436746	0.08	0.32	0.31	1	
rs9436748	0.12	0.03	0.44	0.01	1

c. Pairwise r^2 in African American cases

	rs17097182	rs17412175	rs970467	rs9436746	rs9436748
rs17097182	1				
rs17412175	0.03	1			
rs970467	0.10	0.03	1		
rs9436746	0.20	0.17	0.45	1	
rs9436748	0.06	0.29	0.04	0.34	1

d. Pairwise r^2 in White controls

	rs17097182	rs17412175	rs970467	rs9436746	rs9436748
rs17097182	1				
rs17412175	0.03	1			
rs970467	0.05	0.53	1		
rs9436746	0.03	0.86	0.50	1	
rs9436748	0.00	0.12	0.03	0.10	1

e. Pairwise r^2 in White cases

	rs17097182	rs17412175	rs970467	rs9436746	rs9436748
rs17097182	1				
rs17412175	0.11	1			
rs970467	0.52	0.07	1		
rs9436746	0.87	0.09	0.49	1	
rs9436748	0.47	0.06	0.65	0.45	1

Figure 3.2 Legend

a. Odds ratios for the association between LEPR SNPs and basal-like (red) and luminal A (blue) breast cancer. Odds ratios are adjusted for age, self-identified race, and African ancestry. SNPs rs17412175, rs9436746, and rs9436748 (circled above) show a strong inverse association with basal-like breast cancer. There are two additional SNPs in the region (rs17097182 and rs970467) that are inversely associated with basal-like breast cancer but the estimates were imprecise and are not shown. The 3 SNPs in the plot above are located in intron 2 and span 6786 bp.

b-e. R^2 correlation among LEPR basal-like breast cancer-associated SNPs in (b) African American controls, (c) African American cases, (d) white controls, and (e) white controls.

4. Results Paper 2: Association between CYP19A1, ESR1, HSD3B1, HSD17B2, PGR, and SHBG SNPs and breast cancer: an exploration of tumor subtype-specific effects

4.1 Abstract

Introduction: Previous analyses in the Carolina Breast Cancer Study (CBCS) showed that some reproductive risk factors were differentially associated with basal-like and luminal A breast cancer. This analysis investigated the association between breast cancer and SNPs in genes involved in estrogen and progesterone synthesis and signaling (CYP19A1, ESR1, HSD3B1, HSD17B2, PGR, SHBG). Associations were analyzed with respect to breast cancer overall, and basal-like and luminal A intrinsic molecular subtypes.

Methods: Eligible cases were women aged 20-74, living in North Carolina, who were diagnosed with primary breast cancer between 1993 and 2001. ER, PR, HER2, CK5/6, and EGFR immunohistochemistry was used to determine breast tumor intrinsic molecular subtypes. Controls were cancer-free women living in the same geographic area, and were frequency-matched to cases by age and race (African American/non-African American). 195 candidate gene SNPs were genotyped using the Illumina GoldenGate assay. 1776 controls and 1972 cases (200 basal-like, 679 luminal A) were included in the analysis. Odds ratios and 95% confidence intervals were estimated using unconditional logistic regression for the association between genotypes and all breast cancer cases, basal-like breast cancer, and luminal A breast cancer. Haplotype frequencies and odds ratios were estimated using

HAPSTAT. The likelihood ratio test was used to evaluate multiplicative interaction between genotypes and parity/lactation status. Ancestry informative markers were used to control for population stratification in genotype and haplotype analyses.

Results: The strongest associations for breast cancer overall were in ESR1 (rs2207232, rs6914211, rs985181, rs8052451), HSD17B2 (rs8052451), and CYP19A1 (haplotype 1). Genotypes in ESR1, HSD3B1, HSD17B2, and SHBG were strongly associated with both luminal A and basal-like subtypes. Genotypes in PGR were only strongly associated with the basal-like subtype. Haplotypes in CYP19A1, ESR1, and PGR were associated with breast cancer, and in most cases associations did not differ by subtype.

Conclusion: SNPs in genes related to estrogen and progesterone synthesis and signaling were associated with breast cancer overall and by intrinsic molecular subtype. In some cases, haplotype associations were much stronger than single SNP associations. These results support the hypothesis that variation in genes related to estrogen and progesterone synthesis and signaling are important to different subtypes of breast cancers, including those that do not express ER and/or PR at the time of diagnosis.

Keywords: breast cancer, single nucleotide polymorphism, haplotype, estrogen receptor, progesterone receptor, 17-beta hydroxysteroid dehydrogenase type II, 3-beta hydroxysteroid dehydrogenase type I, sex hormone-binding globulin, cytochrome P450 family 19 type A1, basal-like, luminal A

4.2 Introduction

It has been proposed that common genetic variants contribute to familial breast cancer cases unrelated to known high penetrance mutations, and that because such polymorphisms are common they may play a role in sporadic breast cancer as well (1-3). Genes related to estrogen and progesterone activity are likely candidates for harboring such variants, due to the central role that estrogen, progesterone, and their respective receptors play in breast cancer. In addition to promoting normal growth patterns during puberty and pregnancy (4-6), estrogen and progesterone can also have carcinogenic effects in the breast. When estrogen binds to the estrogen receptor (ER) the ligand-receptor unit translocates to the nucleus where it can act as a transcription factor to genes associated with cell proliferation and survival, such as growth factors, transcription factors, tumor suppressors, pro-apoptotic genes, and growth inhibitors (7). Independent of its receptor, estrogen may contribute to carcinogenesis through the action of its metabolites. Estrogen metabolites can bind to DNA forming adducts; high rates of DNA repair after adduct removal may introduce mutations that initiate carcinogenesis (8).

Exposure to estrogen and progesterone increases proliferation of breast epithelial cells; the higher rate of cell division increases the chance of oncogene activation or tumor suppressor inactivation due to a replication error (9). Furthermore, removing the main source or estrogen and progesterone production via oopherectomy greatly reduces breast cancer risk in high risk premenopausal women (10), underlining the importance of these two hormones in breast tumor formation. The present study focuses on variation in genes that affect estrogen and progesterone production, bioavailability, and signaling.

The estrogen bio-synthesis pathway includes enzymes that catalyze forward and

reverse reactions leading to hormone synthesis from cholesterol. Cytochrome P450 family 19 subfamily A polypeptide 1 (CYP19A1) encodes the enzyme aromatase. Aromatase converts androgens into estrogens, including the conversion of androstenedione to estrone and testosterone to estradiol (11). SNPs in CYP19A1 have been associated with serum estrogen levels in postmenopausal women (12). 17-beta hydroxysteroid dehydrogenase type II (HSD17B2) oxidizes active sex steroids into their inactive precursor forms, including the conversion of estradiol into estrone, a less potent form of estrogen (13). Sex hormone-binding globulin (SHBG) binds estrogen in the bloodstream, controlling its availability to bind to ER. Experimental data supports the hypothesis that SHBG is instrumental in controlling the proliferative and anti-apoptotic effects of estradiol (14, 15). Like estrogen, progesterone is synthesized from cholesterol. 3-beta hydroxysteroid dehydrogenase type I (HSD3B1) is part of the progesterone synthesis pathway and converts pregnolone to progesterone, 17-alpha hydroxypregnenolone to 17-alpha hydroxyprogesterone, and dehydroepiandrosterone to androstenedione (11, 16).

Perou et al. (17) and Sorlie et al. (18, 19) defined five breast cancer 'intrinsic molecular subtypes', which were characterized by hierarchical clustering of breast tumor gene expression profiles. The intrinsic molecular subtypes have been observed independently in other populations, and are associated with breast cancer survival (20, 21-23). Recently, Nordgard et al. (24) reported that associations between SNPs and gene expression levels varied by intrinsic molecular subtype, and Kristensen and Borresen-Dale (25) reported the association of the rs10046 variant homozygote with basal-like breast cancer. We hypothesize that stratification by tumor subtype will allow for the identification of unique associations between CYP19A1, ESR1, HSD3B1, HSD17B2, PGR, and SHBG and the basal-like and

luminal A breast cancer subtypes that would not be detected in a pooled case population.

Luminal A breast cancer is the most common intrinsic molecular subtype and is defined by relatively high expression of the estrogen receptor and related genes (17, 22). Thus, it is reasonable to hypothesize that polymorphisms in genes related to estrogen synthesis or function may affect the incidence of luminal A breast cancer. Basal-like breast cancer is defined in part by low estrogen receptor expression and high expression of cytokeratins 5 and 17, integrin beta 4, and laminin, but with the exception of a high prevalence of basal-like tumors among women with BRCA1 mutations, relatively little is known about the genetic variants that may predispose to this type of breast cancer (17, 19, 26-28). We have reported previously that associations for some traditional breast cancer risk factors differed by intrinsic molecular subtype in the CBCS. Parity (vs. nulliparity) was inversely associated with luminal A breast cancer regardless of breastfeeding status (29). In contrast, parity in women who did not breastfeed was positively associated with basal-like breast cancer, and parity in women who breastfed was unassociated with the basal-like subtype (29). This suggests that factors associated with parity and lactation may influence pathways leading to basal-like and luminal A breast cancer in different ways. Since the candidate gene products are involved in the production and action of hormones directly related to pregnancy, parity, and breastfeeding, we explored the possibility of effect measure modification of genotype and haplotype associations by parity and lactation history.

4.3 Methods

4.3.1 Study population

The Carolina Breast Cancer Study is a population-based case-control study of breast

cancer in North Carolina, and has been described in detail previously (30, 31). Eligible cases included all women ages 20-74 who were diagnosed with primary invasive breast cancer from 1993 to 2001, and who lived within the 24 county study area at the time of diagnosis. Women diagnosed with breast carcinoma *in situ* (CIS) were also enrolled in the study from 1996-2001. Cases were identified through the North Carolina Central Cancer Registry using rapid case ascertainment. Randomized recruitment was used to oversample African American invasive cases and invasive cases younger than 50 years old (32). All eligible CIS cases were asked to participate in the study.

Eligible controls were chosen from the same 24 county study area as cases, and included women between the ages of 20 and 74 who did not have a history of breast cancer. Controls younger than age 65 were identified through Department of Motor Vehicles records, and controls 65 years and older were identified through Health Care Financing Administration records. Controls were frequency-matched to cases by race and 5-year age groups.

Cases and controls were contacted first by mail and then by telephone. Women who agreed to participate in the study were asked to complete an in-home interview conducted by a trained nurse. During the interview, women were asked about social and demographic characteristics, family history of cancer, reproductive history, menstrual history, exogenous hormone use, alcohol use, and occupational history. Waist circumference, hip circumference, height and weight were measured by the nurse following completion of the study questionnaire. Women were also asked to provide a 30 ml blood sample. DNA was extracted from the blood sample and stored at -80^{o}C in TE buffer.

Overall response rates for invasive cases and controls were 76% and 55%,

respectively. Overall response rates for CIS cases and matched controls were 83% and 65%. A total of 2311 cases (894 African American/1417 non-African American) and 2022 controls (788 African American/1234 non-African American) were enrolled in the study.

Tumor subtype was determined by immunohistochemistry (IHC) analysis of archival tumor tissue (22, 33). Cases were asked to provide written consent for access to their medical records and formalin-fixed paraffin-embedded tumor tissue blocks. A centralized pathology review was performed to confirm each breast cancer diagnosis. For invasive breast cancers, estrogen receptor (ER) and progesterone receptor (PR) status were abstracted from the patient's medical record (80% of invasive cases). If ER or PR status was not recorded in the medical record but archival tissue was available, the assay was performed at the UNC Immunohistochemistry Core Laboratory (20% of invasive cases). IHC staining for epidermal growth factor receptor (EGFR), v-erb-b2 erythroblastic leukemia viral oncogene homolog 2 (HER2), and cytokeratin 5/6 (CK 5/6) was performed at the UNC Immunohistochemistry Core Laboratory, and has been described previously (29, 34, 35).

For CIS cases, ER, HER2, CK 5/6, and EGFR IHC was performed by the UNC Immunohistochemistry Core Laboratory, as described by Livasy et al. (33). PR expression was not used in classifying CIS cases due to the high correlation between ER and PR expression, and the need to preserve tissue (22).

Tumor tissue was available for 1845 of 2311 (80%) cases [1446 of 1808 (80%) invasive cases; 399 of 503 (80%) CIS cases]. IHC assays were completed successfully for 1424 of 2311 (62%) cases [1149 of 1808 (64%) invasive cases; 275 of 503 (55%) CIS cases]. Tumors were classified as follows: 796 luminal A (ER+ and/or PR+, HER2-); 225 basal-like (ER-, PR-, HER2-, CK5/6+ and/or EGFR+); 150 unclassified (ER-, PR-, HER2-, CK5/6-,

EGFR-); 137 luminal B (ER+ and/or PR+, HER2+); 116 HER2+/ER- (ER-, PR-, HER2+). Cases with missing subtype data were more likely to be non-African American and to have an earlier stage at diagnosis (29).

4.3.2 Genotyping

4.3.2.1 SNP selection

A two-pronged approach was used in selecting SNPs for genotyping: 1) SNPs previously investigated for an association with breast cancer or an *in vitro* functional effect were selected, based on a review of the literature; and 2) tag SNPs were selected in order to capture any potential unknown variants associated with breast cancer. The following potentially functional SNPs with a minor allele frequency (MAF) of at least 0.05 were genotyped: CYP19A1 rs10046, rs2236722, rs700519; ESR1 rs2077647, rs2234693, rs3798577, rs851982, rs9341070; HSD17B2 rs4445895 and rs8191136; PGR rs10895068; and SHBG rs1799941. No potentially functional SNPs for HSD3B1 were selected from the literature.

Most participants in the CBCS were either white or African American, and so tag SNP selection was performed twice – once using data from the HapMap CEU (European) population and once using data from the HapMap YRI (West African) population (36). The two lists of tag SNPs were combined into a single list. Tag SNPs for ESR1, HSD3B1, HSD17B2, PGR, and SHBG were selected using Haploview Tagger, where a minimum r^2 of 0.80 was used to define tags (37-39). Tag SNPs selection was restricted to SNPs with a minimum MAF of 0.10. For ESR1, PGR, HSD17B2, HSD3B1 and SHBG, tag SNPs were selected initially using the pairwise tagging method. This resulted in a greater number of tag SNPs than there was space available in the genotyping assay. Tags for ESR1, PGR, and

HSD17B2 were re-selected using aggressive tagging with 2-marker haplotypes, reducing the number of tag SNPs by 25, 7, and 8 SNPs, respectively. SNPs selected from the literature were also used as tag SNPs, using the "force include" option in Tagger.

SNPs for CYP19A1 were taken from a list published by Haiman et al. (12, 40) used to tag LD blocks in African American and white Multi-Ethnic Cohort participants.

4.3.2.2 Genotyping results and quality control

SNPs were genotyped by the UNC Mammalian Genotyping Core using the Illumina GoldenGate assay (Illumina, Inc., San Diego, CA) as part of a larger genotyping panel of 1536 SNPs. Assay intensity data and genotype cluster images were reviewed manually, and SNPs were excluded if there was low signal intensity or a lack of clear separation between genotype clusters. 163 of 1536 (11%) SNPs were excluded from the entire dataset based on cluster analysis.

Blind duplicates of 169 study samples were assayed in order to verify the reproducibility of genotype calls from the same sample. 7 SNPs had 1 genotype miscall and 2 SNPs had 2 genotype miscalls. Lab controls (Coriell CEPH trios) were also genotyped in each 96-well plate – each control was repeated between 11 and 14 times over the course of the entire assay. Out of 184 lab control samples, there were 2 instances of genotype disagreement with duplicate samples. These error rates were within our pre-specified range of acceptable values, and no SNPs were excluded from the analysis on the basis of these results.

Exact tests for deviation from Hardy Weinberg equilibrium (HWE) were conducted in controls stratified by self-identified race to determine whether genotype frequencies were distributed as expected given the allele frequencies. Deviations from HWE in controls can

indicate genotyping error, selection bias, the appearance of new mutations, failure of the source population to fulfill HWE assumptions, or random error (41, 42). HWE test statistics and P-values were calculated in Plink v1.05 using methods described by Wigginton et al. (43). Genotyping cluster images were re-reviewed for all SNPs with HWE test P-values less than 0.01 in order to confirm that HWE deviations were not due to erroneous genotype calls. All candidate gene SNPs reviewed during this process were judged to have acceptable signal intensity and genotype cluster definition, and none were excluded for poor quality.

195 of 207 (94%) candidate gene SNPs passed quality control and are included in this analysis (CYP19A1 - 24, ESR1 - 97, HSD3B1 - 7, HSD17B2 - 40, PGR - 26, SHBG – 1). 144 of 158 (91%) ancestry informative markers (AIMs) genotyped in the panel passed quality control. The overall success rate for the panel was 1373 of 1536 (89%).

Blood samples were obtained from 2045 (88%) cases and 1818 (90%) controls, and 2039 (88%) cases and 1818 (90%) controls had a sufficient amount of DNA for genotyping. Of these subjects, 64 cases and 39 controls had genotype calls for less than 95% of SNPs in the Illumina panel and were excluded from the analysis. An additional 2 cases and 3 controls were excluded due to apparent gender mismatch. One case was excluded because of suspected contamination identified though the analysis of non-blind duplicate samples.

1776 of 2022 (88%) controls and 1972 of 2311 (85%) cases were successfully genotyped, including 200 basal-like cases and 679 luminal A cases. Subjects without genotype data were more likely to be cases, recruited during phase 2 of the study, and African American. The presence or absence of genotype data did not differ by any breast cancer risk factors other than African-American race. Among cases the presence of genotype data was not associated with stage at diagnosis, lymph node status, or molecular subtype.

Among all enrolled cases, 978 of 1808 (54%) invasive cases had both genotyping and tumor subtype data and 242 of 503 (48%) CIS cases had genotyping and tumor subtype data, including 200 basal-like cases (182 invasive, 18 CIS) and 679 luminal A cases (528 invasive, 151 CIS). The distribution of intrinsic molecular subtypes did not differ between enrolled cases with and without genotyping data.

4.3.3 Variable definitions and statistical methods

4.3.3.1 Variables

Age was defined as age in years at breast cancer diagnosis for cases, and age in years at the time of sampling for study participation for controls. Self-identified race was reported during the study interview. Less than 2% of CBCS participants reported that they were Native American/Eskimo (N=19), Asian or Pacific Islander (N=18), Hispanic (N=11), or mixed race (N=5). These women and self-described white women were grouped together as non-African American. Self-identified race information was unknown for 2 participants, who were excluded from analyses that adjusted for or stratified by self-identified race.

Individual estimates of African and European ancestry were estimated from 144 AIMs using maximum likelihood estimation (44, 45). The median proportion of African ancestry was 81% among self-reported African Americans and 6% among self-reported non-African Americans. African ancestry was included in regression models as a proportion ranging from 0 to 1.Parity and lactation were reported during the study interview. Women were asked how many times they had been pregnant in their lifetime, including the current pregnancy if they were pregnant at the time of the interview. Women were then asked the duration of each pregnancy and the outcome. Parity was defined as the total number of full-term live births reported by the study subject. For each live birth reported, women were

asked whether they breastfed the baby and for how long.

In previous CBCS analyses, the association between parity/lactation and basal-like breast cancer did not differ among women with 1-2 children compared to women with 3 or more children (where nulliparous women were the referent group) (29). Likewise, the association between parity/lactation and luminal A breast cancer was the same for women with 1-2 children compared to women with 3 or more children (29). Thus, in this analysis parity and lactation was defined as a single 3-category variable: nulliparous (controls N=201; all cases N=301; luminal A N=111; basal-like N=24), parous/never breastfed (controls N=878; all cases N=983; luminal A N=317; basal-like N=124), and parous/ever breastfed (controls N=694; all cases N=686; luminal A N=251; basal-like N=52). 2 cases and 3 controls with genotyping data were missing data on parity and lactation.

4.3.3.2 Genotype associations

Genotype frequencies were calculated for each SNP, and proportions were adjusted for the sampling probabilities used to select eligible participants. Pairwise r^2 was calculated using Haploview (39).

Odds ratios (ORs) and 95% confidence intervals (95% CIs) for the association between genotypes and all breast cancer cases were estimated using unconditional binary logistic regression. ORs and 95% CIs for basal-like and luminal A breast cancer were estimated using unconditional polytomous regression models that simultaneously modeled regression parameters for all breast cancer subtypes. Statistics testing the equality of parameter estimates for basal-like and luminal A subtypes were calculated based on the asymptotic chi-square distribution of the Wald statistic. Confidence limit ratios (CLR, upper 95% confidence limit divided by lower 95% confidence limit) were calculated as a measure

of relative precision. All genotype regression models were run using SAS v9.1.3 (SAS Institute, Cary, NC).

Genotype associations were modeled using the general model with 2 degrees of freedom, unless the rare homozygote cell counts were too small. In that case, the model compared the rare homozygote and heterozygote genotypes to the common homozygote genotype. If the results of the general statistical model indicated that the underlying genetic model form may be recessive, dominant, or additive, additional analyses specific to the genetic model were conducted. Genotype associations with a relatively strong odds ratio (OR ≥ 1.5 or OR ≤ 0.67) or a P-value less than 0.05 and a precise confidence interval (CLR ≤ 5) were considered to be the best candidates for association with luminal A and/or basal-like breast cancer, and are presented in the results section.

4.3.3.3 Haplotype associations

Haplotypes were investigated in order to explore the effects of several alleles in combination. Regions for examining haplotype associations were identified through a two step process. First, a sliding window analysis (window width of 3 SNPs) was conducted for each gene using HAPSCAN, which is based on the HAPSTAT haplotype analysis program. HAPSCAN estimates a global likelihood ratio test statistic and corresponding P-value for the association of all possible haplotypes with the outcome. Sliding window analysis was performed for all cases, as well as stratified by self-identified race and for basal-like and luminal A breast cancer, versus controls. The windows with the highest $-\log_{10}$ P-values were flagged for estimation of specific haplotype ORs. Secondly, regions where consecutive SNPs were positively or inversely associated with luminal A or basal-like breast cancer were selected for haplotype analysis. Haplotype frequencies, ORs, and 95% CIs were estimated

using HAPSTAT (46, 47). HAPSTAT uses a probability distribution to estimate haplotype associations, which yields unbiased parameters estimates with appropriate variance. HAPSTAT software was modified to relax the assumption of independence between genotype and covariate distributions, and to incorporate an offset term into parameter estimation. All haplotype ORs were estimated using the general model.

All SNP and haplotype models were adjusted for the frequency-matching factors, age and self-identified race, and an offset term to account for randomized recruitment methods (32, 48). SNP and haplotype models were also adjusted for the proportion of African ancestry in order to control for residual population stratification.

4.3.3.4 Effect measure modification

In the CBCS, the ORs for the association between parity without lactation and luminal A breast cancer were 0.7 (95% CI, 0.5-0.9) for 1-2 children and 0.7 (95% CI, 0.5-0.9) for ≥ 3 children. The associations between parity without lactation and basal-like breast cancer were 1.8 (95% CI, 1.1-3.0) and 1.9 (95% CI, 1.1-3.3) for ≥ 3 children (29). The ORs for the two subtypes are on opposite sides of the null, suggesting qualitative differences between the two associations. Given the substantial heterogeneity between the subtype-specific main effects, interpretation of effect measure modification for breast cancer overall would be less meaningful and was not evaluated. Genotype-parity and lactation interaction was only examined for basal-like and luminal A breast cancers, and not breast cancer overall. Multiplicative interaction between genotypes and haplotypes and the combined parity and lactation variable was evaluated using the likelihood ratio test (LRT). LRT P-values less than 0.10 were considered consistent with a departure of the joint effects from the multiplicative null. ORs and 95% CIs for SNPs with LRT P-values less than 0.10 are presented in the

results.

4.3.3.5 Interpretation

In the analysis of genotype and haplotype associations, P-values were used in conjunction with point estimates and confidence limit ratios to evaluate the combined strength and precision of estimated associations. Strict hypothesis testing was not performed, and so P-values were not adjusted for multiple comparisons. Decisions to display stratified ORs for the evaluation of multiplicative were based on P-values. However, the intent was to display stratified ORs for the SNPs and haplotypes that showed the strongest evidence of heterogeneity. Interaction P-values were not adjusted for multiple comparisons.

4.3.3.6 Sensitivity analysis

Not all studies use the same set of markers to define 'basal-like', and in studies that have used markers similar to those used by CBCS, there has not been 100% agreement between tumors defined as basal-like using microarray expression profiles and immunohistochemistry definitions (23, 35, 49). Probabilistic sensitivity analyses were conducted to evaluate the potential effects of bias due to molecular subtype misclassification. Simulations of genotype and basal-like vs. luminal A associations were conducted, assuming non-differential misclassification of case status. Sensitivity and specificity ranges were based on previously published data (23, 35, 49). Analyses were conducted using a publicly available program and were run for 5000 simulations (50).

A simple sensitivity analysis was performed to evaluate the effect of including CIS cases in the analysis. There is evidence that CIS is an intermediate step in the progression of cells from hyperplastic to invasive disease [reviewed by (51, 52)]. However, not all CIS progress to invasive cancer, and this is some question as to whether a CIS outcome should be

treated like an invasive breast cancer outcome (53, 54).

CIS were included in this study for several reasons. First, most risk factors for DCIS are similar to risk factors for invasive breast cancer in populations being screened for breast cancer (53, 55, 56). In particular, CIS and invasive breast cancer were both reported to be associated with family history of breast cancer (55, 56).

Intrinsic molecular subtypes have been observed in pure DCIS and DCIS near invasive lesions (33, 57-59). Like invasive tumors, most basal-like DCIS showed strong expression of CK 5/6, vimentin, EGFR, and Ki-67; expression of p63 smooth muscle actin was rare (33, 57). It has not been reported whether intrinsic molecular subtypes are associated with recurrence in CIS. However the similarity between risk factors and molecular features suggest that there are a common set of factors that lead to both types of lesions. For these reasons, we chose to include subjects recruited from the CIS study in this analysis. In a simple sensitivity analysis, the exclusion of CIS cases and matched controls did not systematically change the estimated genotype ORs (data not shown).

4.4 Results

4.4.1 Genotype associations

195 SNPs were genotyped in CYP19A1, ESR1, HSD3B1, HSD17B2, PGR, and SHBG. Genotype counts and proportions are shown in Table 1. Proportions were adjusted using the sampling probabilities used to select study subjects and represent estimates of the population-level genotype prevalence.

Genotype odds ratios for breast cancer overall, the luminal A subtype, and the basal-like subtype are shown in Tables 4.2 and 4.3, selected based upon criteria mentioned in the

Methods section. SNPs that were associated with breast cancer overall are also shown in Table 4.3, and SNPs associated with basal-like or luminal A breast cancer are included in Table 4.2 for comparison of overall and subtype-specific effects. While 2 SNPs in CYP19A1 had small but precise associations with breast cancer overall, the associations were neither strong nor precise by subtype (Table 3). ESR1 SNPs rs2207232, rs6914211, and rs985191 had some of the strongest associations with breast cancer overall (Table 2). ESR1 SNPs were also associated with both luminal A and basal-like subtypes. While rs6914211 and rs985191 were also strongly associated with the luminal A and basal-like subtypes, rs2207232 was unassociated with luminal A breast cancer and the association between rs2207232 and basal-like breast cancer was positive but imprecise. Associations for rs1709183 and rs3020381 were seen only with the luminal A subtype, and rs3020401 was associated with the basal-like subtype, though the differences between the parameters for the two subtypes were not always statistically different (Table 4.3). Several of these associated ESR1 SNPs are located near each other on chromosome 6, and are shown in Figure 1.

HSD3B1 SNPs rs6205 and rs932603 had strong positive associations with breast cancer overall; by subtype these association were much stronger for luminal A compared to basal-like breast cancer. Additionally, the association between rs6428830 and basal-like breast cancer was almost twice as strong as its association with luminal A breast cancer.

Most SNPs in HSD17B2 had moderate associations with breast cancer overall. SNP rs8052451 had the strongest association for a HSD17B2 SNP and breast cancer overall (Table 4.2). When cases were stratified by subtype, the association between rs8052451 and luminal A breast cancer was even stronger than the overall association, and was one of the strongest subtype-specific associations estimated (Table 4.3). HSD17B2 SNPs were

associated with luminal A breast cancer only, basal-like breast cancer only, and both subtypes. In the case of HSD17B2 rs1364287 and rs16956326, the associations for luminal A subtype were in the opposite direction of associations for the basal-like subtype (Table 4.3). HSD17B2 rs7200696, rs8050327, and rs8191072 form a small region of SNPs inversely associated with luminal A breast cancer, and are shown in Figure 2. The 3 HSD17B2 SNPs are located within 1205 bp of each other in intron 1. SNP rs8050327 and rs8191072 were in very strong LD in African Americans and in perfect LD in whites, as is shown in Figure 2. Correlation with rs7200696 was weaker for both African Americans and non-African Americans (Figure 4.2).

PGR associations with breast cancer overall were moderate (Table 4.2). Stratified by subtype, a number of strong associations with basal-like breast cancer were apparent in SNPs that were unassociated with breast cancer overall (rs546763, rs548668, rs596223, and rs660149, Table 4.3). Figure 3a shows 2 groups of PGR SNPs with similar inverse associations with basal-like but not luminal A breast cancer. The first group, rs546763 and rs548668, are within 214 bp of each other in PGR intron 7 and are highly correlated (rs546763 and rs548668: AA cases $r^2 = 0.99$; AA controls $r^2 = 0.99$; white cases $r^2 = 1.00$; white controls $r^2 = 1.00$). The second group, rs660149, rs2124761, and rs503602, are within 2993 bp of each other in PGR intron 3 but are not in strong LD (Figure 4.3b).

Only 1 SNP in SHBG was genotyped, rs1799941 (nt -67 G/A). SHBG rs1799941 was moderately associated with breast cancer overall (Table 2). The association was stronger when cases were divided into luminal A and basal-like subtypes (Table 4.3).

Multiplicative interaction between genotypes and parity and lactation was evaluated for SNPs with a marginal association with basal-like or luminal A breast cancer. ESR1

rs11155818 and rs7759411, HSD17B2 rs2955153, rs2955159, and rs7196087, HSD3B1 rs6428830, PGR rs546763 and rs548668, and SHBG rs1799941 had LRT P-values less than 0.10 for interaction when the outcome was basal-like breast cancer, but the sample size was too limited to calculate stratified estimates that were precise enough to be compared to one another (data not shown). Genotype-outcome associations stratified by parity and lactation status for SNPs with LRT P-values less than 0.10 where the outcome was luminal A breast cancer are shown in Table 4.4. There were few nulliparous luminal A cases with variant genotypes, so in many cases ORs could not be estimated for this group. HSD17B2 rs2911418 and ESR1 rs985191 showed strong differences in effect between parous women who did and did not breastfeed. Associations were also divergent for PGR rs660149 CG genotype comparing association for nulliparous to parous/never breastfed, and parous/never breastfed to parous/ever breastfed.

A series of sensitivity analyses were conducted to determine the effect of breast cancer subtype misclassification on the estimated difference in association between the two subtypes. For some, but not all SNPs analyzed, the bias-corrected odds ratio comparing basal-like to luminal A genotypes was further from the null than the observed odds ratio, indicating that if tumor subtypes were misclassified it would decrease the chances of detecting differences in subtype associations.

4.4.2 Haplotype associations

Associations between common haplotypes and all cases, luminal A, and basal-like cases were somewhat different from single SNP associations. Whereas no individual CYP19A1 SNPs were associated with basal-like breast cancer, CYP19A1 haplotype 1 (rs749292, rs1902586, rs936306, rs2445759, rs28566535) was strongly associated with

basal-like breast cancer, though with a confidence limit ratio of 7.1 it was less precise than most other associations (Table 4.5). Other haplotypes associated with breast cancer were in ESR1 (2a, 2b: rs851984, rs851982, rs2881766; 3: rs1709183, rs9340835, rs9322335; 4: rs6914211, rs9383599, rs3020314, rs3020401, rs985191, rs6557177) and PGR (5a, 5b, 5c: rs1824128, rs660149, rs495997, rs2124761, rs11224579). The strongest associations were for CYP19A1 haplotype 1 and all breast cancer cases, haplotype 1 and basal-like breast cancer, ESR1 haplotype 2b and basal-like breast cancer, and PGR haplotype 5c and luminal A breast cancer (Table 4.5).

There was strong evidence of multiplicative interaction between ESR1 haplotype 4 and parity and lactation with respect to luminal A breast cancer (LRT P = 0.0362). Odds ratios for the association between haplotype 4 and luminal A breast cancer stratified by parity and lactation are shown in Table 4.6. Like the single SNP analysis, haplotype ORs were positively associated with the luminal A subtype in parous women. In nulliparous women, 1 copy of HAP5 was not associated with luminal A breast cancer, and the OR for 2 copies was too imprecise to draw any conclusions.

4.5 Discussion

We investigated whether SNPs in CYP19A1, ESR1, HSD3B1, HSD17B2, PGR, and SHBG were associated with breast cancer. SNPs in most of these genes have been investigated by others, but this analysis using data from the Carolina Breast Cancer Study contributes to the existing literature in the following ways. First, a large proportion of the CBCS population is African American, allowing for the inclusion and potential discovery of SNPs that are relevant to disease risk in this population. Secondly, CBCS cases have been

characterized by IHC-determined intrinsic molecular subtype. Associations were estimated for SNPs and the basal-like and luminal A subtypes in addition to the association for breast cancer cases overall. An increasing number of analyses of SNPs and breast cancer have shown that there are relevant subgroup effects (24, 25, 60-64). Most of these analyses have stratified cases on menopausal status, ER/PR status, and other prognostic factors, except for Nordgard et al. (24) and Kristensen et al.(25), which stratified by intrinsic molecular subtype.

Nordgard et al. (24) examined associations between TNRC9, LSP1, FGFR2, MAP3K1, and H19 gene expression, SNP genotype (SNPs initially identified by GWAS) and breast cancer intrinsic molecular subtype. Gene expression differed by subtype for all five genes, and genotype distribution differed by subtype for TNRC9. These results suggest that there are SNP associations specific to breast tumor types. It follows that a more refined tumor classification, such as the intrinsic molecular subtypes, may improve our ability to detect SNPs associated with specific mechanisms of breast cancer risk.

Tag SNPs in HSD17B2 (rs2955153, rs7196087), HSD3B1 (rs6428830), PGR (rs503602, rs546763, rs548668, rs596223, rs660149), and potentially functional SNP SHBG nt -67 G/A (rs1799941) were strongly associated with luminal A and/or basal-like breast cancer but were not strongly associated with breast cancer overall, supporting the hypothesis that potentially important SNPs with strong associations may be missed if cases are not stratified based on subtype. The most interesting of these associations were those in PGR where all 5 SNPs were strongly associated with the basal-like subtype, but associations with luminal A were close to the null. Progesterone receptor signaling is involved in pregnancy-related breast development and lactation (65). Previously, the CBCS has estimated that the risk of basal-like breast cancer among parous women who did not breastfeed was almost

twice that of nulliparous women (29). Other studies have reported a similar association for parity and basal-like breast cancer, but based on a small number of basal-like cases (66).

Likelihood ratio test P-values for multiplicative genotype-parity/lactation interaction were less than 0.10 for PGR SNPs rs503602 and rs660149 and luminal A breast cancer, and for rs548668 and basal-like breast cancer. PGR contains several splice sites that produce three different isoforms of the progesterone receptor. PR-A and PR-B are structurally similar, but have unique transcriptional activities (65, 67). PR-C lacks the progesterone response element-binding motif and activation domains present in PR-A and PR-B, but PR-C is able to bind PR-B and inhibit PR-B transcriptional activity (67). Thus, SNPs or other genetic variants that affect the ratio of PR isoforms could affect PR signaling. Although the basal-like subtype is defined by low levels of PR expression, SNPs are present in normal cells before tumor formation, and so the effects of SNPs in PGR could influence early stages of basal-like tumor formation.

In addition to the PGR SNPs associated with basal-like breast cancer, there were two other clusters of SNPs that were associated with the breast cancer subtypes. In ESR1, SNPs rs6914211, rs3020401, rs985191, and rs6557177 were positively associated with both luminal A and basal-like breast cancer (Figure 1). Located from 152354108 bp to 152355411 bp in intron 4, these associations may be indicative of a causative locus nearby. The other subtype-associated region was in HSD17B2, where rs8050327, rs8191072, and rs7200696 were inversely associated with luminal A breast cancer, but at or near the null for basal-like breast cancer (Figure 2). Located in intron 1, these SNPs are in LD with each other in CBCS subjects, and are likely marking the effect of the same causal variant. The strength of association and number of SNPs that showed an association within these two regions in this

study makes these SNPs good candidates for genotyping in replication studies. In the case of HSD17B2, replication should be pursued with cases that meet the definition of luminal A subtype (ER+ and/or /PR+).

There were several SNPs that were strongly associated with breast cancer overall, which was unexpected given the theory that associations for common polymorphisms will be modest. The strongest associations for breast cancer overall were for ESR1 rs2207232, rs6914211, and rs985191, HSD17B2 rs8052451. All four of these SNPs are intronic; rs6914211 and rs985191 are approximately 15 kb apart (14918 bp) and are in strong LD (Figure 1b-e). For rs6914211, rs985191, and rs8052451, associations with basal-like and luminal A subtypes were similar to the association for breast cancer overall.

Table 7 compares the odds ratios estimated in the CBCS to those from previously published studies. Our results indicate an association between 2 previously studied candidate SNPs and breast cancer - SHBG rs1799941 (-67 G/A) and CYP19A1 rs700519 (R264C). The SHBG rs1799941 AA genotype was associated moderately with breast cancer overall in the CBCS, but the association was even stronger for the luminal A and basal-like subtypes. Other studies have reported no association for this SNP (68, 69). Thompson et al. (69) reported results for rs1799941 using an additive model, but in the CBCs the mode of inheritance seems to be recessive for breast cancer overall as well as by luminal A or basal-like subtype. When an additive model was used in CBCS, the estimated odds ratio was similar to that reported by Thompson et al. [CBCS all cases, A vs. G, OR (95% CI): 1.02 (0.90, 1.16)].

Dunning et al. (68) reported that the A allele of rs1799941 was associated with elevated levels of circulating SHBG and a reduced estradiol:SHBG ratio in healthy postmenopausal women. The association between the A allele and reduced estradiol:SHBG

levels was most consistent with the additive model (68). Our results do not agree with the biological mechanism that would follow from the work of Dunning et al. (68). SHBG circulates in the blood and sequesters sex hormones (like estrogen and testosterone), limiting their availability to be metabolized or bind to receptors and initiate signaling. In both cases, estrogen signaling and estrogen metabolism can contribute to carcinogenesis. If the rs1799941 A allele increases SHBG levels and sex-hormone binding capacity, it would presumably decrease the amount of estrogen available to contribute to carcinogenesis and reduce breast cancer risk (at least for the proportion of breast cancers caused by this mechanism). But in the CBCS the A allele was associated with increased, not decreased risk. While there could certainly be other biological mechanisms through which rs1799941, SHBG, and estrogen interact, this incongruity highlights the need for further study of this SNP.

The CYP19A1 rs700519 (R264C) TT genotype was inversely associated with breast cancer in CBCS cases overall, but this genotype is uncommon, and was extremely rare in CBCS non-African American subjects (Table 4.1). The minor allele frequency in CBCS subjects was similar to what has been reported in other populations (Table 4.8). Among African Americans only, the association was similar to the association in the overall population [CBCS, AA cases vs. AA controls, TT vs. CC, OR (95% CI): 0.37 (0.17, 0.78)]. In the CBCS, the TT vs. CC association appears to follow a recessive mode of inheritance. Miyoshi et al. (70) reported an inverse association for rs700519 but used a dominant model, likely due to the rarity of TT homozygotes. It is possible that the effect for TT homozygotes using a general or recessive model would be even closer to what was estimated in the CBCS. Other studies that compared the TT to CC genotype found no association (12, 71). The

functional effects of the R264C substitution are unclear – some studies reported that 264C caused reduced the hydrophobicity, reduced expression, and reduced aromatase activity, but another study reported that there was no difference in enzymatic activity between 264R and 264C (72-74). If the 264C polymorphism does actually cause reduced aromatase activity, this would be consistent with the reduced breast cancer risk observed in the CBCS.

Although 2 additional CYP19A1 SNPs were associated with breast cancer overall, these associations were modest and there was no association for the luminal A or basal-like subtype. This is in agreement with results from Haiman et al.(12) who found that although genetic variation in CYP19A1 was related to circulating estrogen levels, SNPs in CYP19A1 were unrelated to breast cancer risk in postmenopausal white women. Talbott et al. (75) reported a positive association between a htSNP (rs1008805) from Haiman et al. (12) and breast cancer, and that the association was modified by menopausal status. Like Haiman et al., there was no association in the CBCS for rs1008805 and breast cancer overall, by subtype, or stratified by menopausal status (Table 7, menopausal status-stratified data not shown). Kristensen et al. (76) reported a positive association between CYP19A1 rs10046 (3' UTR) and breast cancer, and later reported that the rs10046 TT genotype was significantly associated with basal-like breast cancer (25). However, these associations were not replicated in the CBCS or two other studies (68, 77).

Raskin et al. (78) reported that the CYP19A1 V80V (rs700518) polymorphism was associated with an increased risk of breast cancer among BRCA1 carriers less than 50 years old, but not in BRCA2 carriers or BRCA non-carriers. BRCA1 cases exhibit an increased frequency of basal-like tumors (19, 26, 28), but an association for V80V was not seen in CBCS basal-like cases (Table 4.7). When the CBCS analysis was restricted to basal-like

cases less than 50 years old, the OR was elevated (Table 4.7), but with wide confidence intervals that included the null. The prevalence of BRCA1 mutations is low in CBCS cases (prevalence, 2.6%; 95% CI, 0-5.5%)(79), so CBCS results are not directly comparable to Raskin et al. Still, the results are suggestive, and warrant further study in a larger population of young basal-like cases. In a study of postmenopausal women, the V80V GG genotype (vs. AA) was associated with lower serum estradiol (80), which runs contrary to the hypothesized mechanism of increased estradiol being associated with the at risk genotype. However, it is compatible with the fact that basal-like breast cancers do not express the estrogen receptor, and are likely estrogen independent tumors. Further research is needed to determine whether the genotype association is real, and what pro-tumor effects could occur in response to reduced estrogen.

No ESR1 functional SNPs were associated with breast cancer in the CBCS, either for breast cancer overall or by subtype, but other ESR1 SNPS identified in recent comprehensive tag SNP analyses were associated with breast cancer in the CBCS. Using data from the SEARCH and EPIC-Norfolk studies, Mavaddat et al. (62) identified ESR1 SNPs rs3020314, rs3020407, and rs3020401 as 3 of 52 SNPs recommended for further study after a two-stage analysis of 700 candidate gene SNPs and breast cancer. ESR1 rs3020314 was also identified by Dunning et al. (81) as having a very weak but precise association with breast cancer in a 3-stage study of ESR1 haplotype tag SNPs. Subgroup analyses revealed that the rs3020314 association was confined to ER-positive breast cancer and was not present in the two studies with Asian populations (81). Fine-mapping of SNPs in LD with rs3020314 revealed that rs3020314 is likely not responsible for the entire association with breast cancer (81).

In the CBCS, ESR1 rs3020401 was strongly associated basal-like breast cancer, and

more weakly associated with breast cancer overall. Moreover, rs3020401 is in region of associated SNPs, and is part of a breast cancer-associated haplotype in the CBCS, albeit one associated with all cases and luminal A, not basal-like breast cancer. ESR1 SNPs rs3020314 and rs3020407 were correlated with rs3020401 in CBCS subjects, but were not associated with breast cancer overall or by subtype in the CBCS. The OR reported by Dunning et al. (81) is so close to the null that CBCS data provide equal support for no association and the weak association identified by meta-analysis.

In the CBCS, ESR1 rs3020314 was part of a 6-SNP haplotype that was positively associated with breast cancer overall and luminal A breast cancer (Table 5, haplotype 4). There was evidence of interaction between haplotype 4 and parity and lactation with regards to luminal A breast cancer – the excess risk of luminal A breast cancer associated with 1 or 2 copies of haplotype 4 was seen mainly among parous women who breastfed (Table 6). The OR for 2 copies of haplotype 4 was also elevated for parous women who never breastfed (Table 6). That there was sufficient statistical support for this interaction in luminal A breast cancer but not breast cancer overall is consistent with Dunning et al.'s observation that the rs3020314 association was seen for ER-positive breast cancer only. This is in disagreement with the finding by Mavaddat et al. (62) that the rs3020314 association was not specific to any breast cancer subgroup. Despite some inconsistencies in single SNP associations, the repeated identification of SNPs in this region of ESR1 as associated with breast cancer and the CBCS haplotype 4 associations strongly suggests that there is at least one genetic variant of some importance within LD range.

In another study of ESR1 tag SNPs, Einarsdottir et al. (82) identified rs3003925, rs3020318, rs726281, rs3020407, rs2144025 as potentially associated with breast cancer.

SNPs rs3020318, rs3020407, and rs2144025 were genotyped in the CBCS but none were associated with breast cancer, either overall or by subtype. Interestingly, rs3020407 was highlighted by both Einarsdottir et al. (82) and Mavaddat et al. (62).

SNPs in HSD3B1 were associated with breast cancer in the CBCS, a finding that has not been reported before to our knowledge. There have studies on the association between HSD3B1 SNPs and breast density. The codon 367T variant was associated with increased breast density among African American women, decreased breast density among white American women, and decreased breast density in mostly white Australian women (83, 84). Increased breast density is a risk factor for breast cancer (85, 86), so it is possible that the codon 367T variant may be associated with increased breast cancer risk, particularly in African American women.

The strongest haplotype association was for CYP19A1 haplotype 1, a 5-SNP haplotype composed of intronic tag SNPs. Associations for haplotype 1 and breast cancer overall and the basal-like subtype were very strong. This was consistent with a comprehensive haplotype analysis conducted in the MultiEthnic Cohort (MEC) study (40). In the MEC, Haiman and colleagues used dense SNP genotyping to characterize LD and haplotype block structure in CYP19A1. Haplotype associations with breast cancer were observed for MEC Block 2, which contains 2 of the 5 SNPs in CBCS HAP1 (rs2445759 and rs936306) and MEC Block 3, which contains CBCS HAP1 SNP rs749292. Though rs749292 is in intron 2 and the other haplotype SNPs are in intron 1, Haiman et al. reported a stronger odds ratio for the long range haplotype MEC 2b-3c that combined SNPs from the two regions. The OR estimated in the CBCS was much stronger than the OR estimated in the MEC study [MEC 2b-3c, OR (95% CI): 1.31 (1.11, 1.54)] (40). Since the SNPs did not

overlap completely between the MEC and CBCS haplotypes, it is possible that one of the non-overlapping CBCS SNPs tags a region that is in stronger LD with a causal variant.

Associations reported in the MEC were not observed in a follow-up pooled study of haplotype associations with estrogen levels and breast cancer in postmenopausal, mainly white participants from several large cohort studies, where none of the haplotypes were associated with breast cancer (12). CYP19A1 haplotypes in the pooled study block 2 included 4 of 5 SNPs in CBCS CYP19A1 HAP1, but the HAP1 haplotype identified in the CBCS was not analyzed.

Several groups have examined ESR1 haplotypes and breast cancer risk, but the haplotypes identified in CBCS (haplotypes 2, 3, 4, 5)) did not fully correspond to haplotypes identified in these studies. Fernandez et al. (87) reported a protective effect for an ESR1 haplotype that contains functional variants (rs2077647, rs1801132, rs3798577, rs3798758) [C-G-T-G, OR (95% CI): 0.34 (0.14, 0.83)], but this haplotype was not associated with breast cancer in the CBCS (data not shown). Gold et al. (88) identified multiple haplotypes associated with breast cancer, but CBCS results could not be compared to Gold et al. because there were two few SNPs in common between haplotypes in the two studies. Others did not find any association between any haplotypes composed of rs746432, rs2234693, rs9340799, and rs1801132 (identified by Gold et al.) and breast cancer (89).

We investigated the possibility of multiplicative interaction between breast cancer-associated variants and a combined parity and lactation variable. There were several SNPs and one haplotype that demonstrated evidence of multiplicative interaction with parity and lactation. However, estimation of associations stratified by parity/lactation status did not reveal any consistent patterns for luminal A or basal-like breast cancer. This was due in part

to small numbers of basal-like cases and nulliparous women.

Though candidate genes and SNPs were chosen with careful consideration to the biological plausibility of gene effects on breast cancer risk, the results should be replicated in other studies before conclusions are made. It is possible that some associations are due to chance. Since hypothesis testing was neither the goal of this analysis nor was it used as strict criteria for culling results, P-values were not adjusted for multiple comparisons.

There was also some potential for systematic bias. Not all CBCS subjects had DNA available for genotyping. Not all eligible women chose to participate in the CBCS. Furthermore, not all women provided blood for genotyping analyses. 3748 of 4333 (86.5%) CBCS subjects had adequate data for inclusion in the analysis, and subjects without data were more likely to be cases and African American. However, all models were adjusted for self-identified race, which should provide sufficient control for bias that would occur if genotypes for the SNPs in question also differed by case and race status (90).

Additionally, not all of the cases enrolled in the CBCS had sufficient amounts of tumor tissue necessary for determining molecular subtype. Cases with subtype data were more likely to be African American and have a later stage at diagnosis compared to cases without subtype data, but did not differ by other breast cancer risk factors (29). Study results could be biased if the genotypes of cases with sufficient tumor tissue for subtyping were systematically different from the genotypes of women without subtype information. Genotype distributions between cases with and without subtype information differed for HSD17B2 rs3111351 and rs8191136 in African Americans, and for ESR1 rs6557177r and rs985695r in non-African Americans. Selection bias may have affected the associations estimated for these SNPs in the CBCS.

Another potential source of bias was intrinsic molecular subtype misclassification. The intrinsic subtypes were characterized based on gene expression data, but there is not always perfect agreement between gene expression and IHC-based classification (23, 35, 49). In a series of sensitivity analyses that simulated the genotype-subtype association, bias-corrected ORs for some SNPs were further from the null from the observed OR. The difference between observed and bias-corrected ORs was largest when the sample size in variant genotype cells was small. This means that in the case of subtype misclassification between luminal A and basal-like subtypes, observed differences between subtype effects were likely biased towards the null. The differences observed in this study are unlikely to be false positives due to bias; in fact the sensitivity analysis suggests that we may have missed some differences between the subtypes when the number of variant alleles was small.

In conclusion, this analysis has identified SNPs in ESR1, PGR, CYP19A1, HSD17B2, SHBG, and HSD3B1 that were associated with breast cancer. In some cases, these associations were only apparent when cases were stratified by intrinsic molecular subtype. Several associations are consistent with other reports in the literature, including associations for functional SNPs in SHBG and CYP19A1. Further characterization of the effects of these SNPs could help identify novel biological pathways active in breast cancer. The evidence to support interaction between ESR1 haplotype 4, parity and lactation, and luminal A breast cancer is also promising, particularly since comprehensive ESR1 analyses from the SEARCH study also point towards a breast-cancer associated variant in that region. The apparent difference in haplotype 4 association between parous and nulliparous women could have implications for breast cancer prevention strategies and screening among parous women. Fine-mapping of SNPs within this region in the CBCS may help localize the causal

variant.

4.6 References

1. Antoniou AC, Pharoah PD, McMullan G, Day NE, Ponder BA, Easton D. Evidence for further breast cancer susceptibility genes in addition to BRCA1 and BRCA2 in a population-based study. Genet Epidemiol 2001;21(1):1-18.

2. Antoniou AC, Pharoah PD, McMullan G, Day NE, Stratton MR, Peto J, et al. A comprehensive model for familial breast cancer incorporating BRCA1, BRCA2 and other genes. Br J Cancer 2002;86(1):76-83.

3. Pharoah PD, Antoniou A, Bobrow M, Zimmern RL, Easton DF, Ponder BA. Polygenic susceptibility to breast cancer and implications for prevention. Nat Genet 2002;31(1):33-6.

4. Gruber CJ, Tschugguel W, Schneeberger C, Huber JC. Production and actions of estrogens. N Engl J Med 2002;346(5):340-52.

5. Barbieri RL. The Breast. In: Strauss JF, Barbieri RL, editors. Yen and Jaffe's Reproductive Endocrinology. 5th ed. Philadelphia: Elsevier; 2004. p. 307-326.

6. Conneely OM, Jericevic BM, Lydon JP. Progesterone receptors in mammary gland development and tumorigenesis. J Mammary Gland Biol Neoplasia 2003;8(2):205-14.

7. Frasor J, Danes JM, Komm B, Chang KC, Lyttle CR, Katzenellenbogen BS. Profiling of estrogen up- and down-regulated gene expression in human breast cancer cells: insights into gene networks and pathways underlying estrogenic control of proliferation and cell phenotype. Endocrinology 2003;144(10):4562-74.

8. Cavalieri E, Frenkel K, Liehr JG, Rogan E, Roy D. Estrogens as endogenous genotoxic agents--DNA adducts and mutations. J Natl Cancer Inst Monogr 2000(27):75-93.

9. Pike MC, Spicer DV, Dahmoush L, Press MF. Estrogens, progestogens, normal breast cell proliferation, and breast cancer risk. Epidemiol Rev 1993;15(1):17-35.

10. Bermejo-Perez MJ, Marquez-Calderon S, Llanos-Mendez A. Effectiveness of preventive interventions in BRCA1/2 gene mutation carriers: a systematic review. Int J Cancer 2007;121(2):225-31.

11. Payne AH, Hales DB. Overview of steroidogenic enzymes in the pathway from cholesterol to active steroid hormones. Endocr Rev 2004;25(6):947-70.

12. Haiman CA, Dossus L, Setiawan VW, Stram DO, Dunning AM, Thomas G, et al. Genetic variation at the CYP19A1 locus predicts circulating estrogen levels but not breast cancer risk in postmenopausal women. Cancer Res 2007;67(5):1893-7.

13. Luu-The V. Analysis and characteristics of multiple types of human 17beta-hydroxysteroid dehydrogenase. J Steroid Biochem Mol Biol 2001;76(1-5):143-51.

14. Fortunati N, Fissore F, Fazzari A, Becchis M, Comba A, Catalano MG, et al. Sex steroid binding protein exerts a negative control on estradiol action in MCF-7 cells (human breast cancer) through cyclic adenosine 3',5'-monophosphate and protein kinase A. Endocrinology 1996;137(2):686-92.

15. Catalano MG, Frairia R, Boccuzzi G, Fortunati N. Sex hormone-binding globulin antagonizes the anti-apoptotic effect of estradiol in breast cancer cells. Mol Cell Endocrinol 2005;230(1-2):31-7.

16. Duax WL, Thomas J, Pletnev V, Addlagatta A, Huether R, Habegger L, et al. Determining structure and function of steroid dehydrogenase enzymes by sequence analysis, homology modeling, and rational mutational analysis. Ann N Y Acad Sci 2005;1061:135-48.

17. Perou CM, Sorlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, et al. Molecular portraits of human breast tumours. Nature 2000;406(6797):747-52.

18. Sorlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proc Natl Acad Sci U S A 2001;98(19):10869-74.

19. Sorlie T, Tibshirani R, Parker J, Hastie T, Marron JS, Nobel A, et al. Repeated observation of breast tumor subtypes in independent gene expression data sets. Proc Natl Acad Sci U S A 2003;100(14):8418-23.

20. Abd El-Rehim DM, Ball G, Pinder SE, Rakha E, Paish C, Robertson JF, et al. High-throughput protein expression analysis using tissue microarray technology of a large well-characterised series identifies biologically distinct classes of breast cancer confirming recent cDNA expression analyses. Int J Cancer 2005;116(3):340-50.

21. Yu K, Lee CH, Tan PH, Tan P. Conservation of breast cancer molecular subtypes and transcriptional patterns of tumor progression across distinct ethnic populations. Clin Cancer Res 2004;10(16):5508-17.

22. Carey LA, Perou CM, Livasy CA, Dressler LG, Cowan D, Conway K, et al. Race, breast cancer subtypes, and survival in the Carolina Breast Cancer Study. Jama 2006;295(21):2492-502.

23. Calza S, Hall P, Auer G, Bjohle J, Klaar S, Kronenwett U, et al. Intrinsic molecular signature of breast cancer in a population-based cohort of 412 patients. Breast Cancer Res 2006;8(4):R34.

24. Nordgard SH, Johansen FE, Alnaes GI, Naume B, Borresen-Dale AL, Kristensen VN.

Genes harbouring susceptibility SNPs are differentially expressed in the breast cancer subtypes. Breast Cancer Res 2007;9(6):113.

25. Kristensen VN, Borresen-Dale AL. SNPs associated with molecular subtypes of breast cancer: on the usefulness of stratified Genome-wide Association Studies (GWAS) in the identification of novel susceptibility loci. Mol Oncol 2008;2(1):12-5.

26. Foulkes WD, Stefansson IM, Chappuis PO, Begin LR, Goffin JR, Wong N, et al. Germline BRCA1 mutations and a basal epithelial phenotype in breast cancer. J Natl Cancer Inst 2003;95(19):1482-5.

27. Lakhani SR, Reis-Filho JS, Fulford L, Penault-Llorca F, van der Vijver M, Parry S, et al. Prediction of BRCA1 status in patients with breast cancer using estrogen receptor and basal phenotype. Clin Cancer Res 2005;11(14):5175-80.

28. Turner NC, Reis-Filho JS. Basal-like breast cancer and the BRCA1 phenotype. Oncogene 2006;25(43):5846-53.

29. Millikan RC, Newman B, Tse CK, Moorman PG, Conway K, Dressler LG, et al. Epidemiology of basal-like breast cancer. Breast Cancer Res Treat 2008;109(1):123-39.

30. Newman B, Moorman PG, Millikan R, Qaqish BF, Geradts J, Aldrich TE, et al. The Carolina Breast Cancer Study: integrating population-based epidemiology and molecular biology. Breast Cancer Res Treat 1995;35(1):51-60.

31. Millikan R, Eaton A, Worley K, Biscocho L, Hodgson E, Huang WY, et al. HER2 codon 655 polymorphism and risk of breast cancer in African Americans and whites. Breast Cancer Res Treat 2003;79(3):355-64.

32. Weinberg CR, Sandler DP. Randomized recruitment in case-control studies. Am J Epidemiol 1991;134(4):421-32.

33. Livasy CA, Perou CM, Karaca G, Cowan DW, Maia D, Jackson S, et al. Identification of a basal-like subtype of breast ductal carcinoma in situ. Hum Pathol 2007;38(2):197-204.

34. Huang WY, Newman B, Millikan RC, Schell MJ, Hulka BS, Moorman PG. Hormone-related factors and risk of breast cancer in relation to estrogen receptor and progesterone receptor status. Am J Epidemiol 2000;151(7):703-14.

35. Nielsen TO, Hsu FD, Jensen K, Cheang M, Karaca G, Hu Z, et al. Immunohistochemical and clinical characterization of the basal-like subtype of invasive breast carcinoma. Clin Cancer Res 2004;10(16):5367-74.

36. Frazer KA, Ballinger DG, Cox DR, Hinds DA, Stuve LL, Gibbs RA, et al. A second

generation human haplotype map of over 3.1 million SNPs. Nature 2007;449(7164):851-61.

37. Carlson CS, Eberle MA, Rieder MJ, Yi Q, Kruglyak L, Nickerson DA. Selecting a maximally informative set of single-nucleotide polymorphisms for association analyses using linkage disequilibrium. Am J Hum Genet 2004;74(1):106-20.

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

39. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics 2005;21(2):263-5.

40. Haiman CA, Stram DO, Pike MC, Kolonel LN, Burtt NP, Altshuler D, et al. A comprehensive haplotype analysis of CYP19 and breast cancer risk: the Multiethnic Cohort. Hum Mol Genet 2003;12(20):2679-92.

41. Balding DJ. A tutorial on statistical methods for population association studies. Nat Rev Genet 2006;7(10):781-91.

42. Hosking L, Lumsden S, Lewis K, Yeo A, McCarthy L, Bansal A, et al. Detection of genotyping errors by Hardy-Weinberg equilibrium testing. Eur J Hum Genet 2004;12(5):395-9.

43. Wigginton JE, Cutler DJ, Abecasis GR. A note on exact tests of Hardy-Weinberg equilibrium. Am J Hum Genet 2005;76(5):887-93.

44. Tian C, Hinds DA, Shigeta R, Kittles R, Ballinger DG, Seldin MF. A genomewide single-nucleotide-polymorphism panel with high ancestry information for African American admixture mapping. Am J Hum Genet 2006;79(4):640-9.

45. Barnholtz-Sloan JS, Chakraborty R, Sellers TA, Schwartz AG. Examining population stratification via individual ancestry estimates versus self-reported race. Cancer Epidemiol Biomarkers Prev 2005;14(6):1545-51.

46. Lin DY, Zeng D. Likelihood-Based Inference on Haplotype Effects in Genetic Association Studies. Journal of the American Statistical Association 2006;101(473):89-118.

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

48. Weinberg CR, Wacholder S. The design and analysis of case-control studies with biased sampling. Biometrics 1990;46(4):963-75.

49. Livasy CA, Karaca G, Nanda R, Tretiakova MS, Olopade OI, Moore DT, et al. Phenotypic evaluation of the basal-like subtype of invasive breast carcinoma. Mod Pathol 2006;19(2):264-71.

50. Fox MP, Lash TL, Greenland S. A method to automate probabilistic sensitivity analyses of misclassified binary variables. Int J Epidemiol 2005;34(6):1370-6.

51. Lakhani SR. The transition from hyperplasia to invasive carcinoma of the breast. J Pathol 1999;187(3):272-8.

52. Leonard GD, Swain SM. Ductal carcinoma in situ, complexities and challenges. J Natl Cancer Inst 2004;96(12):906-20.

53. Erbas B, Provenzano E, Armes J, Gertig D. The natural history of ductal carcinoma in situ of the breast: a review. Breast Cancer Res Treat 2006;97(2):135-44.

54. Morrow M, O'Sullivan MJ. The dilemma of DCIS. Breast 2007;16 Suppl 2:S59-62.

55. Kerlikowske K, Barclay J, Grady D, Sickles EA, Ernster V. Comparison of risk factors for ductal carcinoma in situ and invasive breast cancer. J Natl Cancer Inst 1997;89(1):76-82.

56. Reinier KS, Vacek PM, Geller BM. Risk factors for breast carcinoma in situ versus invasive breast cancer in a prospective study of pre- and post-menopausal women. Breast Cancer Res Treat 2007;103(3):343-8.

57. Dabbs DJ, Chivukula M, Carter G, Bhargava R. Basal phenotype of ductal carcinoma in situ: recognition and immunohistologic profile. Mod Pathol 2006;19(11):1506-11.

58. Bryan BB, Schnitt SJ, Collins LC. Ductal carcinoma in situ with basal-like phenotype: a possible precursor to invasive basal-like breast cancer. Mod Pathol 2006;19(5):617-21.

59. Hannemann J, Velds A, Halfwerk JB, Kreike B, Peterse JL, van de Vijver MJ. Classification of ductal carcinoma in situ by gene expression profiling. Breast Cancer Res 2006;8(5):R61.

60. Stacey SN, Manolescu A, Sulem P, Rafnar T, Gudmundsson J, Gudjonsson SA, et al. Common variants on chromosomes 2q35 and 16q12 confer susceptibility to estrogen receptor-positive breast cancer. Nat Genet 2007;39(7):865-9.

61. Garcia-Closas M, Hall P, Nevanlinna H, Pooley K, Morrison J, Richesson DA, et al. Heterogeneity of breast cancer associations with five susceptibility loci by clinical and pathological characteristics. PLoS Genet 2008;4(4):e1000054.

62. Mavaddat N, Dunning AM, Ponder BA, Easton DF, Pharoah PD. Common genetic

variation in candidate genes and susceptibility to subtypes of breast cancer. Cancer Epidemiol Biomarkers Prev 2009;18(1):255-9.

63. Stacey SN, Manolescu A, Sulem P, Thorlacius S, Gudjonsson SA, Jonsson GF, et al. Common variants on chromosome 5p12 confer susceptibility to estrogen receptor-positive breast cancer. Nat Genet 2008;40(6):703-6.

64. Tapper W, Hammond V, Gerty S, Ennis S, Simmonds P, Collins A, et al. The influence of genetic variation in 30 selected genes on the clinical characteristics of early onset breast cancer. Breast Cancer Res 2008;10(6):R108.

65. Conneely OM, Mulac-Jericevic B, Lydon JP. Progesterone-dependent regulation of female reproductive activity by two distinct progesterone receptor isoforms. Steroids 2003;68(10-13):771-8.

66. Yang XR, Sherman ME, Rimm DL, Lissowska J, Brinton LA, Peplonska B, et al. Differences in risk factors for breast cancer molecular subtypes in a population-based study. Cancer Epidemiol Biomarkers Prev 2007;16(3):439-43.

67. Mendelson CR, Hardy DB. Role of the progesterone receptor (PR) in the regulation of inflammatory response pathways and aromatase in the breast. J Steroid Biochem Mol Biol 2006;102(1-5):241-9.

68. Dunning AM, Dowsett M, Healey CS, Tee L, Luben RN, Folkerd E, et al. Polymorphisms associated with circulating sex hormone levels in postmenopausal women. J Natl Cancer Inst 2004;96(12):936-45.

69. Thompson DJ, Healey CS, Baynes C, Kalmyrzaev B, Ahmed S, Dowsett M, et al. Identification of common variants in the SHBG gene affecting sex hormone-binding globulin levels and breast cancer risk in postmenopausal women. Cancer Epidemiol Biomarkers Prev 2008;17(12):3490-8.

70. Miyoshi Y, Iwao K, Ikeda N, Egawa C, Noguchi S. Breast cancer risk associated with polymorphism in CYP19 in Japanese women. Int J Cancer 2000;89(4):325-8.

71. Lee KM, Abel J, Ko Y, Harth V, Park WY, Seo JS, et al. Genetic polymorphisms of cytochrome P450 19 and 1B1, alcohol use, and breast cancer risk in Korean women. Br J Cancer 2003;88(5):675-8.

72. Ma CX, Adjei AA, Salavaggione OE, Coronel J, Pelleymounter L, Wang L, et al. Human aromatase: gene resequencing and functional genomics. Cancer Res 2005;65(23):11071-82.

73. Watanabe J, Harada N, Suemasu K, Higashi Y, Gotoh O, Kawajiri K. Arginine-cysteine polymorphism at codon 264 of the human CYP19 gene does not affect aromatase activity. Pharmacogenetics 1997;7(5):419-24.

74. Kagawa N, Hori H, Waterman MR, Yoshioka S. Characterization of stable human aromatase expressed in E. coli. Steroids 2004;69(4):235-43.

75. Talbott KE, Gammon MD, Kibriya MG, Chen Y, Teitelbaum SL, Long CM, et al. A CYP19 (aromatase) polymorphism is associated with increased premenopausal breast cancer risk. Breast Cancer Res Treat 2008;111(3):481-7.

76. Kristensen VN, Harada N, Yoshimura N, Haraldsen E, Lonning PE, Erikstein B, et al. Genetic variants of CYP19 (aromatase) and breast cancer risk. Oncogene 2000;19(10):1329-33.

77. Haiman CA, Hankinson SE, Spiegelman D, Brown M, Hunter DJ. No association between a single nucleotide polymorphism in CYP19 and breast cancer risk. Cancer Epidemiol Biomarkers Prev 2002;11(2):215-6.

78. Raskin L, Lejbkowicz F, Barnett-Griness O, Dishon S, Almog R, Rennert G. BRCA1 breast cancer risk is modified by CYP19 polymorphisms in Ashkenazi Jews. Cancer Epidemiol Biomarkers Prev 2009;18(5):1617-23.

79. Newman B, Mu H, Butler LM, Millikan RC, Moorman PG, King MC. Frequency of breast cancer attributable to BRCA1 in a population-based series of American women. Jama 1998;279(12):915-21.

80. Somner J, McLellan S, Cheung J, Mak YT, Frost ML, Knapp KM, et al. Polymorphisms in the P450 c17 (17-hydroxylase/17,20-Lyase) and P450 c19 (aromatase) genes: association with serum sex steroid concentrations and bone mineral density in postmenopausal women. J Clin Endocrinol Metab 2004;89(1):344-51.

81. Dunning AM, Healey CS, Baynes C, Maia AT, Scollen S, Vega A, et al. Association of ESR1 gene tagging SNPs with breast cancer risk. Hum Mol Genet 2009;18(6):1131-9.

82. Einarsdottir K, Darabi H, Li Y, Low YL, Li YQ, Bonnard C, et al. ESR1 and EGF genetic variation in relation to breast cancer risk and survival. Breast Cancer Res 2008;10(1):R15.

83. Haiman CA, Bernstein L, Berg D, Ingles SA, Salane M, Ursin G. Genetic determinants of mammographic density. Breast Cancer Res 2002;4(3):R5.

84. Stone J, Gurrin LC, Byrnes GB, Schroen CJ, Treloar SA, Padilla EJ, et al. Mammographic density and candidate gene variants: a twins and sisters study. Cancer Epidemiol Biomarkers Prev 2007;16(7):1479-84.

85. Boyd NF, Lockwood GA, Byng JW, Tritchler DL, Yaffe MJ. Mammographic densities and breast cancer risk. Cancer Epidemiol Biomarkers Prev 1998;7(12):1133-

44.

86. McCormack VA, dos Santos Silva I. Breast density and parenchymal patterns as markers of breast cancer risk: a meta-analysis. Cancer Epidemiol Biomarkers Prev 2006;15(6):1159-69.

87. Fernandez LP, Milne RL, Barroso E, Cuadros M, Arias JI, Ruibal A, et al. Estrogen and progesterone receptor gene polymorphisms and sporadic breast cancer risk: a Spanish case-control study. Int J Cancer 2006;119(2):467-71.

88. Gold B, Kalush F, Bergeron J, Scott K, Mitra N, Wilson K, et al. Estrogen receptor genotypes and haplotypes associated with breast cancer risk. Cancer Res 2004;64(24):8891-900.

89. Wang J, Higuchi R, Modugno F, Li J, Umblas N, Lee J, et al. Estrogen receptor alpha haplotypes and breast cancer risk in older Caucasian women. Breast Cancer Res Treat 2007;106(2):273-80.

90. Savitz DA. Interpreting Epidemiologic Evidence. New York: Oxford University Press; 2003.

91. Gulyaeva LF, Mikhailova ON, PustyInyak VO, Kim IVt, Gerasimov AV, Krasilnikov SE, et al. Comparative analysis of SNP in estrogen-metabolizing enzymes for ovarian, endometrial, and breast cancers in Novosibirsk, Russia. Adv Exp Med Biol 2008;617:359-66.

92. Onland-Moret NC, van Gils CH, Roest M, Grobbee DE, Peeters PH. The estrogen receptor alpha gene and breast cancer risk (The Netherlands). Cancer Causes Control 2005;16(10):1195-202.

93. Zheng W, Long J, Gao YT, Li C, Zheng Y, Xiang YB, et al. Genome-wide association study identifies a new breast cancer susceptibility locus at 6q25.1. Nat Genet 2009;41(3):324-8.

94. Iwasaki M, Hamada GS, Nishimoto IN, Netto MM, Motola J, Jr., Laginha FM, et al. Isoflavone, polymorphisms in estrogen receptor genes and breast cancer risk in case-control studies in Japanese, Japanese Brazilians and non-Japanese Brazilians. Cancer Sci 2009;100(5):927-33.

95. Sonestedt E, Ivarsson MI, Harlid S, Ericson U, Gullberg B, Carlson J, et al. The protective association of high plasma enterolactone with breast cancer is reasonably robust in women with polymorphisms in the estrogen receptor alpha and beta genes. J Nutr 2009;139(5):993-1001.

96. Gallicchio L, Berndt SI, McSorley MA, Newschaffer CJ, Thuita LW, Argani P, et al. Polymorphisms in estrogen-metabolizing and estrogen receptor genes and the risk of

developing breast cancer among a cohort of women with benign breast disease. BMC Cancer 2006;6:173.

97. Pooley KA, Healey CS, Smith PL, Pharoah PD, Thompson D, Tee L, et al. Association of the progesterone receptor gene with breast cancer risk: a single-nucleotide polymorphism tagging approach. Cancer Epidemiol Biomarkers Prev 2006;15(4):675-82.

98. Johnatty SE, Spurdle AB, Beesley J, Chen X, Hopper JL, Duffy DL, et al. Progesterone receptor polymorphisms and risk of breast cancer: results from two Australian breast cancer studies. Breast Cancer Res Treat 2008;109(1):91-9.

99. De Vivo I, Hankinson SE, Colditz GA, Hunter DJ. A functional polymorphism in the progesterone receptor gene is associated with an increase in breast cancer risk. Cancer Res 2003;63(17):5236-8.

100. Riancho JA, Valero C, Naranjo A, Morales DJ, Sanudo C, Zarrabeitia MT. Identification of an aromatase haplotype that is associated with gene expression and postmenopausal osteoporosis. J Clin Endocrinol Metab 2007;92(2):660-5.

101. He LN, Xiong DH, Liu YJ, Zhang F, Recker RR, Deng HW. Association study of the oestrogen signalling pathway genes in relation to age at natural menopause. J Genet 2007;86(3):269-76.

102. Kjaergaard AD, Ellervik C, Tybjaerg-Hansen A, Axelsson CK, Gronholdt ML, Grande P, et al. Estrogen receptor alpha polymorphism and risk of cardiovascular disease, cancer, and hip fracture: cross-sectional, cohort, and case-control studies and a meta-analysis. Circulation 2007;115(7):861-71.

103. Sowers MR, Jannausch ML, McConnell DS, Kardia SR, Randolph JF, Jr. Endogenous estradiol and its association with estrogen receptor gene polymorphisms. Am J Med 2006;119(9 Suppl 1):S16-22.

104. Zofkova I, Zajickova K, Hill M. The estrogen receptor alpha gene determines serum androstenedione levels in postmenopausal women. Steroids 2002;67(10):815-9.

105. Pharoah PD, Tyrer J, Dunning AM, Easton DF, Ponder BA. Association between common variation in 120 candidate genes and breast cancer risk. PLoS Genet 2007;3(3):e42.

106. Westberg L, Ho HP, Baghaei F, Nilsson S, Melke J, Rosmond R, et al. Polymorphisms in oestrogen and progesterone receptor genes: possible influence on prolactin levels in women. Clin Endocrinol (Oxf) 2004;61(2):216-23.

107. Garcia-Closas M, Brinton LA, Lissowska J, Richesson D, Sherman ME, Szeszenia-Dabrowska N, et al. Ovarian cancer risk and common variation in the sex hormone-

binding globulin gene: a population-based case-control study. BMC Cancer 2007;7:60.

4.7 Tables

Table 4.1 Genotype frequencies in estrogen-related genes, adjusted for sampling weights[1]

		Controls				Cases			
		African American		Non-African American		African American		Non-African American	
		N	%	N	%	N	%	N	%
CYP19A1									
rs4646	AA	62	9	78	7	72	9	85	6
	AC	295	43	437	45	325	44	533	45
	CC	301	47	601	47	345	47	611	49
	Missing			1					
rs10046	CC	386	55	242	21	395	53	299	24
	CT	235	41	555	54	293	40	625	53
	TT	37	4	320	25	54	7	305	23
rs17601241	AA	1	0	8	1	1	0	9	1
	AG	60	10	184	17	70	9	223	18
	GG	596	90	925	83	671	90	997	82
	Missing	1							
rs700519	CC	458	67	1055	94	537	72	1156	94
	CT	177	29	62	6	193	26	71	6
	TT	23	4			12	2	2	0
rs28757184	CC	606	93	1027	90	662	89	1145	92
	CT	50	7	88	9	78	11	84	8
	TT	2	0	2	1	2	0		
rs700518	AA	425	61	265	23	446	60	327	26
	AG	209	37	558	54	263	36	621	51
	GG	23	2	292	23	31	4	281	22
	Missing	1		2		2			
rs2414096	AA	23	2	290	23	30	4	278	22
	AG	208	37	559	54	264	36	621	52
	GG	427	61	267	23	448	60	327	26
	Missing			1				3	
rs727479	GG	32	4	127	10	36	5	163	13
	GT	208	32	521	51	252	34	575	47
	TT	418	64	469	39	454	61	491	40
rs2236722	TT	658	100	1117	100	742	100	1229	100
rs1008805	CC	30	5	206	20	28	4	227	18
	CT	209	32	512	48	237	32	599	50
	TT	419	64	397	32	477	64	403	33
	Missing			2					
rs6493494	AA	41	5	216	17	64	9	209	16

		Controls				Cases			
		African American		Non-African American		African American		Non-African American	
		N	%	N	%	N	%	N	%
	AG	254	45	527	48	307	41	588	50
	GG	363	51	374	35	371	50	432	34
rs749292	AA	138	18	227	17	188	26	219	17
	AG	313	52	535	48	356	48	602	51
	GG	207	30	355	35	198	26	408	32
rs1902586	AA	160	22	7	1	170	23	4	0
	AG	311	49	92	8	358	48	112	10
	GG	187	29	1018	91	214	29	1113	90
rs936306	CC	117	17	797	71	116	16	854	69
	CT	295	46	281	26	347	47	346	29
	TT	245	37	39	3	279	38	29	2
	Missing	1							
rs2445759	GG	571	84	948	85	613	83	1028	84
	GT	83	16	161	14	124	17	198	16
	TT	3	0	8	1	5	1	3	0
	Missing	1							
rs28566535	AA	171	26	1016	91	202	27	1105	89
	AC	307	49	92	8	353	47	119	10
	CC	179	26	9	1	187	25	5	0
	Missing	1							
rs3751591	CC	1	0	39	3	2	0	41	3
	CT	52	8	303	27	66	9	356	30
	TT	605	92	775	70	674	91	831	67
	Missing							1	
rs1902584	AA	571	83	936	82	655	88	1040	84
	AT	84	17	175	17	86	12	183	15
	TT	3	0	6	1	1	0	6	1
rs1004984	CC	187	26	406	37	212	29	444	37
	CT	344	55	553	46	388	52	573	46
	TT	126	19	158	17	140	19	209	17
	Missing	1				2		3	
rs28757081	CC	458	71	1112	99	525	70	1222	100
	CT	183	26	4	1	203	28	6	0
	TT	17	3	1	0	14	2	1	0
rs2445762	CC	101	13	76	7	97	13	118	10
	CT	315	50	481	41	363	49	492	39
	TT	242	37	560	53	281	38	617	50

		Controls				Cases			
		African American		Non-African American		African American		Non-African American	
		N	%	N	%	N	%	N	%
	Missing					1		2	
rs2470144	AA	26	3	257	19	26	3	286	23
	AG	148	22	584	56	186	25	615	49
	GG	484	75	276	25	530	72	328	27
rs2445765	CC	35	8	32	3	31	4	40	4
	CG	235	42	347	30	285	39	379	31
	GG	388	51	738	67	422	57	808	65
	Missing					4		2	
rs2446405	AA	189	33	34	3	214	29	44	4
	AT	320	48	354	30	356	48	381	31
	TT	148	19	728	66	170	23	804	65
	Missing	1		1		2			
ESR1									
rs851984	CC	384	57	414	33	407	55	418	34
	CT	226	35	518	52	280	38	586	49
	TT	46	7	184	16	55	7	223	17
	Missing	2		1				2	
rs851982	CC	42	7	187	16	41	6	223	17
	CT	215	34	514	51	261	35	585	49
	TT	401	59	415	33	440	59	419	34
	Missing			1				2	
rs2881766	GG	271	44	45	4	297	40	56	5
	GT	293	40	358	31	331	44	373	30
	TT	94	16	714	65	114	15	800	65
rs2077647	AA	155	27	306	26	200	27	333	26
	AG	346	48	566	46	372	50	614	51
	GG	155	25	244	28	170	23	282	23
	Missing	2		1					
rs532010	CC	133	18	170	22	129	18	186	15
	CT	326	49	524	42	354	48	569	48
	TT	197	33	420	36	255	35	474	37
	Missing	2		3		4			
rs17081703	AA	446	65	1111	99	503	68	1221	99
	AG	194	32	6	1	223	30	8	1
	GG	18	3			16	2		
rs12523805	GG	440	68	1103	98	497	67	1212	99
	GT	196	29	14	2	223	30	17	1

		Controls				Cases			
		African American		Non-African American		African American		Non-African American	
		N	%	N	%	N	%	N	%
	TT	22	3			22	3		
rs3866461	CC	113	15	20	1	99	14	29	2
	CT	301	45	246	21	373	50	282	22
	TT	243	40	851	77	270	36	918	75
	Missing	1							
rs7759411	CC	478	74	1086	97	580	78	1199	98
	CT	169	25	30	3	153	21	30	2
	TT	11	1	1	0	9	1		
rs11155813	CC	91	13	16	1	75	10	21	2
	CT	297	43	206	17	360	48	242	19
	TT	270	44	895	81	307	41	966	79
rs17081740	CC	4	0						
	CT	79	14	4	0	100	13	3	0
	TT	574	86	1113	100	637	87	1225	100
	Missing	1				5		1	
rs7761133	CC	256	42	30	2	263	36	47	4
	CT	297	40	300	26	363	49	318	26
	TT	105	18	786	72	116	16	858	70
	Missing			1				6	
rs7775047	CC	256	42	30	2	263	36	47	4
	CG	297	40	300	26	363	49	318	26
	GG	105	18	787	72	116	16	864	70
rs6903763	AA	12	1	3	0	14	2		
	AG	193	28	73	7	179	25	84	7
	GG	453	71	1041	93	548	73	1145	93
	Missing					1			
rs827421	CC	254	40	263	30	274	37	320	27
	CT	304	44	575	45	356	48	608	50
	TT	100	16	279	25	112	15	301	24
rs4870056	AA	159	24	212	18	168	23	266	22
	AG	336	52	575	51	391	53	586	48
	GG	163	24	328	31	182	25	373	30
	Missing			2		1		4	
rs2234693	CC	182	26	216	18	198	27	272	23
	CT	334	52	573	51	396	53	588	48
	TT	142	22	327	31	148	20	369	29
	Missing			1					

		Controls				Cases			
		African American		Non-African American		African American		Non-African American	
		N	%	N	%	N	%	N	%
rs9322332	AA	105	14	217	18	114	16	266	22
	AC	322	50	568	50	358	48	588	48
	CC	231	36	332	32	270	36	375	30
rs9340817	AA	12	2			24	3	2	0
	AC	193	29	52	4	217	29	66	6
	CC	453	70	1065	96	499	67	1161	94
	Missing					2			
rs712221	AA	90	16	387	35	96	13	437	35
	AT	304	44	550	51	338	46	582	48
	TT	264	40	180	14	308	41	210	17
rs1514348	AA	90	16	391	35	96	13	434	35
	AC	304	44	549	51	339	46	585	48
	CC	264	40	177	14	307	41	209	17
	Missing							1	
rs11155818	AA	5	1	2	0	7	1		
	AG	139	21	30	2	120	16	34	3
	GG	514	78	1085	97	615	83	1195	97
rs1709183	AA	263	37	560	48	296	40	612	50
	AG	314	54	486	46	340	46	500	41
	GG	80	10	71	6	106	14	112	9
	Missing	1						5	
rs9340835	AA	61	11	126	11	81	11	154	12
	AG	275	39	516	50	309	42	532	45
	GG	322	50	474	39	350	47	541	43
	Missing			1		2		2	
rs9322335	CC	316	45	626	53	338	46	682	55
	CT	277	42	428	42	317	43	457	37
	TT	65	13	63	5	86	12	89	8
	Missing					1		1	
rs9322336	CC	1	0	41	3	1	0	62	5
	CT	66	10	353	35	58	8	399	32
	TT	591	90	723	62	683	92	768	62
rs9322337	AA	56	6	4	0	72	10	8	1
	AC	287	45	84	7	280	38	93	7
	CC	315	48	1029	93	390	52	1128	93
rs9322338	AA	172	28	4	0	237	32	3	0
	AG	339	54	49	4	333	45	42	3

		Controls				Cases			
		African American		Non-African American		African American		Non-African American	
		N	%	N	%	N	%	N	%
	GG	143	19	1064	96	168	23	1182	97
	Missing	4				4		2	
rs6557170	AA	10	2	59	4	7	1	81	6
	AG	131	20	403	40	116	16	440	36
	GG	517	78	655	56	619	83	708	58
rs6557171	CC	193	29	496	44	225	30	554	45
	CT	323	48	502	47	357	48	533	43
	TT	141	23	119	9	159	22	142	12
	Missing	1				1			
rs9340888	AA	526	78	1109	98	608	82	1228	100
	AC	129	22	8	2	128	17	1	0
	CC	3	0			6	1		
rs12154178	AA	116	14	494	42	129	17	549	45
	AC	326	52	504	49	366	49	526	42
	CC	215	34	118	9	245	33	144	13
	Missing	1		1		2		10	
rs7739274	AA	48	7	1	0	61	8		
	AG	244	41	86	6	284	39	91	8
	GG	364	52	1030	94	393	53	1135	92
	Missing	2				4		3	
rs4583998	AA	204	32	117	9	231	31	143	13
	AG	331	53	504	48	368	50	526	42
	GG	121	15	495	42	142	19	559	45
	Missing	2		1		1		1	
rs6911230	AA	334	52	138	12	349	48	162	14
	AG	261	40	522	49	330	44	553	44
	GG	62	8	457	40	61	8	514	42
	Missing	1				2			
rs1801132	CC	507	76	663	57	599	81	717	58
	CG	144	23	400	39	134	18	433	35
	GG	7	1	54	4	8	1	79	7
	Missing					1			
rs9397459	AA	12	1			6	1		
	AG	135	18	6	0	148	20	5	0
	GG	511	81	1111	100	588	79	1224	100
rs3020410	AA	16	2	21	2	21	3	26	3
	AC	177	27	262	27	176	24	285	23

		Controls				Cases			
		African American		Non-African American		African American		Non-African American	
		N	%	N	%	N	%	N	%
	CC	465	71	831	70	544	73	916	75
	Missing			3		1		2	
rs3003917	AA	272	42	719	61	322	43	774	63
	AG	312	48	356	37	337	46	400	32
	GG	74	11	42	3	81	11	55	5
	Missing					2			
rs6914211	AA	27	4	13	1	45	6	40	4
	AT	229	39	298	24	274	37	278	23
	TT	402	56	806	75	423	57	911	73
rs9383599	CC	561	88	1114	100	622	83	1224	100
	CG	92	11	3	0	115	16	5	0
	GG	5	0			5	1		
rs3020314	CC	315	49	128	9	339	46	181	16
	CT	278	43	523	51	339	45	526	42
	TT	65	8	466	40	63	8	522	42
	Missing					1			
rs3020401	AA	91	12	467	40	86	11	533	43
	AG	312	48	525	51	361	48	524	41
	GG	253	40	124	9	294	40	171	15
	Missing	2		1		1		1	
rs985191	AA	400	58	801	73	411	55	900	72
	AC	233	38	299	26	286	39	288	24
	CC	25	5	17	1	45	6	41	4
rs3003925	AA	438	69	689	60	534	72	772	62
	AG	199	28	384	36	195	26	402	34
	GG	21	3	44	4	13	2	55	5
rs6557177	CC	55	12	40	3	69	9	64	6
	CT	288	44	366	29	328	44	354	29
	TT	315	44	711	68	345	47	811	65
rs2982699	AA	19	3	33	3	17	2	49	4
	AG	198	28	332	33	216	29	349	29
	GG	440	70	752	64	509	69	831	67
	Missing	1							
rs985695	CC	219	28	707	67	233	31	805	65
	CT	329	52	366	30	397	54	358	29
	TT	110	19	44	3	112	15	66	6
rs1884049	CC	253	42	745	64	286	38	819	66

		Controls				Cases			
		African American		Non-African American		African American		Non-African American	
		N	%	N	%	N	%	N	%
	CT	313	46	335	33	361	49	355	30
	TT	90	12	37	3	94	13	49	4
	Missing	2				1		6	
rs3020318	CC	31	4	423	37	29	4	484	39
	CT	213	31	533	51	256	34	544	43
	TT	414	65	161	12	457	62	201	19
rs1884053	CC	414	65	164	12	459	62	203	19
	CT	214	31	535	51	254	34	547	43
	TT	30	4	418	37	29	4	479	38
rs9383951	CC	12	1			13	2		
	CG	151	20	8	0	193	26	6	0
	GG	495	79	1109	100	536	72	1222	100
	Missing							1	
rs3020403	CC	28	4	497	51	30	4	571	45
	CG	225	31	508	41	256	35	513	42
	GG	405	65	112	8	456	61	144	13
	Missing							1	
rs3020404	AA	591	91	593	48	657	88	655	54
	AG	66	9	442	44	83	12	483	39
	GG	1	0	82	8	2	0	90	7
	Missing							1	
rs9397462	AA	24	3			20	3	1	0
	AT	172	25	11	1	205	28	14	1
	TT	462	72	1106	99	517	70	1214	99
rs9397463	CC	407	59	772	73	438	59	867	69
	CT	215	34	320	25	261	35	314	26
	TT	35	7	25	2	42	6	48	5
	Missing	1				1			
rs926777	AA	204	36	65	6	237	32	101	9
	AC	308	45	465	36	350	47	456	38
	CC	145	19	587	59	154	21	672	53
	Missing	1				1			
rs2982684	AA	34	6	5	1	27	4	14	1
	AC	188	25	207	15	197	26	205	18
	CC	436	70	905	84	518	70	1010	81
rs9371236	AA	382	60	1063	94	454	61	1177	96
	AG	237	36	51	5	248	34	48	3

		Controls				Cases			
		African American		Non-African American		African American		Non-African American	
		N	%	N	%	N	%	N	%
	GG	39	4	2	0	37	5	3	0
	Missing			1		3		1	
rs3020407	AA	30	5	483	46	33	4	549	43
	AG	205	27	520	45	247	33	523	43
	GG	423	68	113	9	461	62	155	13
	Missing			1		1		2	
rs2144025	CC	219	32	811	76	255	34	870	70
	CT	323	48	281	23	367	49	326	27
	TT	115	20	25	2	120	16	33	3
	Missing	1							
rs12212176	CC	608	92	747	68	694	94	835	68
	CT	50	8	335	27	48	6	350	29
	TT			35	5			44	3
rs11964865	AA	502	78	1062	94	582	78	1161	95
	AT	146	21	55	6	149	20	68	5
	TT	10	1			11	2		
rs7754762	AA	79	18	13	1	89	12	23	2
	AT	294	40	228	17	309	42	249	22
	TT	285	42	875	82	343	46	957	76
	Missing			1		1			
rs13192976	AA	393	55	879	83	450	61	957	77
	AT	224	35	221	16	250	34	242	22
	TT	41	10	12	1	42	6	22	2
	Missing			5				8	
rs9340944	AA	9	1	28	2	11	1	29	2
	AG	135	19	275	22	160	22	321	26
	GG	514	80	814	76	571	77	879	72
rs6905370	AA	226	37	88	7	244	33	115	9
	AG	312	44	458	36	381	52	496	42
	GG	120	19	571	58	117	15	618	49
rs9340971	AA	529	77	1084	98	603	81	1201	98
	AG	119	18	33	2	131	18	28	2
	GG	10	4			8	1		
rs7755185	AA	184	29	514	54	178	24	570	45
	AG	322	49	495	38	394	54	529	45
	GG	152	22	108	8	170	23	130	10
rs17082028	CC	8	1			7	1		

		Controls				Cases			
		African American		Non-African American		African American		Non-African American	
		N	%	N	%	N	%	N	%
	CT	120	14	5	0	147	20		
	TT	530	85	1112	100	588	79	1229	100
rs2207232	CC	20	3	9	1	28	4	21	2
	CT	184	31	227	17	197	27	241	22
	TT	453	67	881	83	517	69	967	77
	Missing	1							
rs9397472	AA	490	71	943	88	559	75	1029	82
	AG	156	28	167	11	165	22	186	17
	GG	12	1	7	1	18	2	14	1
rs2982712	CC	509	79	225	16	560	75	259	21
	CT	136	19	541	52	173	23	585	48
	TT	13	2	350	32	9	1	385	31
	Missing			1					
rs3020434	CC	489	76	774	74	524	71	851	69
	CT	161	23	304	24	203	27	344	29
	TT	8	1	39	2	15	2	34	2
rs3020364	AA	114	14	478	51	122	16	528	42
	AG	304	43	511	40	348	47	551	46
	GG	240	43	128	9	272	37	150	12
rs6901451	AA	130	26	12	1	147	20	25	2
	AG	326	49	230	17	356	48	253	23
	GG	202	25	875	82	239	32	951	75
rs3020368	CC	495	77	907	85	537	73	999	81
	CT	154	22	194	14	189	25	213	17
	TT	9	1	16	1	16	2	17	1
rs9341008	AA	6	0			5	1		
	AG	94	14	2	0	99	14		
	GG	558	85	1115	100	638	86	1229	100
rs2207396	AA	43	5	58	3	58	8	59	4
	AG	255	35	390	33	280	38	426	34
	GG	360	60	669	64	404	55	744	61
rs3020371	CC	108	13	490	53	115	16	532	42
	CT	326	50	497	37	353	47	550	46
	TT	224	37	130	9	274	37	147	12
rs3798569	AA	126	23	13	1	146	20	26	2
	AG	324	51	230	16	354	48	250	23
	GG	208	25	874	83	242	32	953	75

		Controls				Cases			
		African American		Non-African American		African American		Non-African American	
		N	%	N	%	N	%	N	%
rs3778082	AA	107	20	14	1	134	18	28	2
	AG	316	51	231	17	351	47	251	23
	GG	235	29	872	82	257	35	950	75
rs3020375	AA	54	7	467	51	50	7	515	41
	AC	267	39	514	40	315	43	558	47
	CC	337	53	135	10	377	51	156	12
	Missing			1					
rs12055837	AA	272	38	875	82	299	40	951	75
	AT	298	47	228	17	340	46	249	23
	TT	88	15	14	1	103	14	29	2
rs2459107	AA	283	41	132	9	308	42	152	12
	AG	292	45	501	39	353	48	548	46
	GG	82	13	484	52	81	11	529	42
	Missing	1							
rs11155833	AA	257	37	875	82	282	39	949	75
	AG	293	46	228	17	342	47	248	23
	GG	101	17	13	1	102	14	27	2
	Missing	7		1		16		5	
rs13192678	CC	258	36	875	82	283	38	951	75
	CT	295	46	227	17	350	47	249	23
	TT	105	17	15	1	109	15	29	2
rs3020381	AA	236	33	133	10	272	37	157	12
	AT	317	50	508	39	349	47	554	47
	TT	105	17	476	51	121	16	518	41
rs2474148	GG	262	45	535	55	299	40	588	47
	GT	318	44	474	38	337	45	522	43
	TT	78	11	107	7	106	14	118	10
	Missing			1				1	
rs9341052	AA	638	97	976	89	721	97	1080	88
	AG	20	3	136	11	21	3	146	12
	GG			5	0			3	0
rs3020383	CC	1	0	15	1	1	0	15	1
	CG	43	6	168	12	50	7	184	15
	GG	614	94	934	87	691	93	1030	84
rs3778099	CC	18	2	10	1	26	3	18	1
	CT	192	28	202	15	211	29	220	19
	TT	443	69	904	84	503	68	989	79

		Controls				Cases			
		African American		Non-African American		African American		Non-African American	
		N	%	N	%	N	%	N	%
	Missing	5		1		2		2	
rs9341062	AA	569	86	1053	95	625	85	1148	93
	AG	84	13	63	5	112	15	81	7
	GG	5	0	1	0	4	1		
	Missing					1			
rs9341070	CC	658	100	1109	100	742	100	1223	99
	CT			8	0			6	1
rs3798577	CC	130	23	268	21	140	19	277	22
	CT	309	47	541	51	367	49	609	51
	TT	219	31	308	28	235	32	342	27
	Missing							1	
rs3798758	GG	513	81	1036	94	586	79	1136	92
	GT	130	17	79	6	147	20	90	8
	TT	15	2	2	0	9	1	1	0
	Missing							2	
rs2813543	AA	3	0	49	3	3	0	61	5
	AG	61	8	359	33	62	8	385	31
	GG	594	92	709	64	677	91	782	64
	Missing							1	
HSD3B1									
rs932603	CC	68	13	2	0	105	14	2	0
	CT	287	44	15	1	325	44	19	1
	TT	303	43	1100	99	312	42	1208	99
rs6671149	GG	499	75	1111	100	549	74	1218	99
	GT	143	22	5	0	175	24	11	1
	TT	16	2	1	0	17	2		
	Missing					1			
rs3765945	CC	260	42	131	13	328	44	157	13
	CT	304	42	506	45	328	44	538	42
	TT	94	16	479	42	86	12	532	45
	Missing			1				2	
rs6428830	AA	6	0	91	8	7	1	119	10
	AG	114	14	462	44	149	20	484	38
	GG	538	85	563	48	586	79	626	52
	Missing			1					
rs6205	CC	94	17	2	0	132	18	1	0
	CT	297	45	21	1	346	46	20	1

		Controls				Cases			
		African American		Non-African American		African American		Non-African American	
		N	%	N	%	N	%	N	%
	TT	267	38	1094	98	263	36	1208	99
	Missing					1			
rs6203	CC	541	84	375	42	626	84	374	31
	CT	111	15	527	38	106	14	622	50
	TT	6	1	215	20	10	1	233	20
rs10754400	GG	267	43	118	11	323	43	134	11
	GT	299	43	480	44	338	46	527	41
	TT	92	14	519	45	81	11	568	48
HSD17B2									
rs4445895	CC	306	47	427	41	328	44	460	38
	CT	281	43	525	47	340	46	607	49
	TT	71	9	165	12	74	10	162	13
rs8052451	CC	99	19	1	0	84	11	2	0
	CT	295	43	10	1	310	41	12	1
	TT	264	38	1106	99	348	47	1215	99
rs8059915	CC	408	68	177	14	421	57	187	15
	CG	211	27	546	57	264	36	620	51
	GG	33	5	392	29	57	8	419	34
	Missing	6		2				3	
rs16956274	CC	456	66	1112	99	540	73	1219	99
	CT	184	33	4	0	187	25	10	1
	TT	18	2	1	0	15	2		
rs11648233	AA	20	3	371	29	34	5	397	33
	AC	176	23	543	55	206	27	620	50
	CC	462	74	203	16	502	68	212	17
rs8049423	AA	424	62	1110	99	511	69	1222	100
	AC	208	35	6	1	207	28	7	0
	CC	25	3	1	0	24	3		
	Missing	1							
rs8050327	AA	281	41	1104	99	331	45	1219	99
	AC	290	38	12	1	332	45	9	1
	CC	87	21	1	0	79	11	1	0
rs8191072	AA	269	40	1104	99	319	43	1218	99
	AG	298	39	12	1	340	46	10	1
	GG	91	21	1	0	83	11	1	0
rs7200696	CC	166	30	2	0	150	20	2	0
	CG	300	44	46	3	374	50	45	3

		Controls				Cases			
		African American		Non-African American		African American		Non-African American	
		N	%	N	%	N	%	N	%
	GG	190	25	1068	97	217	29	1176	97
	Missing	2		1		1		6	
rs8045494	AA	32	3	1	0	30	4		
	AG	221	39	7	1	224	30	10	1
	GG	405	58	1109	99	488	66	1219	99
rs4243229	AA	6	1			10	1	1	0
	AG	115	18	29	3	135	18	25	2
	GG	537	82	1088	97	597	80	1203	98
rs7196087	CC	410	57	1093	98	456	61	1213	99
	CT	221	37	23	2	248	34	16	1
	TT	27	6	1	0	38	5		
rs16956326	CC	452	64	1113	100	519	70	1225	100
	CT	186	30	4	0	201	27	4	0
	TT	20	6			22	3		
rs8191102	CC	394	53	1110	100	461	62	1221	99
	CT	219	41	7	0	247	33	7	1
	TT	45	6			34	5	1	0
rs8191136	AG	22	4			16	2		
	GG	636	96	1117	100	726	98	1229	100
rs8191167	AA	318	43	1104	99	407	56	1207	99
	AG	277	50	12	1	270	36	15	1
	GG	61	7			59	8		
	Missing	2		1		6		7	
rs2966245	CC	76	10	199	23	95	13	220	18
	CT	315	47	549	48	347	47	617	51
	TT	267	43	369	28	300	41	390	31
	Missing							2	
rs2955159	AA	62	12	1	0	72	10	2	0
	AG	254	39	26	2	294	40	23	2
	GG	339	49	1090	98	374	50	1204	98
	Missing	3				2			
rs2042429	AA	263	41	306	25	291	39	338	27
	AG	317	48	578	47	353	47	630	51
	GG	78	10	233	28	98	13	261	21
rs3111351	AA	163	24	1085	98	206	28	1200	98
	AG	329	46	30	2	348	47	26	2
	GG	166	30	2	0	188	25	3	0

		Controls				Cases			
		African American		Non-African American		African American		Non-African American	
		N	%	N	%	N	%	N	%
rs2966244	CC	437	60	1055	94	475	64	1152	94
	CT	197	33	60	6	240	32	76	6
	TT	24	6	2	0	27	4	1	0
rs9889094	CC	476	71	1029	92	555	75	1127	92
	CT	166	26	87	8	172	23	101	8
	TT	16	3	1	0	15	2	1	0
rs7196807	CC	393	58	502	42	456	62	548	44
	CT	241	39	497	49	246	33	526	44
	TT	24	3	116	9	40	5	155	12
	Missing			2					
rs9319572	CC	197	34	161	21	238	32	209	16
	CG	341	47	532	45	368	50	559	47
	GG	120	19	424	34	136	18	461	36
rs1364287	CC	488	77	507	50	521	71	557	45
	CT	158	21	494	41	202	27	528	44
	TT	10	2	115	9	16	2	142	11
	Missing	2		1		3		2	
rs723013	GG	34	7	123	10	66	9	160	12
	GT	254	39	493	49	272	36	526	44
	TT	370	54	501	41	404	55	543	43
rs6564962	AA	166	24	420	34	174	24	459	36
	AG	340	50	536	45	380	51	561	48
	GG	151	26	160	21	188	26	209	16
	Missing	1		1					
rs2911418	CC	19	2			23	3		
	CT	166	23	31	11	214	29	35	3
	TT	473	74	1086	89	505	68	1194	97
rs2966250	AA	458	70	1076	88	499	67	1185	96
	AG	182	28	41	12	218	29	43	3
	GG	18	2			25	4	1	0
rs2955153	AA	353	52	1082	89	399	54	1190	97
	AG	264	43	35	11	283	38	37	3
	GG	41	5			60	8	2	0
rs2911420	AA	20	3	129	10	28	4	157	12
	AC	212	37	515	45	235	32	548	46
	CC	425	61	472	46	473	64	521	42
	Missing	1		1		6		3	

		Controls				Cases			
		African American		Non-African American		African American		Non-African American	
		N	%	N	%	N	%	N	%
rs8191232	CC	16	4			12	2		
	CT	145	20	9	1	164	22	5	0
	TT	497	76	1108	99	566	76	1224	100
rs6564964	GG	89	14	143	12	111	15	175	14
	GT	304	49	529	52	343	46	555	46
	TT	264	37	444	36	288	39	498	40
	Missing	1		1				1	
rs2966248	GG	311	44	468	45	336	45	517	42
	GT	282	45	520	45	321	43	547	46
	TT	65	11	128	10	85	12	161	13
	Missing			1				4	
rs1364286	AA	172	25	443	36	196	27	497	40
	AG	343	51	531	52	362	49	550	46
	GG	142	24	142	12	182	25	181	14
	Missing	1		1		2		1	
rs2955163	CC	439	64	1012	83	536	72	1107	90
	CG	199	33	103	17	183	25	117	10
	GG	20	3	2	0	23	3	5	0
rs1364285	CC	261	35	453	36	285	39	504	40
	CG	309	51	527	52	346	46	551	46
	GG	88	14	137	11	111	15	174	14
rs2955162	AA	9	4	62	5	5	1	61	5
	AG	245	39	426	35	312	42	459	37
	GG	404	57	629	61	424	57	708	58
	Missing					1		1	
rs2966246	CC	46	6	2	0	61	8	4	0
	CG	264	43	109	17	250	34	124	10
	GG	348	51	1006	82	431	58	1101	89
rs1424151	CC	4	0	1	0	8	1	2	0
	CT	122	23	83	7	92	12	98	8
	TT	532	77	1033	93	641	87	1129	92
	Missing					1			
PGR									
rs11224565	AA	11	4			15	2		
	AG	147	24	8	1	179	25	8	0
	GG	500	72	1109	99	548	73	1221	100
rs11224566	CC	2	0			1	0		

		Controls				Cases			
		African American		Non-African American		African American		Non-African American	
		N	%	N	%	N	%	N	%
	CG	63	13	4	0	71	10	4	0
	GG	593	87	1113	100	670	90	1225	100
rs546763	AA	105	16	78	8	92	12	83	7
	AC	302	44	409	35	340	46	479	39
	CC	251	40	629	57	309	42	665	54
	Missing			1		1		2	
rs548668	CC	105	16	78	8	92	12	83	7
	CT	304	45	409	35	341	46	480	39
	TT	248	39	629	57	309	42	666	54
	Missing	1		1					
rs492827	CC	186	31	79	8	191	26	84	7
	CT	312	47	415	35	356	48	485	39
	TT	160	22	623	57	195	26	659	54
	Missing							1	
rs11224570	AA	6	3			8	1		
	AT	109	18	5	0	141	19	2	0
	TT	543	79	1112	100	593	80	1227	100
rs11571247	AA	495	74	1112	100	593	80	1227	100
	AG	150	24	5	0	140	19	2	0
	GG	13	2			8	1		
	Missing					1			
rs578029	AA	73	11	89	15	87	12	77	7
	AT	321	50	453	35	339	46	480	40
	TT	263	40	575	50	316	43	668	53
	Missing	1						4	
rs11224575	AA	11	2	41	4	18	2	44	4
	AG	124	14	300	24	197	27	355	28
	GG	517	84	776	72	517	71	830	69
	Missing	6				10			
rs543936	AA	213	29	605	52	235	32	698	56
	AG	315	49	438	43	362	49	457	38
	GG	130	22	73	5	144	20	74	6
	Missing			1		1			
rs679275	AA	122	22	73	5	137	19	73	6
	AG	323	49	439	43	369	50	459	38
	GG	213	29	605	52	236	32	697	56
rs693765	GG	526	80	1112	100	620	84	1223	100

		Controls				Cases			
		African American		Non-African American		African American		Non-African American	
		N	%	N	%	N	%	N	%
	GT	123	18	5	0	118	16	6	0
	TT	9	2			4	1		
rs1824128	GG	527	84	820	75	532	72	862	71
	GT	121	14	266	22	193	26	327	26
	TT	10	2	31	3	17	2	39	3
	Missing							1	
rs660149	CC	331	50	579	51	398	54	667	53
	CG	283	42	451	35	281	38	488	40
	GG	44	8	87	15	63	9	74	6
rs495997	AA	215	36	394	38	264	36	397	33
	AG	329	49	556	45	357	48	615	49
	GG	114	15	167	16	121	16	217	18
rs2124761	GG	428	65	1095	98	541	73	1201	97
	GT	198	31	21	2	185	25	28	3
	TT	32	4	1	0	16	2		
rs11224579	CC	506	74	924	84	527	71	1004	82
	CT	144	22	181	15	203	28	207	16
	TT	8	4	12	1	12	2	17	2
	Missing							1	
rs503602	AA	87	14	19	2	69	9	27	3
	AC	268	39	272	23	286	39	302	24
	CC	303	47	826	75	387	52	900	74
rs653752	CC	160	22	150	10	217	29	163	14
	CG	338	53	524	52	363	49	576	47
	GG	160	25	443	38	161	22	489	39
	Missing					1		1	
rs538915	CC	480	72	841	76	561	76	923	76
	CT	166	25	260	22	165	22	284	22
	TT	12	3	15	2	16	2	22	2
	Missing			1					
rs555653	AA	216	29	280	23	264	36	306	25
	AG	331	50	594	57	366	49	625	51
	GG	110	21	243	20	111	15	298	24
	Missing	1				1			
rs11224590	GG	565	82	1113	100	637	86	1228	100
	GT	90	16	4	0	101	14	1	0
	TT	3	3			4	1		

		Controls				Cases			
		African American		Non-African American		African American		Non-African American	
		N	%	N	%	N	%	N	%
rs11224591	AA	565	82	1113	100	636	86	1228	100
	AG	90	16	4	0	101	14	1	0
	GG	3	3			4	1		
	Missing					1			
rs596223	AA	514	82	494	42	582	79	554	45
	AG	132	17	508	49	147	20	557	46
	GG	9	1	113	9	12	2	113	9
	Missing	3		2		1		5	
rs501732	CC	118	17	35	2	161	22	36	3
	CT	325	52	329	29	355	48	349	28
	TT	215	32	753	68	225	30	844	69
	Missing					1			
rs10895068	AA			3	0	1	0	4	1
	AG	18	3	117	10	18	2	146	11
	GG	640	97	996	89	723	97	1079	88
	Missing			1					
SHBG									
rs1799941	AA	3	0	68	5	1	0	88	8
	AG	96	13	420	34	91	12	435	36
	GG	559	86	628	61	650	88	706	56
	Missing			1					

1- counts (N) reflect raw data. Genotype percentages are adjusted for study sampling probabilities.

Table 4.2 Association between estrogen-related candidate gene SNPs and breast cancer

	Cases	Controls	OR (95% CI)[1]	CLR[2]	P-value
CYP19A1					
rs4646					
AA	157	140	1.00 (0.77 , 1.29)	1.7	0.9699
AC	859	733	1.15 (1.00 , 1.33)	1.3	0.0499
CC	956	902	Referent		
rs2445759					
TT+GT	330	255	1.24 (1.03 , 1.50)	1.5	0.0213
GG	1642	1520	Referent		
ESR1					
rs7759411					
TT+CT	192	211	0.77 (0.61 , 0.97)	1.6	0.0255
CC	1780	1565	Referent		
rs827421					
CC+CT	1558	1397	1.06 (0.90 , 1.25)	1.4	0.5009
TT	414	379	Referent		
rs11155818					
AA+AG	161	176	0.77 (0.60 , 0.98)	1.6	0.0318
GG	1811	1600	Referent		
rs1709183					
GG	218	151	1.37 (1.08 , 1.74)	1.6	0.0099
AG	841	801	0.93 (0.81 , 1.07)	1.3	0.3182
AA	908	823	Referent		
rs9340835					
AA	235	187	1.19 (0.96 , 1.48)	1.5	0.1027
AG+GG	1733	1588	Referent		
rs9322335					
TT	175	128	1.32 (1.02 , 1.71)	1.7	0.0333
CT	775	706	0.99 (0.86 , 1.14)	1.3	0.8759
CC	1020	942	Referent		
rs9322338					
AA	240	176	1.18 (0.88 , 1.60)	1.8	0.2718
AG	375	388	0.81 (0.63 , 1.03)	1.6	0.0876
GG	1351	1208	Referent		
rs9340888					
CC+AC	135	140	0.76 (0.58 , 1.00)	1.7	0.0485
AA	1837	1636	Referent		
rs6914211					
AA	85	40	2.24 (1.51 , 3.33)	2.2	0.0001
AT+TT	1887	1736	Referent		

	Cases	Controls	OR (95% CI)[1]	CLR[2]	P-value
rs3020314					
TT	585	532	0.92 (0.75 , 1.12)	1.5	0.3924
CT	866	801	0.84 (0.70 , 1.00)	1.4	0.0442
CC	520	443	Referent		
rs3020401					
GG	465	377	1.25 (1.05 , 1.48)	1.4	0.0102
AA+AG	1505	1396	Referent		
rs985191					
CC	86	42	2.11 (1.43 , 3.13)	2.2	0.0002
AC	574	532	0.99 (0.86 , 1.16)	1.4	0.9447
AA	1312	1202	Referent		
rs3003925					
GG	68	65	0.95 (0.66 , 1.37)	2.1	0.7779
AG	598	583	0.86 (0.74 , 0.99)	1.3	0.0382
AA	1306	1128	Referent		
rs6557177					
CC	133	95	1.38 (1.04 , 1.83)	1.8	0.0258
CT+TT	1839	1681	Referent		
rs9397463					
TT	90	60	1.38 (0.97 , 1.95)	2.0	0.0731
CC+CT	1881	1715	Referent		
rs2982684					
AA	41	39	0.87 (0.54 , 1.39)	2.6	0.5566
AC	402	395	0.89 (0.75 , 1.05)	1.4	0.1672
CC	1529	1342	Referent		
rs9340944					
AA+AG	521	448	1.10 (0.94 , 1.29)	1.4	0.2185
GG	1451	1328	Referent		
rs2207232					
CC	49	29	1.77 (1.09 , 2.86)	2.6	0.0202
CT	438	411	0.96 (0.81 , 1.12)	1.4	0.5872
TT	1485	1335	Referent		
rs3020381					
AA	429	369	1.14 (0.94 , 1.40)	1.5	0.1927
AT	904	826	1.06 (0.91 , 1.25)	1.4	0.4430
TT	639	581	Referent		
rs9341052					
AG+GG	171	161	1.00 (0.78 , 1.27)	1.6	0.9819
AA	1801	1615	Referent		
rs3778099					

	Cases	Controls	OR (95% CI)[1]	CLR[2]	P-value
CC	44	28	1.51 (0.92 , 2.48)	2.7	0.1058
CT	431	394	0.96 (0.82 , 1.14)	1.4	0.6648
TT	1493	1348	Referent		
HSD3B1					
rs932603					
CC	107	70	1.48 (1.06 , 2.06)	1.9	0.0207
CT+TT	1865	1706	Referent		
rs3765945					
TT	618	574	0.83 (0.68 , 1.02)	1.5	0.0738
CT	867	810	0.83 (0.70 , 1.00)	1.4	0.0460
CC	485	391	Referent		
rs6428830					
AA	126	97	1.21 (0.90 , 1.62)	1.8	0.2047
AG+GG	1846	1678	Referent		
rs6205					
CC	133	96	1.48 (1.07 , 2.05)	1.9	0.0178
CT	366	318	1.20 (0.95 , 1.51)	1.6	0.1188
TT	1472	1362	Referent		
rs10754400					
TT	649	612	0.86 (0.70 , 1.05)	1.5	0.1459
GT	866	779	0.93 (0.77 , 1.11)	1.4	0.4243
GG	457	385	Referent		
HSD17B2					
rs8052451					
CC	86	100	0.56 (0.40 , 0.80)	2.0	0.0012
CT	322	305	0.81 (0.64 , 1.02)	1.6	0.0688
TT	1564	1371	Referent		
rs8059915					
GG	477	425	1.23 (1.00 , 1.52)	1.5	0.0473
CG	884	758	1.22 (1.03 , 1.45)	1.4	0.0229
CC	608	585	Referent		
rs16956274					
CT+TT	212	207	0.85 (0.67 , 1.08)	1.6	0.1794
CC	1760	1569	Referent		
rs8049423					
CC+AC	238	240	0.79 (0.63 , 1.00)	1.6	0.0480
AA	1734	1535	Referent		
rs8050327					
CC	80	88	0.67 (0.47 , 0.95)	2.0	0.0266
AC	341	302	0.94 (0.75 , 1.18)	1.6	0.6179

	Cases	Controls	OR (95% CI)[1]	CLR[2]	P-value
AA	1551	1386	Referent		
rs8191072					
GG	84	92	0.67 (0.47 , 0.96)	2.0	0.0286
AG	350	310	0.96 (0.76 , 1.20)	1.6	0.6953
AA	1538	1374	Referent		
rs7200696					
CC	152	168	0.69 (0.53 , 0.90)	1.7	0.0061
CG+GG	1813	1605	Referent		
rs7196087					
TT	38	28	1.15 (0.68 , 1.96)	2.9	0.5917
CT	264	244	0.94 (0.76 , 1.18)	1.6	0.6051
CC	1670	1504	Referent		
rs16956326					
TT	22	20	0.89 (0.46 , 1.69)	3.7	0.7149
CT	205	190	0.90 (0.71 , 1.15)	1.6	0.4087
CC	1745	1566	Referent		
rs8191102					
TT	35	45	0.63 (0.39 , 1.03)	2.6	0.0632
CT	254	226	0.90 (0.72 , 1.14)	1.6	0.3911
CC	1683	1505	Referent		
rs8191167					
GG	59	61	0.71 (0.48 , 1.07)	2.2	0.0994
AG	285	289	0.71 (0.56 , 0.88)	1.6	0.0024
AA	1615	1423	Referent		
rs2955159					
AA	74	63	1.05 (0.72 , 1.53)	2.1	0.8121
AG	317	280	1.03 (0.83 , 1.28)	1.5	0.8001
GG	1579	1430	Referent		
rs3111351					
GG	191	168	0.85 (0.63 , 1.16)	1.8	0.3080
AG	374	359	0.84 (0.66 , 1.08)	1.6	0.1705
AA	1407	1249	Referent		
rs9319572r					
CC	448	358	1.20 (1.02 , 1.42)	1.4	0.0311
GG+CG	1524	1418	Referent		
rs1364287					
TT	159	125	1.27 (0.97 , 1.66)	1.7	0.0861
CT	730	653	1.05 (0.90 , 1.21)	1.3	0.5484
CC	1078	995	Referent		
rs723013					

	Cases	Controls	OR (95% CI)[1]	CLR[2]	P-value
GG	227	157	1.41 (1.11 , 1.78)	1.6	0.0044
GT	798	748	0.97 (0.84 , 1.12)	1.3	0.7231
TT	947	871	Referent		
rs6564962					
GG	398	311	1.24 (1.02 , 1.51)	1.5	0.0313
AG	941	877	1.00 (0.85 , 1.16)	1.4	0.9596
AA	633	586	Referent		
rs2911418					
CC+CT	272	216	1.22 (0.98 , 1.51)	1.5	0.0752
TT	1700	1560	Referent		
rs2955153					
GG	62	41	1.43 (0.94 , 2.19)	2.3	0.0929
AA+AG	1910	1735	Referent		
rs1364286					
GG	363	284	1.24 (1.04 , 1.49)	1.4	0.0182
AA+AG	1606	1490	Referent		
rs1364285					
GG	285	225	1.24 (1.02 , 1.51)	1.5	0.0325
CC+CG	1687	1551	Referent		
rs1424151					
CC+CT	200	210	0.82 (0.66 , 1.02)	1.5	0.0789
TT	1771	1566	Referent		
PGR					
rs546763					
AA	175	183	0.81 (0.64 , 1.02)	1.6	0.0671
AC+CC	1794	1592	Referent		
rs548668					
CC	175	183	0.80 (0.64 , 1.01)	1.6	0.0645
CT+TT	1797	1591	Referent		
rs11571247					
GG+AG	150	168	0.77 (0.59 , 1.00)	1.7	0.0467
AA	1821	1608	Referent		
rs11224575					
AA	62	52	1.00 (0.67 , 1.50)	2.2	0.9888
AG	553	424	1.28 (1.10 , 1.50)	1.4	0.0018
GG	1347	1294	Referent		
rs693765					
TT+GT	128	137	0.76 (0.57 , 1.00)	1.7	0.0491
GG	1844	1639	Referent		
rs1824128					

	Cases	Controls	OR (95% CI)[1]	CLR[2]	P-value
TT+GT	577	428	1.33 (1.14 , 1.55)	1.4	0.0003
GG	1394	1348	Referent		
rs660149					
GG	137	131	0.90 (0.69 , 1.18)	1.7	0.4494
CG	770	735	0.93 (0.81 , 1.07)	1.3	0.2955
CC	1065	910	Referent		
rs2124761					
TT+GT	229	252	0.74 (0.59 , 0.92)	1.6	0.0073
GG	1743	1524	Referent		
rs11224579					
TT+CT	439	345	1.21 (1.03 , 1.43)	1.4	0.0237
CC	1532	1431	Referent		
rs503602					
AA	96	106	0.74 (0.54 , 1.01)	1.9	0.0590
AC	588	540	0.92 (0.79 , 1.07)	1.4	0.2866
CC	1288	1130	Referent		
rs596223					
GG	125	122	0.90 (0.68 , 1.19)	1.7	0.4754
AA+AG	1841	1649	Referent		
SHBG					
rs1799941					
AA	89	71	1.35 (0.96 , 1.89)	2.0	0.0830
AG+GG	1883	1704	Referent		

1 - odds ratio, 95% confidence interval, adjusted for age, self-identified race, African ancestry, offset term
2- CLR – confidence limit ratio, upper 95% confidence limit divided by lower 95% confidence limit

Table 4.3 Association between estrogen-related candidate SNPs and luminal A and basal-like breast cancer

	Luminal A					Basal-like					Luminal A vs. basal-like[3]
	Case	Control	OR (95% CI)[1]	CLR[2]	P-value	Case	Control	OR (95% CI)[1]	CLR[2]	P-value	P-value
CYP19A1											
rs4646											
AA	47	140	0.85 (0.59, 1.23)	2.1	0.3995	18	140	1.00 (0.58, 1.74)	3.0	0.9947	0.5987
AC	305	733	1.21 (1.00, 1.46)	1.5	0.0554	87	733	1.13 (0.83, 1.55)	1.9	0.4381	0.7114
CC	327	902	Referent			95	902	Referent			
rs2445759											
GT+TT	112	255	1.22 (0.95, 1.57)	1.7	0.1185	30	255	1.09 (0.72, 1.66)	2.3	0.6787	0.6209
GG	567	1520	Referent			170	1520	Referent			
ESR1											
rs7759411											
CT+TT	56	211	0.70 (0.50, 0.98)	2.0	0.0393	22	211	0.63 (0.38, 1.03)	2.7	0.0656	0.6933
CC	623	1565	Referent			178	1565	Referent			
rs827421											
CC+CT	550	1397	1.27 (1.01, 1.61)	1.6	0.0419	163	1397	1.14 (0.77, 1.67)	2.2	0.5207	0.5827
TT	129	379	Referent			37	379	Referent			
rs11155818											
AA+AG	51	176	0.76 (0.53, 1.07)	2.0	0.1182	18	176	0.60 (0.35, 1.02)	2.9	0.0587	0.4294
GG	628	1600	Referent			182	1600	Referent			
rs1709183											
GG	81	151	1.55 (1.14, 2.13)	1.9	0.0058	16	151	0.83 (0.47, 1.47)	3.1	0.5211	0.0384
AG	288	801	0.93 (0.76, 1.13)	1.5	0.4590	91	801	0.90 (0.65, 1.23)	1.9	0.4931	0.8354
AA	309	823	Referent			93	823	Referent			
rs9340835											
AA	88	187	1.35 (1.02, 1.79)	1.8	0.0374	25	187	1.36 (0.86, 2.14)	2.5	0.1901	0.9819

	Luminal A					Basal-like					Luminal A vs. basal-like[3]
	Case	Control	OR (95% CI)[1]	CLR[2]	P-value	Case	Control	OR (95% CI)[1]	CLR[2]	P-value	P-value
AG+GG	589	1588	Referent			175	1588	Referent			
rs9322335											
TT	62	128	1.38 (0.98, 1.95)	2.0	0.0622	22	128	1.38 (0.82, 2.30)	2.8	0.2236	0.9814
CT	267	706	0.99 (0.82, 1.21)	1.5	0.9329	73	706	0.88 (0.64, 1.21)	1.9	0.4263	0.4851
CC	350	942	Referent			104	942	Referent			
rs9322338											
AA	72	176	0.99 (0.66, 1.50)	2.3	0.9697	26	176	0.82 (0.45, 1.49)	3.3	0.5094	0.5572
AG	109	388	0.66 (0.47, 0.92)	2.0	0.0147	61	388	0.92 (0.56, 1.49)	2.6	0.7321	0.2126
GG	496	1208	Referent			113	1208	Referent			
rs9340888											
AC +CC	39	140	0.73 (0.49, 1.09)	2.2	0.1276	20	140	0.81 (0.48, 1.37)	2.9	0.4308	0.7479
AA	640	1636	Referent			180	1636	Referent			
rs6914211											
AA	33	40	2.70 (1.66, 4.39)	2.6	0.0001	10	40	2.19 (1.06, 4.53)	4.3	0.0349	0.5770
AT+TT	646	1736	Referent			190	1736	Referent			
rs3020314											
TT	209	532	0.95 (0.72, 1.26)	1.7	0.7386	45	532	0.79 (0.50, 1.25)	2.5	0.3112	0.4482
CT	310	801	0.89 (0.70, 1.14)	1.6	0.3602	88	801	0.80 (0.56, 1.15)	2.1	0.2284	0.5870
CC	159	443	Referent			67	443	Referent			
rs3020401											
GG	145	377	1.19 (0.94, 1.50)	1.6	0.1495	65	377	1.61 (1.14, 2.28)	2.0	0.0063	0.1084
AA+AG	534	1396	Referent			135	1396	Referent			
rs985191											
CC	32	42	2.47 (1.51, 4.04)	2.7	0.0003	9	42	1.98 (0.92, 4.24)	4.6	0.0801	0.5750
AC	203	532	1.08 (0.88, 1.33)	1.5	0.4525	74	532	1.34 (0.97, 1.84)	1.9	0.0749	0.2265

	Luminal A					Basal-like					Luminal A vs. basal-like[3]
	Case	Control	OR (95% CI)[1]	CLR[2]	P-value	Case	Control	OR (95% CI)[1]	CLR[2]	P-value	P-value
AA	444	1202	Referent			117	1202	Referent			
rs3003925											
GG	24	65	0.94 (0.57, 1.55)	2.7	0.8122	5	65	0.67 (0.26, 1.71)	6.6	0.3988	0.4939
AG	205	583	0.84 (0.69, 1.03)	1.5	0.0895	60	583	0.91 (0.65, 1.26)	1.9	0.5582	0.6693
AA	450	1128	Referent			135	1128	Referent			
rs6557177											
CC	47	95	1.48 (1.02, 2.15)	2.1	0.0397	19	95	1.72 (1.01, 2.93)	2.9	0.0455	0.6029
CT+TT	632	1681	Referent			181	1681	Referent			
rs9397463											
TT	35	60	1.62 (1.04, 2.53)	2.4	0.0329	10	60	1.31 (0.65, 2.66)	4.1	0.4463	0.5732
CC+CT	644	1715	Referent			189	1715	Referent			
rs2982684											
AA	9	39	0.59 (0.28, 1.26)	4.6	0.1742	7	39	1.12 (0.48, 2.63)	5.5	0.7929	0.2185
AC	132	395	0.85 (0.67, 1.07)	1.6	0.1592	39	395	0.80 (0.54, 1.16)	2.1	0.2393	0.7667
CC	538	1342	Referent			154	1342	Referent			
rs9340944											
AA+AG	196	448	1.26 (1.03, 1.55)	1.5	0.0262	50	448	1.14 (0.81, 1.61)	2.0	0.4634	0.5760
GG	483	1328	Referent			150	1328	Referent			
rs2207232											
CC	9	29	0.98 (0.45, 2.13)	4.7	0.9648	6	29	1.58 (0.63, 3.99)	6.4	0.3309	0.3802
CT	160	411	1.02 (0.82, 1.28)	1.5	0.8310	42	411	0.82 (0.56, 1.18)	2.1	0.2823	0.2574
TT	510	1335	Referent			152	1335	Referent			
rs3020381											
AA	154	369	1.39 (1.06, 1.83)	1.7	0.0171	47	369	0.99 (0.64, 1.54)	2.4	0.9632	0.1561
AT	309	826	1.15 (0.93, 1.43)	1.5	0.2016	91	826	0.95 (0.66, 1.36)	2.1	0.7904	0.3305

	Luminal A					Basal-like					Luminal A vs. basal-like[3]
	Case	Control	OR (95% CI)[1]	CLR[2]	P-value	Case	Control	OR (95% CI)[1]	CLR[2]	P-value	P-value
TT	216	581	Referent			62	581	Referent			
rs9341052											
AG+GG	61	161	1.01 (0.73, 1.40)	1.9	0.9512	9	161	0.63 (0.31, 1.27)	4.1	0.1925	0.1995
AA	618	1615	Referent			191	1615	Referent			
rs3778099											
CC	8	28	0.84 (0.38, 1.89)	5.0	0.6777	5	28	1.42 (0.53, 3.82)	7.3	0.4894	0.3728
CT	147	394	0.95 (0.75, 1.18)	1.6	0.6237	49	394	1.05 (0.73, 1.49)	2.0	0.8042	0.6018
TT	523	1348	Referent			145	1348	Referent			
HSD3B1											
rs932603											
CC	36	70	1.77 (1.14, 2.77)	2.4	0.0115	12	70	1.13 (0.58, 2.19)	3.8	0.7231	0.2083
CT+TT	643	1706	Referent			188	1706	Referent			
rs3765945											
TT	223	574	0.71 (0.54, 0.93)	1.7	0.0125	42	574	0.72 (0.45, 1.15)	2.6	0.1641	0.9662
CT	279	810	0.68 (0.53, 0.86)	1.6	0.0015	105	810	1.11 (0.76, 1.60)	2.1	0.5980	0.0162
CC	175	391	Referent			53	391	Referent			
rs6428830											
AA	50	97	1.36 (0.93, 1.98)	2.1	0.1099	18	97	2.37 (1.35, 4.15)	3.1	0.0026	0.0636
AG+GG	629	1678	Referent			182	1678	Referent			
rs6205											
CC	45	96	1.72 (1.10, 2.69)	2.4	0.0169	15	96	1.27 (0.65, 2.47)	3.8	0.4839	0.4053
CT	107	318	1.12 (0.81, 1.56)	1.9	0.4956	60	318	1.50 (0.96, 2.36)	2.5	0.0762	0.2531
TT	527	1362	Referent			125	1362	Referent			
rs10754400											
TT	236	612	0.73 (0.55, 0.96)	1.7	0.0255	43	612	0.73 (0.45, 1.18)	2.6	0.2013	0.9954

	Luminal A					Basal-like					Luminal A vs. basal-like[3]
	Case	Control	OR (95% CI)[1]	CLR[2]	P-value	Case	Control	OR (95% CI)[1]	CLR[2]	P-value	P-value
GT	279	779	0.75 (0.59, 0.96)	1.6	0.0249	107	779	1.25 (0.85, 1.83)	2.1	0.2499	0.0153
GG	164	385	Referent			50	385	Referent			
HSD17B2											
rs8052451											
CC	21	100	0.39 (0.23, 0.67)	2.9	0.0006	11	100	0.50 (0.25, 1.03)	4.2	0.0601	0.5505
CT	93	305	0.68 (0.49, 0.94)	1.9	0.0182	47	305	0.81 (0.52, 1.26)	2.4	0.3465	0.4691
TT	565	1371	Referent			142	1371	Referent			
rs8059915											
GG	167	425	1.24 (0.93, 1.65)	1.8	0.1432	38	425	1.12 (0.70, 1.81)	2.6	0.6288	0.7062
CG	320	758	1.29 (1.02, 1.64)	1.6	0.0324	81	758	1.07 (0.75, 1.54)	2.1	0.7108	0.3421
CC	192	585	Referent			81	585	Referent			
rs16956274											
CT+TT	54	207	0.65 (0.46, 0.93)	2.0	0.0172	25	207	0.65 (0.40, 1.06)	2.6	0.0861	0.9867
CC	625	1569	Referent			175	1569	Referent			
rs8049423											
AC+CC	64	240	0.66 (0.47, 0.92)	2.0	0.0157	28	240	0.60 (0.37, 0.95)	2.6	0.0303	0.7014
AA	615	1535	Referent			172	1535	Referent			
rs8050327											
CC	20	88	0.49 (0.28, 0.85)	3.0	0.0106	14	88	0.91 (0.46, 1.78)	3.8	0.7718	0.1231
AC	99	302	0.81 (0.59, 1.11)	1.9	0.1878	53	302	1.25 (0.80, 1.94)	2.4	0.3311	0.0811
AA	560	1386	Referent			133	1386	Referent			
rs8191072											
GG	21	92	0.48 (0.28, 0.83)	2.9	0.0082	14	92	0.86 (0.44, 1.68)	3.8	0.6555	0.1478
AG	101	310	0.80 (0.58, 1.09)	1.9	0.1599	54	310	1.23 (0.79, 1.91)	2.4	0.3696	0.0827
AA	557	1374	Referent			132	1374	Referent			

	Luminal A					Basal-like					Luminal A vs. basal-like[3]
	Case	Control	OR (95% CI)[1]	CLR[2]	P-value	Case	Control	OR (95% CI)[1]	CLR[2]	P-value	P-value
rs7200696											
CC	38	168	0.52 (0.35, 0.78)	2.2	0.0014	29	168	1.01 (0.63, 1.63)	2.6	0.9711	0.0205
CG+GG	639	1605	Referent			170	1605	Referent			
rs7196087											
TT	10	28	1.00 (0.47, 2.16)	4.6	0.9965	10	28	2.23 (1.00, 4.97)	4.9	0.0487	0.0940
CT	87	244	0.97 (0.71, 1.33)	1.9	0.8569	34	244	0.88 (0.56, 1.37)	2.4	0.5626	0.6791
CC	582	1504	Referent			156	1504	Referent			
rs16956326											
TT	5	20	0.71 (0.39, 1.27)	3.3	0.2467	7	20	2.02 (1.01, 4.07)	4.0	0.0479	0.0103
CT	62	190	0.84 (0.63, 1.13)	1.8	0.2467	38	190	1.42 (1.00, 2.02)	2.0	0.0479	0.0103
CC	612	1566	Referent			155	1566	Referent			
rs8191102											
TT	9	45	0.48 (0.22, 1.03)	4.6	0.0583	6	45	0.97 (0.39, 2.45)	6.3	0.9553	0.2019
CT	77	226	0.85 (0.61, 1.17)	1.9	0.3150	49	226	1.46 (0.95, 2.24)	2.4	0.0829	0.0252
CC	593	1505	Referent			145	1505	Referent			
rs8191167											
GG	14	61	0.48 (0.25, 0.90)	3.6	0.0231	10	61	0.92 (0.44 , 1.94)	4.4	0.8307	0.1455
AG	83	289	0.61 (0.44, 0.85)	1.9	0.0031	38	289	0.63 (0.40, 0.98)	2.4	0.0415	0.9259
AA	577	1423	Referent			151	1423	Referent			
rs2955159											
AA	25	63	1.15 (0.68, 1.94)	2.9	0.6029	14	63	1.51 (0.77, 2.95)	3.8	0.2257	0.4702
AG	96	280	0.99 (0.72, 1.34)	1.9	0.9266	45	280	1.13 (0.73, 1.73)	2.4	0.5837	0.5786
GG	558	1430	Referent			141	1430	Referent			
rs3111351											
GG	56	168	0.70 (0.46, 1.06)	2.3	0.0915	33	168	0.92 (0.53, 1.59)	3.0	0.7631	0.3692

	Luminal A					Basal-like					Luminal A vs. basal-like[3]
	Case	Control	OR (95% CI)[1]	CLR[2]	P-value	Case	Control	OR (95% CI)[1]	CLR[2]	P-value	P-value
AG	109	359	0.68 (0.48, 0.95)	2.0	0.0245	43	359	0.62 (0.38, 1.00)	2.6	0.0507	0.7223
AA	514	1249	Referent			124	1249	Referent			
rs9319572											
CC	146	358	1.20 (0.95, 1.51)	1.6	0.1178	48	358	1.08 (0.75, 1.55)	2.1	0.6726	0.5922
CG+GG	533	1418	Referent			152	1418	Referent			
rs1364287											
TT	61	125	1.45 (1.02, 2.06)	2.0	0.0387	7	125	0.74 (0.33, 1.66)	5.1	0.4608	0.1102
CT	259	653	1.08 (0.89, 1.32)	1.5	0.4390	83	653	1.38 (1.00, 1.89)	1.9	0.0500	0.1687
CC	358	995	Referent			109	995	Referent			
rs723013											
GG	78	157	1.46 (1.07, 2.00)	1.9	0.0174	17	157	1.27 (0.73, 2.23)	3.1	0.4023	0.6343
GT	285	748	1.02 (0.84, 1.25)	1.5	0.8161	96	748	1.32 (0.96, 1.80)	1.9	0.0860	0.1421
TT	316	871	Referent			87	871	Referent			
rs6564962											
GG	134	311	1.31 (1.00, 1.72)	1.7	0.0471	40	311	1.18 (0.76, 1.84)	2.4	0.4687	0.6516
AG	331	877	1.04 (0.84, 1.28)	1.5	0.7199	104	877	1.09 (0.77, 1.55)	2.0	0.6213	0.7939
AA	214	586	Referent			56	586	Referent			
rs2911418											
CC+CT	93	216	1.35 (1.00, 1.82)	1.8	0.0478	37	216	1.20 (0.78, 1.83)	2.3	0.4054	0.6066
TT	586	1560	Referent			163	1560	Referent			
rs2955153											
GG	25	41	1.96 (1.15, 3.35)	2.9	0.0141	11	41	1.78 (0.87, 3.65)	4.2	0.1138	0.8094
AA+AG	654	1735	Referent			189	1735	Referent			
rs1364286											
GG	130	284	1.35 (1.06, 1.72)	1.6	0.0145	34	284	0.98 (0.65, 1.46)	2.2	0.9093	0.1355

	Luminal A					Basal-like					Luminal A vs. basal-like[3]
	Case	Control	OR (95% CI)[1]	CLR[2]	P-value	Case	Control	OR (95% CI)[1]	CLR[2]	P-value	P-value
AA+AG	549	1490	Referent			166	1490	Referent			
rs1364285											
GG	104	225	1.35 (1.04, 1.75)	1.7	0.0251	27	225	1.12 (0.72, 1.74)	2.4	0.6066	0.4364
CC+CG	575	1551	Referent			173	1551	Referent			
rs1424151											
CC+CT	58	210	0.70 (0.51, 0.96)	1.9	0.0255	22	210	0.74 (0.46, 1.20)	2.6	0.2217	0.8154
TT	621	1566	Referent			178	1566	Referent			
PGR											
rs546763											
AA	64	183	0.87 (0.63, 1.19)	1.9	0.3673	12	183	0.47 (0.25, 0.86)	3.4	0.0151	0.0613
AC+CC	614	1592	Referent			187	1592	Referent			
rs548668											
CC	64	183	0.86 (0.63, 1.18)	1.9	0.3601	12	183	0.46 (0.25, 0.86)	3.4	0.0142	0.0592
CT+TT	615	1591	Referent			188	1591	Referent			
rs11571247											
AG+GG	49	168	0.81 (0.56, 1.18)	2.1	0.2724	20	168	0.67 (0.40, 1.13)	2.9	0.1364	0.5184
AA	630	1608	Referent			180	1608	Referent			
rs11224575											
AA	23	52	0.96 (0.56, 1.63)	2.9	0.8799	6	52	0.88 (0.36, 2.15)	5.9	0.7860	0.8631
AG	180	424	1.18 (0.95, 1.46)	1.5	0.1339	46	424	1.04 (0.73, 1.49)	2.0	0.8351	0.5191
GG	474	1294	Referent			146	1294	Referent			
rs693765											
GT+TT	38	137	0.69 (0.46, 1.04)	2.3	0.0781	19	137	0.75 (0.44, 1.29)	2.9	0.3015	0.7935
GG	641	1639	Referent			181	1639	Referent			
rs1824128											

	Luminal A					Basal-like					Luminal A vs. basal-like[3]
	Case	Control	OR (95% CI)[1]	CLR[2]	P-value	Case	Control	OR (95% CI)[1]	CLR[2]	P-value	P-value
GT+TT	190	428	1.22 (0.99, 1.50)	1.5	0.0609	52	428	1.16 (0.83, 1.64)	2.0	0.3852	0.7969
GG	489	1348	Referent			148	1348	Referent			
rs660149											
GG	47	131	0.88 (0.61, 1.28)	2.1	0.5039	8	131	0.53 (0.25, 1.13)	4.5	0.0985	0.2040
CG	265	735	0.92 (0.76, 1.12)	1.5	0.4217	91	735	1.17 (0.86, 1.59)	1.8	0.3122	0.1572
CC	367	910	Referent			101	910	Referent			
rs2124761											
GT+TT	74	252	0.75 (0.54, 1.02)	1.9	0.0689	29	252	0.64 (0.40, 1.00)	2.5	0.0519	0.5330
GG	605	1524	Referent			171	1524	Referent			
rs11224579											
TT+CT	152	345	1.24 (0.99, 1.55)	1.6	0.0644	48	345	1.20 (0.84, 1.72)	2.0	0.3053	0.8885
CC	527	1431	Referent			152	1431	Referent			
rs503602											
AA	34	106	0.85 (0.55, 1.30)	2.4	0.4566	10	106	0.58 (0.29, 1.17)	4.1	0.1290	0.3255
AC	204	540	0.96 (0.78, 1.18)	1.5	0.6693	63	540	0.87 (0.62, 1.21)	2.0	0.3972	0.5868
CC	441	1130	Referent			127	1130	Referent			
rs596223											
GG	43	122	0.82 (0.56, 1.20)	2.1	0.3130	18	122	1.73 (0.99, 3.01)	3.0	0.0538	0.0153
AA+AG	635	1649	Referent			180	1649	Referent			
SHBG											
rs1799941											
AA	36	71	1.57 (1.02, 2.43)	2.4	0.0408	9	71	1.87 (0.89, 3.92)	4.4	0.0976	0.6600
AG+GG	643	1704	Referent			191	1704	Referent			

1 - odds ratio, 95% confidence interval, adjusted for age, self-identified race, African ancestry, offset term

2- confidence limit ratio, upper 95% confidence limit divided by lower 95% confidence limit

3 - H_0: β (luminal A)= β(basal-like)

Table 4.4 Stratified odds ratios for SNPs with multiplicative interaction between genotype and parity/lactation, comparing luminal A cases and controls

	Nulliparous	Parous, never breastfed	Parous, ever breastfed	
	Luminal A case/control	Luminal A case/control	Luminal A case/control	
	OR (95%CI)[1]	OR (95%CI)[1]	OR (95%CI)[1]	LRT[2] P-value
ESR1				
rs6914211				0.0954
AA	4/8	15/18	14/14	
	---[3]	2.96 (1.43, 6.14)	3.55 (1.58, 7.99)	
AT+TT	107/193	302/860	237/680	
	Referent	Referent	Referent	
rs985191				0.0153
CC	4/8	15/20	13/14	
	---[3]	2.46 (1.20, 5.06)	---[3]	
AC	35/66	82/269	86/195	
	1.13 (0.66, 1.93)	0.76 (0.56, 1.04)	1.53 (1.10, 2.12)	
AA	72/127	220/589	152/485	
	Referent	Referent	Referent	
rs6557177				0.0471
CC	7/19	22/45	18/31	
	---[3]	1.58 (0.91, 2.76)	1.78 (0.94, 3.35)	
CT+TT	104/182	295/833	233/663	
	Referent	Referent	Referent	
HSD17B2				
rs2911418				0.0017
CC+CT	16/20	58/114	19/82	
	---[3]	1.82 (1.21, 2.74)	0.61 (0.34, 1.09)	
TT	95/181	259/764	232/612	
	Referent	Referent	Referent	
PGR				
rs660149				0.0026
GG	4/17	21/61	22/53	
	---[3]	0.75 (0.43, 1.30)	1.09 (0.61, 1.94)	
CG	56/80	107/379	102/276	
	1.69 (1.01, 2.82)	0.63 (0.47, 0.83)	1.15 (0.83, 1.59)	
CC	51/104	189/438	127/365	
	Referent	Referent	Referent	

rs503602				0.0597
AA	3/15	17/47	14/43	
	---[3]	1.25 (0.67, 2.31)	0.95 (0.47, 1.90)	
AC	31/71	106/270	67/198	
	0.61 (0.35, 1.07)	1.14 (0.84, 1.55)	0.87 (0.62, 1.24)	
CC	77/115	194/561	170/453	
	Referent	Referent	Referent	

1- odds ratio and 95% confidence interval, adjusted for self-identified race, age, African ancestry, and offset term

2 - likelihood ratio test

3 - parameters not estimated due to sample size

Table 4.5 Association between estrogen-related gene haplotypes and breast cancer

Haplotype		All Cases			Luminal A			Basal-like		
	No. Copies	OR (95% CI)[1]	P-Value	CLR[2]	OR (95% CI) [1]	P-Value	CLR[2]	OR (95% CI) [1]	P-Value	CLR[2]
CYP19A1										
1: rs749292, rs1902586, rs936306, rs2445759, rs28566535										
A-A-T-G-C	0	Referent			Referent			Referent		
	1	1.29 (1.00, 1.66)	0.0513	1.7	1.22 (0.84, 1.77)	0.2944	2.1	1.22 (0.72, 2.06)	0.4520	2.8
	2	4.65 (2.60, 8.31)	< 0.0001	3.2	1.79 (0.54, 5.93)	0.3431	11.0	9.01 (3.39, 23.98)	< 0.0001	7.1
ESR1										
2: rs851984, rs851982, rs2881766										
2a: C-T-G	0	Referent			Referent			Referent		
	1	0.61 (0.52, 0.71)	< 0.0001	1.4	0.55 (0.43, 0.69)	< 0.0001	1.6	0.52 (0.35, 0.79)	0.0019	2.3
	2	1.04 (0.81, 1.34)	0.7489	1.7	0.94 (0.66, 1.35)	0.7367	2.0	0.68 (0.39, 1.17)	0.1630	3.0
2b: T-C-G	0	Referent			Referent			Referent		
	1	1.00 (0.82, 1.23)	0.9796	1.5	0.80 (0.58, 1.10)	0.1692	1.9	1.10 (0.69, 1.76)	0.6970	2.6
	2	1.67 (0.99, 2.81)	0.0562	2.8	1.63 (0.73, 3.61)	0.2339	4.9	3.27 (1.27, 8.44)	0.0141	6.6
3: rs1709183, rs9340835, rs9322335										
A-A-T	0	Referent			Referent			Referent		
	1	0.85 (0.75, 0.97)	0.0130	1.3	0.75 (0.62, 0.92)	0.0065	1.5	0.82 (0.57, 1. 18)	0.2767	2.1
	2	0.92 (0.75, 1.12)	0.4092	1.5	0.77 (0.56, 1.05)	0.1019	1.9	1.07 (0.64, 1.77)	0.8023	2.7
4: rs6914211, rs9383599, rs3020314, rs3020401, rs985191, rs6557177										
A-C-C-G-C-C	0	Referent			Referent			Referent		
	1	1.07 (0.93, 1.23)	0.3627	1.3	1.20 (0.98, 1.46)	0.0799	1.5	1.31 (0.94, 1.82)	0.1049	1.9
	2	1.62 (1.17, 2.25)	0.0041	1.9	1.83 (1.13, 2.97)	0.0134	2.6	1.27 (0.51, 3.19)	0.6095	6.3
PGR										
5: rs1824128, rs660149, rs495997, rs2124761, rs11224579										
5a: G-C-A-G-C	0	Referent			Referent			Referent		
	1	1.04 (0.92, 1.17)	0.5100	1.3	0.96 (0.80, 1.14)	0.6165	1.4	1.24 (0.91, 1.68)	0.1787	1.9
	2	1.06 (0.85,1.33)	0.6024	1.6	1.04 (0.76, 1.42)	0.8232	1.9	1.63 (0.97, 2.73)	0.0647	2.8
5b: G-C-G-T-C	0	Referent			Referent			Referent		

	1	0.66 (0.54, 0.82)	0.0001	1.5	0.67 (0.49, 0.91)	0.0112	1.8	0.57 (0.36, 0.90)	0.0170	2.5
	2	1.35 (0.77, 2.36)	0.3007	3.1	1.95 (0.87, 4.35)	0.1035	5.0	1.79 (0.55, 5.85)	0.3376	10.7
5c: T-C-G-G-T	0	Referent			Referent			Referent		
	1	1.25 (1.04, 1.49)	0.0150	1.4	1.31 (1.03, 1.67)	0.0253	1.6	1.11 (0.72, 1.71)	0.6312	2.4
	2	1.89 (1.12, 3.18)	0.0163	2.8	2.29 (1.09, 4.81)	0.0295	4.4	2.11 (0.51, 8.77)	0.3053	17.3

1 - odds ratio, 95% confidence interval, adjusted for age, self-identified race, African ancestry, and offset term

2 - confidence limit ratio, upper 95% confidence limit divided by lower 95% confidence limit

Table 4.6 Association between estrogen-related gene haplotypes and breast cancer

Haplotype		Nulliparous			Parous/never breastfed			Parous/ever breastfed		
	No. Copies	OR (95% CI)[1]	P-Value	CLR[2]	OR (95% CI) [1]	P-Value	CLR[2]	OR (95% CI) [1]	P-Value	CLR[2]
A-C-C-G-C-C	0	Referent			Referent			Referent		
	1	0.91 (0.54, 1.54)	0.7337	2.9	0.92 (0.68, 1.25)	0.6131	1.8	1.73 (1.26, 2.38)	0.0007	1.9
	2	0.68 (0.15, 3.09)	0.6150	20.9	1.86 (0.95, 3.64)	0.0681	3.8	2.45 (1.12, 5.35)	0.0254	4.8

1 - odds ratio, 95% confidence interval, adjusted for age, self-identified race, African ancestry, and offset term
2 - confidence limit ratio, upper 95% confidence limit divided by lower 95% confidence limit
3 - rs6914211, rs9383599, rs3020314, rs3020401, rs985191, rs6557177
4 - LRT P-value = 0.0362

Table 4.7 Estrogen-related functional SNP odds ratios[1] and 95% confidence intervals in the CBCS and previously published studies

		CBCS all cases[2]	CBCS luminal A[2]	CBCS basal-like[2]	CBCS basal-like[2], < 50 yrs old	Raskin et al. (78) BRCA1 carriers, < 50 yrs old	Raskin et al. (78) non-carriers, <50 yrs old
CYP19A1							
rs700518	GG	0.87 (0.70, 1.08)	0.93 (0.70, 1.25)	1.05 (0.64, 1.70)	1.46 (0.81, 2.64)	2.81 (1.09, 7.22)	1.20 (0.51, 3.21)
V80V	AG	1.07 (0.92, 1.25)	1.08 (0.87, 1.34)	0.93 (0.66, 1.32)	0.88 (0.56, 1.37)	1.41 (0.61, 3.26)	1.22 (0.59, 2.38)
	AA	Referent	Referent	Referent	Referent	Referent	Referent
		CBCS all cases[2]	CBCS luminal A[2]	CBCS basal-like[2]	Miyoshi et al. (70)	Lee et al. (71)	Haiman et al. (12)
rs700519	TT	0.44 (0.21, 0.89)	0.67 (0.26, 1.71)	0.48 (0.11, 2.13)		1.0 (0.3, 3.9)	1.05(0.94, 1.17)
R264C	CT	0.99 (0.81, 1.21)	1.17 (0.89, 1.56)	1.06 (0.70, 1.60)		1.5 (1.1, 2.2)	1.05 (0.78, 1.41)
	CT+TT	0.94 (0.77, 1.15)	1.13 (0.86, 1.49)	0.99 (0.66, 1.50)	0.75 (0.50, 1.12)	1.5 (1.1, 2.2)	
	CC	Referent	Referent	Referent	Referent	Referent	Referent
					Gulyaeva et al. (91)		
rs700519	TT						
(cont.)	CT				1.34 (0.51, 3.51)		
	CT+TT				0.96 (0.39, 2.36)		
	CC				Referent		
		CBCS all cases[2]	CBCS luminal A[2]	CBCS basal-like[2]	Talbott et al. (75), premenopausal	Talbott et al. (75), postmenopausal	Haiman et al. (12)
rs1008805	TT	0.98 (0.78, 1.22)	0.91 (0.68, 1.23)	0.86 (0.50, 1.46)	1.61[5]	1.14[5]	Referent
Intron 2 A/G	CT	1.10 (0.95, 1.28)	1.07 (0.87, 1.31)	1.02 (0.73, 1.41)	1.27 (1.02, 1.58)	1.07 (0.91, 1.25)	1.02 (0.95, 1.10)
	CC	Referent	Referent	Referent	Referent	Referent	0.98 (0.88, 1.08)
		CBCS all cases[2]	CBCS luminal A[2]	CBCS basal-like[2]	Kristensen et al. (76)	Haiman et al. (77)	Dunning et al. (68)
rs10046	TT	0.90 (0.73, 1.10)	0.90 (0.68, 1.20)	0.98 (0.61, 1.56)	2.00 (1.28, 3.11)	0.87 (0.60, 1.27)	1.07 (0.93, 1.23)
3' UTR	CT	1.11 (0.95, 1.30)	1.16 (0.94, 1.44)	1.00 (0.71, 1.41)	1.53 (1.04, 2.24)	0.96 (0.69, 1.34)	1.03 (0.91, 1.16)
	CC	Referent	Referent	Referent	Referent	Referent	Referent
		CBCS all cases[2]	CBCS luminal A[2]	CBCS basal-like[2]	Onland Moret et al.[3] (92)	Onland Moret et al. [4] (92)	Wang et al. (89)

ESR1							
rs2234693	TT	0.90 (0.74, 1.09)	0.79 (0.61, 1.04)	0.93 (0.60, 1.45)	1.43 (0.93, 2.22)	1.23 (1.08, 1.42)	
+397 C/T	CT	0.92 (0.77, 1.09)	0.98 (0.78, 1.23)	1.09 (0.75, 1.59)	1.32 (0.90, 1.95)	1.14 (1.00, 1.32)	0.98 (0.74, 1.32)
PvuII	CC	Referent	Referent	Referent	Referent	Referent	Referent
					Iwasaki et al. (94), Japanese (Nagano)	Iwasaki et al. (94), Japanese-Brazilian (Sao Paolo)	Iwasaki et al. (94), non-Japanese Brazilian (Sao Paolo)
rs2234693	TT				Referent	Referent	Referent
(cont.)	CT				0.70 (0.49, 1.00)	0.66 (0.29, 1.47)	0.99 (0.68, 1.43)
	CC				0.64 (0.40, 1.02)	0.93 (0.31, 2.86)	1.51 (0.98, 2.31)
					Zheng et al. (93)	Sonestedt. et al. (95)	Dunning et al. (81)
rs2234693	TT				0.79 (0.63, 1.00)	Referent	Referent
(cont.)	CT				0.95 (0.80, 1.12)	1.03 (0.81, 1.31)	1.01 (0.88, 1.16)
	CC				Referent	1.00 (0.74, 1.34)	1.09 (0.92, 1.28)
		CBCS all cases[2]	CBCS all cases[2], postmenopausal only	CBCS luminal A[2]	CBCS basal-like[2]	Gallicchio et al. (96)	Gallicchio et al. (96), postmenopausal only
rs2077647	CC	0.99 (0.81, 1.20)	1.17 (0.90, 1.52)	1.12 (0.85, 1.46)	1.10 (0.71, 1.69)	1.14 (0.65, 1.99)	1.88 (0.94, 3.75)
+29 T/C	CT	0.96 (0.82, 1.13)	0.98 (0.79, 1.22)	1.10 (0.88, 1.38)	1.08 (0.75, 1.55)	0.71 (0.42, 1.20)	0.90 (0.46, 1.77)
S10S	TT	Referent	Referent	Referent	Referent	Referent	Referent
MspI							
					Fernandez et al. (87)	Dunning et al. (81)	
rs2077647	CC				0.74 (0.53, 1.02)	1.07 (0.91, 1.27)	
(cont.)	CT				0.76 (0.57, 1.00)	0.92 (0.80, 1.06)	
	TT				Referent	Referent	
		CBCS all cases[2]	CBCS luminal A[2]	CBCS basal-like[2]	Fernandez et al. (87)	Tapper et al. (64)	Dunning et al. (81)

rs3798577	CC	0.94 (0.78, 1.14)	1.06 (0.82, 1.37)	0.97 (0.62, 1.50)	1.04 (0.75, 1.46)	1.23 [6]	1.03 (0.87, 1.22)
3' UTR	CT	1.08 (0.92, 1.26)	1.08 (0.87, 1.35)	1.27 (0.90, 1.80)	1.09 (0.83, 1.44)	1.11 (1.00, 1.24)	0.98 (0.85, 1.12)
	TT	Referent	Referent	Referent	Referent	Referent	Referent
		CBCS all cases[2]	CBCS luminal A[2]	CBCS basal-like[2]	Gallicchio et al. (96)		
rs851982	TT	0.93 (0.75, 1.15)	1.08 (0.80, 1.46)	0.80 (0.48, 1.32)	0.88 (0.47, 1.67)		
-104062 C/T	CT	1.03 (0.83, 1.28)	1.14 (0.85, 1.54)	1.08 (0.66, 1.77)	0.65 (0.40, 1.05)		
	CC	Referent	Referent	Referent	Referent		
		CBCS all cases[2]	CBCS luminal A[2]	CBCS basal-like[2]	Pooley et al. (97)	Fernandez et al. (87)	Johnatty et al. (98)
PGR	AA	1.44 (0.31, 6.72)			0.87 (0.40, 1.88)		
rs10895068	AG	1.10 (0.85, 1.41)			0.97 (0.85, 1.10)	0.84 (0.53, 1.33)	
+331G/A	AA+AG	1.10 (0.86, 1.42)	0.94 (0.66, 1.33)	0.99 (0.53, 1.83)			1.06 (0.76, 1.49)
	GG	Referent	Referent	Referent	Referent	Referent	Referent
					DeVivo et al. (99)		
rs10895068	AA						
(cont.)	AG						
	AA+AG				1.26 (0.97, 1.63)		
	GG				Referent		
		CBCS all cases[2]	CBCS luminal A[2]	CBCS basal-like[2]	Dunning et al. (68)	Thompson et al. (69)	
SHBG	AA	1.31 (0.93, 1.85)	1.52 (0.98, 2.37)	1.92 (0.90, 4.07)	1.02 (0.75, 1.38)	1.06 [6]	
rs1799941	AG	0.93 (0.80, 1.09)	0.92 (0.74, 1.14)	1.07 (0.74, 1.54)	1.08 (0.92, 1.27)	1.03 (0.94, 1.13)	
-67 G>A	GG	Referent	Referent	Referent	Referent	Referent	

1 - unless specified otherwise
2 - odds ratio, 95% confidence interval, adjusted for age, self-identified race, African ancestry, offset term
3 - risk ratio, 95% confidence interval
4 - meta-analysis
5 - additive model
6 - calculated based on additive model OR given in paper

Table 4.8 Minor allele frequencies in functional SNPs

Gene	Group	Subjects	Minor allele frequency				
CYP19A1			rs10046 (3'UTR) minor allele: T	rs700518 (V80V) minor allele: G	rs700519 (R264C) minor allele: T	rs2236722 (W39R) minor allele: C	rs28757184 (T201M) minor allele: T
	CBCS	AA and white female population-based controls from North Carolina	nonAA: 0.52 AA: 0.25	nonAA: 0.50 AA: 0.21	nonAA: 0.03 AA: 0.19	nonAA: 0 AA: 0	nonAA: 0.04 AA: 0.04
	International HapMap Project		CEU: 0.57 ASW:0.22 YRI: 0.17	CEU: 0.58 ASW: 0.16 YRI: 0.13	CEU: 0.03 ASW: 0.17 YRI: 0.21	CEU: 0 ASW: --- YRI: 0	CEU: 0.04 ASW: 0.04 YRI: 0.06
	Perlegen	AFD_AFR PANEL		EUR: 0.52 AA: 0.22	EUR: 0 AA: 0.23	EUR: 0 AA: 0	
	SNP500	Anonymized samples from individuals of self-described African/African American or Caucasian heritage	CAUC1: 0.50 AFR1: 0.17	CAUC1: 0.50 AFR1: 0.10	CAUC1: 0 AFR1: 0.15		CAUC1: 0.06 AFR1: 0
	Ma et al. (72)	Anonymous Caucasian-American and African American samples from Coriell Cell Repository	white: 0.56 AA: 0.19	white: 0.54 AA: 0.17	white: 0.03 AA: 0.22	white: 0 AA: 0	
	Haiman et al. (12)	70 white subjects in the MultiEthnic Cohort study	0.44	white: 0.46	white: 0.46		white: 0.01
	Dunning et al. (68)	Postmenopausal women from EPIC-Norfolk (>98% white)	0.53				
	Riancho et al. (100)	Postmenopausal women living in northern Spain		0.48			
	Somner et al. (80)	White, postmenopausal women living in London		0.33			
	Haiman et al.	White and AA controls	white: 0.48	white: 0.46	white: 0.04		

Gene	Group	Subjects	Minor allele frequency				
	(40)	in the MEC cohort study	AA: 0.26	AA: 0.81	AA: 0.15		
	Raskin et al. (78)	Ashkenazi Jewish BRCA1/2 non-carriers enrolled in Clalit Health Services in Israel		0.49			
	He et al. (101)	American women of European origin, recruited for genetic studies		0.52			
	Gulyaeva et al. (91)	Control women, without gynecological disease, living in Novosibirsk, Russia			0.05		
ESR1			rs2077647 (S10S) minor allele: C	rs2234693 (*PvuII*/-397T/C) minor allele: C	rs3798577 (C/T) minor allele: T(A)	rs851982 (-104062 C/T) minor allele: T	
	CBCS	AA and white female population-based controls from North Carolina	nonAA: 0.51 AA: 0.49	nonAA: 0.44 AA: 0.52	nonAA: 0.54 AA: 0.54	nonAA: 0.59 AA: 0.76	
	International HapMap Project		CEU: 0.43 ASW: 0.58 YRI: 0.50	CEU: 0.41 ASW: 0.56 YRI: 0.50	CEU: 0.56 ASW: 0.52 YRI: 0.56	CEU: 0.54 ASW: 0.85 YRI: 0.86	
	Perlegen	AFD_AFR PANEL		EUR: 0.38 AA: 0.46	EUR: 0.50 AA: 0.57	EUR: 0.58 AA: 0.76	
	SNP500	Anonymized samples from individuals of self-described African/African American or Caucasian heritage	CAUC1: 0.52 AFR1: 0.50	CAUC1: 0.52 AFR1: 0.71	CAUC1: 0.44 AFR1: 0.58		
	Tapper et al. (64)	Controls of western European ancestry from the Wellcome			0.53		

Gene	Group	Subjects	Minor allele frequency				
		Trust Case Control Consortium					
	Sonestedt et al. (95)	Female controls from the Malmo Diet and Cancer in Sweden	0.47	0.45			
	Kjaergaard et al. (102)	Male and female participants in the Copenhagen City Heart Study in Denmark		0.46			
	Onland-Moret et al. (92)	Participants in population-based case-cohort study in Utrecht, Holland		0.51			
	Fernandez et al. (87)	Pre-/postmenopausal female volunteers from Madrid and Oviedo, Spain (recruited through Madrid College of Lawyers and Menopause Research Centre)	0.52		0.55		
	Gallicchio et al. (96)	White women from Washington County, Maryland with benign breast disease	0.51			0.39	
	Sowers et al. (103)	Participants in the Study of Women's Health Across the Nation, a prospective cohort study of women's health		white: 0.45 AA: 0.55	white: 0.52 AA: 0.56		
	Wang et al. (89)	Controls, age 65 and older, in the Study of Osteoporotic Fractures		0.48			
	Zofkova et al. (104)	Postmenopausal Czech women		0.43			
	Pharoah et al.	EPIC-Norfolk/Anglian	0.47	0.46	0.46		

Gene	Group	Subjects	Minor allele frequency				
	(105), Dunning et al. (81)	Breast Cancer Study					
	Iwasaki et al. (94)	Non-Japanese Brazilian hospital-based cancer-free controls living in Sao Paolo, Brazil		0.42			
PGR			rs10895068 (+331G/A) minor allele: A				
	CBCS	AA and white female population-based controls from North Carolina	nonAA: 0.05 AA: 0.02				
	International HapMap Project		CEU: 0.06 ASW: 0.01 YRI: ---				
	SNP500	Unrelated controls of self-described African/African American or Caucasian heritage from the Human Diversity Panel	CAUC3: 0.05 AFR3: 0				
	DeVivo et al. (99)	Cancer-free Nurses' Health Study participants	0.10				
	Westberg et al. (106)	Women ages 41-42 living in Göteborg, Sweden	0.07				
	Fernandez et al. (87)	Pre-/postmenopausal female volunteers from Madrid and Oviedo, Spain (recruited through Madrid	0.04				

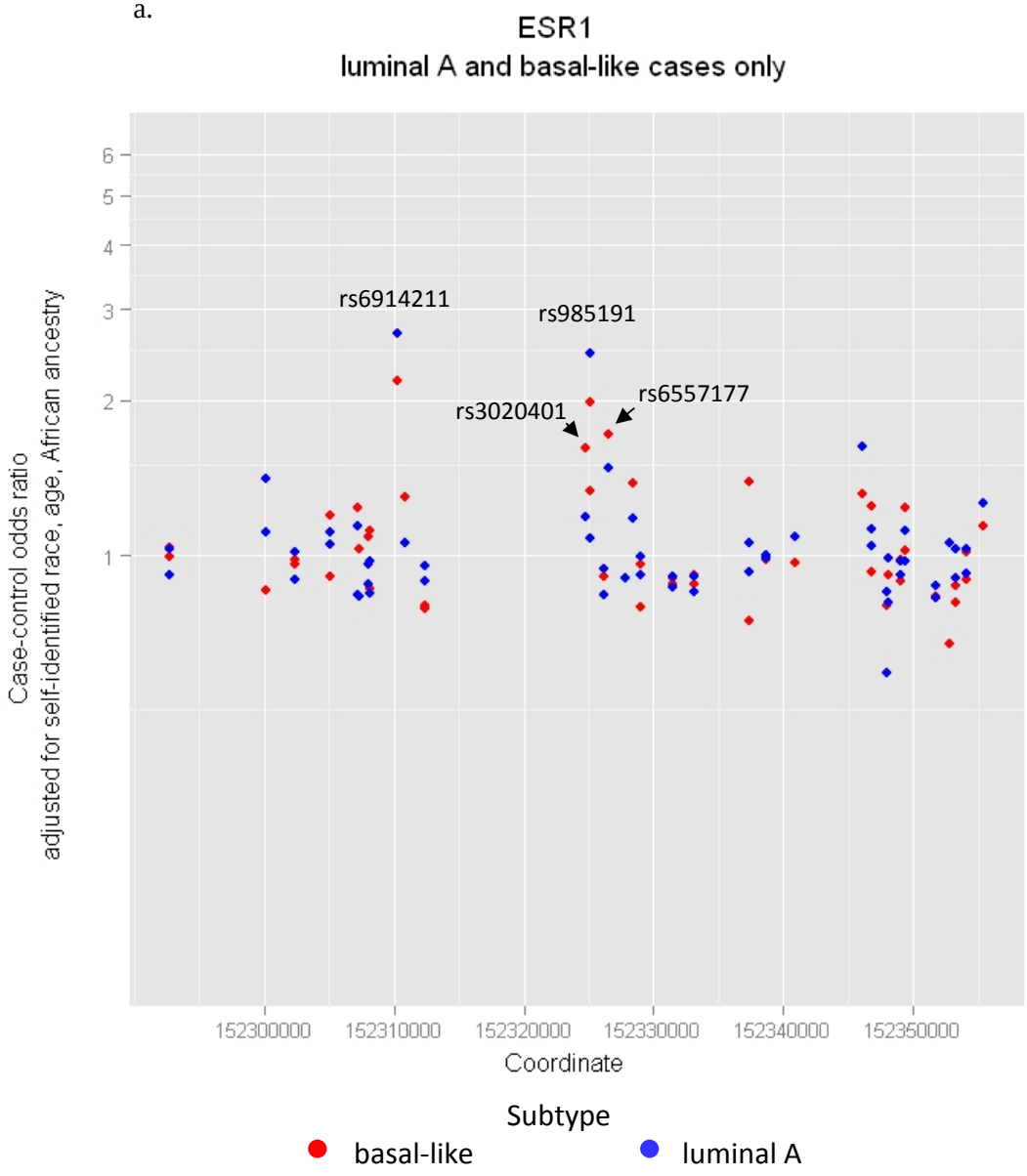
Gene	Group	Subjects	Minor allele frequency				
		College of Lawyers and Menopause Research Centre)					
	Johnatty et al. (98)	Population-based controls less than 59 yrs, from the Australian Breast Cancer Family Study	0.05				
	Pooley et al. (97)	EPIC-Norfolk/Anglian Breast Cancer Study	0.06				
SHBG			rs1799941 (-67 G/A) minor allele: A				
	CBCS	AA and white female population-based controls from North Carolina	non-AA: 0.22 AA: 0.07				
	International HapMap Project		CEU: 0.23 ASW: 0.07 YRI: 0.03				
	SNP500	Anonymized samples from individuals of self-described African/African American or Caucasian heritage	CAUC1: 0.39 AFR1: 0.08				
	Dunning et al. (68)	EPIC-Norfolk/Anglian Breast Cancer Study	0.25				
	Garcia-Closas et al. (107)	Population controls from Warsaw and Lodz, Poland	0.68				

Abbreviations: AA-African American, AC – Afro-Caribbean.

4.8 Figures

Figure 4.1 ESR1 breast cancer-associated SNPs

a.



b. R^2 in African American controls

	rs6914211	rs3020401	rs985191	rs6557177
rs6914211	1.			
rs3020401	0.12	1.		
rs985191	0.70	0.15	1.	
rs6557177	0.32	0.15	0.47	1.

c. R^2 in African American cases

	rs6914211	rs3020401	rs985191	rs6557177
rs6914211	1.			
rs3020401	0.17	1.		
rs985191	0.76	0.17	1.	
rs6557177	0.37	0.18	0.49	1.

d. R^2 in White controls

	rs6914211	rs3020401	rs985191	rs6557177
rs6914211	1.			
rs3020401	0.32	1.		
rs985191	0.95	0.32	1.	
rs6557177	0.38	0.16	0.40	1.

e. R^2 in White cases

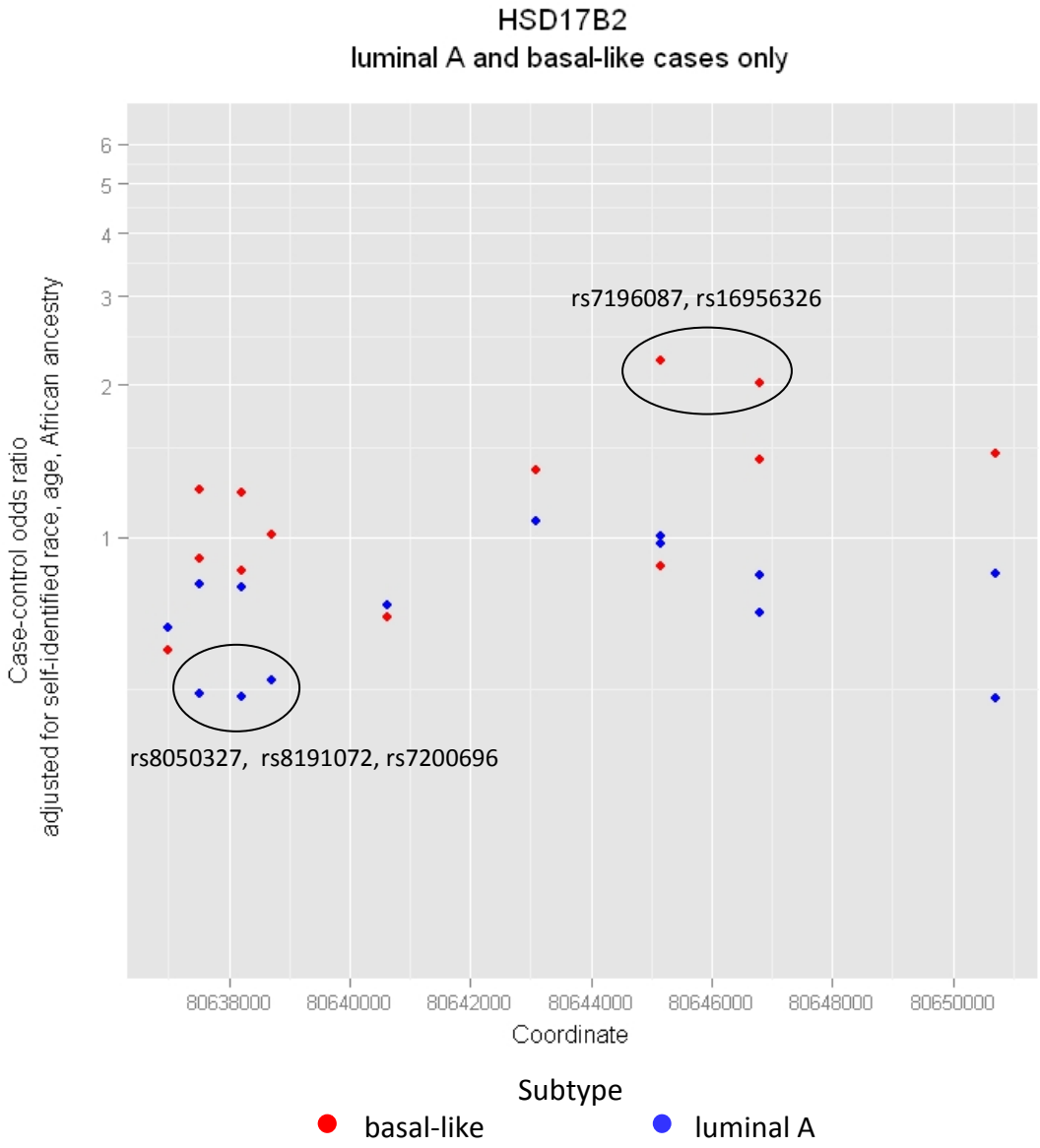
	rs6914211	rs3020401	rs985191	rs6557177
rs6914211	1.			
rs3020401	0.31	1.		
rs985191	0.96	0.30	1.	
rs6557177	0.45	0.19	0.46	1.

Figure 4.1 Legend

a. ORs for luminal A (blue) and basal-like (red) breast cancer-associated SNPs in ESR1. SNPs shown are from 152354108 bp to 152355411 bp on chromosome 6. While there were several SNPs in ESR1 associated with luminal A and basal-like breast cancer, SNPs in this region seemed to show a consistent, positive association within a defined area. SNPs rs6914211 and rs985191 are 14918 bp apart.

b. Pairwise r^2 among SNPs rs6914211, rs3020401, rs985191, and rs6557177.

Figure 4.2 HSD17B2 breast cancer-associated SNPs
a.



b. R^2 in African American controls

	rs8050327	rs8191072	rs7200696
rs8050327	1.		
rs8191072	0.94	1.	
rs7200696	0.58	0.62	1.

c. R^2 in African American cases

	rs8050327	rs8191072	rs7200696
rs8050327	1.		
rs8191072	0.94	1.	
rs7200696	0.59	0.61	1.

d. R^2 in White controls

	rs8050327	rs8191072	rs7200696
rs8050327	1.		
rs8191072	1.00	1.	
rs7200696	0.14	0.14	1.

e. R^2 in White cases

	rs8050327	rs8191072	rs7200696
rs8050327	1.		
rs8191072	1.00	1.	
rs7200696	0.22	0.22	1.

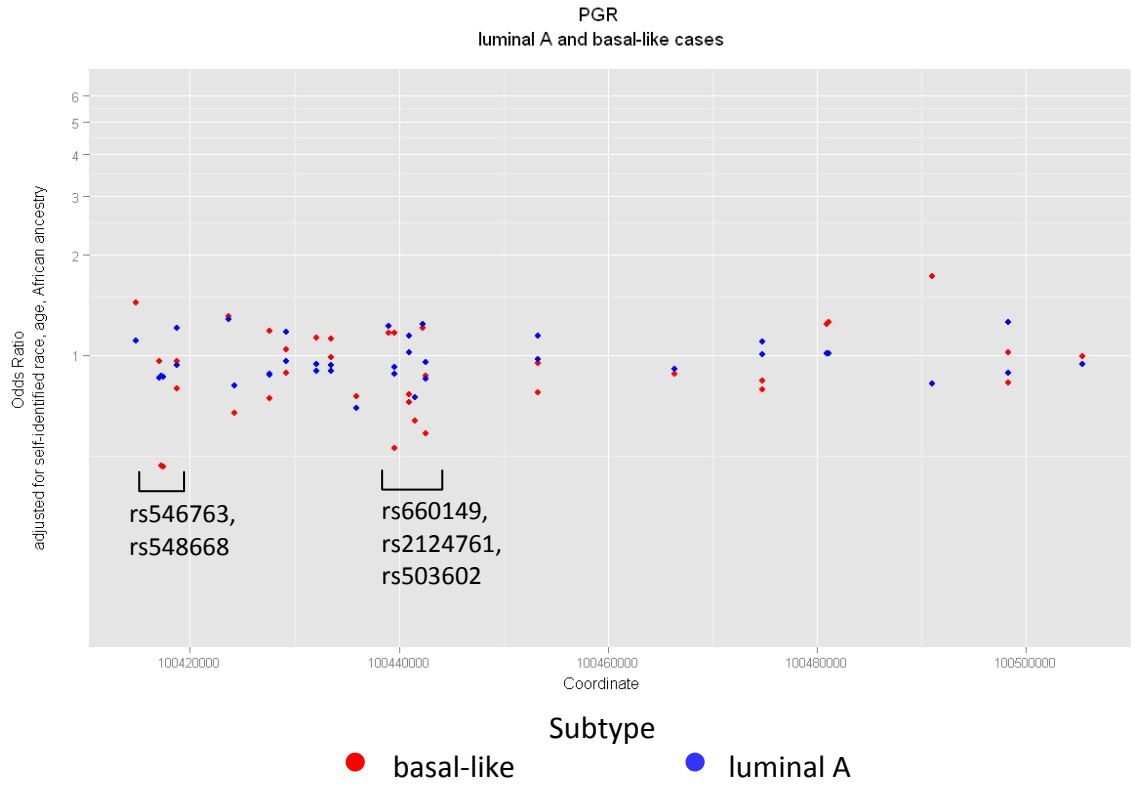
Figure 4.2 Legend

a. HSD17B2 SNPs rs8050327, rs8191072, and rs7200696 had strong inverse associations with luminal A breast cancer. These 3 SNPs span 1205 bp in intron 1. Nearby SNPs rs7196087 and rs16956326 were strongly associated with basal-like breast cancer. These SNPs are 1624 bp apart in intron 1.

b-e. Pairwise r^2 among SNPs rs8050327, rs8191072, and rs7200696.

Figure 4.3 PGR breast cancer-associated SNPs

a.



b. R^2 in African American controls

	rs660149	rs2124761	rs503602
rs660149	1.		
rs2124761	0.09	1.	
rs503602	0.18	0.22	1.

c. R^2 in African American cases

	rs660149	rs2124761	rs503602
rs660149	1.		
rs2124761	0.06	1.	
rs503602	0.15	0.18	1.

d. R^2 in White controls

	rs660149	rs2124761	rs503602
rs660149	1.		
rs2124761	0.00	1.	
rs503602	0.06	0.04	1.

e. R^2 in White cases

	rs660149	rs2124761	rs503602
rs660149	1.		
rs2124761	0.00	1.	
rs503602	0.05	0.06	1.

Figure 4.3 Legend

a. ORs for the association between luminal A (blue) and basal-like (red) breast cancer. The first two SNPs, rs546763 and rs548668, are 214 bp apart and were in strong LD in CBCS African Americans and whites. The second group of SNPs, rs660149, rs2124761, and rs503602, are 2993 bp apart. They were moderately correlated in CBCS African Americans, and showed little to no correlation in whites.

b-e. Pairwise r^2 for rs660149, rs2124761, and rs503602.

5. Summary and Conclusions

5.1 Main findings

The purpose of this dissertation was to explore the association between common SNPs and breast cancer, with the hypothesis that the associations for some SNPs differ by breast cancer molecular subtype. Breast cancer molecular subtypes were determined in the CBCS using immunohistochemistry, where the two principal subtypes were luminal A (ER+ and/or PR+ and HER2-) and basal-like (ER-, PR-, HER2-, and CK 5/6+ or EGFR+). The hypothesis that SNP associations differ by molecular subtype is based on the fact that previous CBCS analyses found that molecular subtypes exhibit different tumor characteristics and clinical outcomes, as well as distinct associations with non-genetic breast cancer risk factors (1-3). The present analysis was driven by a candidate gene approach, focusing on genes involved in biological pathways identified by non-genetic risk factors for basal-like and luminal A breast cancer in the CBCS. Additive and multiplicative interaction was evaluated in order to determine whether genetic associations were modified in the presence of non-genetic risk factors.

Chapter 3 focused on a set of genes chosen based on the differing association for WHR and basal-like vs. luminal A breast cancer in the CBCS. WHR was positively associated with both tumor subtypes, but the association was approximately twice as strong for the basal-like compared to luminal A subtype (3). ADIPOQ, IL6, LEP, LEPR, and TNF were chosen for this study because of their relationship to central obesity, for which WHR is a proxy measure. Plasma levels of interleukin-6, TNF-alpha, and leptin increase with

increasing amounts of obesity and serum adiponectin decreases with increasing obesity (4). Studies have also shown that these gene products affect pathways potentially involved in breast cancer [reviewed in (4-8)].

The genes in Chapter 4 were chosen based on the differing association of parity and lactation with the basal-like and luminal A subtypes in the CBCS, specifically the association for parity without lactation. Compared to nulliparous women, parous women who did not breastfeed were at a decreased risk of luminal A breast cancer but a strong increased risk of basal-like breast cancer (3). CYP19A1, ESR1, HSD3B1, HSD17B2, PGR, and SHBG were selected for their roles in the synthesis and action of estrogen and progesterone, two hormones instrumental in breast carcinogenesis, at least for a subset of breast tumors [reviewed in (9-11)]. CYP19A1, HSD17B2, and HSD3B1 encode enzymes that convert cholesterol into estrogen and progesterone (and several other hormones) (12). SHBG binds estrogen in the blood stream and affects bioavailability for estrogen signaling (13). The progesterone receptor is a nuclear transcription factor that binds progesterone and affects the transcription of genes whose sequence includes progesterone response elements [reviewed in (14, 15)]. The estrogen receptor binds estrogen, and can act as a nuclear transcription factor, though it can also be activated without estrogen binding [reviewed in (16-18)].

A total of 117 adipocytokine-related SNPs and 195 SNPs in the estrogen-related genes were genotyped using the Illumina GoldenGate assay. The SNPs included previously reported SNPs of interest as well as tag SNPS selected to cover the region from the first to last SNP recorded for each gene in the NCBI dbSNP database. The tag SNPs were chosen using allele frequency and LD information for populations of European and West African descent from Phases 1 and 2 of the International HapMap Project. Overall, 89% of the SNPs

were genotyped successfully and were included in the analysis (adipocytokine-related – 82%, estrogen-related– 94%), a success rate identical to the overall success rate for the 1536 SNPs in the assay.

African and European ancestry in CBCS subjects was estimated from 144 AIMs. Individual ancestry estimates were used to control for any residual confounding due to population stratification still present after adjustment for self-identified race. Overall, ancestry adjustment did not affect the ORs, though there were some differences by subgroups. For example, AIM adjustment affected |lnORs| by more than 0.10 when comparing luminal A cases to controls, but not when comparing basal-like cases to controls. AIM adjustment also caused |lnORs| to differ by 0.10 or more when comparing white cases to controls in a subset of SNPs, but adjustment had little effect on ORs when comparing African-American cases and controls. These two results were expected, considering that 54% of genotyped basal-like cases were African American and 65% of genotyped luminal A cases were white. To maintain model consistency all estimates were adjusted for ancestry.

The first step of the analysis followed a traditional approach, estimating odds ratios for the association between individual genotypes and all breast cancer and controls. The majority of single SNP associations were close to the null, though some displayed moderate associations with breast cancer. Notably, the strongest genotype associations for breast cancer overall (ORs > 2) were estimated for SNPs in ESR1 (rs6914211 and rs985191). A handful of other SNPs from the estrogen-related genes pathway had ORs between 1.5 and 2, or between 0.67 and 0.5 (ESR1 rs2207232 and rs3778099; HSD17B2 rs8052451 and rs8191102). All of the single SNP associations for adipocytokine SNPs were between 1.50 and 0.67.

Next, genotype associations were estimated separately for the basal-like and luminal A breast cancer subtypes. The results of this analysis supported the hypothesis that some SNP-breast cancer associations are unique to breast cancer subtypes. SNPs that were not strongly associated with breast cancer overall showed associations of greater magnitude when the analysis was stratified by subtype. For example, small groups of SNPs in LEPR, HSD17B2, and PGR were strongly associated with basal-like breast cancer. When ORs differed by subtype it was unusual for the associations to be in opposite directions. More commonly, when ORs differed by subtype the OR was weak or close to the null for one subtype, and elevated (or decreased) for the other subtype. There were also SNPs for which a strong association with breast cancer was seen in breast cancer overall and in both subtypes, including both ESR1 rs6914211 and rs985191.

The majority of SNPs associated with breast cancer, either overall or by subtype, were intronic SNPs that have not been evaluated previously for an association with breast cancer, emphasizing the importance of a systematic tagging approach as opposed to genotyping a few individual SNPs per gene. Comprehensive tagging of whole candidate genes has become more common, in large part due to the increased efficiency and decreased cost of genotyping in large studies. CBCS ORs for many SNPs selected from the literature indicated no association with breast cancer. Exceptions were LEPR K109R (rs1137100), CYP19A1 R264C (rs700519), TNF -863 C/A (rs1800630) and SHBG -67 G/A (rs1799941), which were associated with breast cancer overall. By subtype, LEPR K109R, Q223R (rs1137101), IL6 -572 G/C (rs1800796) and SHBG -67 G/A were associated with luminal A breast cancer, and SHBG -67 G/A was associated with the basal-like subtype.

The magnitude of the strongest genotype ORs was unexpected, given the hypothesis

that common variant associations will be modest, and it is the accumulation of multiple variants that impact disease risk (19). Furthermore, breast cancer-associated SNPs identified by GWAS, most of which have been validated in multi-stage analyses, are in the range of 1.2 to 2.0 for 2 copies of the variant allele (20-26). In the literature, the highest ORs were estimated for the FGFR2 SNPs identified by Hunter et al. (21); most other SNP associations were lower. CBCS ORs for breast cancer overall are similar in magnitude to the GWAS ORs, but several of the subtype-specific ORs are stronger. This could be interpreted as better estimation of the association due to refining the outcome definition. For example, if a SNP was associated with the basal-like subtype but completely unassociated with other subtypes, performing a subtype-specific regression model would be equivalent to removing non-cases from the case group. However, the novel associations estimated in the CBCS should be interpreted with caution until they can be replicated. It is likely that most true associations are of a lower magnitude than the ones reported here (27).

Associations between haplotypes and breast cancer were estimated using maximum likelihood estimation in HAPSTAT. Haplotypes in IL6, LEP, LEPR, CYP19A1, ESR1, and PGR were associated with breast cancer (all cases, luminal A, and/or basal-like). In many cases, haplotype ORs were stronger in magnitude than single SNP ORs. This is consistent with the findings of some investigations of the advantages of haplotype analysis but not others (28-30). Additionally, by using a sliding, overlapping 3-SNP window to estimate associations for all possible 3-SNP haplotypes, associated haplotypes were identified where the single SNP ORs indicated no association. For example, the SNPs in LEP haplotype 3 were largely unassociated with breast cancer when considered individually (the strongest ORs were for rs11763517 CC vs. TT [all cases OR: 0.86; luminal A OR: 0.86; basal-like OR:

0.88]), but 2 copies of LEP haplotype 3 were associated with more than twice the risk of breast cancer overall and luminal A breast cancer. This finding further supports the hypothesis that alleles inherited together on the same chromosome may have greater effects together compared to when they are inherited on different chromosomes. One caveat to this finding is that the haplotype associations involve a likelihood-based estimation of haplotypes in the population, but the true haplotype phase in individuals is unknown. This is a problem that is inherent to all studies with unrelated subjects where haplotypes are not sequenced directly. The algorithms used in HAPSCAN and HAPSTAT were designed to incorporate the uncertainty attached to haplotype estimation in unphased data into the effect estimates and confidence intervals, yielding unbiased ORs with proper standard errors. Thus, though the haplotype associations should be replicated, the estimates provided by HAPSTAT should be an accurate estimation of the association between measured haplotypes in the CBCS population and breast cancer assuming no other sources of bias.

Statistical interaction between WHR and genotypes was evaluated on the additive and multiplicative scales for breast cancer-associated SNPs in ADIPOQ, IL6, LEP, LEPR, and TNFA. Interaction was evaluated on the multiplicative scale only for breast cancer-associated SNPs in estrogen-related genes and parity and lactation. Additive effects were not evaluated due to sample size limitations. There was evidence to suggest interaction between WHR and SNPs in LEPR (breast cancer overall), IL6 (basal-like), and ADIPOQ and LEPR (luminal A). Overall the pattern of interaction was less than multiplicative and less than additive, suggesting antagonism between genotypes and WHR (31). These results suggest that the predominant type of interaction in the study population for these SNPs and WHR involves the presence of one factor blocking the effect of the other factor, although other mechanisms

of interaction may also be present in the population (31). In the case of the parity and lactation interaction, additional studies would benefit from a larger study population. Both of the genotype and parity/lactation variables were 3-level variables, leading to imprecise stratified ORs.

5.2 Strengths and Limitations

5.2.1 Strengths

One of the mains strengths of this dissertation is the high proportion of African-American women in the CBCS. Most previous studies involving the SNPs evaluated in this study and breast cancer have consisted mainly of women of European descent, and may not be representative of the general US population. There was some diversity. One study population was based in Tunisia (32), one consisted of Ashkenazi Jews (33), one consisted of women of Indo-Aryan ethnicity recruited in New Delhi, India, several were studies of Asian women (24, 34-38), and another had a large proportion of Hispanic women from the southwestern US (39). However, with the exception of the MultiEthnic Cohort (40, 41), these studies included few African Americans. African-American women have higher age-adjusted breast cancer mortality than any other racial group in the US (42), and so inclusion of African Americans in studies of breast cancer risk is essential to identifying risk factors relevant to disease prevention in this subgroup. The CBCS used randomized recruitment to oversample African-American (and younger) breast cancer cases. This increases the likelihood that associations for risk alleles important to African-American women will be identified and estimated with sufficient precision.

A second strength of this study is that molecular subtypes were determined for cases,

allowing the distinction of basal-like cases from other triple-negative tumors (ER-, PR-, HER2-). The triple-negative designation groups basal-like tumors along with a less well characterized group of tumors that gene expression studies have shown are biologically distinct (43-47). This study identified SNPs and haplotypes associated with the basal-like subtype that were not associated with breast cancer overall. It is possible that these associations would not have been detected in a mixed group of triple-negative tumors.

Thirdly, SNPs were selected using a combination of approaches in order to increase the chance of identifying SNPs associated with the basal-like and luminal A breast cancer subtypes. Regions of the genome that were more likely to be associated with the basal-like or luminal A subtypes, based on gene function and subtype risk factors, were pre-selected for analysis by focusing on candidate genes. Focusing the study on genetic variation in regions where there is evidence that part of the genome has an effect on breast cancer has the potential to increase the efficiency with which causal alleles are identified (48). The selection of tag SNPs in addition to potentially functional SNPs allowed for the identification of previously unreported breast cancer associations. Although the majority of tag SNPs are unlikely to have a direct effect on gene function, they may be in linkage disequilibrium with one or more untyped variants which do have a functional effect. The efficiency of the tagging method is dependent on how well HapMap SNPs covered the candidate genes, and on LD between genotyped and untyped SNPs being as high in the CBCS as in the CEU and YRI HapMap populations. LD similarities cannot be compared because non-tag SNPs remained untyped in the CBCS. Nevertheless, this SNP selection method has enabled the identification of breast cancer-associated SNPs, including SNPs with hypothesized functional effects and those with no known function.

Finally, this analysis was strengthened by the estimation of African and European ancestry in CBCS subjects. The proportion of African ancestry followed a continuous distribution in self-identified African Americans and non-African Americans, suggesting that there could be residual confounding after adjustment for self-identified race. Ancestry estimates were similar using maximum likelihood and structured association, demonstrating that the results were robust to the statistical methods used. Analyses confirmed that ancestry adjustment does not affect SNP associations with breast cancer overall and only affects a few SNP associations with the luminal A subtype. The results of ancestry analyses provide evidence that admixture of African and European populations was unlikely to have biased the majority of genetic associations in the CBCS.

5.2.2 Limitations

One of the main limitations of this analysis was the potential for selection bias related to study participation. Savitz (49) discusses case-control study selection in terms of the comparability of case exposure and control exposure to the source populations from which they were sampled. In terms of this analysis, if case study participant genotypes are representative of genotypes among all cases sampled for the study, and control participant genotypes are representative of genotypes among all sampled controls, then there will be no selection bias due to non-participation. This assumes that the sampling procedure used by the CBCS was unbiased.

An in-depth analysis of response patterns among CBCS subjects recruited from 1993 to 1996 reported that the most common reason for non-participation in cases and controls was subject refusal (50). Other reasons for non-participation included being deceased, ineligible, not able to be located, and physician refusal to grant permission for the study to

contact the patient (cases only) (50). Analysis of women who responded to a partial telephone interview suggests that women who did not complete the in-person study interview differed from full participants for several non-genetic breast cancer risk factors, but most of these risk factors do not influence genotype. Race has the potential to affect study participation and genotype frequencies (through correlation with ancestry), and this could result in selection bias if genotype frequencies differed by race in the source population.

In cases, if genotypes were related to factors downstream of diagnosis, such as severity of illness, response to treatment, or survival, then the genotype distribution among case participants would differ from that in all eligible cases. Sampled CBCS cases who refused to participate in the study were more likely to be African American, whereas among controls rates of refusal were similar (50).

The same scenario applies to selection into the study at the genotyping phase. African Americans were less likely to donate blood for genotyping, but within self-identified race groups blood donation did not differ by case status. Potential reasons for not donating blood could include unwillingness to donate and inability to donate because of illness. In addition to women who did not donate DNA, 103 women were excluded from the analysis because of poor genotyping results. Subjects with low call rates were similar to the overall CBCS population with DNA submitted for genotyping with respect to case status, self-identified race, AJCC stage at diagnosis (invasive cases only), and molecular subtype distribution (cases only). Thus, the potential for selection bias at the genotyping phase again depends on whether genotype distributions in the source population differ by race group. All models in this study were adjusted for self-identified race and African ancestry, which should control for potential selection bias at the enrollment and genotyping stages of the study (49).

Non-African Americans and women with earlier stage tumors were less likely to have tumor subtype information, which could lead to additional bias of the molecular subtype-specific analyses. It is difficult to evaluate whether the tumors without subtype information are a random sample of all eligible cases. Other studies of breast tumor molecular subtypes that defined basal-like rather than triple-negative tumors have been conducted in various countries (47, 51-54). It is unlikely that those any of those case groups are comparable to the CBCS case base, and so valid inferences cannot be made based about subtype identity. Efforts are currently underway to determine molecular subtypes for African-American breast cancer cases in the MultiEthnic Cohort, Women's Circle of Health Study, and the Black Women's Health Study (40, 41, 55). While it will be useful to compare subtype distribution in the CBCS to that of other large case series in the US, other studies are also subject to possible selection bias due to an inability to subtype all eligible cases.

Another limitation is the reduction in sample size due to missing subtype information, particularly the small number of basal-like cases. Because of the small sample size, the standard error for basal-like parameter estimates will always be higher than the standard error for luminal A parameter estimates, leading to wider confidence intervals and higher P-values for similar point estimates. The imbalance between basal-like and luminal A cases is a natural function of the lower prevalence of basal-like tumors in the population (assuming no selection bias), but the low absolute number of basal-like cases was problematic when estimating the association for a rare genotype or haplotype. Unstable parameter estimates were avoided by eliminating estimates with wide confidence limit ratios, which indicated imprecision. Small sample size also limited the ability to evaluate associations for the HER2+/ER-, luminal B, unclassified molecular subtypes.

WHR was measured after cases were diagnosed, raising the possibility that physical changes related to breast cancer may cause WHR to be systematically different among cases compared to WHR in those cases prior to diagnosis. A sensitivity analysis was conducted to evaluate the possible effects WHR misclassification on the WHR-basal-like breast cancer association. The sensitivity analysis addressed the possible effects of disease and treatment-related changes in WHR. Treatment information was not collected for CBCS cases, so estimates of the proportion of cases who would have received chemotherapy were based on data from North Carolina Central Cancer Registry data (56). The validity of this sensitivity analysis depends on the extent to which treatment patterns recorded by the Cancer Registry from 2001 to 2002 reflect the treatment patterns of CBCS cases in 1993 to 2001. Bias-corrected and observed ORs were similar, even under some of the more extreme misclassification scenarios. Overall, the results suggest that if chemotherapy-related WHR misclassification occurred in the CBCS, the effects on this analysis were minimal.

The tagging procedure used in this analysis was designed to tag other SNPs. So, although efforts were made to capture the associations for all CEU and YRI HapMap SNPs in the candidate genes through genotyping or high pairwise LD, this study did not evaluate all genetic variation in the gene regions. There may be residual unmeasured genetic variation. Additionally, this study did not systematically capture the effects of non-SNP variation such as copy number variants, insertion-deletion polymorphisms, and repeat polymorphisms, though by chance some of these variants may be in LD with genotyped tag SNPs. Methods are currently available to analyze SNP and copy number polymorphism information from a single genotyping assay using the Affymetrix SNP 6.0 array (57). It is likely that similar software will be developed for use with other commonly used genotyping chips in the future.

5.3 Public health impact

Some researchers have hypothesized that racial and ethnic differences in the prevalence of genetic variants contribute to racial differences in breast cancer incidence and survival (58, 59). Basal-like tumors are more prevalent among premenopausal African-American women, and are associated with poorer survival compared to hormone receptor positive subtypes (1, 3). The identification of additional risk factors for basal-like breast cancer will help improve identification of women at high risk for poor prognosis breast cancer. It may also help explain some of the correlation between African-American race and poor prognosis. Several SNPs with associations specific to the basal-like subtype were identified in this analysis. Through further research, the identification causal variants near these SNPs could be used to define a subgroup of susceptible women who may benefit from increased breast cancer surveillance.

Interaction between genotypes and/or haplotypes and other breast cancer risk factors is an important aspect of the investigation between genetic variation and breast cancer. Evaluation of additive and multiplicative interaction has the potential to further define subgroups with elevated or decreased breast cancer risk. Using LEPR rs1137100 as an example, the antagonism detected in the present study suggests that the population of women carrying the GG genotype who also have elevated WHR would experience fewer cases of breast cancer than would be expected based on the independent associations of WHR and rs1137100 with breast cancer. In such a situation, population-level interventions to reduce WHR may not result in the expected reduction in breast cancer if the variant is highly prevalent. Knowledge of the patterns of biological interaction in the population may help policy makers tailor primary and secondary prevention strategies so they are more efficient.

This study also has the potential to impact public health by helping identify biological mechanisms underlying the intrinsic molecular subtypes. Identification of key genes, proteins, and pathways active in basal-like breast cancer could provide new drug targets. This type of research is necessary because women with basal-like breast cancer have limited treatment options. The development of a long-term breast cancer drug to be taken after initial treatment, similar to how selective estrogen receptor modulators and aromatase inhibitors are used now, has the potential to increase survival among basal-like cases.

5.4 Future directions

The results of this dissertation investigation support two main conclusions: 1) there is evidence that some of the genotypes and haplotypes studied are associated with breast cancer; and 2) some genotype and haplotype associations differ between the basal-like and luminal A breast cancer subtypes. With respect to the first point, replication of the most promising SNP and haplotype associations is necessary before strong conclusions can be made about the magnitude of candidate gene associations. In addition to the potential selection bias that has been discussed, OR estimates are also subject to random sampling error. In the absence of systematic error, consideration of CBCS SNP associations in the context of associations estimated in other populations will provide a more precise estimate of the SNP or haplotype association in the general population. It may also be useful to replicate the results using a different genotyping platform to insure that the genotype calls are robust to the laboratory methods.

Another logical step in this investigation is to fine map regions in ESR1, HSD17B2, LEPR, and PGR where multiple breast cancer-associated SNPs were located in close

proximity. Fine mapping involves genotyping a dense group of SNPs within a specific region in order to localize the SNP(s) most strongly associated with the outcome (60). Fine mapping in the CBCS could be accomplished by genotyping a second group of SNPs concentrated in the regions of interest. Alternatively, this analysis could be conducted within a proposed GWAS of African-American cases and controls from the CBCS, MultiEthnic Cohort, and Women's Circle of Health Study.

Finally, future research should include a comprehensive evaluation of the genomic elements in the regions where the breast cancer-associated SNPs are located. As discussed previously, the majority of the SNPs associated with breast cancer (overall and by subtype) were intronic SNPs with no known function, suggesting that either the typed SNP is in LD with the true causal variant and is indeed functioning as a tag, or that the SNP itself is associated with breast cancer. There are currently several publicly available databases containing genome sequence information (ENCODE database, 1000 Genomes Project), promoter sequence information (Eukaryotic Promoter Database, Transcriptional Element Search System, Transcriptional Regulatory Element Database), and estimates of evolutionary conservation across species (University of California Santa Cruz genome browser, PANTHER classification system) that can be used to evaluate the possible effects of SNPs. Identification and characterization of the causal variant are important steps in identifying the biological mechanism through which SNPs act. Identifying the biological mechanism will be instrumental in further exploring interaction between genetic variants and environmental factors in the etiology of breast cancer subtypes.

5.5 References

1. Carey LA, Perou CM, Livasy CA, Dressler LG, Cowan D, Conway K, et al. Race, breast cancer subtypes, and survival in the Carolina Breast Cancer Study. Jama 2006;295(21):2492-502.

2. Livasy CA, Perou CM, Karaca G, Cowan DW, Maia D, Jackson S, et al. Identification of a basal-like subtype of breast ductal carcinoma in situ. Hum Pathol 2007;38(2):197-204.

3. Millikan RC, Newman B, Tse CK, Moorman PG, Conway K, Dressler LG, et al. Epidemiology of basal-like breast cancer. Breast Cancer Res Treat 2008;109(1):123-39.

4. Rose DP, Komninou D, Stephenson GD. Obesity, adipocytokines, and insulin resistance in breast cancer. Obes Rev 2004;5(3):153-65.

5. Schaffler A, Scholmerich J, Buechler C. Mechanisms of disease: adipokines and breast cancer - endocrine and paracrine mechanisms that connect adiposity and breast cancer. Nat Clin Pract Endocrinol Metab 2007;3(4):345-54.

6. Vona-Davis L, Rose DP. Angiogenesis, adipokines and breast cancer. Cytokine Growth Factor Rev 2009;20(3):193-201.

7. Knupfer H, Preiss R. Significance of interleukin-6 (IL-6) in breast cancer (review). Breast Cancer Res Treat 2007;102(2):129-35.

8. Cirillo D, Rachiglio AM, la Montagna R, Giordano A, Normanno N. Leptin signaling in breast cancer: an overview. J Cell Biochem 2008;105(4):956-64.

9. Pike MC, Spicer DV, Dahmoush L, Press MF. Estrogens, progestogens, normal breast cell proliferation, and breast cancer risk. Epidemiol Rev 1993;15(1):17-35.

10. Dickson RB, Stancel GM. Estrogen receptor-mediated processes in normal and cancer cells. J Natl Cancer Inst Monogr 2000(27):135-45.

11. Russo J, Russo IH. The role of estrogen in the initiation of breast cancer. J Steroid Biochem Mol Biol 2006;102(1-5):89-96.

12. Payne AH, Hales DB. Overview of steroidogenic enzymes in the pathway from cholesterol to active steroid hormones. Endocr Rev 2004;25(6):947-70.

13. Fortunati N, Becchis M, Catalano MG, Comba A, Ferrera P, Raineri M, et al. Sex hormone-binding globulin, its membrane receptor, and breast cancer: a new approach to the modulation of estradiol action in neoplastic cells. J Steroid Biochem Mol Biol 1999;69(1-6):473-9.

14. Mendelson CR, Hardy DB. Role of the progesterone receptor (PR) in the regulation of inflammatory response pathways and aromatase in the breast. J Steroid Biochem Mol Biol 2006;102(1-5):241-9.

15. Conneely OM, Mulac-Jericevic B, Lydon JP. Progesterone-dependent regulation of female reproductive activity by two distinct progesterone receptor isoforms. Steroids 2003;68(10-13):771-8.

16. Gruber CJ, Tschugguel W, Schneeberger C, Huber JC. Production and actions of estrogens. N Engl J Med 2002;346(5):340-52.

17. Osborne CK, Schiff R. Estrogen-receptor biology: continuing progress and therapeutic implications. J Clin Oncol 2005;23(8):1616-22.

18. Speirs V, Walker RA. New perspectives into the biological and clinical relevance of oestrogen receptors in the human breast. J Pathol 2007;211(5):499-506.

19. Pharoah PD, Antoniou A, Bobrow M, Zimmern RL, Easton DF, Ponder BA. Polygenic susceptibility to breast cancer and implications for prevention. Nat Genet 2002;31(1):33-6.

20. Easton DF, Pooley KA, Dunning AM, Pharoah PD, Thompson D, Ballinger DG, et al. Genome-wide association study identifies novel breast cancer susceptibility loci. Nature 2007;447(7148):1087-93.

21. Hunter DJ, Kraft P, Jacobs KB, Cox DG, Yeager M, Hankinson SE, et al. A genome-wide association study identifies alleles in FGFR2 associated with risk of sporadic postmenopausal breast cancer. Nat Genet 2007;39(7):870-4.

22. Stacey SN, Manolescu A, Sulem P, Rafnar T, Gudmundsson J, Gudjonsson SA, et al. Common variants on chromosomes 2q35 and 16q12 confer susceptibility to estrogen receptor-positive breast cancer. Nat Genet 2007;39(7):865-9.

23. Stacey SN, Manolescu A, Sulem P, Thorlacius S, Gudjonsson SA, Jonsson GF, et al. Common variants on chromosome 5p12 confer susceptibility to estrogen receptor-positive breast cancer. Nat Genet 2008;40(6):703-6.

24. Zheng W, Long J, Gao YT, Li C, Zheng Y, Xiang YB, et al. Genome-wide association study identifies a new breast cancer susceptibility locus at 6q25.1. Nat Genet 2009;41(3):324-8.

25. Thomas G, Jacobs KB, Kraft P, Yeager M, Wacholder S, Cox DG, et al. A multistage genome-wide association study in breast cancer identifies two new risk alleles at 1p11.2 and 14q24.1 (RAD51L1). Nat Genet 2009;41(5):579-84.

26. Ahmed S, Thomas G, Ghoussaini M, Healey CS, Humphreys MK, Platte R, et al.

Newly discovered breast cancer susceptibility loci on 3p24 and 17q23.2. Nat Genet 2009;41(5):585-90.

27. Ioannidis JP. Common genetic variants for breast cancer: 32 largely refuted candidates and larger prospects. J Natl Cancer Inst 2006;98(19):1350-3.

28. Long AD, Langley CH. The power of association studies to detect the contribution of candidate genetic loci to variation in complex traits. Genome Res 1999;9(8):720-31.

29. Akey J, Jin L, Xiong M. Haplotypes vs single marker linkage disequilibrium tests: what do we gain? Eur J Hum Genet 2001;9(4):291-300.

30. Morris RW, Kaplan NL. On the advantage of haplotype analysis in the presence of multiple disease susceptibility alleles. Genet Epidemiol 2002;23(3):221-33.

31. Greenland S, Lash TL, Rothman KJ. Concepts of Interaction. In: Rothman KJ, Greenland S, Lash TL, editors. Modern Epidemiology. Third ed. Philadelphia: Lippincott Williams & Wilkins; 2008. p. 71-86.

32. Snoussi K, Strosberg AD, Bouaouina N, Ben Ahmed S, Helal AN, Chouchane L. Leptin and leptin receptor polymorphisms are associated with increased risk and poor prognosis of breast carcinoma. BMC Cancer 2006;6:38.

33. Raskin L, Lejbkowicz F, Barnett-Griness O, Dishon S, Almog R, Rennert G. BRCA1 breast cancer risk is modified by CYP19 polymorphisms in Ashkenazi Jews. Cancer Epidemiol Biomarkers Prev 2009;18(5):1617-23.

34. Lee KM, Abel J, Ko Y, Harth V, Park WY, Seo JS, et al. Genetic polymorphisms of cytochrome P450 19 and 1B1, alcohol use, and breast cancer risk in Korean women. Br J Cancer 2003;88(5):675-8.

35. Miyoshi Y, Iwao K, Ikeda N, Egawa C, Noguchi S. Breast cancer risk associated with polymorphism in CYP19 in Japanese women. Int J Cancer 2000;89(4):325-8.

36. Woo HY, Park H, Ki CS, Park YL, Bae WG. Relationships among serum leptin, leptin receptor gene polymorphisms, and breast cancer in Korea. Cancer Lett 2006;237(1):137-42.

37. Iwasaki M, Hamada GS, Nishimoto IN, Netto MM, Motola J, Jr., Laginha FM, et al. Isoflavone, polymorphisms in estrogen receptor genes and breast cancer risk in case-control studies in Japanese, Japanese Brazilians and non-Japanese Brazilians. Cancer Sci 2009;100(5):927-33.

38. Han CZ, Du LL, Jing JX, Zhao XW, Tian FG, Shi J, et al. Associations among lipids, leptin, and leptin receptor gene Gin223Arg polymorphisms and breast cancer in China. Biol Trace Elem Res 2008;126(1-3):38-48.

39. Slattery ML, Curtin K, Baumgartner R, Sweeney C, Byers T, Giuliano AR, et al. IL6, aspirin, nonsteroidal anti-inflammatory drugs, and breast cancer risk in women living in the southwestern United States. Cancer Epidemiol Biomarkers Prev 2007;16(4):747-55.

40. Kolonel LN, Henderson BE, Hankin JH, Nomura AM, Wilkens LR, Pike MC, et al. A multiethnic cohort in Hawaii and Los Angeles: baseline characteristics. Am J Epidemiol 2000;151(4):346-57.

41. Lee SA, Haiman CA, Burtt NP, Pooler LC, Cheng I, Kolonel LN, et al. A comprehensive analysis of common genetic variation in prolactin (PRL) and PRL receptor (PRLR) genes in relation to plasma prolactin levels and breast cancer risk: the Multiethnic Cohort. BMC Med Genet 2007;8(1):72.

42. Horner MJ RL, Krapcho M, Neyman N, Aminou R, Howlader N, Altekruse SF, Feuer EJ, Huang L, Mariotto A, Miller BA, Lewis DR, Eisner MP, Stinchcomb DG, Edwards BK. SEER Cancer Statistics Review, 1975-2006. Bethesda, MD: National Cancer Institute; 2009.

43. Abd El-Rehim DM, Ball G, Pinder SE, Rakha E, Paish C, Robertson JF, et al. High-throughput protein expression analysis using tissue microarray technology of a large well-characterised series identifies biologically distinct classes of breast cancer confirming recent cDNA expression analyses. Int J Cancer 2005;116(3):340-50.

44. Perou CM, Sorlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, et al. Molecular portraits of human breast tumours. Nature 2000;406(6797):747-52.

45. Sorlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proc Natl Acad Sci U S A 2001;98(19):10869-74.

46. Sorlie T, Tibshirani R, Parker J, Hastie T, Marron JS, Nobel A, et al. Repeated observation of breast tumor subtypes in independent gene expression data sets. Proc Natl Acad Sci U S A 2003;100(14):8418-23.

47. Calza S, Hall P, Auer G, Bjohle J, Klaar S, Kronenwett U, et al. Intrinsic molecular signature of breast cancer in a population-based cohort of 412 patients. Breast Cancer Res 2006;8(4):R34.

48. Tabor HK, Risch NJ, Myers RM. Candidate-gene approaches for studying complex genetic traits: practical considerations. Nat Rev Genet 2002;3(5):391-7.

49. Savitz DA. Interpreting Epidemiologic Evidence. New York: Oxford University Press; 2003.

50. Moorman PG, Newman B, Millikan RC, Tse CK, Sandler DP. Participation rates in a

case-control study: the impact of age, race, and race of interviewer. Ann Epidemiol 1999;9(3):188-95.

51. Adebamowo CA, Famooto A, Ogundiran TO, Aniagwu T, Nkwodimmah C, Akang EE. Immunohistochemical and molecular subtypes of breast cancer in Nigeria. Breast Cancer Res Treat 2007;Aug 3 - epub ahead of print.

52. Yang XR, Sherman ME, Rimm DL, Lissowska J, Brinton LA, Peplonska B, et al. Differences in risk factors for breast cancer molecular subtypes in a population-based study. Cancer Epidemiol Biomarkers Prev 2007;16(3):439-43.

53. Kim MJ, Ro JY, Ahn SH, Kim HH, Kim SB, Gong G. Clinicopathologic significance of the basal-like subtype of breast cancer: a comparison with hormone receptor and Her2/neu-overexpressing phenotypes. Hum Pathol 2006;37(9):1217-26.

54. Kurebayashi J, Moriya T, Ishida T, Hirakawa H, Kurosumi M, Akiyama F, et al. The prevalence of intrinsic subtypes and prognosis in breast cancer patients of different races. Breast 2007;16 Suppl 2:S72-7.

55. Rosenberg L, Adams-Campbell L, Palmer JR. The Black Women's Health Study: a follow-up study for causes and preventions of illness. J Am Med Womens Assoc 1995;50(2):56-8.

56. Ali S. Female Breast Cancer Incidence, Stage at Diagnosis, Treatment, and Mortality in North Carolina. In: State Center for Health Statistics, North Carolina Department of Health and Human Services; 2006. p. 1-6.

57. Korn JM, Kuruvilla FG, McCarroll SA, Wysoker A, Nemesh J, Cawley S, et al. Integrated genotype calling and association analysis of SNPs, common copy number polymorphisms and rare CNVs. Nat Genet 2008;40(10):1253-60.

58. Ademuyiwa FO, Olopade OI. Racial differences in genetic factors associated with breast cancer. Cancer Metastasis Rev 2003;22(1):47-53.

59. Berger FG. The interleukin-6 gene: a susceptibility factor that may contribute to racial and ethnic disparities in breast cancer mortality. Breast Cancer Res Treat 2004;88(3):281-5.

60. Ioannidis JP, Thomas G, Daly MJ. Validating, augmenting and refining genome-wide association signals. Nat Rev Genet 2009;10(5):318-29.