COMMON GENETIC VARIATION IN CELL CYCLE REGULATORY GENES AND ETIOLOGY OF INTRINSIC BREAST CANCER SUBTYPE: A CANDIDATE GENE APPROACH

Nicholas J. Taylor

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 in the Gillings School of Global Public Health.

Chapel Hill
2013

Approved by:

Andrew F. Olshan

Charles Poole

Jeannette T. Bensen

Marilie D. Gammon

Melissa A. Troester

ABSTRACT

Nicholas J. Taylor: Common Genetic Variation in Cell Cycle Regulatory Genes and Etiology of Intrinsic Breast Cancer Subtype: A Candidate Gene Approach
(Under the direction of Andrew F. Olshan)

A large proportion of unexplained risk for breast cancer remains to be accounted for. Contributing factors may be environmental, genetic, or a combination of both and there is considerable debate about which factors are most important. However, the scope and magnitude of the genetic contribution to the causation of breast cancer remains unclear. Genetic risk factors for breast cancer remain to be discovered, and with heterogeneity of breast cancer being characterized into intrinsic molecular subtypes, the difficulty in identifying these risk factors is diminishing.

This dissertation used a candidate gene approach based on factors involved in cell cycle regulation to identify single nucleotide polymorphisms (SNPs) associated with overall rate of breast cancer and intrinsic breast cancer subtype in the Carolina Breast Cancer Study (CBCS). A total of 65 SNPs on five genes were genotyped in 1,946 cases and 1,747 controls in African American and Caucasian participants of the CBCS. Additionally, 144 ancestry informative markers were genotyped in these individuals to estimate individual ancestry and adjust logistic models for potential population stratification. Race-stratified odds ratios were calculated, as estimates of rate ratios, along with 95% confidence intervals for the associations between SNP genotypes and breast cancer using logistic regression and adjusting for age and ancestry. These

associations were also estimated by intrinsic subtype of breast cancer in a similarly adjusted combined race group.

The intronic SNP rs6092309 on *AURKA* showed an inverse association with rate of breast cancer among African Americans (OR=0.69, 95%CI=0.53-0.90), with inverse associations also noted across all strata of intrinsic subtype. Exploratory race-stratified, subtype-specific analyses for some *AURKA* SNPs suggested race-specific effects. Three SNPs in high LD on *BRCA1* (rs16941, rs16942, and rs1799966) had positive associations with overall rate of breast cancer among Caucasians. One SNP on *BARD1* (rs28997576: OR=1.42, 95%CI: 1.00-2.03) showed a positive association with rate of breast cancer among Caucasians.

These results suggest that associations between genetic exposures and rate of breast cancer may differ by intrinsic subtype and possibly by race within subtype. Replication of these findings in larger populations of African American and Caucasian women will be required to make more accurate interpretations.

ACKNOWLEDGEMENTS

I would like to think the members of my dissertation committee for their guidance and patience through this process. I would like to thank David Richardson for his support and advice. I would also like to thank Dr. Danyu Lin for his consultation and direction regarding HAPSTAT. Many thanks to Dr. Patricia Basta and the UNC Biospecimen Processing Facility for their work handling CBCS biological specimens. I would also like to acknowledge the late Dr. Robert C. Millikan, my academic advisor upon my entrance in the PhD program at UNC, without whom I would not have gained the rich appreciation I have for cancer epidemiology. Finally, I would like to thank the Department of Epidemiology and the Lineberger Comprehensive Cancer Center for financial support during the course of my matriculation.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AA – African American(s)

AIM – ancestry informative marker

BMI – body mass index

Cau – Caucasian(s)

CBCS – Carolina Breast Cancer Study

CEU – Utah residents with ancestry from northern and western Europe genotyped by the International HapMap Project

CI – confidence interval

CIS – carcinoma *in situ*

CK – cytokeratin

CLR – confidence limit ratio

DCIS – ductal carcinoma *in situ*

EM – expectation maximization

ER – estrogen receptor

GWAS – genome-wide association study

HER1 – human epidermal growth factor receptor 1

HER2 – human epidermal growth factor receptor 2

HRT – hormone replacement therapy

HWE – Hardy-Weinberg equilibrium

IHC – immunohistochemistry

LD – linkage disequilibrium

LFS – Li-Fraumeni Syndrome

MAF – minor allele frequency

OR – odds ratio

PCM – pericentriolar material

PJS – Peutz-Jeghers Syndrome

PR – progesterone receptor

RERI – relative excess risk due to interaction

SNP – single nucleotide polymorphism

r^2 – pairwise correlation coefficient

UNC – University of North Carolina at Chapel Hill

YRI – Individuals of Yoruban descent from Idaban, Nigeria, genotyped in the International HapMap Project

Chapter 1. Background and Significance

1.1 The Public Health Burden of Breast Cancer

Breast cancer continues to represent a tremendous health burden in the United States. The American Cancer Society estimates that 30% of all cancers diagnosed among American women in 2010 will be breast cancers, making them the most commonly diagnosed cancers among women in the U.S. [1]. After cancers of the lung and bronchus, breast cancer is the leading cause of cancer death in American women [1]. Although recent data indicate a decline in incidence and mortality, a consistent disparity between African American and Caucasian women persists [1, 2].

Incident cases of breast cancer have been and continue to be more frequent in Caucasian women (126.5 per 100,000) compared to African American women (118.3 per 100,000) [2]. However, age-adjusted trends have been consistent, if not convergent since 1975 [1] (Figure 2.2). The racial disparity in incidence is highlighted in women aged 40 and above (Figure 2.1). Notably however, this trend is reversed in women under 35, with African American women displaying a higher incidence rate.

Despite a modest difference in the rates of newly diagnosed cases, African American women show a significantly higher age-adjusted mortality rate (Figure 2.3). This disparity is even more pronounced when age-specific mortality rates are examined (Figure 2.4). African American women under the age of 50 are 77% more likely to die of breast cancer when compared to Caucasian women in the same age range [3]. Breast cancer in African

American women is distinguished by larger, higher-grade tumors that are diagnosed at later stages [3-6]. Even after controlling for stage at diagnosis, African American women still exhibit poorer survival when compared to Caucasian women [3, 6-8]. It has been suggested that differences in survival may be attributed to socioeconomic factors [5, 9-12] or differences in access to care [8-13]. However, recent studies have reported that trends in screening by mammography among African Americans and Caucasians are similar [14-16]. In fact, controlling for socioeconomic factors, access to healthcare and co-morbidities does not diminish the racial disparity in mortality [11, 17-21]. This may suggest potential differences in tumor biology among African American and Caucasian women.

1.2 Genetic Risk Factors for Breast Cancer

A family history of breast cancer is a strong risk factor; women having a single first-degree relative with breast cancer are nearly twice as likely to develop the disease, while having two first-degree relatives with breast cancer approximately triples a woman's risk [1, 22, 23]. Still, the vast majority of women who develop breast cancer (~85%) have no family history of the disease [22].

Hereditary breast cancers constitute between 5 and 10% of all cases [24]. The most common predisposing factors contributing to these cases are highly penetrant mutations in *BRCA1* and *BRCA2*. However, population-based epidemiologic studies have demonstrated that only 15-20% of familial breast cancers exhibit a mutation in either of these genes [24, 25]. The large proportion of unexplained familial risk may be explained by unidentified genetic traits, environmental risk factors, or a combination of both. There is considerable debate as to which predominates, but the scope and magnitude of the genetic contribution to

the causation of breast cancer remains unclear [26, 27]. Twin-studies and studies of familial inheritance have suggested that common, low penetrance genetic factors may account for the observed residual familial risk [26, 28]. This so-called polygenic model proposes that genetic susceptibility to breast cancer is not entirely predicted by rare, highly penetrant genes but more often stems from several common loci that each confer smaller independent increases in risk [28-31]. Acting multiplicatively, this aggregate of common risk variants may contribute a significant proportion of familial risk. Under this model it would be rare to observe multiple-case families (as is the case for those demonstrating mutations in highly penetrant genes such as *BRCA1*) since an individual would have to inherit each of several different variants.

Results from genome-wide association studies (GWAS) seem to support the polygenic model with respect to breast cancer. GWAS take advantage of technological advances allowing for hundreds of thousands to millions of single nucleotide polymorphisms (SNPs) to be analyzed as potential risk modifying loci without information regarding function. A large GWAS conducted by Easton *et al.* identified significant associations between SNPs on *FGFR2*, *TNRC9*, *MAP3K1, LSP1, H19* and breast cancer in European women from the United Kingdom [32]. These findings were supported by results from the Shanghai Breast Cancer Study (A GWAS conducted in Chinese women) [33]. GWAS have primarily been conducted in populations of European descent. Recently, Hutter *et al.* examined 22 previously identified breast cancer GWAS susceptibility loci in a study of 7,800 African American women from the Women's Health Initiative SNP Health Association Resource [34]. SNPs in *FGFR2* and *TOX3* were associated breast cancer risk [34].

Additionally, there are several rare conditions that substantially increase the risk of breast cancer in a small proportion of the population. Li-Fraumeni Syndrome (LFS) is caused by a mutation in *TP53* and is thought to account for approximately 1% of hereditary breast cancers [34]. LFS is characterized by early-onset cancers, including: breast cancer, soft-tissue sarcoma and leukemia [35]. LFS families experience an increased risk of cancer up to 90% by age 60 [36]. Cowden Syndrome is also associated with increased risk for breast cancer [37]. Cowden Syndrome is generally defined by germline mutations in *PTEN*, a putative tumor suppressor gene [38, 39]. Women with Cowden Syndrome have a lifetime risk of breast cancer between 25-50% [40]. Peutz-Jeghers Syndrome (PJS) has also been associated with an increased risk of cancer. PJS is characterized by germline mutations in the tumor suppressor gene *STK11* [41]. Women with PJS have demonstrated increased risk for breast cancer of up to 30% by age 60 [42]. Ataxia telangiectasia, a rare childhood condition characterized by neurological deterioration and hypersensitivity to ionizing radiation, has also been associated with an increased risk for breast cancer [37].

1.3 Other Risk Factors for Breast Cancer

1.3.1 Age

Age is one of the strongest risk factors for breast cancer, with incidence rates nearly six times as high in American women aged 75 or older compared to those aged 20-49 [43]. Based on SEER data from 1975-2007, the Centers for Disease Control and Prevention estimated an 8-fold difference in 10-year risk of developing breast cancer between women currently age 60 and women currently age 30 [44, 45].

1.3.2 BMI

Studies investigating the relationship between BMI and risk of breast cancer have been inconsistent. Reports have suggested that increased BMI is associated with an increased risk of breast cancer; in a pooled multivariate analysis of prospective cohorts, van den Brandt *et al*. reported increased risk of breast cancer with increasing BMI and weight only in postmenopausal women, but the trend was not linear. Women between 75-80kg showed a higher relative risk than women ≥80kg. Likewise, BMI results demonstrated the same trend, with postmenopausal women having a BMI of 31-33 demonstrating a higher relative risk than women of BMI ≥33 [46]. In premenopausal women, an inverse trend in risk was noted in both weight and BMI. Possible explanations for this inverse trend include more frequent anovulatory menstrual cycles resultant from decreased concentrations of estrogen and progesterone exhibited in obese women [46-49]. In contrast, several case-control studies have found both inverse and direct associations between BMI and odds of breast cancer among premenopausal women [50-52].

As a result of inconsistent findings for associations between BMI and risk of breast cancer, it has been suggested that distribution of adiposity may be an important factor in explaining the relationship between BMI, weight, and risk of breast cancer in premenopausal women [53-55]. In a European cohort of women, Lahmann *et al*. reported a significant increase in breast cancer risk among premenopausal women in the highest quintiles of both waist circumference and hip circumference after adjusting for BMI (RR=1.81, 95%CI: 1.11-2.97; RR=1.70, 95%CI: 1.05-2.77 respectively). However, Lahmann's findings based on waist-hip ratio (WHR) were consistent with no association after controlling for BMI (RR=1.05, 95%CI: 0.74-1.50) [53]. Similarly, an IARC review

found no association between WHR and risk of breast cancer in premenopausal women [52]. In contrast, a meta-analysis of case-control and cohort studies performed by Connolly *et al*. reported significant associations between WHR and risk of breast cancer, regardless of menopausal status, after controlling for BMI [56].

1.3.3 Physical Activity

Studies of potential associations between physical activity and risk of breast cancer have been equivocal, probably due in part to the lack of any clear standardized instrument for measuring exposure and a failure to thoroughly evaluate confounding and effect measure modification [57]. Nevertheless, Monninkhoff's systematic review of 29 case-control and 19 cohort studies found strong evidence for risk reductions with increased physical activity among postmenopausal women; evidence for risk reduction among physically active premenopausal women was weaker [58].

1.3.4 Menarche

Reproductive factors such as age at menarche and regularity of menstrual cycles have also been associated with breast cancer risk. Early age at menarche (12 years or earlier) has been associated with an increased risk of breast cancer, with modest declines in risk accompanying each year of delayed onset [59, 60]. Moreover, there is evidence that menstrual cycle regularity is also an important risk factor; studies have demonstrated a doubling of risk among women who experienced earlier menarche with predictable menstrual cycles compared to women who had irregular cycles [61]. These associations have been attributed to earlier exposures to and higher concentrations of estrogen in the

adolescent years [59, 62]. Estrogen is known to influence normal breast epithelial cell growth by promoting cellular proliferation [59]. The increased exposure to estrogen during adolescence provides an environment of rapid cell proliferation that is thought to increase the risk of random mutations in the genome [59, 60, 63]. Supporting the role of estrogen in tumorigenesis, early menopause has been shown to decrease a woman's risk for breast cancer [60].

1.3.5 Breast Density

Breast density based on parenchymal patterns has also been strongly and consistently associated with breast cancer risk. Mammographic studies have demonstrated increased risks among women with large nodular densities and/or extensive areas of homogenous density (i.e. high proportions of connective and epithelial tissues) compared to women whose breasts were largely composed of less dense fat tissue [64-67].

1.3.6 Breast Feeding

Bernstein *et al*. reported that breast feeding decreased risk of breast cancer in premenopausal women, but only in those who had a full-term pregnancy within 5 years [68]. Adjusting for age at first pregnancy, Newcomb *et al*. reported similar decreases in risk among premenopausal women based on lifetime months of breast feeding; findings among postmenopausal women were consistent with no association [69]. A subsequent age-matched, population-based case-control study of breast cancer among postmenopausal women conducted by Enger *et al*. found an overall inverse association between breast feeding (OR=0.79, 95%CI: 0.66-0.96) and risk of breast cancer after controlling for more

than a dozen known and purported risk factors. This association was monotonically strengthened with increasing number of children breastfed [70]. In a 2000 meta-analysis of the effects of breast feeding on risk for breast cancer, Bernier *et al.* reported a slight protective effect in women who ever breast fed (Pooled OR=0.88, 95%CI: 0.84-0.92) [71].

1.3.7 Exogenous Hormone Use

Exogenous hormone use has been associated with a modest increased risk of breast cancer. Previous studies have provided substantial evidence for a modest increased risk of breast cancer among young women who are currently taking oral contraceptives (OC) or who have discontinued the use of OC within 10 years [72-75]. However, evidence from the large population-based Women's CARE Study showed no association between past or present use of OC and breast cancer in women 35-64 years old [76]. Recognizing that formulations of more modern OC have changed since the 1970's, Hunter *et al*. examined newer OC use by analyzing data from the Nurses' Health Study II. Overall findings were consistent with previous literature, indicating a small increased risk of breast cancer among current OC users (RR=1.33, 95%CI: 1.03-1.73). However, a substantially increased risk was observed among current users of triphasic preparations with levonorgestrel (a progestin) (RR=3.05, 95%CI: 2.00-4.66) [77]. Likewise, studies of combined hormone replacement therapy (HRT) have also been examined with respect to breast cancer. Contemporary dosing of combined HRT (estrogen + progestin) has been associated with an increased risk of breast cancer in postmenopausal women, but not HRT containing estrogen alone [78, 79].

1.3.8 Other Reproductive Factors

Other reproductive factors that influence endogenous estrogen exposure, such as parity and early age at first birth, have shown inverse associations with risk of breast cancer. Parity and early age at first full-term pregnancy are associated with an overall decreased risk of breast cancer [59, 80]. Lifetime risk decreases with increasing number of full-term pregnancies, but only among those women who experienced their first full-term pregnancy before the age of 20 [59, 80]. This reduction in risk observed among younger women at first birth is an *overall* reduction. In actuality, the short term effects of term pregnancies on breast cancer risk appear to increase risk [59, 81]. Bruzzi *et al.* found that full term pregnancy at any age is followed by a short increase in risk of breast cancer, irrespective of the increase associated with aging alone, that distorts the long term inverse association between parity and risk of breast cancer [81]. One explanation for this short term increase in risk is the increased level of bioavailable estradiol during the first trimester of pregnancy. Exposure to high levels of estradiol is suspected to increase risk for breast cancer [47, 50, 59, 82].

Women who experience their first full-term pregnancy after the age of 35 have been shown to experience a 20% increased risk for breast cancer compared to nulliparous women and a 70% increased risk compared to women whose first full-term pregnancy occurred before age 20 [80]. Multiparity among women experiencing their first full-term pregnancy after the age of 35 has been shown to confer additional modest increases in risk [80, 83].

1.3.9 Height

Associations between height and breast cancer risk have also been investigated, yielding conflicting results with respect to menopausal status. Several studies have found an association between height and risk of breast cancer only among postmenopausal women [84-86]. However, Ahlgren *et al*. reported a significant increase in risk among women who were in the highest quintile of height at age 14 [87]. A pooled analysis conducted by van den Brandt *et al*. also found a significant association between height and risk for breast cancer, irrespective of menopausal status [46].

1.3.10 Ionizing Radiation

Exposure to ionizing radiation has also been associated with increased risk of breast cancer. A review of evidence from Japanese survivors of the atomic bomb supports a linear relationship between radiation dose and risk of breast cancer, with age modifying this relationship [88]. Using genotype data from the Women's Environmental, Cancer, and Radiation Epidemiology study (WECARE), Brooks et al. investigated the effects of ionizing radiation on 152 SNPs involved in DNA double-strand break repair pathways in women with contralateral breast cancer [89]. None of the variants were found to interact with radiation dosage, however one haplotype in RAD50 was associated with increased risk of contralateral breast cancer [89].

1.4 Intrinsic Breast Cancer Subtypes

1.4.1 Gene Expression Patterns & Hormonal Receptor Status

In an effort to improve molecular taxonomy and targeted therapies for breast cancers, Perou *et al*. identified four distinct subtypes of breast cancer based on differences in gene expression patterns using cDNA microarrays and hierarchical clustering [90]. Each subtype can also be described by immunohistochemical staining profiles based on hormonal receptor and cellular cytokeratin status, which are surrogates for the gene expression profiles [3, 90, 91]. Of these, estrogen receptor positive (ER+) tumors are characterized by high expression of genes expressed by luminal breast cells. ER+ tumors were also distinguished immunohistochemically by staining with antibodies against luminal cytokeratins 8 and 18 [90]. Recent studies showed that ER+/luminal tumors can be further classified into luminal-A and luminal-B subtypes based on expression of human epidermal growth factor receptor-2 (HER2) (Figure 2.5) [3, 92, 93]. In comparison, luminal-A tumors are more common, express higher levels of estrogen receptor and little to no expression of HER2, and generally render a better prognosis [3, 92]. A second subtype was characterized by high expression of genes expressed by breast basal epithelial cells [90]. Support for this finding was evidenced by immunohistochemical staining of basal cell keratins 5/6 and 17 [90]. Basal-like tumors are also distinguished by the absence or low expression of estrogen receptor (ER-) and human epidermal growth factor receptor-2 (HER2-) [3, 90]. Basal-like tumors, often referred to as “triple-negative” breast cancers, are among the least responsive to hormonal and targeted therapies, and usually result in poorer prognoses [3]. A third subtype, HER2+/ER-, is characterized by low to no expression of genes that are highly expressed among luminal breast cells, low expression of estrogen receptor (ER-) and low expression of

nearly all genes associated with ER expression [90]. HER2+/ER- tumors exhibit gene expression patterns similar to those of basal-like cancers [3]. However, the availability of Herceptin treatment renders a more favorable prognosis in women with HER2+ tumors. The final subtype includes those remaining tumors whose gene expression profiles are characteristic of basal epithelial and adipose cells. Tumors of this subtype are denoted "normal-like" due to their low expression of genes typified by ER+/luminal tumors and cannot be identified via immunohistochemistry [90].

In early 2006, Carey *et al*. used immunohistochemical surrogates for expression profiling to identify subtypes, including ER and progesterone receptor (PR) status, and also to further distinguish between those tumors expressing HER2 [3]. PR was included in the definitions because it is a commonly used breast tumor marker that is regulated by ER and is associated with response to hormonal therapy [3]. The HER2+ tumors were further categorized by ER expression due to the propensity for HER2+/ER- tumors to express genes that cluster closer to those of basal-like tumors, while HER2+/ER+ more closely resembled the clustering pattern of luminal cancers (Figure 2.5) [3]. Luminal-A tumors were defined as (ER+ and/or PR+, HER2-); luminal-B tumors were (ER+ and/or PR+, HER2+); HER2+/ER- tumors were further defined by PR status as (ER-, HER2+, PR-); basal-like tumors were defined as (ER-, PR-, HER2-, cytokeratin 5/6+, and/or HER1+) [3].

1.4.2 Epidemiologic Findings

It is well established that breast cancer subtypes differ in their responsiveness to endocrine therapies as well as prognoses [3, 92-94]. Studies of invasive breast tumors have reported the best survival among women diagnosed with the most commonly diagnosed

luminal-A subtype, while women exhibiting HER2+/ER- tumors and basal-like tumors demonstrated the worst survival [3, 95, 96]. In a population-based study of African American and Caucasian women, Carey *et al.* reported women with HER2+/ER- tumors and luminal-B tumors were more likely to have lymph node metastases, while those women with basal-like tumors were not [3]. Several studies have reported a tendency for younger, premenopausal women to develop basal-like tumors when compared to older, postmenopausal women [3, 97-102]. Basal-like tumors are associated with poor prognosis, often characterized by higher grade, higher mitotic index, and significant DNA mutations [3, 95, 100, 103-106]. Basal-like tumors are also characterized by aneuploidy [107, 108]. Other research has reported that basal-like tumors are more likely to be larger and exhibit a greater tendency to metastasize [95, 106, 109-112]. Basal-like breast cancers are also more likely to be associated with *BRCA1* mutations compared to other subtypes, suggesting a distinct biological mechanism [92, 104, 113]. Since *BRCA1* mutation carriers tend to develop basal-like breast tumors, there may be other inherited genetic variants that predispose to developing specific subtypes of breast cancer [3, 92, 113]. In addition to relatively worse prognoses and fewer treatment options, basal-like breast cancers tend to develop in younger African American women disproportionately [3, 97, 98, 114]. Carey *et al.* found a high prevalence of basal-like tumors in African American women, all of whom were negative for *BRCA1* mutations, suggesting genes other than *BRCA1* may be associated with basal-like breast cancers as well [3]. In line with the polygenic model, increasing evidence suggests that common risk variants differ by intrinsic subtype of breast cancer [115, 116].

1.4.3 Etiology

As more epidemiologic evidence supports the biological heterogeneity of breast cancer, assessing risk factors by distinct breast cancer subtypes may reveal more accurate associations. Lacroix *et al.* suggest that molecular tumor characteristics do not change appreciably over the progression from *in situ* carcinoma to invasive carcinoma [117]. As such, exposures that are associated with breast cancer etiology may show different associations according to molecular subtype. Several studies have found varying associations between common risk factors for breast cancer (age, parity, age at first birth, age at menarche, race) and hormone receptor status [114, 118].

The Carolina Breast Cancer Study (CBCS), a population-based case-control study of African American and Caucasian women in North Carolina, reported increased odds of basal-like breast cancer as opposed to luminal breast cancer among women who were younger at first pregnancy [97]. On the other hand, a reduced odds of basal-like breast cancer was noted among women who breastfed more children for a longer duration, but not among luminal cases [97]. This finding is in contrast to other study findings indicating a reduced risk of breast cancer among Chinese women who breast fed, however those studies did not stratify by intrinsic subtype and were based on study populations that are not comparable to CBCS [119, 120]. In addition to finding the highest prevalence of basal-like tumors in younger African American women, Millikan *et al.* also reported increased odds of basal-like breast cancer associated with higher waist-hip ratio in both pre- and post-menopausal women [97].

A case control study of invasive breast cancer in Polish women (805 cases, 2,502 controls) reported higher BMI was associated with decreased odds of luminal breast tumors

among premenopausal women (OR=0.71, 95% CI: 0.57-0.88 per five-unit increase), while a slightly increased odds for basal-like breast cancer was noted among women with higher BMI (OR=1.18, 95% CI: 0.86-1.64) [114]. The same study also noted a significantly reduced risk of basal-like breast cancer with increasing age at menarche (OR=0.78, 95% CI: 0.68-0.89 per 2-year increase) [114].

Contrasting data from two centers participating in the Cancer and Steroid Hormone Study (CASH) suggests clear differences in risk associated with late age at first birth between African American and Caucasian women. A stronger association between ER negative tumors and late age at first birth was noted among African American women, whereas a strong association was noted with ER positive tumors in Caucasian women [118, 121, 122]. Significant heterogeneity of associations by subtype was also reported in a case only study of 2,544 breast cancer cases classified by ER, PR, and HER2 status [123]. Notable risk factors that may be related to the development of particular molecular subtypes of breast cancer included: BMI, alcohol consumption, and history of breastfeeding [123].

1.5 Cell Cycle Regulation and Cancer

Cancers are characterized by aberrations in cell cycle regulation, leading to inappropriate cell replication. This unchecked cell proliferation is associated with reduction in or loss of sensitivity to normal signals to either differentiate or initiate apoptosis. Many genes are responsible for adherence to proper cell cycle function, and interpreting the changes that can disrupt this process is integral for understanding the etiology of cancer [124].

Two general types of genetic mutations have been shown to contribute to abnormal cell proliferation and the development of cancer: gain-of-function mutations and loss-of-function

mutations. Gain-of-function mutations are characterized by the transformation of proto-oncogenes into oncogenes (mutated genes that once performed normal cellular functions as proto-oncogenes, and now contribute to aberrant cell proliferation) [125]. Proto-oncogenes perform important functions within the cell, from signal transduction to programmed cell death [125]. Conversion of proto-oncogenes into oncogenes can result in unregulated cell growth [125]. Studies have shown that individual oncogenes can have identical effects leading to gain-of-function or can be cell-type specific, suggesting different genetic pathways resulting in cancer [124]. Gain-of-function mutations only require one copy of the mutant allele for transformation to the oncogene [124].

Loss-of-function mutations occur in tumor suppressor genes and are far more common. Only individuals who are homozygous for the mutant allele will exhibit loss-of-function, with heterozygotes demonstrating the normal wild-type phenotype. However, heterozygotes will bear an increased risk for developing cancer due to the fact that a subsequent deleterious mutation will prevent normal gene function [124]. Loss-of-function mutations in tumor suppressor genes have been shown to result in circumvention of normal negative regulation that controls entry into the cell cycle [126]. One such example is a loss-of-function mutation in the tumor suppressor gene p53; p53 normally functions to arrest cell cycle progression in response to DNA damage [127]. Loss of normal p53 function allows for unchecked cellular proliferation of mutant DNA.

1.6 The Centrosome and Centrosome Cycle

The Centrosome is a membrane-less organelle whose function is necessary to maintain cell cycle fidelity [128]. It is composed of a pair of barrel shaped centrioles, surrounded by

an amorphous pericentriolar material (PCM) [129, 130]. Often referred to as the microtubule organizing center of the cell, centrosomes determine the number, length and distribution of microtubules. Animal cells normally contain one centrosome which is duplicated once and only once per cell cycle. Centrosomal duplication involves centriolar duplication in G_1 of interphase and culminates in dual centrosomes by G_2/mitosis. As this is a semi-conservative process, one of the centrioles present in the centrosome will be more "mature" than the other and is denoted the mother centriole since it has experienced more cell cycles. Likewise a centriole that has not yet completed a full cell cycle is referred to as a daughter centriole. The distinction is in the number of microtubules each centriole can nucleate; the more mature centriole can be identified by appendages protruding from its distal end and is capable of nucleating more microtubules [129, 131]. During mitosis, the centrosomes nucleate microtubules in a polarized array with their positive ends directed outward from the electron-dense PCM [129, 130]. This polymerization of microtubules toward either pole of the cell forms the spindle apparatus that will facilitate alignment of the chromosomes in preparation for cell division. The centrosomes are also important for cytokinesis and in establishing a midpoint at metaphase for the cleavage furrow to form; studies have shown that removal of the centrosome from cells resulted in failure to complete cytokinesis [129, 132]. Further studies of individual centriole removal provided evidence of the same; after removing one of two centrioles from a cell's centrosome, Piel *et al.* demonstrated that the daughter cell lacking a centriole failed to complete cytokinesis whereas the daughter cell containing a centriole pair went on to complete the cell cycle normally [129, 133].

Cells completing cytokinesis and exiting mitosis are characterized by a single centrosome, composed of two orthogonally positioned centrioles. During G_1, the centrioles separate in preparation for duplication. Centriole duplication is distinguished by the formation of procentrioles on either parental centriole, a process referred to as centriole engagement [134]. Formation and orientation of the procentrioles and the duplication process are thought to be tightly regimented to prevent more than one replication in the same cell cycle. Tsou and Stearns hypothesize that the physical presence and positioning of the procentrioles blocks reduplication [134]. During S phase and throughout G_2, the procentrioles grow until they achieve their maximum length. Maturation of the previously immature centriole begins during G_2 and culminates in the development of distal appendages. The maturation process requires approximately 1.5 cell cycles to complete [129, 131]. As the cell transitions from G_2 into mitosis, it contains two centrosomes which will separate and migrate to either pole of the dividing cell to establish the mitotic spindle. A concomitant centrosomal and cell cycle are integral to ensure two independent centrosomes at mitosis, bipolar spindle formation, and equal segregation of chromosomes. The separation and migration of the centrosomes during mitosis is regulated by several kinases, including the serine-threonine kinases of the Aurora family of proteins [129]. Glover *et al.* found that mutations in the single *Aurora* gene of Drosophila result in several deleterious manifestations. In early metaphase, failure of the centrosomes to separate leads to monopolar spindles and aneuploidy accompanied by centrosomal amplification in daughter cells [135-137], suggesting a loss-of-function of the serine-threonine kinase [136, 137]. After chromosomal alignment and segregation, cytokinesis ensues resulting in

identical daughter cells each containing a single centrosome. The centrosome cycle is then repeated.

1.7 Centrosomal Amplification and Breast Cancer

It is important that a cell undergoing mitosis contain two independent centrosomes, each located at either pole. Since the centrosome acts as a microtubule organizing center in the dividing cell, the presence of more than two could result in improper formation of the spindle apparatus, aberrant segregation of chromosomes, or failure of cytokinesis [138]. Pihan *et al.* found amplified centrosomes (more than two centrosomes or more than four centrioles) as commonly characteristic of solid malignant tumors [139, 140]. In a study of high grade human breast tumors, Lingle *et al.* had similar findings, reporting increased microtubule nucleation in addition to amplified centrosomes [138].

Seven hundred eighty-two SNPs from 101 centrosomal genes were analyzed in a population-based study of 798 invasive breast cancer cases and 843 controls from the Mayo Clinic Breast Cancer Study. Findings indicated that genes involved in the centrosome regulatory pathway were highly enriched with SNPs associated with risk of breast cancer ($p=4.6x10^{-50}$) [141]. Amplified centrosomes are suspected of contributing to aneuploidy by increasing the rate of aberrant mitoses resulting in chromosomal missegregation [135, 142]. Furthermore, Lingle *et al.* found evidence to support the hypothesis that centrosomal amplification occurs early in the tumorigenesis process by demonstrating supernumerary centrosomes in ductal carcinoma *in situ* (DCIS) [142]. Centrosome amplification can result from multiple mechanisms, including dissociation of the centrosomal and cell cycles [143] and overexpression of Aurora A serine-threonine kinase (*AURKA*) [135]. Balczon *et al.*

demonstrated unchecked centrosome replication in cells arrested at the G_1/S boundary, supporting the hypothesis that centrosome replication is driven by activation and inactivation of centrosomal regulatory genes during the cell cycle [143]. An investigation by Zhou *et al.* showed that the *AURKA* locus encoding a serine-threonine kinase associated with centrosome regulation was implicated in causing centrosome duplication abnormalities and aneuploidy in mammalian cells; overexpression of *AURKA* was associated with centrosome amplification and chromosomal instability [135, 144]. Chromosomal instability is the rate of gains or losses of chromosomes, whereas aneuploidy is the cross-sectional disposition of the cell with respect to chromosome number [129]. Although aneuploidy is a common characteristic of cancer cells [129, 145-148], it is unclear as to whether or not it causes or results from disease progression.

1.8 *AURKA*

The *Aurora A* gene, also known as *AURKA*, encodes a serine/threonine kinase and is located on the q arm of chromosome 20 at amplicon 13.2, a region commonly amplified in human breast cancers [149, 150]. Isola *et al.* reported poorer prognosis and survival among breast cancer cases exhibiting highly amplified 20q13 [151]. Likewise, Tanner *et al.* found high amplification of 20q13 in primary breast carcinomas to be significantly associated with high histological grade, aneuploidy, short disease-free survival, and poor clinical outcome suggesting this region contains a gene involved in breast cancer progression [152]. *AURKA* functions in centrosomal maturation and separation, mitotic spindle formation and stabilization, and proper chromosome segregation [153]. *AURKA* activity is localized at the centrosome throughout all phases of the cell cycle [129] and is necessary for cell cycle

progression [154-157]. In a study of mammary tumorigenesis in mice, Wang *et al.* demonstrated centrosome amplification and aneuploidy in transgenic mice overexpressing *AURKA* in mammary epithelium [158]. Notably, severe chromosomal abnormalities failed to trigger apoptosis in cells overexpressing *AURKA*, allowing for continued proliferation of abnormal karyotypes [158]. Tanaka *et al.* showed overexpression of *AURKA* in 94% of invasive ductal carcinomas of the breast in a cohort of Japanese women [159].

Few population-based studies of genetic variation in *AURKA* have been conducted, and those that have been conducted have focused on a few functional variants in European and Asian populations. The T/A coding region polymorphism (F31I) on *AURKA* that results in an amino acid substitution (Phe→Ile) has been studied extensively in European and Asian populations. Functional evidence for a biochemical difference between the proteins encoded by the Ile31 variant and the more common Phe31 variant has been reported. The Ile allele is more efficient in inducing cell growth *in vitro*, which facilitates the oncogenic effect of *AURKA* [160]. The Ile-Ile genotype has been associated with increased aneuploidy in human colon tumors, compared to the more common Phe-Phe genotype [161]. Sun T. *et al.* found increased risk for breast carcinoma associated with the Ile/Ile genotype in a case-control study of unrelated Han Chinese women [162]. Additional studies of (F31I) in both Chinese [161, 163] and European [164] populations failed to replicate the finding. A 2011 meta-analysis conducted by Sun H. *et al.*, which included 11 case-control studies, reported a slight inverse association between the Ile/Ile genotype and odds of breast cancer (OR=0.857, 95% CI: 0.742-0.991), but only in Asian populations [165]. Overall, the evidence for an association between the F31I polymorphism and risk for breast cancer is summarized in Figure 2.14. Another coding region polymorphism (V57I) on *AURKA* resulting in a

valine→isoleucine substitution has been investigated for potential association with risk of breast cancer. Egan *et al.* reported no association with breast cancer risk among European women with the Ile/Ile genotype [164]. However, examining a relatively common genotype combining the two polymorphisms (31I-57V/31I-57V), Egan *et al.* reported a 2-fold increase in risk of breast cancer (OR=1.96; 95% CI: 1.01-3.79) [164]. A study by Dai *et al.* investigating both (F31I) and (V57I) reported no association between the combined (31I-57V/31I-57V) genotype and risk of breast cancer in a population-based study of Chinese women [161]. Evidence for purported breast risk loci on *AURKA* are summarized in Table 5.

Because *AURKA* is strongly involved in centrosomal regulation, and aberrations in the centrosomal cycle lead to aneuploidy, polymorphisms on *AURKA* may also be more strongly associated with subtypes of breast cancer that are known to demonstrate relatively high levels of aneuploidy; namely, triple-negative and basal like breast cancers [108, 113]. Relative to all other subtypes, basal like and triple-negative breast cancers are more likely to demonstrate higher levels of aneuploidy, with other subtypes demonstrating a more variable degree of aneuploidy [113]. This proposal will be the first to investigate *AURKA*'s association with intrinsic subtype of breast cancer.

1.9 *BRCA1* and Interacting Genes: *BARD1, BRIP1,* and *ZNF350*

AURKA is not the only centrosomal regulatory gene that has been associated with breast cancer. *BRCA1* proteins are involved in centrosomal regulation as well, and are known to interact with *AURKA* proteins [166]. The Aurora-A kinase phosphorylates *BRCA1* and regulates *BRCA1* mediated inhibition of centrosome-dependent microtubule nucleation

[167]. Ruan *et al.* reported a significant interaction between haplotypes on *AURKA* and *BRCA1* in a Han Chinese population, warranting further investigation in different populations [166]. Other purported risk loci on *BRCA1* are summarized in Table 6.

Genetic variants conferring high risk for breast cancer are not sufficient to cause breast cancer. Even for carriers of mutations in the highly penetrant *BRCA1* and *BRCA2*, the distribution of risk varies suggesting possible gene-gene or gene-environment interactions that affect overall risk [29]. Evidence supporting the multiplicative joint effects of low-penetrance genes on breast cancer risk has been reported by Antoniou *et al.* [28]. *BARD1*, *BRIP1* and *ZNF350* are three putative low penetrance breast cancer susceptibility genes.

The *BARD1* protein interacts with *BRCA1* to form a heterodimer complex *BRCA1/BARD1* [168]. By itself, *BRCA1* confers an ubiquitin ligase activity that is essential for its normal tumor suppression action, namely in coordinating DNA repair [168]. However, when complexed with *BARD1*, the ubiquitin ligase activity is markedly enhanced [168]. This interaction with *BRCA1* suggests a possible role for *BARD1* in DNA repair processes. Mutations in *BRCA1* are known to deactivate the ubiquitin ligase activity of the *BRCA1/BARD1* heterodimer complex [168, 169], which has been correlated with its tumor suppression function [170].

Germline mutations in *BARD1* were demonstrated in a cohort of 40 Italian families with a history of breast and breast/ovarian cancer whose probands were chosen due to their lack of *BRCA1* mutations [171]. The Nordic collaborative study has reported that a specific *BARD1* variant (Cys557Ser) may represent a common breast cancer susceptibility allele [172, 173]. In contrast, a Japanese case-control study found little to no evidence supporting a role for mutations in *BARD1* contributing to familial breast cancer risk [174].

BRIP1 encodes a helicase that binds directly to the C-terminus of *BRCA1* and directly contributes to the double-strand break repair function of *BRCA1* [175]. Cantor *et al.* found germline mutations in *BRIP1* affecting the helicase domain among breast cancer patients with normal *BRCA1* but not among controls, implicating *BRIP1* as a potential low-penetrance gene that contributes to familial breast cancer risk. Further evidence of *BRIP1's* important association with *BRCA1* was noted: missense and deletion mutations in the C-terminus of *BRCA1*, which inactivated its normal double-strand break repair function, also inhibited *BRIP1* binding, suggesting a functional role in DNA repair [176].

ZNF350 is a corepressor of *GADD45*, which is involved in cell cycle arrest at the G2/M checkpoint subsequent to DNA damage [177]. *GADD45* is regulated by both *ZNF350* and *BRCA1* [177, 178]. The interaction of *ZNF350* with *BRCA1* is necessary for the modulation of *GADD45* [179]. *ZNF350* has been shown to negatively regulate overexpression of *BRCA1* [177], but its role in human carcinomas is largely unknown [180].

Due to their interactive roles with *BRCA1*, *BARD1*, *BRIP1* and/or *ZNF350* may show gene-gene interactions in association with odds of breast cancer with other *BRCA1*-interacting genes, such as *AURKA*.

1.10 Summary—Background and Significance

Breast cancer still represents a public health burden in the United States. Growing etiologic understanding of the heterogeneity of breast cancer will be crucial to future prevention and treatment, especially in light of racial disparities in intrinsic breast cancer subtype. As our understanding of breast cancer heterogeneity improves, environmental and genetic risk factor investigations may become more focused and informative. Given the unique CBCS population, this study is innovative in its attempt to further characterize racial

differences in breast cancer subtype using a candidate gene approach. This study investigated important genes involved in cell cycle regulation that were carefully chosen for their suspected role in oncogenesis.

Due to its function as an important centrosomal and cell cycle regulator, *AURKA* may play a significant role in oncogenesis, especially with respect to the proliferation of aneuploid cells—a common feature of basal-like tumors. Although *AURKA* has been investigated in large population-based studies, these studies have been largely conducted among European and Chinese individuals. In light of a disproportionate number of basal-like cases among African American women, and due to previous investigations' findings that basal-like and triple negative breast tumors demonstrate higher degrees of aneuploidy, the CBCS offers a novel opportunity to evaluate *AURKA* in African American women.

This study was the first to investigate statistical interactions between *BRCA1* and *AURKA* in African American women with breast cancer. Since *BRCA1* is known to be an important regulator of the cell cycle, and since *BRCA1* proteins are known to interact with *AURKA* proteins, it's reasonable to suspect gene—gene interactions between them. In addition, three other *BRCA1*-interacting genes (*BARD1, BRIP1,* and *ZNF350*), which have not been investigated heavily with respect to breast cancer rate, were also investigated for statistical interaction with *AURKA.*

REFERENCES

1. American Cancer Society. *Cancer Facts and Figures 2011*. In. Atlanta: American Cancer Society.

2. National Cancer Institute--Surveillance Epidemiology and End Results (SEER) Program 2010. In. Edited by Incidence - SEER 9 Regs Research Data NS-.

3. Carey LA, Perou CM, Livasy CA, Dressler LG, Cowan D, Conway K, Karaca G, Troester MA, Tse CK, Edmiston S *et al*: Race, breast cancer subtypes, and survival in the Carolina Breast Cancer Study. *JAMA* 2006, 295(21):2492-2502.

4. 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.

5. Porter PL, Lund MJ, Lin MG, Yuan X, Liff JM, Flagg EW, Coates RJ, Eley JW: Racial differences in the expression of cell cycle-regulatory proteins in breast carcinoma. *Cancer* 2004, 100(12):2533-2542.

6. Eley JW, Hill HA, Chen VW, Austin DF, Wesley MN, Muss HB, Greenberg RS, Coates RJ, Correa P, Redmond CK *et al*: Racial differences in survival from breast cancer. Results of the National Cancer Institute Black/White Cancer Survival Study. *JAMA* 1994, 272(12):947-954.

7. Clegg LX, Li FP, Hankey BF, Chu K, Edwards BK: Cancer survival among US whites and minorities: a SEER (Surveillance, Epidemiology, and End Results) Program population-based study. *Arch Intern Med* 2002, 162(17):1985-1993.

8. Shavers VL, Brown ML: Racial and ethnic disparities in the receipt of cancer treatment. *J Natl Cancer Inst* 2002, 94(5):334-357.

9. Li CI, Malone KE, Daling JR: Differences in breast cancer stage, treatment, and survival by race and ethnicity. *Arch Intern Med* 2003, 163(1):49-56.

10. O'Malley CD, Le GM, Glaser SL, Shema SJ, West DW: Socioeconomic status and breast carcinoma survival in four racial/ethnic groups: a population-based study. *Cancer* 2003, 97(5):1303-1311.

11. Chen VW, Correa P, Kurman RJ, Wu XC, Eley JW, Austin D, Muss H, Hunter CP, Redmond C, Sobhan M *et al*: Histological characteristics of breast carcinoma in blacks and whites. *Cancer Epidemiol Biomarkers Prev* 1994, 3(2):127-135.

12. Cunningham JE, Butler WM: Racial disparities in female breast cancer in South Carolina: clinical evidence for a biological basis. *Breast Cancer Res Treat* 2004, 88(2):161-176.

13. del Carmen MG, Hughes KS, Halpern E, Rafferty E, Kopans D, Parisky YR, Sardi A, Esserman L, Rust S, Michaelson J: Racial differences in mammographic breast density. *Cancer* 2003, 98(3):590-596.

14. Smigal C, Jemal A, Ward E, Cokkinides V, Smith R, Howe HL, Thun M: Trends in breast cancer by race and ethnicity: update 2006. *CA Cancer J Clin* 2006, 56(3):168-183.

15. Yankaskas BC, Gill KS: Diagnostic mammography performance and race: outcomes in Black and White women. *Cancer* 2005, 104(12):2671-2681.

16. Gill KS, Yankaskas BC: Screening mammography performance and cancer detection among black women and white women in community practice. *Cancer* 2004, 100(1):139-148.

17. 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):8327-8330.

18. Miller BA, Hankey BF, Thomas TL: Impact of sociodemographic factors, hormone receptor status, and tumor grade on ethnic differences in tumor stage and size for breast cancer in US women. *Am J Epidemiol* 2002, 155(6):534-545.

19. Yood MU, Johnson CC, Blount A, Abrams J, Wolman E, McCarthy BD, Raju U, Nathanson DS, Worsham M, Wolman SR: Race and differences in breast cancer survival in a managed care population. *J Natl Cancer Inst* 1999, 91(17):1487-1491.

20. Chlebowski RT, Chen Z, Anderson GL, Rohan T, Aragaki A, Lane D, Dolan NC, Paskett ED, McTiernan A, Hubbell FA *et al*: Ethnicity and breast cancer: factors influencing differences in incidence and outcome. *J Natl Cancer Inst* 2005, 97(6):439-448.

21. Smith-Bindman R, Miglioretti DL, Lurie N, Abraham L, Barbash RB, Strzelczyk J, Dignan M, Barlow WE, Beasley CM, Kerlikowske K: Does utilization of screening mammography explain racial and ethnic differences in breast cancer? *Ann Intern Med* 2006, 144(8):541-553.

22. American Cancer Society--Breast Cancer Risk Factors (2010) [http://www.cancer.org/cancer/breastcancer/detailedguide/breast-cancer-risk-factors]

23. 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):1389-1399.

24. Bradbury AR, Olopade OI: Genetic susceptibility to breast cancer. *Rev Endocr Metab Disord* 2007, 8(3):255-267.

25. Balmain A, Gray J, Ponder B: The genetics and genomics of cancer. *Nat Genet* 2003, 33 Suppl:238-244.

26. Lichtenstein P, Holm NV, Verkasalo PK, Iliadou A, Kaprio J, Koskenvuo M, Pukkala E, Skytthe A, Hemminki K: 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.

27. Peto J: Cancer epidemiology in the last century and the next decade. *Nature* 2001, 411(6835):390-395.

28. Antoniou AC, Pharoah PD, McMullan G, Day NE, Stratton MR, Peto J, Ponder BJ, Easton DF: A comprehensive model for familial breast cancer incorporating BRCA1, BRCA2 and other genes. *Br J Cancer* 2002, 86(1):76-83.

29. 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):33-36.

30. 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.

31. Antoniou AC, Easton DF: Polygenic inheritance of breast cancer: Implications for design of association studies. *Genet Epidemiol* 2003, 25(3):190-202.

32. Easton DF, Pooley KA, Dunning AM, Pharoah PD, Thompson D, Ballinger DG, Struewing JP, Morrison J, Field H, Luben R *et al*: Genome-wide association study identifies novel breast cancer susceptibility loci. *Nature* 2007, 447(7148):1087-1093.

33. Zheng W, Long J, Gao YT, Li C, Zheng Y, Xiang YB, Wen W, Levy S, Deming SL, Haines JL *et al*: Genome-wide association study identifies a new breast cancer susceptibility locus at 6q25.1. *Nat Genet* 2009, 41(3):324-328.

34. Hutter CM, Young AM, Ochs-Balcom HM, Carty CL, Wang T, Chen CT, Rohan TE, Kooperberg C, Peters U: Replication of breast cancer GWAS susceptibility loci in the Women's Health Initiative African American SHARe Study. *Cancer Epidemiol Biomarkers Prev* 2011, 20(9):1950-1959.

35. Shulman LP: Hereditary breast and ovarian cancer (HBOC): clinical features and counseling for BRCA1 and BRCA2, Lynch syndrome, Cowden syndrome, and Li-Fraumeni syndrome. *Obstet Gynecol Clin North Am* 2010, 37(1):109-133, Table of Contents.

36. Lustbader ED, Williams WR, Bondy ML, Strom S, Strong LC: Segregation analysis of cancer in families of childhood soft-tissue-sarcoma patients. *Am J Hum Genet* 1992, 51(2):344-356.

37. Hankinson S HD: Breast Cancer. In: *Textbook of Cancer Epidemiology.* Edited by Adami H HD, Trichopoulos D. New York: Oxford University Press, Inc.; 2002: 301-309.

38. Orloff MS, Eng C: Genetic and phenotypic heterogeneity in the PTEN hamartoma tumour syndrome. *Oncogene* 2008, 27(41):5387-5397.

39. Lynch ED, Ostermeyer EA, Lee MK, Arena JF, Ji H, Dann J, Swisshelm K, Suchard D, MacLeod PM, Kvinnsland S *et al*: Inherited mutations in PTEN that are associated with breast cancer, cowden disease, and juvenile polyposis. *Am J Hum Genet* 1997, 61(6):1254-1260.

40. Hampel H, Panescu J, Lockman J, Sotamaa K, Fix D, Comeras I, LaJeunesse J, Nakagawa H, Westman JA, Prior TW *et al*: Comment on: Screening for Lynch Syndrome (Hereditary Nonpolyposis Colorectal Cancer) among Endometrial Cancer Patients. *Cancer Res* 2007, 67(19):9603.

41. Hemminki A, Avizienyte E, Roth S, Loukola A, Aaltonen LA, Jarvinen H, de la Chapelle A: [A serine/threonine kinase gene defective in Peutz-Jeghers syndrome]. *Duodecim* 1998, 114(7):667-668.

42. Hearle N, Schumacher V, Menko FH, Olschwang S, Boardman LA, Gille JJ, Keller JJ, Westerman AM, Scott RJ, Lim W *et al*: Frequency and spectrum of cancers in the Peutz-Jeghers syndrome. *Clin Cancer Res* 2006, 12(10):3209-3215.

43. Altekruse SF KC, Krapcho M, Neyman N, Aminou R, Waldron W, Ruhl J, Howlader N, Tatalovich Z, Cho H, Mariotto A, Eisner MP, Lewis DR, Cronin K, Chen HS, Feuer EJ, Stinchcomb DG, Edwards BK (eds). : SEER Cancer Statistics Review, 1975-2007. In. Bethesda: National Cancer Institute; 2010.

44. SEER Stat Fact Sheets: Breast [http://seer.cancer.gov/statfacts/html/breast.html]

45. Breast Cancer Rates by Age [http://www.cdc.gov/cancer/breast/statistics/age.htm]

46. van den Brandt PA, Spiegelman D, Yaun SS, Adami HO, Beeson L, Folsom AR, Fraser G, Goldbohm RA, Graham S, Kushi L *et al*: Pooled analysis of prospective cohort studies on height, weight, and breast cancer risk. *Am J Epidemiol* 2000, 152(6):514-527.

47. Sherman BM, Korenman SG: Measurement of serum LH, FSH, estradiol and progesterone in disorders of the human menstrual cycle: the inadequate luteal phase. *J Clin Endocrinol Metab* 1974, 39(1):145-149.

48. Stoll BA: Breast cancer: the obesity connection. *Br J Cancer* 1994, 69(5):799-801.

49. Key TJ, Pike MC: The role of oestrogens and progestagens in the epidemiology and prevention of breast cancer. *Eur J Cancer Clin Oncol* 1988, 24(1):29-43.

50. Ursin G, Longnecker MP, Haile RW, Greenland S: A meta-analysis of body mass index and risk of premenopausal breast cancer. *Epidemiology* 1995, 6(2):137-141.

51. Hunter DJ, Willett WC: Diet, body size, and breast cancer. *Epidemiol Rev* 1993, 15(1):110-132.

52. IARC: Weight Control and Physical Activity. In., vol. 6. Lyon; 2002.

53. Lahmann PH, Hoffmann K, Allen N, van Gils CH, Khaw KT, Tehard B, Berrino F, Tjonneland A, Bigaard J, Olsen A *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):762-771.

54. Cleary MP, Maihle NJ: The role of body mass index in the relative risk of developing premenopausal versus postmenopausal breast cancer. *Proc Soc Exp Biol Med* 1997, 216(1):28-43.

55. Kaaks R: Nutrition, hormones, and breast cancer: is insulin the missing link? *Cancer Causes Control* 1996, 7(6):605-625.

56. 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):127-138.

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

58. Monninkhof EM, Elias SG, Vlems FA, van der Tweel I, Schuit AJ, Voskuil DW, van Leeuwen FE: Physical activity and breast cancer: a systematic review. *Epidemiology* 2007, 18(1):137-157.

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

60. 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.

61. Henderson BE, Pike MC, Casagrande JT: Breast cancer and the oestrogen window hypothesis. *Lancet* 1981, 2(8242):363-364.

62. Vihko R, Apter D: Endocrine characteristics of adolescent menstrual cycles: impact of early menarche. *J Steroid Biochem* 1984, 20(1):231-236.

63. Henderson BE, Feigelson HS: Hormonal carcinogenesis. *Carcinogenesis* 2000, 21(3):427-433.

64. 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):428-437.

65. Boyd NF, Lockwood GA, Martin LJ, Knight JA, Byng JW, Yaffe MJ, Tritchler DL: Mammographic densities and breast cancer risk. *Breast Dis* 1998, 10(3-4):113-126.

66. 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):2550-2556.

67. 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):1159-1169.

68. Enger SM, Ross RK, Henderson B, Bernstein L: Breastfeeding history, pregnancy experience and risk of breast cancer. *Br J Cancer* 1997, 76(1):118-123.

69. Newcomb PA, Storer BE, Longnecker MP, Mittendorf R, Greenberg ER, Clapp RW, Burke KP, Willett WC, MacMahon B: Lactation and a reduced risk of premenopausal breast cancer. *N Engl J Med* 1994, 330(2):81-87.

70. Enger SM, Ross RK, Paganini-Hill A, Bernstein L: Breastfeeding experience and breast cancer risk among postmenopausal women. *Cancer Epidemiol Biomarkers Prev* 1998, 7(5):365-369.

71. Bernier MO, Plu-Bureau G, Bossard N, Ayzac L, Thalabard JC: Breastfeeding and risk of breast cancer: a metaanalysis of published studies. *Hum Reprod Update* 2000, 6(4):374-386.

72. 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. Collaborative Group on Hormonal Factors in Breast Cancer. *Lancet* 1996, 347(9017):1713-1727.

73. Romieu I, Willett WC, Colditz GA, Stampfer MJ, Rosner B, Hennekens CH, Speizer FE: Prospective study of oral contraceptive use and risk of breast cancer in women. *J Natl Cancer Inst* 1989, 81(17):1313-1321.

74. Breast cancer and hormonal contraceptives: further results. Collaborative Group on Hormonal Factors in Breast Cancer. *Contraception* 1996, 54(3 Suppl):1S-106S.

75. Brinton LA, Daling JR, Liff JM, Schoenberg JB, Malone KE, Stanford JL, Coates RJ, Gammon MD, Hanson L, Hoover RN: Oral contraceptives and breast cancer risk among younger women. *J Natl Cancer Inst* 1995, 87(11):827-835.

76. Davidson NE, Helzlsouer KJ: Good news about oral contraceptives. *N Engl J Med* 2002, 346(26):2078-2079.

77. Hunter DJ, Colditz GA, Hankinson SE, Malspeis S, Spiegelman D, Chen W, Stampfer MJ, Willett WC: Oral contraceptive use and breast cancer: a prospective study of young women. *Cancer Epidemiol Biomarkers Prev* 2010, 19(10):2496-2502.

78. Chlebowski RT, Hendrix SL, Langer RD, Stefanick ML, Gass M, Lane D, Rodabough RJ, Gilligan MA, Cyr MG, Thomson CA *et al*: Influence of estrogen plus progestin on breast cancer and mammography in healthy postmenopausal women: the Women's Health Initiative Randomized Trial. *JAMA* 2003, 289(24):3243-3253.

79. Martin LJ, Minkin S, Boyd NF: Hormone therapy, mammographic density, and breast cancer risk. *Maturitas* 2009, 64(1):20-26.

80. MacMahon B, Cole P, Lin TM, Lowe CR, Mirra AP, Ravnihar B, Salber EJ, Valaoras VG, Yuasa S: Age at first birth and breast cancer risk. *Bull World Health Organ* 1970, 43(2):209-221.

81. Bruzzi P, Negri E, La Vecchia C, Decarli A, Palli D, Parazzini F, Del Turco MR: Short term increase in risk of breast cancer after full term pregnancy. *BMJ* 1988, 297(6656):1096-1098.

82. Henderson BE, Ross RK, Judd HL, Krailo MD, Pike MC: Do regular ovulatory cycles increase breast cancer risk? *Cancer* 1985, 56(5):1206-1208.

83. Yuan JM, Yu MC, Ross RK, Gao YT, Henderson BE: Risk factors for breast cancer in Chinese women in Shanghai. *Cancer Res* 1988, 48(7):1949-1953.

84. 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.

85. 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):217-224.

86. 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.

87. Ahlgren M, Melbye M, Wohlfahrt J, Sorensen TI: Growth patterns and the risk of breast cancer in women. *N Engl J Med* 2004, 351(16):1619-1626.

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

89. Brooks JD, Teraoka SN, Reiner AS, Satagopan JM, Bernstein L, Thomas DC, Capanu M, Stovall M, Smith SA, Wei S *et al*: Variants in activators and downstream targets of ATM, radiation exposure, and contralateral breast cancer risk in the WECARE study. *Hum Mutat* 2011.

90. Perou CM, Sorlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, Pollack JR, Ross DT, Johnsen H, Akslen LA *et al*: Molecular portraits of human breast tumours. *Nature* 2000, 406(6797):747-752.

91. Nielsen TO, Hsu FD, Jensen K, Cheang M, Karaca G, Hu Z, Hernandez-Boussard T, Livasy C, Cowan D, Dressler L *et al*: Immunohistochemical and clinical characterization of the basal-like subtype of invasive breast carcinoma. *Clin Cancer Res* 2004, 10(16):5367-5374.

92. Sorlie T, Tibshirani R, Parker J, Hastie T, Marron JS, Nobel A, Deng S, Johnsen H, Pesich R, Geisler S *et al*: Repeated observation of breast tumor subtypes in independent gene expression data sets. *Proc Natl Acad Sci U S A* 2003, 100(14):8418-8423.

93. Sorlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, Hastie T, Eisen MB, van de Rijn M, Jeffrey SS *et al*: Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci U S A* 2001, 98(19):10869-10874.

94. Sotiriou C, Neo SY, McShane LM, Korn EL, Long PM, Jazaeri A, Martiat P, Fox SB, Harris AL, Liu ET: Breast cancer classification and prognosis based on gene expression profiles from a population-based study. *Proc Natl Acad Sci U S A* 2003, 100(18):10393-10398.

95. 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):1217-1226.

96. Kurebayashi J, Moriya T, Ishida T, Hirakawa H, Kurosumi M, Akiyama F, Kinoshita T, Takei H, Takahashi K, Ikeda M *et al*: The prevalence of intrinsic subtypes and prognosis in breast cancer patients of different races. *Breast* 2007, 16 Suppl 2:S72-77.

97. Millikan RC, Newman B, Tse CK, Moorman PG, Conway K, Dressler LG, Smith LV, Labbok MH, Geradts J, Bensen JT *et al*: Epidemiology of basal-like breast cancer. *Breast Cancer Res Treat* 2008, 109(1):123-139.

98. Olopade OI, Grushko TA, Nanda R, Huo D: Advances in breast cancer: pathways to personalized medicine. *Clin Cancer Res* 2008, 14(24):7988-7999.

99. Calza S, Hall P, Auer G, Bjohle J, Klaar S, Kronenwett U, Liu ET, Miller L, Ploner A, Smeds J *et al*: Intrinsic molecular signature of breast cancer in a population-based cohort of 412 patients. *Breast Cancer Res* 2006, 8(4):R34.

100. 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):179-185.

101. Huo D, Ikpatt F, Khramtsov A, Dangou JM, Nanda R, Dignam J, Zhang B, Grushko T, Zhang C, Oluwasola O *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):4515-4521.

102. Ihemelandu CU, Leffall LD, Jr., Dewitty RL, Naab TJ, Mezghebe HM, Makambi KH, Adams-Campbell L, Frederick WA: Molecular breast cancer subtypes in premenopausal and postmenopausal African-American women: age-specific prevalence and survival. *J Surg Res* 2007, 143(1):109-118.

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

104. Livasy CA, Karaca G, Nanda R, Tretiakova MS, Olopade OI, Moore DT, Perou CM: Phenotypic evaluation of the basal-like subtype of invasive breast carcinoma. *Mod Pathol* 2006, 19(2):264-271.

105. Bergamaschi A, Kim YH, Wang P, Sorlie T, Hernandez-Boussard T, Lonning PE, Tibshirani R, Borresen-Dale AL, Pollack JR: 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):1033-1040.

106. Rodriguez-Pinilla SM, Sarrio D, Honrado E, Hardisson D, Calero F, Benitez J, Palacios J: Prognostic significance of basal-like phenotype and fascin expression in node-negative invasive breast carcinomas. *Clin Cancer Res* 2006, 12(5):1533-1539.

107. Foulkes WD, Smith IE, Reis-Filho JS: Triple-negative breast cancer. *N Engl J Med* 2010, 363(20):1938-1948.

108. Anders CK, Winer EP, Ford JM, Dent R, Silver DP, Sledge GW, Carey LA: Poly(ADP-Ribose) polymerase inhibition: "Targeted" therapy for triple-negative breast cancer. *Clin Cancer Res* 2010, 16(19):4702-4710.

109. Maegawa RO, Tang SC: Triple-negative breast cancer: unique biology and its management. *Cancer Invest* 2010, 28(8):878-883.

110. Kennecke H, Yerushalmi R, Woods R, Cheang MC, Voduc D, Speers CH, Nielsen TO, Gelmon K: Metastatic behavior of breast cancer subtypes. *J Clin Oncol* 2010, 28(20):3271-3277.

111. Honrado E, Benitez J, Palacios J: The molecular pathology of hereditary breast cancer: genetic testing and therapeutic implications. *Mod Pathol* 2005, 18(10):1305-1320.

112. Foulkes WD, Reis-Filho JS, Narod SA: Tumor size and survival in breast cancer--a reappraisal. *Nat Rev Clin Oncol* 2010, 7(6):348-353.

113. Foulkes WD, Stefansson IM, Chappuis PO, Begin LR, Goffin JR, Wong N, Trudel M, Akslen LA: Germline BRCA1 mutations and a basal epithelial phenotype in breast cancer. *J Natl Cancer Inst* 2003, 95(19):1482-1485.

114. Yang XR, Sherman ME, Rimm DL, Lissowska J, Brinton LA, Peplonska B, Hewitt SM, Anderson WF, Szeszenia-Dabrowska N, Bardin-Mikolajczak A *et al*: Differences in risk factors for breast cancer molecular subtypes in a population-based study. *Cancer Epidemiol Biomarkers Prev* 2007, 16(3):439-443.

115. Garcia-Closas M, Hall P, Nevanlinna H, Pooley K, Morrison J, Richesson DA, Bojesen SE, Nordestgaard BG, Axelsson CK, Arias JI *et al*: Heterogeneity of breast cancer associations with five susceptibility loci by clinical and pathological characteristics. *PLoS Genet* 2008, 4(4):e1000054.

116. Reeves GK, Travis RC, Green J, Bull D, Tipper S, Baker K, Beral V, Peto R, Bell J, Zelenika D *et al*: Incidence of breast cancer and its subtypes in relation to individual and multiple low-penetrance genetic susceptibility loci. *JAMA* 2010, 304(4):426-434.

117. Lacroix M, Toillon RA, Leclercq G: Stable 'portrait' of breast tumors during progression: data from biology, pathology and genetics. *Endocr Relat Cancer* 2004, 11(3):497-522.

118. 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):1558-1568.

119. Liu YT, Gao CM, Ding JH, Li SP, Cao HX, Wu JZ, Tang JH, Qian Y, Tajima K: Physiological, reproductive factors and breast cancer risk in Jiangsu province of China. *Asian Pac J Cancer Prev* 2011, 12(3):787-790.

120. Tao P, Hu YY, Huang Y, Li JY: [Risk factors of breast cancer in Asian women: a Meta-analysis.]. *Zhonghua Liu Xing Bing Xue Za Zhi* 2011, 32(2):164-169.

121. Stanford JL, Szklo M, Boring CC, Brinton LA, Diamond EA, Greenberg RS, Hoover RN: A case-control study of breast cancer stratified by estrogen receptor status. *Am J Epidemiol* 1987, 125(2):184-194.

122. McTiernan A, Thomas DB, Johnson LK, Roseman D: Risk factors for estrogen receptor-rich and estrogen receptor-poor breast cancers. *J Natl Cancer Inst* 1986, 77(4):849-854.

123. Kwan ML, Kushi LH, Weltzien E, Maring B, Kutner SE, Fulton RS, Lee MM, Ambrosone CB, Caan BJ: Epidemiology of breast cancer subtypes in two prospective cohort studies of breast cancer survivors. *Breast Cancer Res* 2009, 11(3):R31.

124. Collins K, Jacks T, Pavletich NP: The cell cycle and cancer. *Proc Natl Acad Sci U S A* 1997, 94(7):2776-2778.

125. Definition of Proto-oncogene [http://www.medterms.com/script/main/art.asp?articlekey=5088]

126. Jacks T, Weinberg RA: Cell-cycle control and its watchman. *Nature* 1996, 381(6584):643-644.

127. Lowe SW, Bodis S, McClatchey A, Remington L, Ruley HE, Fisher DE, Housman DE, Jacks T: p53 status and the efficacy of cancer therapy in vivo. *Science* 1994, 266(5186):807-810.

128. Zyss D, Gergely F: Centrosome function in cancer: guilty or innocent? *Trends Cell Biol* 2009, 19(7):334-346.

129. Lukasiewicz KB, Lingle WL: Aurora A, centrosome structure, and the centrosome cycle. *Environ Mol Mutagen* 2009, 50(8):602-619.

130. Lingle WLaS, Jeffrey L.: Analysis of Centrosome Amplification in Cancer. In: *Molecular Genetics: Liver and Pancreatic Carcinomas.* vol. 3. Burlington: Elsevier Academic Press; 2005.

131. Nigg EA: Centrosome aberrations: cause or consequence of cancer progression? *Nat Rev Cancer* 2002, 2(11):815-825.

132. Khodjakov A, Rieder CL: Centrosomes enhance the fidelity of cytokinesis in vertebrates and are required for cell cycle progression. *J Cell Biol* 2001, 153(1):237-242.

133. Piel M, Nordberg J, Euteneuer U, Bornens M: Centrosome-dependent exit of cytokinesis in animal cells. *Science* 2001, 291(5508):1550-1553.

134. Tsou MF, Stearns T: Controlling centrosome number: licenses and blocks. *Curr Opin Cell Biol* 2006, 18(1):74-78.

135. Zhou H, Kuang J, Zhong L, Kuo WL, Gray JW, Sahin A, Brinkley BR, Sen S: Tumour amplified kinase STK15/BTAK induces centrosome amplification, aneuploidy and transformation. *Nat Genet* 1998, 20(2):189-193.

136. Glover DM, Leibowitz MH, McLean DA, Parry H: Mutations in aurora prevent centrosome separation leading to the formation of monopolar spindles. *Cell* 1995, 81(1):95-105.

137. Vessey KB, Ludwiczak RL, Briot AS, Underwood EM: abnormal chromatin (abc), a maternal-effect locus in Drosophila melanogaster. *J Cell Sci* 1991, 98 (Pt 2):233-243.

138. Lingle WL, Lutz WH, Ingle JN, Maihle NJ, Salisbury JL: Centrosome hypertrophy in human breast tumors: implications for genomic stability and cell polarity. *Proc Natl Acad Sci U S A* 1998, 95(6):2950-2955.

139. Pihan GA, Purohit A, Wallace J, Knecht H, Woda B, Quesenberry P, Doxsey SJ: Centrosome defects and genetic instability in malignant tumors. *Cancer Res* 1998, 58(17):3974-3985.

140. Pihan GA, Wallace J, Zhou Y, Doxsey SJ: Centrosome abnormalities and chromosome instability occur together in pre-invasive carcinomas. *Cancer Res* 2003, 63(6):1398-1404.

141. Olson JE, Wang X, Pankratz VS, Fredericksen ZS, Vachon CM, Vierkant RA, Cerhan JR, Couch FJ: Centrosome-related genes, genetic variation, and risk of breast cancer. *Breast Cancer Res Treat* 2011, 125(1):221-228.

142. Lingle WL, Barrett SL, Negron VC, D'Assoro AB, Boeneman K, Liu W, Whitehead CM, Reynolds C, Salisbury JL: Centrosome amplification drives chromosomal instability in breast tumor development. *Proc Natl Acad Sci U S A* 2002, 99(4):1978-1983.

143. Balczon R, Bao L, Zimmer WE, Brown K, Zinkowski RP, Brinkley BR: Dissociation of centrosome replication events from cycles of DNA synthesis and mitotic division in hydroxyurea-arrested Chinese hamster ovary cells. *J Cell Biol* 1995, 130(1):105-115.

144. Hoque A, Carter J, Xia W, Hung MC, Sahin AA, Sen S, Lippman SM: Loss of aurora A/STK15/BTAK overexpression correlates with transition of in situ to invasive ductal carcinoma of the breast. *Cancer Epidemiol Biomarkers Prev* 2003, 12(12):1518-1522.

145. Holland AJ, Cleveland DW: Boveri revisited: chromosomal instability, aneuploidy and tumorigenesis. *Nat Rev Mol Cell Biol* 2009, 10(7):478-487.

146. Gisselsson D: Classification of chromosome segregation errors in cancer. *Chromosoma* 2008, 117(6):511-519.

147. Lengauer C, Kinzler KW, Vogelstein B: Genetic instabilities in human cancers. *Nature* 1998, 396(6712):643-649.

148. Brinkley BR: Managing the centrosome numbers game: from chaos to stability in cancer cell division. *Trends Cell Biol* 2001, 11(1):18-21.

149. Tanner MM, Grenman S, Koul A, Johannsson O, Meltzer P, Pejovic T, Borg A, Isola JJ: Frequent amplification of chromosomal region 20q12-q13 in ovarian cancer. *Clin Cancer Res* 2000, 6(5):1833-1839.

150. Sen S, Zhou H, White RA: A putative serine/threonine kinase encoding gene BTAK on chromosome 20q13 is amplified and overexpressed in human breast cancer cell lines. *Oncogene* 1997, 14(18):2195-2200.

151. Isola JJ, Kallioniemi OP, Chu LW, Fuqua SA, Hilsenbeck SG, Osborne CK, Waldman FM: Genetic aberrations detected by comparative genomic hybridization predict outcome in node-negative breast cancer. *Am J Pathol* 1995, 147(4):905-911.

152. Tanner MM, Tirkkonen M, Kallioniemi A, Holli K, Collins C, Kowbel D, Gray JW, Kallioniemi OP, Isola J: Amplification of chromosomal region 20q13 in invasive breast cancer: prognostic implications. *Clin Cancer Res* 1995, 1(12):1455-1461.

153. Saskova A, Solc P, Baran V, Kubelka M, Schultz RM, Motlik J: Aurora kinase A controls meiosis I progression in mouse oocytes. *Cell Cycle* 2008, 7(15):2368-2376.

154. Marumoto T, Hirota T, Morisaki T, Kunitoku N, Zhang D, Ichikawa Y, Sasayama T, Kuninaka S, Mimori T, Tamaki N *et al*: Roles of aurora-A kinase in mitotic entry and G2 checkpoint in mammalian cells. *Genes Cells* 2002, 7(11):1173-1182.

155. Marumoto T, Honda S, Hara T, Nitta M, Hirota T, Kohmura E, Saya H: Aurora-A kinase maintains the fidelity of early and late mitotic events in HeLa cells. *J Biol Chem* 2003, 278(51):51786-51795.

156. Hirota T, Kunitoku N, Sasayama T, Marumoto T, Zhang D, Nitta M, Hatakeyama K, Saya H: Aurora-A and an interacting activator, the LIM protein Ajuba, are required for mitotic commitment in human cells. *Cell* 2003, 114(5):585-598.

157. Liu Q, Ruderman JV: Aurora A, mitotic entry, and spindle bipolarity. *Proc Natl Acad Sci U S A* 2006, 103(15):5811-5816.

158. Wang X, Zhou YX, Qiao W, Tominaga Y, Ouchi M, Ouchi T, Deng CX: Overexpression of aurora kinase A in mouse mammary epithelium induces genetic instability preceding mammary tumor formation. *Oncogene* 2006, 25(54):7148-7158.

159. Tanaka T, Kimura M, Matsunaga K, Fukada D, Mori H, Okano Y: Centrosomal kinase AIK1 is overexpressed in invasive ductal carcinoma of the breast. *Cancer Res* 1999, 59(9):2041-2044.

160. Polymorphisms in the BRCA1 and ABCB1 genes modulate menopausal hormone therapy associated breast cancer risk in postmenopausal women. *Breast Cancer Res Treat* 2010, 120(3):727-736.

161. Dai Q, Cai QY, Shu XO, Ewart-Toland A, Wen WQ, Balmain A, Gao YT, Zheng W: Synergistic effects of STK15 gene polymorphisms and endogenous estrogen exposure in the risk of breast cancer. *Cancer Epidemiol Biomarkers Prev* 2004, 13(12):2065-2070.

162. Sun T, Miao X, Wang J, Tan W, Zhou Y, Yu C, Lin D: Functional Phe31Ile polymorphism in Aurora A and risk of breast carcinoma. *Carcinogenesis* 2004, 25(11):2225-2230.

163. Lo YL, Yu JC, Chen ST, Yang HC, Fann CS, Mau YC, Shen CY: Breast cancer risk associated with genotypic polymorphism of the mitosis-regulating gene Aurora-A/STK15/BTAK. *Int J Cancer* 2005, 115(2):276-283.

164. Egan KM, Newcomb PA, Ambrosone CB, Trentham-Dietz A, Titus-Ernstoff L, Hampton JM, Kimura MT, Nagase H: STK15 polymorphism and breast cancer risk in a population-based study. *Carcinogenesis* 2004, 25(11):2149-2153.

165. Sun H, Bai J, Chen F, Jin Y, Yu Y, Fu S: Lack of an association between AURKA T91A polymorphisms and breast cancer: a meta-analysis involving 32,141 subjects. *Breast Cancer Res Treat* 2011, 125(1):175-179.

166. Ruan Y, Song AP, Wang H, Xie YT, Han JY, Sajdik C, Tian XX, Fang WG: Genetic polymorphisms in AURKA and BRCA1 are associated with breast cancer susceptibility in a Chinese Han population. *J Pathol* 2011.

167. Sankaran S, Crone DE, Palazzo RE, Parvin JD: Aurora-A kinase regulates breast cancer associated gene 1 inhibition of centrosome-dependent microtubule nucleation. *Cancer Res* 2007, 67(23):11186-11194.

168. Chen A, Kleiman FE, Manley JL, Ouchi T, Pan ZQ: Autoubiquitination of the BRCA1*BARD1 RING ubiquitin ligase. *J Biol Chem* 2002, 277(24):22085-22092.

169. Brzovic PS, Rajagopal P, Hoyt DW, King MC, Klevit RE: Structure of a BRCA1-BARD1 heterodimeric RING-RING complex. *Nat Struct Biol* 2001, 8(10):833-837.

170. Ruffner H, Joazeiro CA, Hemmati D, Hunter T, Verma IM: Cancer-predisposing mutations within the RING domain of BRCA1: loss of ubiquitin protein ligase activity and protection from radiation hypersensitivity. *Proc Natl Acad Sci U S A* 2001, 98(9):5134-5139.

171. Ghimenti C, Sensi E, Presciuttini S, Brunetti IM, Conte P, Bevilacqua G, Caligo MA: Germline mutations of the BRCA1-associated ring domain (BARD1) gene in breast and

breast/ovarian families negative for BRCA1 and BRCA2 alterations. *Genes Chromosomes Cancer* 2002, 33(3):235-242.

172. Stacey SN, Sulem P, Johannsson OT, Helgason A, Gudmundsson J, Kostic JP, Kristjansson K, Jonsdottir T, Sigurdsson H, Hrafnkelsson J *et al*: The BARD1 Cys557Ser variant and breast cancer risk in Iceland. *PLoS Med* 2006, 3(7):e217.

173. Karppinen SM, Barkardottir RB, Backenhorn K, Sydenham T, Syrjakoski K, Schleutker J, Ikonen T, Pylkas K, Rapakko K, Erkko H *et al*: Nordic collaborative study of the BARD1 Cys557Ser allele in 3956 patients with cancer: enrichment in familial BRCA1/BRCA2 mutation-negative breast cancer but not in other malignancies. *J Med Genet* 2006, 43(11):856-862.

174. Ishitobi M, Miyoshi Y, Hasegawa S, Egawa C, Tamaki Y, Monden M, Noguchi S: Mutational analysis of BARD1 in familial breast cancer patients in Japan. *Cancer Lett* 2003, 200(1):1-7.

175. Zhang F, Fan Q, Ren K, Auerbach AD, Andreassen PR: FANCJ/BRIP1 recruitment and regulation of FANCD2 in DNA damage responses. *Chromosoma* 2010, 119(6):637-649.

176. Cantor SB, Bell DW, Ganesan S, Kass EM, Drapkin R, Grossman S, Wahrer DC, Sgroi DC, Lane WS, Haber DA *et al*: BACH1, a novel helicase-like protein, interacts directly with BRCA1 and contributes to its DNA repair function. *Cell* 2001, 105(1):149-160.

177. Desjardins S, Belleau P, Labrie Y, Ouellette G, Bessette P, Chiquette J, Laframboise R, Lepine J, Lesperance B, Pichette R *et al*: Genetic variants and haplotype analyses of the ZBRK1/ZNF350 gene in high-risk non BRCA1/2 French Canadian breast and ovarian cancer families. *Int J Cancer* 2008, 122(1):108-116.

178. Harkin DP, Bean JM, Miklos D, Song YH, Truong VB, Englert C, Christians FC, Ellisen LW, Maheswaran S, Oliner JD *et al*: Induction of GADD45 and JNK/SAPK-dependent apoptosis following inducible expression of BRCA1. *Cell* 1999, 97(5):575-586.

179. Liao CC, Tsai CY, Chang WC, Lee WH, Wang JM: RB.E2F1 complex mediates DNA damage responses through transcriptional regulation of ZBRK1. *J Biol Chem* 2010, 285(43):33134-33143.

180. Garcia V, Garcia JM, Pena C, Silva J, Dominguez G, Rodriguez R, Maximiano C, Espinosa R, Espana P, Bonilla F: The GADD45, ZBRK1 and BRCA1 pathway: quantitative analysis of mRNA expression in colon carcinomas. *J Pathol* 2005, 206(1):92-99.

181. Liu, C., The association between *AURKA* T91A polymorphism and breast cancer risk. Breast Cancer Res. Treat, 2011. 129(1): p. 281-3.

Chapter 2. Study Design and Methods

2.1 Specific Aims

Breast cancer continues to represent a tremendous health burden in the United States. The American Cancer Society estimates that 30% of all cancers diagnosed among American women in 2011 will be invasive breast cancers, making them the most commonly diagnosed cancers in the U.S. [1]. After cancers of the lung and bronchus, breast cancer is the leading cause of cancer death in American women [1]. Traditionally, incidence and mortality statistics have been reported by race/ethnicity under the assumption that breast cancer is a single disease. However, recent findings have established the significant heterogeneity in breast cancer by identifying at least five distinct subtypes that vary in their gene expression profiles and in their responsiveness to endocrine therapies [9,11,12,31]. The implications may be clinically important: differences in gene expression patterns suggest differences in tumor biology.

Differences in gene expression and tumor biology may contribute to the significant disparity in breast cancer incidence and mortality that exists between African Americans and Caucasians. Although incidence rates are higher among Caucasian women, African American women are 38% more likely to die from breast cancer [32]. Previous research has shown that risk factors for breast cancer may differ by subtype [23], suggesting variable molecular pathways of carcinogenesis by subtype. Of particular significance is the higher

prevalence of basal-like breast tumors among younger African American women [2,13,23]. Basal-like breast cancers have poor prognoses and relative insensitivity to hormonal or targeted therapies [11,12,33] due to the absence of expression of estrogen receptor (ER), progesterone receptor (PR), and human epidermal growth factor receptor (HER2) in these "triple negative breast cancers". Compounding the concern for treatment options is the tendency for basal-like breast cancers to form distant metastases [18-21].

Although the relative prognosis of basal-like breast cancer is poor, environmental and genetic risk factors contributing to the etiology of this subtype remain unknown. Hereditary breast cancers only account for 5-10% of all breast cancer cases, while 12-20% of all breast cancer cases are triple-negative or basal-like [2,3,13,23,34-40]. Variation in the proportion of triple-negative or basal-like cases relative to other intrinsic subtypes may stem from differences in population characteristics. Mutations in the highly penetrant genes *BRCA1* and *BRCA2*, in addition to all other known breast cancer susceptibility genes, are only likely to account for 20-25% of hereditary cases [3,34,41-43]. Other hereditary breast cancer susceptibility genes remain to be discovered [4,5-8].

It has been suggested that common low-penetrance susceptibility genes may play an important role in the etiology of breast cancer, individually conferring small increases in risk [4,5-8]. In aggregate, these independently minor increases in risk may become substantial [4,5-8]. *AURKA*, encoding a serine/threonine kinase, is a putative low-penetrance tumor susceptibility gene due to its prominent role in cell cycle regulation [25]. *AURKA* overexpression has been demonstrated in several types of cancer and correlated with poor prognosis [24,44,45]. Previous studies of *AURKA* and risk of breast cancer have yielded conflicting results, and have been largely limited to investigations of a single functional

polymorphism in Asian [27-29] and Caucasian [26,30,46-49] populations. The main objective of this study was to determine the association between SNPs on *AURKA* and breast cancer among African Americans and Caucasians.

It has been suggested that overexpression of *AURKA* in conjunction with inactivation of *BRCA1* could be associated with tumor development and progression [50]. Due to its prominent role in cell cycle regulation and the relatively high risk for breast cancer conferred by mutations in *BRCA1*, *BRCA1* and lesser penetrant genes encoding *BRCA1*-interacting proteins are also logical targets for further investigation [34]. The secondary objective of this study was to determine if hereditary genetic variation in each of three *BRCA1*-interacting genes is associated with odds of breast cancer. The tertiary aim of this study was to evaluate whether a gene-gene interaction exists between SNPs on *AURKA* and SNPs on *BRCA1* and the *BRCA1*-interacting genes: *BARD1*, *BRIP1*, and *ZNF350* in association with breast cancer.

To address these objectives, a case-control analysis of data from a population-based study of breast cancer (The Carolina Breast Cancer Study, CBCS) was performed. CBCS data was collected on 1,972 cases of primary invasive breast cancer (742 African American cases, 1,230 Caucasian cases) and 1,776 controls (658 African American controls, 1,118 Caucasian controls) residing in North Carolina. Because basal-like breast cancers render a relatively poor prognosis [2] and luminal A breast cancers are the most commonly diagnosed [13], the main objectives of this investigation focus on those subtypes. However, luminal B, HER2+/ER-, and unclassified breast cancer subtypes were also examined for associations with *AURKA*, *BRCA1* and the *BRCA1*-inrteracting genes.

Summary—Aims

1.) Determine the association between genetic variation in *AURKA* and breast cancer rate (all cases and controls) and intrinsic breast cancer subtype rate.

a.) Determine the association between *AURKA* SNPs and breast cancer in a race-stratified analysis using case and control subjects enrolled in CBCS, with adjustment for population stratification using ancestry informative markers.

b.) Explore the association between *AURKA* SNPs and breast cancer subtype in a pooled analysis of African Americans and Caucasians, with adjustment for population stratification using ancestry informative markers.

Hypothesis 1. There will be a significant positive association between SNPs on AURKA and breast cancer overall.

2.) Determine the association between genetic variation in *BRCA1, BARD1, BRIP1,* and *ZNF350* and breast cancer rate (all cases and controls) and intrinsic breast cancer subtype rate.

a.) Determine the association between SNPs on *BRCA1, BARD1, BRIP1,* and *ZNF350* and breast cancer in a race-stratified analysis using case and control subjects enrolled

in CBCS, with adjustment for population stratification using ancestry informative markers.

b.) Explore the association between SNPs on *BRCA1, BARD1, BRIP1,* and *ZNF350* and breast cancer subtype in a pooled analysis of African Americans and Caucasians, with adjustment for population stratification using ancestry informative markers.

Hypothesis 2. There will be significant positive associations between SNPs on each of BRCA1, BARD1, BRIP1 and *ZNF35 and breast cancer overall.*

3.) Explore whether a gene-gene additive interaction exists between *AURKA* and *BRCA1* and *BRCA1*-interacting genes (*BARD1, BRIP1,* and *ZNF350*) in association with all breast cancers, with adjustment for population stratification using ancestry informative markers.

Hypothesis 3. There will be gene-gene additive interactions between AURKA and each of BRCA1, BARD1, BRIP1 and ZNF350 in association with all breast cancer.

2.2 Purpose

The primary purpose of this study was to estimate the association between genetic variation in the cell cycle regulatory gene *AURKA* and rate of breast cancer. Specifically, the outcomes of primary interest were the luminal A and basal-like subtypes, but in order for this investigation to be more comparable to existing literature, and to estimate an overall association, all breast cancer without respect to subtype was also an outcome of interest.

Since the highly penetrant *BRCA1* gene is a known risk factor for breast cancer, and since *BRCA1* proteins are known to interact with *AURKA* proteins, a secondary aim of this dissertation was to estimate the association between genetic variations in *BRCA1* and select *BRCA1*-interacting genes and rate of breast cancer. The final aim of this study was to estimate gene-gene additive interactions between tag and candidate polymorphisms in *AURKA* and tag and candidate polymorphisms in each of *BRCA1* and select *BRCA1*-interacting genes in association with breast cancer. Data from the CBCS, a population-based case-control study of breast cancer in African American and Caucasian women residing in North Carolina, was used to conduct this investigation.

Germline DNA collected from CBCS participants provided biallelic genotype data on tag and candidate single nucleotide polymorphisms (SNPs) for the genes of interest. Breast cancer subtype data for case participants was acquired from medical records and/or paraffin-embedded tumor tissue. Logistic regression was employed to estimate the association between genotype and rate of breast cancer (all breast cancer, intrinsic subtype of breast cancer). In addition to potential main effects demonstrated by SNPs on *BRCA1* and *BRCA1*-interacting genes: *BARD1, BRIP1,* and *ZNF350,* gene-gene additive interaction was examined by considering interactions between SNPs on each of *BRCA1, BARD1, BRIP1,* and *ZNF350* and SNPs on *AURKA*. Due to the large number of SNPs under investigation, interactions were examined based on main effects and/or potential functionality. Those SNPs showing statistically significant main effects on *AURKA* were further examined for statistical interaction with any SNPs showing statistically significant findings on each of the *BRCA1*-interacting genes. Additionally, candidate SNPs (potentially functional) were

identified on each of the candidate genes of interest using a decision tree (Figure 2.7), and were included in the interaction study.

2.3 The Carolina Breast Cancer Study

The Carolina Breast Cancer Study (CBCS) is a population-based, case-control study of genetic and environmental risk factors for breast cancer among African American and Caucasian women residing in North Carolina [51]. CBCS study design and methods have been previously outlined by Newman *et al.* [51]. Study participants were recruited and selected from 24 contiguous counties in central and eastern North Carolina in an effort to accrue a representative sample of African American and rural participants (Figure 2.6) [51].

CBCS recruitment was conducted in two phases—from 1993 through 1995 (Phase 1) and from 1996 through 2001 (Phase 2). Women living in the study area between the ages of 20 and 74 and diagnosed with invasive breast cancer for the first time were eligible cases in Phase 1. Phase 2 included women diagnosed with *in situ* breast cancer as well as those diagnosed with invasive breast cancer. Cases were identified using a rapid case ascertainment system via the North Carolina Central Cancer Registry (NCCCR). After eligibility criteria were met, randomized recruitment case sampling was undertaken to ensure adequate representation of African American and younger women. Case sampling probabilities were as follows: 100% of African American women between the ages of 20 and 49, 75% of African American women between the ages of 50 and 74, 67% of Caucasian women between the ages of 20 and 49, and 20% of Caucasian women between the ages of 50 and 74. Phase 2 *in situ* cases did not undergo random recruitment sampling; all eligible

in situ cases were enrolled. After selection via sampling, potential case participants were contacted only after requesting and receiving permission from the patient's physician [51].

Controls were selected from two sources: women younger than 65 were selected from a list maintained by the North Carolina Division of Motor Vehicles; women between the ages of 65 and 74 were selected from Health Care Financing Administration records. Controls were sampled from these lists using modified randomized recruitment, and sampling fractions were designed to ensure frequency-matching of cases to controls by race and five-year age interval [52,53].

Potential cases and controls were contacted first by letter and then by telephone, if available. Women agreeing to participate were scheduled for an in-home visit by a registered nurse interviewer. The interviewer administered a study questionnaire and collected anthropometric measurements in addition to a 30cc blood sample. Germline DNA was extracted from peripheral blood lymphocytes and stored for future analysis [51]. Written consent was obtained from cases to retrieve medical records and paraffin-embedded tumor tissue. The CBCS pathologist performed a standardized review of all breast tissue received to confirm the diagnosis of breast cancer and to characterize histology [51]. Slides were cut from paraffin blocks for molecular and immunohistochemical (IHC) assays, procedures for which have been described previously [2,54,55].

2.4 Immunohistochemistry

2.4.1 Receptor Status

For invasive cases, ER and PR status was primarily obtained from medical records (80%). Various clinical laboratories determined ER/PR results on these cases. Approximately half used IHC on paraffin-embedded tissue, and employed cutoffs for receptor positivity from more than 0% to more than 20%. The other half performed biochemical assays on frozen tissue with cutoffs for receptor positivity of 10-15 fmol/mg [55]. For approximately 11% of invasive cases, ER/PR status was not available in the medical record; however, paraffin-embedded tissue was available and ER/PR status was ascertained by the UNC Immunohistochemistry Core laboratory. For these cases, IHC scoring was based on UNC Hospitals Department of Pathology standards, using a cutoff of 5% positive nuclei staining in invasive breast cancer cells [2]. A random sample of ER+ and ER- cases based on medical record abstraction was drawn to compare with IHC performed by the UNC Immunohistochemistry Core laboratory. A kappa statistic of 0.62 and concordance of 81% resulted from the comparison, indicating good agreement [2,56]. The remaining 9% of invasive cases had missing data for ER/PR status [2].

HER2 status in invasive cases was determined using the CB11 monoclonal antibody as previously described by Millikan *et al.* [54]. HER2 positivity was defined by weak to strong staining of membrane or membrane plus cytoplasm in at least 10% of tumor cells [2]. Interscorer agreement was evaluated on a subset of cases, yielding a kappa statistic of 0.58 and concordance of 82% resulted from the comparison overall concordance of 81% [2]. HER1 and cytokeratin (CK) 5/6 characterization have been previously described [10,57],

and invasive cases demonstrating any staining were classified as positive [2]. All assays were performed by the UNC Immunohistochemistry Core laboratory.

ER, HER2, CK5/6, and HER1 classification and determination for *in situ* cases were described in detail by Livasy *et al.* [58]. ER+ was defined by an Allred score of above 2 for ER nuclear staining; HER2 membrane positivity was defined by 3+ intensity with DAB chromogen staining and 2+ or 3+ intensity with SG chromogen staining in >10% of cells [58]. CK5/6 positivity was determined by the presence of any membrane staining. HER1 positivity was defined by any cytoplasmic staining [58]. All assays were performed by the UNC Immunohistochemistry Core laboratory. PR status was not determined for *in situ* cases due to its high correlation with ER expression and to preserve tissue [13].

2.4.2 Intrinsic Breast Cancer Subtypes

CBCS intrinsic breast cancer subtypes were based on expression of ER, PR, HER2, CK 5/6, and HER1 according to Table 2.1 [2]. Tumors that were negative for expression of all five markers were unclassified. Negative staining for all markers is not necessarily indicative of receptor negativity in the tumor, and can result from poor tumor block quality or inadequate tissue present in the tumor block [2].

2.5 CBCS Participation

Detailed CBCS participation rates are presented in Table 2.2. The overall response rate was ~77% for cases and ~57% for controls. Higher participation rates were noted among Caucasians, regardless of case status or age. The lowest participation rates were found in younger African American controls, aged 20-39 years (40.5%). Among cases, the lowest

rates were noted in African American women older than 64 years (65.8%), while the highest rates were demonstrated in Caucasians aged 20-39 years (83.3%). Among controls, the highest participation rates were found in Caucasian women older than 64 years (68.7%). Among cases, ~6% were ineligible mostly due to a prior history of breast cancer [59]. Approximately 10% of controls did not meet eligibility criteria, primarily due to current residence outside the study area [59]. A total of 2,279 incident cases and 1,988 controls were enrolled.

DNA was acquired and successfully genotyped for 2,013 of 2,279 enrolled case participants (88%) and 1,787 of 1,988 control participants (90%). Among participants who were successfully genotyped, 38% of both cases and controls were African American. IHC intrinsic breast cancer subtype data was available for 1,412 of 2,279 enrolled case participants (62%). IHC intrinsic breast cancer subtype data was successfully acquired for 1,250 (502 African American, 748 Caucasian) of 2,013 enrolled cases who were successfully genotyped (62%). Of the 2,279 cases enrolled, inadequate tumor tissue and/or incomplete IHC data was available for 867 participants, who were excluded from analyses. Included cases were more likely than excluded cases to be stage II (40% vs. 25%) and less likely to be stage I (30% vs. 37%). There was little difference between included and excluded cases with respect to stage III (8% vs. 7%) or stage IV (2% vs. 3%). There was no statistical difference in age or menopausal status between included and excluded cases. African American women were more likely to have adequate tumor tissue and/or complete IHC subtype data compared to Caucasian women (36% of excluded cases were African American). African American case participants were more likely to have larger tumors and later stage at diagnosis compared to Caucasian case participants [13].

2.6 Characteristics of CBCS Case Participants

Characteristics of 1,412 CBCS case participants with IHC data (but not necessarily successful genotype data) are presented in Table 2.3. Case participants with basal-like tumors were younger than women with other tumors and more likely to be African American. A higher prevalence of basal-like tumors was also noted among premenopausal participants. The prevalence of luminal-A and luminal-B tumors was higher among Caucasians and postmenopausal women.

2.7 CBCS Genotyping

SNPs in this study were genotyped by the University of North Carolina Mammalian Genotyping Core using the Illumina Golden Gate Assay (Illumina, San Diego, CA). A combination of tag and candidate SNPs were selected for genotyping (Table 2.4). Tag SNPs are single nucleotide polymorphisms that are highly correlated with SNPs on the same gene or chromosome, but don't necessarily have any functionality. The high degree of correlation allows for more efficient genotyping in order to characterize genetic variation for a specific chromosomal region. Candidate SNPs are those that alter gene function or expression, or are suspected of altering gene function or expression. Tag SNPs were identified for Caucasians and African Americans from CEU (Utah residents with ancestry from northern and western Europe) and YRI (individuals of Yoruban descent from Idaban, Nigeria) HapMap populations respectively [60], and selected using the Tagger program developed by de Bakker *et al.* [61]. Tag SNPs were selected based on a LD threshold of r^2=0.80 and a minimum minor allele frequency (MAF) of 0.10 separately for both CEU and YRI populations. Tag SNPs in each population were then combined and CBCS participants were

genotyped for the pooled list. Inclusion of suspected functional SNPs identified from the literature review was based on several criteria: relevance to plausible molecular pathways related to breast cancer, consistency of results related to breast cancer, and a minor allele frequency of at least 0.05 in CBCS participants. Assay intensity data and genotype cluster images for all SNPs were reviewed individually. To ensure quality control of genetic data, SNPs with low signal intensity or SNPs that were unable to be distinguished by genotype cluster were excluded. Detailed genotyping procedures and quality control measures were described previously [62,63].

The overall genotyping rate in CBCS was ~66% for case participants and ~56% for control participants. Among case and control participants, the highest genotyping rates were noted in young Caucasians (~77%, ~68% respectively) while African American women older than 64 years exhibited the lowest rates (~50%, ~45% respectively) (Table 2.2).

2.8 Population Stratification and Ancestry

Population stratification is a form of confounding caused by differences in allele frequencies between cases and controls that result from ancestral disparities, as opposed to real associations between a genetic marker and disease [64,65]. If the genetic marker under study shows significant variation across ancestral groups, and if these ancestral groups also differ in their baseline risk for the outcome, then false positive associations could arise between genotypes in a particular subgroup and the outcome of interest, regardless of whether the locus is in LD with the true risk allele [64].

Several methods for assessing population stratification have been developed. Genomic control, developed by Devlin *et al.*, employs the testing of multiple unrelated (null)

polymorphisms, in addition to candidate loci, throughout the genome to estimate the degree of population substructure. X^2 test statistics are computed for both null and candidate loci. Population stratification increases the variability and magnitude of the test statistics observed in the null loci. Based on this variance, a multiplier describing the degree of population stratification can be derived and used to adjust significance tests for candidate loci [66]. The genomic control approach works under the assumption that the inflation of variance due to population stratification is constant for all null loci [66]. However, SNP allele frequencies can differ markedly across ancestral populations. Employing genomic control could result in an overly conservative adjustment for markers showing high variation across populations and/or an excessive adjustment at loci with low variability across populations [67]. A decrease in power may be noted under such circumstances.

Another commonly employed method for controlling population stratification is the principal components analysis (PCA). The first step in this process is to examine the sample covariance matrix. If the covariance between any two genetic markers is not equal to zero, then a linear relationship exists between them and the strength of that relationship is represented by the correlation coefficient. Via principal axis transformation, correlated markers are transformed into new uncorrelated markers known as principal components. Depending on the degree of ancestral variability in the population, the number of principal components will be less than or equal to the number of original markers under consideration. The principal axis transformation is defined in order that the first principal component should account for the maximum variability in the population, with each subsequent principal component accounting for the maximum residual variability in the population while maintaining no correlation with previous principal components [68]. After

determining the principal components in the population, methods can be employed to adjust for population stratification. Price *et al.* have developed a method employed in their software package, EIGENSTRAT, which identifies axes of variation in a sample population and then continuously adjusts genotypes and phenotypes by amounts attributable to ancestry along each axis. These ancestry-adjusted genotypes and phenotypes are then used in association analyses [67]. Principal component-based adjustment for population stratification is useful when analyzing hundreds of thousands of markers due to its efficiency in identifying population structure [67]. The disadvantage is that there is ambiguity in how each axis is defined without a standard reference sample with which to compare them.

The CBCS used the software package Structure to infer the number of distinct ancestral populations (K) present in the study population [69]. Results indicated the most likely number of distinct populations was K=2. CBCS then employed maximum likelihood estimation to determine individual ancestry using a predetermined set of 144 ancestry informative markers (AIMS) that were selected to maximally distinguish between African and European ancestry [62]. AIMs are a set of polymorphisms exhibiting substantially different allele frequencies between different populations. They can be used to more accurately group people who share similar markers and phenotypes, like self-reported race. CBCS employed AIMs chosen to distinguish between African and European ancestries by maximizing differences in allele frequencies between the two ancestral populations and by maximizing Fisher's information criterion (FIC) for distinguishing between African and European ancestries [62,70]. FIC is the inverse of the maximum likelihood estimation of the ancestral proportion and can be used to increase the efficiency of AIM selection [71]. FIC was based on allele frequencies in HapMap Project populations from Yoruba in Ibadan,

Nigeria (YRI) and Utah residents with ancestry from northern and western Europe (CEU) [60,62]. Each participant's proportion European or African ancestry is computed and this continuous measure of individual ancestry is then used to adjust association analyses. Detailed statistical methods have been previously described [72].

The median proportion of African ancestry was 81% among participants self-reporting as African American and 6% among those self-reporting as non-African American. Proportion African ancestry will be used as a variable in regression models and will range from 0 to 0.96 based on the maximum individual proportion African ancestry in the study population [63].

2.9 Modeling Genotype Effects

SNP allele and genotype frequencies were calculated for the CBCS study population. All SNPs genotyped for this study are biallelic, meaning a participant may have one of three potential genotypes at any locus. For example, if a particular locus is comprised of major allele X and minor allele Y, then a participant may be homozygous for the major allele (XX), heterozygous (XY,YX), or homozygous for the minor allele (YY). Three genetic models were considered for SNP main effect analyses. Since the mode of inheritance for SNPs genotyped in CBCS participants is unknown, and to maximize our power to detect an association, we employed 1-degree of freedom dominant models for SNPs which assume that a single risk allele (usually the minor allele) is sufficient to impact risk of the outcome. Under this model, homozygotes for the major allele act as the referent group and are coded as 0. Both heterozygotes and homozygotes for the minor allele are assumed to have the same risk for the outcome, so they are grouped together and coded as index.

Unconditional logistic regression was used to estimate associations between SNP genotypes and all breast cancer (all cases and controls) and intrinsic subtype of breast cancer. Specifying all breast cancer as an outcome provides an overall estimate of effect that is comparable to other case-control studies that do not distinguish between subtypes of breast cancer. Although the primary interests of this study are luminal A and basal-like breast cancers, all intrinsic subtypes were examined for associations with SNP genotypes. Odds ratios and 95% confidence intervals were calculated to estimate the association between genotype and each of the outcomes of interest.

The dependent variable Y can take on two possible values (Y=1 if the outcome is present, and Y=0 if otherwise). The outcome probability that is modeled is $\pi = \Pr(Y = 1 \mid x)$, where $x = (x_1, ..., x_s)$ is the vector of s independent variables.

The binary logistic model function then has the form:

$$\text{logit}(\pi) \equiv \log\left(\frac{\pi}{1-\pi}\right) = \alpha + \beta' x$$

where α is the intercept parameter and $\beta' = (\beta_1, ..., \beta_s)$ is the vector of s regression coefficients [223]. For example, the dominant binary logistic model for each SNP will be:

$$logit\ (Y = 1 | x = X) = \ \alpha + \beta_1 X_1 + \beta' X'$$

where β_1 is the regression coefficient for the heterozygous or homozygous minor allele genotype, X_1 indicates presence or absence of the heterozygous or homozygous minor allele genotype ($X_1 = 1$ for heterozygous or homozygous minor allele genotype, $X_1 = 0$

otherwise), $\beta^{'} = (\beta_3,...,\beta_s)$ is the vector of s regression coefficients corresponding to confounders, and $X^{'} = (X_3,...,X_s)$ is the vector of s confounders.

2.10 Gene-gene Interaction

It has been suggested that overexpression of *AURKA* in conjunction with inactivation of *BRCA1* could be associated with tumor development and progression [50]. Due to its prominent role in cell cycle regulation and the relatively high risk for breast cancer conferred by mutations in *BRCA1*, *BRCA1* and lesser penetrant genes encoding *BRCA1*-interacting proteins are also logical targets for further investigation [34]. The secondary objective of this study was to determine if hereditary genetic variation in each of three *BRCA1*-interacting genes is associated with odds of breast cancer. The tertiary aim of this study was to evaluate whether a gene-gene additive interaction exists between SNPs on *AURKA* and SNPs on *BRCA1* and the *BRCA1*-interacting genes: *BARD1*, *BRIP1*, and *ZNF350* in association with breast cancer.

First, main effects were determined for SNPs on *AURKA*, *BRCA1*, *BARD1*, *BRIP1*, and *ZNF350*. SNPs demonstrating statistical significance (determined by exclusion of 1 from OR confidence intervals) were further scrutinized for additive interaction. There was potential for none of the genotyped SNPs on each of the candidate genes of interest to show statistically significant main effects. For this reason, potentially functional variants on each of the candidate genes under investigation were chosen for interaction investigation. The primary criterion for inclusion was based on the likelihood that a SNP was functional (Tables 2.11-2.15). Likelihood of SNP functionality was determined using the FS Score, an integrative *in silico* scoring system for assessing potential SNP functionality based on

protein coding, splicing regulation, transcriptional regulation, and post-translation [89]. SNPs demonstrating FS Scores of ≥0.50 were included in the interaction study (Table 2.10).

Second, gene-gene interaction was evaluated on an additive scale for selected SNPs by calculating the relative excess risk due to interaction (RERI) based on the formula $RERI=OR_{11} - OR_{01} - OR_{10} + 1$ [224], with 95% confidence intervals calculated based on the method proposed by Hosmer and Lemeshow [75].

2.11 Methodological Considerations

Before association analyses were conducted, Fisher's exact tests of HWE were performed for SNPs on the genes under study using SAS ver. 9.3. HWE tests were performed on genotype data provided by controls and stratified by self-reported race, since deviations in cases can be indicative of an association between a SNP and the disease or a SNP in LD with a SNP associated with the disease [76]. Deviations from HWE in controls can occur due to genotyping error, violations of the Hardy-Weinberg principle [77], or chance. For SNPs that deviate from HWE ($p<0.05$), genotype cluster images were reviewed to rule out any artifact in genotype calling. SNPs that did not meet the following criteria among controls were excluded from analysis: (a) minor allele frequency ≥5%, (b) genotyping call rate ≥80% for Illumina and ≥95% for TaqMan, (c) HWE $p>0.05$ and genotype cluster plot indicating distinction between genotypes.

Previous sensitivity analyses to assess outcome misclassification were conducted by Nyante *et al.* to evaluate the effect of molecular subtype misclassification, which was found to be minimal. However, there are varying definitions for characterizing the basal-like

subtype, and even those investigators employing similar classification criteria do not report complete agreement [10,14,15].

Selection bias is another methodological concern and occurs when the study population does not represent the target population [79]. Selection bias is usually introduced during the recruitment of study participants and/or during the process of following participants up [79]. Selection bias can also be introduced if missing data is related to case or exposure status. In this investigation, cases and controls were frequency matched on race and five-year age interval which should mitigate the effects of selection bias since both age and race were included in regression models. Also, any potential residual confounding by ancestry within race strata was adjusted for by including a model term accounting for proportion of African ancestry. It is also possible that those enrolled participants who did not contribute genotype data for the analysis may be systematically different than those participants who did contribute genotype data with respect to race and age. Since race and age were both adjusted for in regression models, selection bias that could occur should be minimized.
It is important to consider the introduction of selection bias from missing values for the exposure or the outcome in CBCS cases and controls. If the number of case and/or control participants in the analytic cohort with missing values for either exposure or outcome is small, then it should be acceptable to assume those values are missing at random and excluding them from the analysis should not introduce any significant bias. However, if a significant proportion of participants are excluded due to missing values there is the potential for bias to be introduced if those participants are selected out of the analysis based on a factor associated with either the outcome (IHC subtype) or the exposure (genotype); that is to say, if participation is related to IHC subtype, genotype, or some integral factor that

influences IHC subtype or genotype. It is implausible for genotype to be associated with participation, so missing genotype data will be missing at random. On the other hand, IHC subtype could be associated with other tumor characteristics that could be associated with selection into the analytic cohort. For example, if those case participants missing subtype data are excluded from analysis *and* the distribution of tumor size is different between the excluded group and the included group (i.e. the included group being those case participants providing both IHC subtype data and genotype data) then bias could be introduced. The bias could be introduced if "missingness" is associated with tumor size. The most likely reasons for missing IHC subtype data are lack of hospital participation with respect to procurement of tumor tissue, lack of patient consent, and lost specimens. However, it is possible that smaller tumors are less likely to be procured due to the higher probability of exhausting the tissue, and smaller tumors may be associated with a particular subtype. However, subtype distributions were similar between cases with and without genotyping data. Likewise, genotyping distributions were similar between cases with and without subtype data. This suggests that the subtype distribution in cases with genotype data is likely representative of the subtype distribution in all cases. Similarly, the genotype distribution in cases with subtype data is likely representative of the genotype distribution in all cases.

For controls, genotype would have to be the factor by which participants are preferentially selected into the study in order for selection bias to be a problem. It is possible that controls may be less likely than cases to donate blood for DNA procurement. In the CBCS, participation with respect to DNA procurement between cases and controls is

comparable. Since cases were matched to controls, the factors affecting selection should be the same or similar between cases and controls.

Results from this study were interpreted and reported based on magnitude of effect estimate, precision, and observed patterns. Precision of estimates was evaluated based on relative values of confidence limit ratios, with lower ratios indicating higher precision. Null hypothesis testing and P-values were not used interpret any results or draw conclusions about SNP associations with breast cancer.

2.12 Statistical Power

Power was calculated using Episheet [80] based on a distribution of binary genotype prevalences ranging from 5% to 30% and main effects odds ratios of 1.25 and 1.50 at an α level of 0.05. Table 2.8 contains power estimates for overall breast cancer by race. Figures 2.8-2.13 provide visual context for the change in power as the estimate of effect changes in each race group. Power was also calculated by intrinsic subtype in the combined race group and is reported in Table 2.9. Due to the exploratory nature of the additive interaction analysis, power was not calculated for this analysis.

There was low power to detect an odds ratio of 1.25 for the association between binary coded genotype and overall breast cancer in African Americans at all ranges of genotype prevalence. Generally, the same was true for Caucasians although there was moderate power to detect an odds ratio of 1.25 at a genotype prevalence of 30%. There was moderate to high power to detect an odds ratio of 1.50 in African Americans when genotype

prevalence was between 10% and 30% and high power in Caucasians within the same genotype prevalence range. Overall, there was low power to detect subtype-specific effects.

Previous case control studies of *AURKA* and overall breast cancer have reported odds ratios ranging from 1.35 to 2.56, which suggests adequate power in CBCS to replicate similar findings [50,81].

2.13 Public Health Impact and Scientific Significance

There is a clear disparity in age-adjusted mortality between African American women and Caucasian women with respect to breast cancer [2]. Reducing the incidence of breast cancer in general is a significant public health concern, since breast cancer is the second leading cause of cancer death in American women. Aside from differences in mortality, younger African American women are more likely to be diagnosed with basal-like tumors, which are associated with poorer survival [2,13]. Recently, O'Brien *et al.* reported that basal-like breast tumors were equally aggressive among African American and Caucasian women [83]. This suggests the disproportionate number of basal-like cases among African American women may be related to a particular genetic profile. Identifying genotypes that are associated with breast cancer subtype could help to further explain biological differences between Caucasian women, who are more likely to present with luminal breast tumors [83], and African American women. In addition, the identification of SNP associations with subtypes of breast cancer could lead to distinct preventive measures for women demonstrating genetic profiles associated with higher risk for a particular subtype of breast cancer.

This study conducted a novel investigation of genetic variation in biologically plausible candidate genes and association with breast cancer among African American and Caucasian women. The CBCS provides a distinctly under studied population of African American participants, who are prone to the basal-like subtype of breast cancer. The candidate genes examined were carefully chosen for their potential roles in oncogensis and cell cycle regulation. A better understanding of common genetic exposures involved in specific breast cancer subtype etiology could lead to more specific treatments in the future.

2.14 Strengths and Limitations

The Carolina Breast Cancer Study is the largest population-based study of breast cancer in African American women in the United States. In addition to the collection of genetic data on nearly 70% of cases and 60% of controls, molecular subtype of breast cancer was measured using immunohistochemistry in all cases included in the data set under analysis (~1,400 cases). Because the CBCS oversampled the African American population in North Carolina, genetic loci that may increase the risk for particular subtypes of breast cancer in African American women are more likely to be accurately identified. With specific regard to the main gene of interest, this study was well powered to address a wide range of loci on *AURKA* in relation to breast cancer among African American women.

Another advantage of this study is the candidate gene approach, focused on biological plausibility. The candidate gene approach allows us to directly test the effects of genetic variants on a particular gene in an association study. A candidate gene study may be limited if our understanding of the biology of the disease under investigation is lacking. However, this proposal is guided by the role *AURKA* is known to play in cell cycle regulation.

AURKA and the other genes chosen for this investigation are responsible for distinct biological mechanisms that are likely to play a role in cancer etiology and/or progression. *AURKA* in particular was chosen for its prominent role in cell cycle regulation and its potential to contribute to the proliferation of aneuploid cells, a common characteristic of basal-like breast tumors that may distinguish them from other subtypes [16,17]. *BRCA1* was chosen due to its role in breast cancer etiology, its involvement in cell cycle regulation, and its tendency to be associated with basal-like breast cancers more often than other subtypes [12,15,22].

This study collected and incorporated adjustment for ancestry informative markers in addition to self-reported race to diminish the potentially confounding effects of population stratification. A sensitivity analysis was conducted to evaluate potential misclassification of the breast cancer subtypes (Nyante, 2010); misclassification is likely to be minimal and non-differential.

I did not expect that any of the SNPs under investigation would be associated with a decreased odds of breast cancer, and for this reason power calculations were based on expected odds ratios >1. Although the Carolina Breast Cancer Study is the largest of its kind, these analyses were generally underpowered if the true association between locus and overall rate of breast cancer is closer to the null value (Table 2.8). This lack of power will be more pronounced in the African American stratum because there are relatively fewer African American cases compared to Caucasian cases. Power to detect subtype-specific effects was generally low (Table 2.9), but the innovative nature of the CBCS in distinguishing intrinsic subtypes makes this analysis a worthwhile endeavor in an effort to generate new hypotheses about the relationships between genetic exposures and breast

tumor heterogeneity. Given the hypothesis that a single locus may contribute only a small proportion of the change in risk, I expected that individual risk loci would demonstrate odds ratios on the order of 1.2-2.0, with the majority of meaningful risk loci demonstrating odds ratios of between 1.5 and 2.0.

A limitation of this study is that only certain candidate and tag SNPs were evaluated in the genes of interest. This study did not capture all of the genetic variability in these genes, however—tag SNPs were chosen by the CBCS primary investigators to maximize interrogation of the selected genes in two HapMap populations (CEU and YRI). It is possible that these tag SNPs may not be representative of the CBCS population, however this strategy was chosen because specific data on African Americans and Caucasians living in North Carolina are not available. Candidate SNPs were selected based on an exhaustive literature review. Nevertheless, potentially functional loci may have been ignored or missed which limits the ability to fully characterize the risk conferred by the genes of interest.

2.15 Summary—Study Design and Methods

The Carolina Breast Cancer Study offers an excellent opportunity to further characterize potential racial differences in genetic exposures associated with intrinsic subtype of breast cancer. The CBCS recruited a large proportion of African American women in an effort to better understand racial disparities in risk factors for breast cancer. By subtyping case participants, the CBCS affirms the heterogeneity within breast cancer and allows for more distinct etiologic pathways to be investigated.

This study takes advantage of the CBCS study population by investigating a key racial disparity in breast cancer—younger African American women are more likely to be

diagnosed with basal-like breast cancer. This could be due in part to polymorphic differences in *AURKA* between African Americans and Caucasians. Since *AURKA* is a key regulator of the cell cycle, and aberrations in its function have been shown to lead to aneuploidy, its potential as an oncogene should be considered. *AURKA* is further implicated in playing a specific role in the etiology of basal-like breast cancer, which commonly demonstrates higher degrees of aneuploidy.

To investigate *AURKA*'s association with breast cancer, logistic regression was employed to calculate odds ratios and 95% confidence intervals as estimates of rate ratios.

Acknowledging the potential for gene-gene interaction among cell cycle regulatory genes, additive interactions between *AURKA* and *BRCA1* (a known cell cycle regulatory gene) were also evaluated. *BRCA1* was chosen in particular due to its role as a known risk factor for breast cancer and due to its known protein interaction with the *AURKA* protein. Additionally, *BARD1, BRIP1,* and *ZNF350* were also investigated for additive interaction with *AURKA* due to their established roles as *BRCA1*-interacting genes.

2.16 Tables

Table 2.1 Breast Cancer Intrinsic Subtype Classification by Immunohistochemistry (IHC)*

	IHC Receptor Status				
Intrinsic Subtype	ER	PR	HER2	CK 5/6	HER1
Luminal-A	+	+	-	N/A	N/A
Luminal-A	+	-	-	N/A	N/A
Luminal-A	-	+	-	N/A	N/A
Luminal-B[a]	+	+	+	N/A	N/A
Luminal-B[a]	+	-	+	N/A	N/A
Luminal-B[a]	-	+	+	N/A	N/A
HER+/ER-	-	-	+	N/A	N/A
Basal-like	-	-	-	+	+
Basal-like	-	-	-	+	-
Basal-like	-	-	-	-	+
Unclassified	-	-	-	-	-

* *in situ* cases were classified without respect to PR status

[a] Definition does not identify all luminal-B tumors, since only 30-50% are HER2+

Table 2.2 Response/Participation Rates of Women Selected as Potential Participants for the CBCS by Case Status, Race, and Age

		Cases*							
		African Americans				Caucasians			
	Total	20-39 yrs	40-49 yrs	50-64 yrs	> 64 yrs	20-39 yrs	40-49 yrs	50-64 yrs	> 64 yrs
Sampled (N)	3360	192	432	471	323	236	680	603	423
Ineligible (%)[a]	6.3	3.6	6.5	5.3	9.9	3.4	5.0	6.8	8.5
Deceased (%)	1.2	2.6	0.9	2.3	2.2	0.4	0.6	0.8	0.5
Uncontactable (%)	2.0	2.6	3.0	2.3	3.1	2.5	1.9	1.7	0
Physician Refusal (%)	6.5	4.2	5.8	3.8	5.0	7.2	7.1	8.8	8.3
Participant Refusal (%)	12.2	12.0	13.0	17.8	22.0	6.4	7.5	8.3	14.2
Interviewed (%)[b]	71.8	75.0	70.8	68.3	57.9	80.1	77.9	73.6	68.5
Contact Rate[c]	98.0%	97.4%	97.0%	97.7%	96.9%	97.5%	98.1%	98.3%	100.0%
Cooperation Rate[d]	79.3%	82.3%	79.1%	75.9%	68.2%	85.5%	84.3%	81.2%	75.3%
Overall Response Rate[e]	77.6%	80.0%	76.5%	74.0%	65.8%	83.3%	82.6%	79.7%	75.3%
Genotyping Rate[f]	66.3%	70.9%	63.0%	62.0%	49.6%	77.4%	71.2%	69.8%	64.7%

		Controls							
		African Americans				Caucasians			
	Total	20-39 yrs	40-49 yrs	50-64 yrs	> 64 yrs	20-39 yrs	40-49 yrs	50-64 yrs	> 64 yrs
Sampled (N)	4465	297	663	631	452	272	788	805	557
Ineligible (%)[a]	9.9	9.1	5.0	4.1	17.7	15.1	9.0	8.1	17.6
Deceased (%)	2.3	0.3	1.2	2.4	6.6	0.4	0.4	2.4	4.3
Uncontactable (%)	17.0	36.0	27.6	22.0	14.4	21.7	15.2	8.7	2.9
Participant Refusal (%)	20.6	17.8	18.4	25.8	24.6	12.1	18.1	21.9	21.5
Interviewed (%)[b]	50.2	36.7	47.8	45.7	36.7	50.7	57.2	59.0	53.6
Contact Rate[c]	83.0%	64.0%	72.4%	78.0%	85.6%	78.3%	84.8%	91.3%	97.1%
Cooperation Rate[d]	70.9%	67.3%	72.2%	63.9%	59.9%	80.7%	75.9%	73.0%	71.4%
Overall Response Rate[e]	57.2%	40.5%	51.0%	48.8%	48.5%	60.0%	63.2%	65.9%	68.7%
Genotyping Rate[f]	56.4%	56.2%	54.0%	50.3%	44.8%	67.8%	64.3%	57.6%	55.4%

* Includes *in situ* cases

[a] Eligibility criteria include age (20-74 years), female gender, residence in 24-county study area, able to complete an interview in English, and no prior history of breast cancer

[b] Includes women who did not complete a full interview

[c] Contact rate= # of women contacted divded by # of women identified as potential cases or controls

[d] Cooperation rate= # of completed interviews divided by # of women contacted and eligible

[e] Overall response rate= # of completed interviews divided by # of women selected for study minus ineligible and deceased women

[f] Genotyping rate= # of women who were successfully genotyped divided by # of women selected for study minus ineligible, uncontactable, and deceased women

Table 2.3 Attributes of CBCS Case Participants with IHC Subtype Data

Attribute	All cases (N=1,412)	Luminal A (n=790)	Luminal B (n=135)	Her2+/ER- (n=116)	Basal-like (n=224)	Unclassified (n=147)
	No. (%)					
Age (yrs), mean (SD)	52 (11)	53 (11)	51 (11)	51 (12)	48 (11)	50 (12)
Race						
African American	581 (41)	287 (36)	45 (33)	48 (41)	122 (54)	79 (54)
European American	831 (59)	503 (64)	90 (67)	68 (59)	102 (46)	68 (46)
Menopausal Status						
Premenopausal	632 (45)	322 (41)	64 (47)	46 (40)	124 (55)	76 (52)
Postmenopausal	780 (55)	468 (59)	71 (53)	70 (60)	100 (45)	71 (48)
AJCC Stage						
in situ	272 (20)	170 (22)	24 (18)	43 (38)	19 (9)	16 (11)
I	414 (30)	263 (34)	37 (28)	17 (15)	47 (22)	50 (34)
II	559 (40)	277 (36)	59 (45)	37 (32)	125 (57)	61 (42)
III	108 (8)	49 (6)	10 (8)	12 (11)	21 (10)	16 (11)
IV	28 (2)	11 (1)	2 (2)	5 (4)	6 (3)	4 (3)
Missing	31	20	3	2	6	0
ER Status						
Positive	841 (60)	716 (91)	125 (93)	0	0	0
Negative	571 (40)	74 (9)	10 (7)	116 (100)	224 (100)	147 (100)
PR Status*						
Positive	614 (54)	524 (85)	90 (81)	0	0	0
Negative	526 (46)	96 (15)	21 (19)	73 (100)	205 (100)	131 (100)
Combined ER/PR Status*						
ER+/PR+	530 (46)	450 (73)	80 (72)	0	0	0
ER+/PR-	117 (10)	96 (15)	21 (19)	0	0	0
ER-/PR+	84 (7)	74 (12)	10 (9)	0	0	0
ER-/PR-	409 (36)	0	0	73 (100)	205 (100)	131 (100)
HER2 Status						
Positive	251 (18)	0	135 (100)	116 (100)	0	0
Negative	1,161 (82)	790 (100)	0	0	224 (100)	147 (100)
CK 5/6 Status						
Positive	206 (15)	55 (7)	7 (5)	20 (17)	124 (55)	0
Negative	1,206 (85)	735 (93)	128 (95)	96 (83)	100 (45)	147 (100)
HER1 Status						
Positive	289 (20)	44 (6)	17 (13)	45 (39)	183 (82)	0
Negative	1,123 (80)	746 (94)	118 (87)	71 (61)	41 (18)	147 (100)

Abbreviations: AJCC, American Joint Committee on Cancer; ER, estrogen receptor; HER2, human epidermal growth factor receptor-2; PR, progesterone receptor.

*PR status not collected for *in situ* cases (n=272)

Table 2.4 Single Nucleotide Polymorphisms (SNPs) Genotyped in CBCS Participants

Gene	SNP	Gene	SNP	Gene	SNP
AURKA	rs1047972	*BRCA1*	rs4986850	*BARD1*	rs10932568
	rs34987347		rs1799950		rs10221582
	rs1468056		rs16941		rs10932573
	rs16979826		rs16942		rs12474696
	rs16979829		rs1799966		rs12477063
	rs16979865		rs799917		rs1542173
	rs2064863		rs4986852		rs16852761
	rs2180691		rs3737559		rs16852798
	rs2236207		rs799923		rs16852799
	rs2273535	*BRIP1*	rs4986764		rs17487827
	rs2298016		rs7213430		rs1979028
	rs6014711		rs4988350		rs2075622
	rs6014712		rs4988346		rs2888294
	rs6024840		rs4988351		rs3768704
	rs6092309		rs2048718		rs3768707
	rs6099120		rs1978111		rs3768708
	rs6099122	*ZNF350*	rs4986773		rs3820727
	rs6099126		rs2278420		rs4672729
	rs6099127		rs3764538		rs6706777
	rs6099128		rs4986771		rs6712055
	rs1468055		rs2278415		rs6749828
	rs6024836		rs11879758		rs6751923
	rs33923703		rs2278417		rs6753417
	rs6099119		rs4986770		rs6756902
	rs911162		rs4988334		rs7557557
			rs8102072		rs7566806
					rs7585356
					rs1048108
					rs3738888
					rs28997576
					rs2229571

Table 2.5 Previous Study Results of the Aassociations Between Polymorphisms on *AURKA* and Odds/Hazard of Breast Cancer

AURKA polymorphism	Author	Year	Study Population	Design	Cases	Controls	Effect Estimate	95% CI
rs2273535 (T>A)	Dai	2004	Asian	Population-based	1102	1186	OR*=1.2	0.90-1.60
	Egan	2004	Caucasian	Population-based	940	830	OR*=1.56	0.96-2.47
	Sun	2004	Asian	Hospital-based	520	520	OR*=1.76	1.16-2.66
	Lo	2005	Asian	Hospital-based	707	1969	OR†=1.08	0.81-1.46
	Ewart-Toland	2005	Mixed	Population-based	898	448	OR†=1.54	0.92-2.59
	Fletcher	2006	Caucasian	Population-based	507	875	OR*=0.78	0.57-1.04
	BCAC	2006	Caucasian	Pooled case-control	7816	9285	OR*=1.04	0.91-1.20
	Cox	2006	Caucasian	Nested case-control‡	1241	1711	OR*=1.43	0.99-2.06
	Couch	2007	Mixed	Population-based	3884	3303	HR*=0.91	0.77-1.06
	Vidarsdottir	2007	Caucasian	Hospital-based	759	653	OR*=1.87	1.09-3.21
	Guenard	2009	Caucasian	Cohort-familial	96	96	OR†=1.38	0.42-4.57
	The MARIE-GENICA Consortium	2010	Caucasian	Population-based	3136	5466	OR*=1.17	0.94-1.44
	Ruan	2011	Asian	Population-based	1334	1568	OR*=1.99	1.10-3.61
	Shi	2011	Caucasian	Population-based	763	1516	OR*=0.72	0.45-1.13
rs6064391 (T>G)	Ruan	2011	Asian	Population-based	1326	1569	OR*=1.17	0.71-1.95
rs6064389 (T>A)	Shi	2011	Caucasian	Population-based	765	1529	OR*=0.80	0.62-1.03
rs16979877 (G>A)	Shi	2011	Caucasian	Population-based	765	1530	OR*=2.59	0.70-9.75
rs8173 (G>C)	Shi	2011	Caucasian	Population-based	762	1524	OR*=0.79	0.55-1.16
rs911162 (G>A)	Ruan	2011	Asian	Population-based	1334	1568	OR*=1.62	0.62-4.24
rs1047972 (G>A)	The MARIE-GENICA Consortium	2010	Caucasian	Population-based	3139	5469	OR*=0.70	0.52-0.94
rs2064863 (T>G)	Ruan	2011	Asian	Population-based	1323	1568	OR*=0.63	0.40-1.01
rs2298016 (C>G)	Ruan	2011	Asian	Population-based	1331	1568	OR*=0.43	0.24-0.78
rs8117896 (T>C)	Ruan	2011	Asian	Population-based	1334	1568	OR*=2.05	0.93-4.50
rs10485805 (G>A)	Ruan	2011	Asian	Population-based	1330	1568	OR*=0.68	0.46-1.00
rs6024836 (A>G)	Ruan	2011	Asian	Population-based	1332	1568	OR*=1.59	1.20-2.11

* Based on a codominant model; i.e. index= homozygous for the rare allele, referent=homozygous for the common allele; adjusted for study covariates

† Based on a codominant model; i.e. index= homozygous for the rare allele, referent=homozygous for the common allele; unadjusted

‡ Nested case-control study within th Nurses' Health Study

Table 2.6 Previous Study Results of the Associations Between Polymorphisms on *BRCA1* and Odds of Breast Cancer

BRCA1 polymorphism	Author	Year	Study Population Ethnicity	Design	Cases	Controls	Effect Estimate	95% CI
rs3737559 (C>T)	Ruan	2011	Asian	Population-based	1330	1568	OR* =1.35	1.11-1.64
rs4986850 (A>G)	Bhatti--USRT	2008	Caucasian	US Radiologic Technologists Cohort	859	1083	OR** =0.97	0.75-1.25
rs16942 (C>T)†	Cox--CIMBA	2011	Mixed	Population-based Consortium	2980	2672	HR=0.85	0.74-0.96
rs799917 (C>T)	Huo	2008	Asian	Hospital-based	568	624	OR* =0.98	0.66-1.44

* Based on a codominant model; i.e. index= homozygous for the rare allele, referent=homozygous for the common allele; adjusted for study covariates

** Based on a dominant model; i.e. index= homozygous for the rare allele + heterozygotes, referent=homozygous for the common allele; unadjusted

† Data is from the Consortium of Investigators of Modifiers of *BRCA1/2* (CIMBA); a total of 9,874 *BRCA1* mutation carriers were available; "cases" and "controls" refer to "breast cancer case" and "unaffected by breast cancer" respectively; HR describes the association between rs16942 genotypes on the 'wild-type' (non-mutant) allele of *BRCA1* and breast cancer risk given a total of 235,488 person-years contributed.

Table 2.7 Candidate Gene Single Nucleotide Polymorphisms (SNPs) with Extreme Hardy-Weinberg Equilibrium (HWE) P-values

	Gene	SNP	Exact HWE P-value
Caucasian Controls	*AURKA*	rs6099127	0.01
	BRIP1	rs4988346	0.01
	BARD1	rs6712055	0.03
	BRCA1	rs4986850	0.02
	BRCA1	rs3737559	0.02
	ZNF350	rs4986771	0.04
African American Controls	*AURKA*	rs2236207	0.03
	AURKA	rs33923703	0.01
	BRIP1	rs2048718	0.01
	BARD1	rs6706777	0.04
	ZNF350	rs3764538	0.001
	ZNF350	rs2278415	<0.001
	ZNF350	rs8102072	0.03

Table 2.8 Study Power for Main Effects of Genotype on All Breast Cancer in CBCS Participants by Race (α=0.05)*

Genotype† prevalence	Minimal Detectable Odds Ratio	Power in Caucasians (N_{cases} = 1,204, $N_{controls}$ = 1,089)	Power in African Americans (N_{cases} = 742, $N_{controls}$ = 658)
5%	1.25	0.23	0.16
	1.50	0.63	0.43
10%	1.25	0.39	0.25
	1.50	0.88	0.69
20%	1.25	0.60	0.40
	1.50	0.98	0.89
30%	1.25	0.70	0.49
	1.50	0.99	0.95

* Power calculations performed using Episheet

† Genotype prevalence in controls, assuming a dominant genetic model

Table 2.9 Study Power for Main Effects of Genotype on Intrinsic Subtype of Breast Cancer in CBCS Participants (α=0.05)*

Subtype	Genotype† prevalence	Minimal Detectable Odds Ratio	Power
Luminal A (Cases=674)	5%	1.25	0.22
		1.50	0.59
	10%	1.25	0.35
		1.50	0.84
	20%	1.25	0.54
		1.50	0.97
	30%	1.25	0.64
		1.50	0.99
Luminal B (Cases=114)	5%	1.25	0.10
		1.50	0.23
	10%	1.25	0.13
		1.50	0.33
	20%	1.25	0.18
		1.50	0.47
	30%	1.25	0.20
		1.50	0.53
HER2+/ER- (Cases=94)	5%	1.25	0.09
		1.50	0.20
	10%	1.25	0.12
		1.50	0.29
	20%	1.25	0.16
		1.50	0.41
	30%	1.25	0.18
		1.50	0.47
Basal-like (Cases=199)	5%	1.25	0.12
		1.50	0.31
	10%	1.25	0.18
		1.50	0.47
	20%	1.25	0.25
		1.50	0.66
	30%	1.25	0.30
		1.50	0.74

* Power calculations performed using Episheet and an estimated 1,747 controls

† Genotype prevalence in controls, assuming a dominant genetic model

Table 2.10 Single nucleotide polymorphisms (SNPs) included in additive interaction analysis

Gene	SNP
AURKA	rs34987347
	rs2273535
	rs2298016
	rs6024836
ZNF350	rs4986773
	rs2278420
	rs11879758
	rs4986770
BARD1	rs1048108
	rs3738888
	rs2229571
BRCA1	rs1799950
	rs16941
	rs16942
	rs1799966
	rs799917
	rs4986852
BRIP1	rs4986764
	rs4988350

Table 2.11 Assessment of potential functionality of single nucleotide polymorphisms (SNPs) on candidate gene *ZNF350*genotyped in the CBCS

SNP	Status	FS Score[1]	Polyphen prediction[2]	SIFT prediction[3]	Description
rs4986773	Possibly functional	0.319	N/A	N/A	Predicted to impact splicing regulation by ESEfinder[5] and ESRSearch[6]; previous study reported joint effects of variants of rs4986773 with rs799917 (*BRCA1*) were found to increase breast cancer risk in a population of Chinese women (OR=2.03, 95%CI= 1.02-4.05, P(int)=0.059) PMID=19484476
rs2278420	Possibly functional	0.599	Benign	Tolerated	Predicted to be deleterious by SNPeffect[4]; predicted to impact splicing regulation by ESEfinder[5] and ESRSearch[6]; predicted to impact post translation by OGPET[7]
rs3764538	Probably functional	0.623	N/A	N/A	Predicted to impact splicing regulation by ESRSearch[6], PESX[8] and RESCUE_EXE[9]
rs4986771	Functional	0.906	Benign	Tolerated	Predicted to be deleterious by SNPeffect; predicted to impact splicing regulation by ESEfinder[5], ESRSearch[6], PESX[8] and RESCUE_EXE[9]; predicted to impact post translation by OGPET[7]
rs2278415	Functional	0.902	Possibly damaging	Damaging	Predicted to be deleterious by PolyPhen[2], SIFT[3], SNPeffect[4], and SNPs3D[10]; predicted to impact splicing regulation by ESEfinder[5], ESRSearch[6], PESX[8] and RESCUE_EXE[9]
rs11879758	Possibly functional	0.5	N/A	N/A	Predicted to impact transcriptional regulation by TFSearch[11]; No previously reported literature results.
rs2278417	Not functional	0	N/A	N/A	

rs4986770	Possibly functional	0.5	N/A	N/A	Predicted to impact transcriptional regulation by TFSearch[11]; 1845 C>T variant previously studied in a 2004 kin-cohort study of familial breast cancer risk. Among 2,430 female first-degree relatives of women with a history of breast cancer, 190 cases of breast cancer arose. The 1845 C>T variant was associated with an increased risk for breast cancer up to age 50 in this cohort (OR=2.2, 95%CI=0.5-4.3). PMID=15113441
rs4988334	Possibly functional	0.365	N/A	N/A	Predicted to impact splicing regulation by ESEfinder[5], ESRSearch[6] and RESCUE_EXE[9]; no previously reported literature results.
rs8102072	Possibly functional	0.5	N/A	N/A	Predicted to impact transcriptional regulation by TFSearch[11]; No previously reported literature results.

1 F-SNP database and algorithms are the work of Phil H. Lee and Hagit Shatkay of Queen's University; the database provides information about the functional effects of SNPs by integrating results from 16 bioinformatics tools and databases. SNPs are evaluated on the basis of four integral functions: protein coding, splicing regulation, transcriptional regulation and post-translation. SNPs then receive an FS Score which ranges from 0-1, with scores above 0.5 indicating high probability of functionality. A detailed explanation of how the algorithm scores SNPs can be found at http://compbio.cs.queensu.ca/F-SNP/

2 PolyPhen is a website and method for predicting SNP variant impact on protein structure and function developed by Ramensky V., Bork P., and Sunyaev S. and described in Human non-synonymous SNPs: server and survey. Nucleic Acids Res. (2002) 30(17): 3894-900 (PMID: 12202775)

3 SIFT predicts whether an amino acid substitution affects protein function and is based on the conservation of amino acid residues in the genome. SIFT was developed by Ng, P. and Henikoff, S. and described in Predicting deleterious amino acid substitutions. Genome Research, (2001) 11, 863-874. The SIFT database can be accessed at http://blocks.fhcrc.org/sift/SIFT.html

4 SNPeffect predicts deleterious missense SNPs using methods described by Reumers, J., Schymkowitz, J., Ferkinghoff-Borg, J., Stricher, F., Serrano, L., and Rousseau, F. in SNPeffect: a database mapping molecular phenotypic effects of human non-synonymous coding SNPs. Nucleic Acid Research, (2005) 33 (Database issue), D527–532. The SNPeffect database can be accessed at http://snpeffect.switchlab.org/index.php

5 ESEfinder identifies exonic splice sites using methods described by Cartegni, L., Wang, J., Zhu, Z., Zhang, M. Q., and Krainer, A. R. in ESEfinder: A web resource to identify exonic splicing enhancers. Nucleic Acids Research, (2003), 31(13), 3568–3571. The ESEfinder database can be accessed at http://rulai.cshl.edu/cgi-bin/tools/ESE3/esefinder.cgi

6 ESRSearch identifies exonic splice sites using methods described by Fairbrother, W. G., Yeh, R. F., Sharp, P. A., and Burge, C. B. in Predictive identification of exonic splicing enhancers in human genes.. Science, (2002), 297, 1007–1013. The ESRSearch database can be accessed at http://www.tau.ac.il/lifesci/

7 OGPET identifies O-glycosylation sites, which are associated with post-translational functionality. Methods employed by OGPET are described by Gerken, T., Tep, C., and Rarick, J. in The role of peptide sequence and neighboring residue glycosylation on the substrate specificity of the uridine 5 diphosphate-alpha-n-acetylgalactosamine:polypeptide n-acetylgalactosaminyl transferases t1 and t2: kinetic modeling of the porcine and canine submaxillary gland mucin tandem repeats. (2004) Biochemistry, 43, 9888–9900. The OGPET database may be accessed at http://ogpet.utep.edu/main.php

8 PESX identifies exonic splice sites using methods described by Zhang et al. in Exon inclusion is dependent on predictable exonic splicing enhancers. Molecular and Cellular Biology, (2005), 25(16), 7323–7332. The PESX database may be accessed at http://cubweb.biology.columbia.edu/pesx/

9 RESCUE_EXE identifies exonic splice sites using methods described by Yeo, G. and Burge, C. B. in Variation in sequence and organization of splicing regulatory elements in vertebrate genes. In the Proceeding of Proc. Natl. Acad. Sci., (2004), 101(44), 15700–15705. 5. The RESCUE_EXE database can be accessed at http://genes.mit.edu/burgelab/rescue-ese/

10 SNPs3D predicts deleterious missense mutations using methods described by Yue, P., Melamud, E., and Moult, J. in SNPs3D: candidate gene and SNP selection for association studies. BMC Bioinformatics, (2006), 7, 166. The SNPs3D database can be accessed at http://www.snps3d.org/modules.php?name=SNPtargets

11 TFSearch identifies transcription factor binding sites using methods described by Akiyama, Y. in TFSEARCH: Searching Transcription Factor Binding Sites (1998). The TFSearch database can be accessed at http://www.cbrc.jp/research/db/TFSEARCH.html

12 The MARIE-GENICA Consortium on Genetic Susceptibility for Menopausal Hormone Therapy Related Breast Cancer Risk Polymorphisms in the BRCA1 and ABCB1 genes modulate menopausal hormone therapy associated breast cancer risk in postmenopausal women, Breast Cancer Res Treat (2010) 120:727–736

Table 2.12 Assessment of potential functionality of single nucleotide polymorphisms (SNPs) on candidate gene *BARD1* genotyped in the CBCS

SNP	Status	FS Score[1]	Polyphen prediction[2]	SIFT prediction[3]	Description
rs10932568	Probably not functional	0.268	N/A	N/A	Predicted to impact transcriptional regulation by TFSearch[11].
rs10221582	Probably not functional	0.268	N/A	N/A	Predicted to impact transcriptional regulation by TFSearch[11].
rs10932573	Probably not functional	0.268	N/A	N/A	Predicted to impact transcriptional regulation by TFSearch[11].
rs12474696	Not functional	0	N/A	N/A	
rs12477063	Probably not functional	0.242	N/A	N/A	Predicted to impact transcriptional regulation by TFSearch[11].
rs1542173	Probably not functional	0.242	N/A	N/A	Predicted to impact transcriptional regulation by TFSearch[11].
rs16852761	Probably not functional	0.109	N/A	N/A	
rs16852798	Probably not functional	0.176	N/A	N/A	Predicted to impact transcriptional regulation by TFSearch[11].
rs16852799	Probably not functional	0.176	N/A	N/A	Predicted to impact transcriptional regulation by TFSearch[11].
rs17487827	Probably not functional	0.176	N/A	N/A	Predicted to impact transcriptional regulation by TFSearch[11].
rs1979028	Probably not functional	0.144	N/A	N/A	
rs2075622	Probably not functional	0.242	N/A	N/A	Predicted to impact transcriptional regulation by TFSearch[11].
rs2888294	Probably not functional	0.217	N/A	N/A	Predicted to impact transcriptional regulation by TFSearch[11].
rs3768707	Probably not functional	0.242	N/A	N/A	Predicted to impact transcriptional regulation by TFSearch[11].
rs3768708	Probably not functional	0.242	N/A	N/A	Predicted to impact transcriptional regulation by TFSearch[11].
rs3820727	Probably not functional	0.242	N/A	N/A	Predicted to impact transcriptional regulation by TFSearch[11].
rs4672729	Probably not functional	0.109	N/A	N/A	

rs6706777	Probably not functional	0.242	N/A	N/A	Predicted to impact transcriptional regulation by TFSearch[11].
rs6712055	Probably not functional	0.176	N/A	N/A	Predicted to impact transcriptional regulation by TFSearch[11]; common variation in *BARD1* was studied in a case-control GWAS of high risk neuroblastoma (397 cases, 2,043 controls). rs6712055 was associated with an increased odds of neuroblastoma (OR_{CMH}=1.56, 95%CI=1.37-1.78) PMID=19412175.
rs6749828	Not functional	0	N/A	N/A	
rs6751923	Probably not functional	0.268	N/A	N/A	Predicted to impact transcriptional regulation by TFSearch[11].
rs6753417	Probably not functional	0.208	N/A	N/A	Predicted to impact transcriptional regulation by TFSearch[11].
rs6756902	Not functional	0	N/A	N/A	
rs7557557	Probably not functional	0.268	N/A	N/A	Predicted to impact transcriptional regulation by TFSearch[11].
rs7566806	Probably not functional	0.208	N/A	N/A	Predicted to impact transcriptional regulation by TFSearch[11].
rs7585356	Probably not functional	0.242	N/A	N/A	Predicted to impact transcriptional regulation by TFSearch[11]; common variation in *BARD1* was studied in a case-control GWAS of high risk neuroblastoma (397 cases, 2,043 controls). rs7585356 was associated with a decreased odds of neuroblastoma (OR_{hom}=0.36, 95%CI=0.22-0.58) PMID=19412175.
rs1048108	Functional	0.774	Probably damaging	Tolerated	Predicted to be deleterious by PolyPhen[2]; predicted to impact splicing regulation by ESEfinder[5]; predicted to impact post-translation by OGPET[7]. rs1048108 has been studied in relation to cervical cancer in Chinese women. PMID=19482343.

rs3738888	Probably functional	0.69	Possibly damaging	Damaging	Predicted to be deleterious by PolypPhen[2], SIFT[3], SNPeffect[4] and SNPs3D[10]; predicted to impact splicing regulation by ESEfinder[5], ESRSearch[6] and PESX[8]. No significant literature results reported.
rs28997576	Possibly functional	0.33	Benign	N/A	Predicted to impact splicing regulation by ESEfinder[5], ESRSearch[6] and PESX[8]; predicted to impact post-translation by OGPET[7]. rs28997576 has been previously associated with risk of schizophrenia in a 3-cohort meta-analysis (RR=1.655, 95%CI=1.095-2.502) PMID=19435634.
rs2229571	Probably functional	0.649	Benign	Damaging	Predicted to be deleterious by SIFT[3] and SNPs3D[10]; predicted to impact splicing regulation by ESEfinder[5], ESRSearch[6] and RESCUE_EXE[9]; predicted to impact post-translation by OGPET[7]. No significant literature results reported.

1 F-SNP database and algorithms are the work of Phil H. Lee and Hagit Shatkay of Queen's University; the database provides information about the functional effects of SNPs by integrating results from 16 bioinformatics tools and databases. SNPs are evaluated on the basis of four integral functions: protein coding, splicing regulation, transcriptional regulation and post-translation. SNPs then receive an FS Score which ranges from 0-1, with scores above 0.5 indicating high probability of functionality. A detailed explanation of how the algorithm scores SNPs can be found at http://compbio.cs.queensu.ca/F-SNP/

2 PolyPhen is a website and method for predicting SNP variant impact on protein structure and function developed by Ramensky V., Bork P., and Sunyaev S. and described in Human non-synonymous SNPs: server and survey. Nucleic Acids Res. (2002) 30(17): 3894-900 (PMID: 12202775)

3 SIFT predicts whether an amino acid substitution affects protein function and is based on the conservation of amino acid residues in the genome. SIFT was developed by Ng, P. and Henikoff, S. and described in Predicting deleterious amino acid substitutions. Genome Research, (2001) 11, 863-874. The SIFT database can be accessed at http://blocks.fhcrc.org/sift/SIFT.html

4 SNPeffect predicts deleterious missense SNPs using methods described by Reumers, J., Schymkowitz, J., Ferkinghoff-Borg, J., Stricher, F., Serrano, L., and Rousseau, F. in SNPeffect: a database mapping molecular phenotypic effects of human non-synonymous coding SNPs. Nucleic Acid Research, (2005) 33 (Database issue), D527–532. The SNPeffect database can be accessed at http://snpeffect.switchlab.org/index.php

5 ESEfinder identifies exonic splice sites using methods described by Cartegni, L., Wang, J., Zhu, Z., Zhang, M. Q., and Krainer, A. R. in ESEfinder: A web resource to identify exonic splicing enhancers. Nucleic Acids Research, (2003), 31(13), 3568–3571. The ESEfinder database can be accessed at http://rulai.cshl.edu/cgi-bin/tools/ESE3/esefinder.cgi

6 ESRSearch identifies exonic splice sites using methods described by Fairbrother, W. G., Yeh, R. F., Sharp, P. A., and Burge, C. B. in Predictive identification of exonic splicing enhancers in human genes.. Science, (2002), 297, 1007–1013. The ESRSearch database can be accessed at http://www.tau.ac.il/lifesci/

7 OGPET identifies O-glycosylation sites, which are associated with post-translational functionality. Methods employed by OGPET are described by Gerken, T., Tep, C., and Rarick, J. in The role of peptide sequence and neighboring residue glycosylation on the substrate specificity of the uridine 5 diphosphate-alpha-n-acetylgalactosamine:polypeptide n-acetylgalactosaminyl transferases t1 and t2: kinetic modeling of the porcine and canine submaxillary gland mucin tandem repeats. (2004) Biochemistry, 43, 9888–9900. The OGPET database may be accessed at http://ogpet.utep.edu/main.php

8 PESX identifies exonic splice sites using methods described by Zhang et al. in Exon inclusion is dependent on predictable exonic splicing enhancers. Molecular and Cellular Biology, (2005), 25(16), 7323–7332. The PESX database may be accessed at http://cubweb.biology.columbia.edu/pesx/

9 RESCUE_EXE identifies exonic splice sites using methods described by Yeo, G. and Burge, C. B. in Variation in sequence and organization of splicing regulatory elements in vertebrate genes. In the Proceeding of Proc. Natl. Acad. Sci., (2004), 101(44), 15700–15705. 5. The RESCUE_EXE database can be accessed at http://genes.mit.edu/burgelab/rescue-ese/

10 SNPs3D predicts deleterious missense mutations using methods described by Yue, P., Melamud, E., and Moult, J. in SNPs3D: candidate gene and SNP selection for association studies. BMC Bioinformatics, (2006), 7, 166. The SNPs3D database can be accessed at http://www.snps3d.org/modules.php?name=SNPtargets

11 TFSearch identifies transcription factor binding sites using methods described by Akiyama, Y. in TFSEARCH: Searching Transcription Factor Binding Sites (1998). The TFSearch database can be accessed at http://www.cbrc.jp/research/db/TFSEARCH.html

12 The MARIE-GENICA Consortium on Genetic Susceptibility for Menopausal Hormone Therapy Related Breast Cancer Risk Polymorphisms in the BRCA1 and ABCB1 genes modulate menopausal hormone therapy associated breast cancer risk in postmenopausal women, Breast Cancer Res Treat (2010) 120:727–736

Table 2.13 Assessment of potential functionality of single nucleotide polymorphisms (SNPs) on candidate gene *BRCA1* genotyped in the CBCS

SNP	Status	FS Score[1]	Polyphen prediction[2]	SIFT prediction[3]	Description
rs4986850	Probably functional	0.684	Benign	Tolerated	Predicted to be deleterious by SNPs3D[10]; predicted to impact splicing regulation by ESEfinder[5], ESRSearch[6] and RESCUE_EXE[9]
rs1799950	Functional	0.892	Possibly damaging	Damaging	Predicted to be deleterious by PolyPhen[2], SIFT[3], and SNPs3D[10]; predicted to impact splicing regulation by ESEfinder[5], ESRSearch[6], PESX[8] and RESCUE_EXE[9].
rs16941	Functional	0.945	Possibly damaging	Damaging	Predicted to be deleterious by SIFT[3], SNPeffect[4] and SNPs3D[10]; predicted to impact splicing regulation by ESEfinder[5], ESRSearch[6], PESX[8] and RESCUE_EXE[9]
rs16942	Functional	0.934	Benign	Tolerated	Predicted to be deleterious by SNPeffect[4] and SNPs3D[10]; predicted to impact splicing regulation by ESEfinder[5], ESRSearch[6], PESX[8] and RESCUE_EXE[11]
rs1799966	Possibly functional	0.5	Benign	Damaging	Predicted to be deleterious by SIFT[3], SNPeffect[4] and SNPs3D[10]; predicted to impact splicing regulation by ESEfinder[5] and ESRSearch[6]; predicted to impact post-translation by OGPET[7]. A 2007 case control study of functional variants on *BRCA1* found rs1799966 to be associated with a first primary breast tumor (OR= 1.17, 95%CI=1.00-1.36) in cases with more than one occurrence of primary breast cancer. PMID=17341484. A 2008 pathway analysis of SNPs associated with Glioblastoma Multiforme susceptibility found a statistically significant interaction between rs1799966 and rs1047840 (*EXO1*) (OR=0.06, 95%CI=0.01-0.41, P(int)=0.01) PMID=18559551.

rs799917	Possibly functional	0.518	Benign	Damaging	Predicted to be deleterious by SIFT[3], SNPeffect[4] and SNPs3D[10]; predicted to impact splicing regulation by ESEfinder[5], ESRSearch[6] and PESX[8]. Per the above referenced Globlastoma Multiforme study, rs799917 is in linkage disequilibrium with rs1799966. rs799917 is also in near complete linkage disequilibrium with rs 16942 (D' = 0.97; r^2 = 0.93)[12]. A 2009 case control study of *BRCA1*-interacting genes in Chinese women with breast cancer found a statistically significant interaction between rs799917 and rs4986773 (*ZNF350*) (OR=2.03, 95%CI= 1.02-4.05, P(int)=0.059) PMID=19484476.
rs4986852	Possibly functional	0.576	Benign	N/A	Predicted to be deleterious by SNPeffect[4] and SNPs3D[10]; predicted to impact splicing regulation by ESEfinder[5], ESRSearch[6] and PESX[8]; predicted to impact post-translation by OGPET[7].
rs3737559	Probably not functional	0.208	N/A	N/A	Predicted to impact transcriptional regulation by TFSearch[11]. Was associated with breast cancer in a 2011 case-control study of Han Chinese women. (dominant model OR = 1.35, 95% CI = 1.11-1.64)
rs799923	Probably not functional	0.176	N/A	N/A	Predicted to impact transcriptional regulation by TFSearch[11].

1 F-SNP database and algorithms are the work of Phil H. Lee and Hagit Shatkay of Queen's University; the database provides information about the functional effects of SNPs by integrating results from 16 bioinformatics tools and databases. SNPs are evaluated on the basis of four integral functions: protein coding, splicing regulation, transcriptional regulation and post-translation. SNPs then receive an FS Score which ranges from 0-1, with scores above 0.5 indicating high probability of functionality. A detailed explanation of how the algorithm scores SNPs can be found at http://compbio.cs.queensu.ca/F-SNP/

2 PolyPhen is a website and method for predicting SNP variant impact on protein structure and function developed by Ramensky V., Bork P., and Sunyaev S. and described in Human non-synonymous SNPs: server and survey. Nucleic Acids Res. (2002) 30(17): 3894-900 (PMID: 12202775)

3 SIFT predicts whether an amino acid substitution affects protein function and is based on the conservation of amino acid residues in the genome. SIFT was developed by Ng, P. and Henikoff, S. and described in Predicting deleterious amino acid substitutions. Genome Research, (2001) 11, 863-874. The SIFT database can be accessed at http://blocks.fhcrc.org/sift/SIFT.html

4 SNPeffect predicts deleterious missense SNPs using methods described by Reumers, J., Schymkowitz, J., Ferkinghoff-Borg, J., Stricher, F., Serrano, L., and Rousseau, F. in SNPeffect: a database mapping molecular phenotypic effects of human non-synonymous coding SNPs. Nucleic Acid Research, (2005) 33 (Database issue), D527–532. The SNPeffect database can be accessed at http://snpeffect.switchlab.org/index.php

5 ESEfinder identifies exonic splice sites using methods described by Cartegni, L., Wang, J., Zhu, Z., Zhang, M. Q., and Krainer, A. R. in ESEfinder: A web resource to identify exonic splicing enhancers. Nucleic Acids Research, (2003), 31(13), 3568–3571. The ESEfinder database can be accessed at http://rulai.cshl.edu/cgi-bin/tools/ESE3/esefinder.cgi

6 ESRSearch identifies exonic splice sites using methods described by Fairbrother, W. G., Yeh, R. F., Sharp, P. A., and Burge, C. B. in Predictive identification of exonic splicing enhancers in human genes.. Science, (2002), 297, 1007–1013. The ESRSearch database can be accessed at http://www.tau.ac.il/lifesci/

7 OGPET identifies O-glycosylation sites, which are associated with post-translational functionality. Methods employed by OGPET are described by Gerken, T., Tep, C., and Rarick, J. in The role of peptide sequence and neighboring residue glycosylation on the substrate specificity of the uridine 5 diphosphate-alpha-n-acetylgalactosamine:polypeptide n-acetylgalactosaminyl transferases t1 and t2: kinetic modeling of the porcine and canine submaxillary gland mucin tandem repeats. (2004) Biochemistry, 43, 9888–9900. The OGPET database may be accessed at http://ogpet.utep.edu/main.php

8 PESX identifies exonic splice sites using methods described by Zhang et al. in Exon inclusion is dependent on predictable exonic splicing enhancers. Molecular and Cellular Biology, (2005), 25(16), 7323–7332. The PESX database may be accessed at http://cubweb.biology.columbia.edu/pesx/

9 RESCUE_EXE identifies exonic splice sites using methods described by Yeo, G. and Burge, C. B. in Variation in sequence and organization of splicing regulatory elements in vertebrate genes. In the Proceeding of Proc. Natl. Acad. Sci., (2004), 101(44), 15700–15705. 5. The RESCUE_EXE database can be accessed at http://genes.mit.edu/burgelab/rescue-ese/

10 SNPs3D predicts deleterious missense mutations using methods described by Yue, P., Melamud, E., and Moult, J. in SNPs3D: candidate gene and SNP selection for association studies. BMC Bioinformatics, (2006), 7, 166. The SNPs3D database can be accessed at http://www.snps3d.org/modules.php?name=SNPtargets

11 TFSearch identifies transcription factor binding sites using methods described by Akiyama, Y. in TFSEARCH: Searching Transcription Factor Binding Sites (1998). The TFSearch database can be accessed at http://www.cbrc.jp/research/db/TFSEARCH.html

12 The MARIE-GENICA Consortium on Genetic Susceptibility for Menopausal Hormone Therapy Related Breast Cancer Risk Polymorphisms in the BRCA1 and ABCB1 genes modulate menopausal hormone therapy associated breast cancer risk in postmenopausal women, Breast Cancer Res Treat (2010) 120:727–736

Table 2.14 Assessment of potential functionality of single nucleotide polymorphisms (SNPs) on candidate gene *BRIP1* genotyped in the CBCS

SNP	Status	FS Score[1]	Polyphen prediction[2]	SIFT prediction[3]	Description
rs4986764	Possibly functional	0.58	Benign	Tolerated	Predicted to deleterious by SNPeffect[4]; predicted to impact splicing regulation by ESEfinder[5], ESRSearch[6] and PESX[8]; predicted to impact post-translation by OGPET[7].
rs7213430	Probably not functional	0.208	N/A	N/A	Predicted to impact transcriptional regulation by TFSearch[11].
rs4988350	Possibly functional	0.557	Benign	N/A	Predicted to be deleterious by SNPs3D[10]; predicted to impact splicing regulation by ESEfinder[5], ESRSearch[6] and PESX[8].
rs4988346	Probably not functional	0.237	Benign	N/A	Predicted to impact splicing regulation by ESRSearch[6] and PESX[8].
rs4988351	Probably not functional	0.176	N/A	N/A	Predicted to impact transcriptional regulation by TFSearch[11].
rs2048718	Probably not functional	0.208	N/A	N/A	Predicted to impact transcriptional regulation by TFSearch[11].
rs1978111	Probably not functional	0.176	N/A	N/A	Predicted to impact transcriptional regulation by TFSearch[11].

1 F-SNP database and algorithms are the work of Phil H. Lee and Hagit Shatkay of Queen's University; the database provides information about the functional effects of SNPs by integrating results from 16 bioinformatics tools and databases. SNPs are evaluated on the basis of four integral functions: protein coding, splicing regulation, transcriptional regulation and post-translation. SNPs then receive an FS Score which ranges from 0-1, with scores above 0.5 indicating high probability of functionality. A detailed explanation of how the algorithm scores SNPs can be found at http://compbio.cs.queensu.ca/F-SNP/

2 PolyPhen is a website and method for predicting SNP variant impact on protein structure and function developed by Ramensky V., Bork P., and Sunyaev S. and described in Human non-synonymous SNPs: server and survey. Nucleic Acids Res. (2002) 30(17): 3894-900 (PMID: 12202775)

3 SIFT predicts whether an amino acid substitution affects protein function and is based on the conservation of amino acid residues in the genome. SIFT was developed by Ng, P. and Henikoff, S. and described in Predicting deleterious amino acid substitutions. Genome Research, (2001) 11, 863-874. The SIFT database can be accessed at http://blocks.fhcrc.org/sift/SIFT.html

4 SNPeffect predicts deleterious missense SNPs using methods described by Reumers, J., Schymkowitz, J., Ferkinghoff-Borg, J., Stricher, F., Serrano, L., and Rousseau, F. in SNPeffect: a database mapping molecular phenotypic effects of human non-synonymous coding SNPs. Nucleic Acid Research, (2005) 33 (Database issue), D527–532. The SNPeffect database can be accessed at http://snpeffect.switchlab.org/index.php

5 ESEfinder identifies exonic splice sites using methods described by Cartegni, L., Wang, J., Zhu, Z., Zhang, M. Q., and Krainer, A. R. in ESEfinder: A web resource to identify exonic splicing enhancers. Nucleic Acids Research, (2003), 31(13), 3568–3571. The ESEfinder database can be accessed at http://rulai.cshl.edu/cgi-bin/tools/ESE3/esefinder.cgi

6 ESRSearch identifies exonic splice sites using methods described by Fairbrother, W. G., Yeh, R. F., Sharp, P. A., and Burge, C. B. in Predictive identification of exonic splicing enhancers in human genes.. Science, (2002), 297, 1007–1013. The ESRSearch database can be accessed at http://www.tau.ac.il/lifesci/

7 OGPET identifies O-glycosylation sites, which are associated with post-translational functionality. Methods employed by OGPET are described by Gerken, T., Tep, C., and Rarick, J. in The role of peptide sequence and neighboring residue glycosylation on the substrate specificity of the uridine 5 diphosphate-alpha-n-acetylgalactosamine:polypeptide n-acetylgalactosaminyl transferases t1 and t2: kinetic modeling of the porcine and canine submaxillary gland mucin tandem repeats. (2004) Biochemistry, 43, 9888–9900. The OGPET database may be accessed at http://ogpet.utep.edu/main.php

8 PESX identifies exonic splice sites using methods described by Zhang et al. in Exon inclusion is dependent on predictable exonic splicing enhancers. Molecular and Cellular Biology, (2005), 25(16), 7323–7332. The PESX database may be accessed at http://cubweb.biology.columbia.edu/pesx/

9 RESCUE_EXE identifies exonic splice sites using methods described by Yeo, G. and Burge, C. B. in Variation in sequence and organization of splicing regulatory elements in vertebrate genes. In the Proceeding of Proc. Natl. Acad. Sci., (2004), 101(44), 15700–15705. 5. The RESCUE_EXE database can be accessed at http://genes.mit.edu/burgelab/rescue-ese/

10 SNPs3D predicts deleterious missense mutations using methods described by Yue, P., Melamud, E., and Moult, J. in SNPs3D: candidate gene and SNP selection for association studies. BMC Bioinformatics, (2006), 7, 166. The SNPs3D database can be accessed at http://www.snps3d.org/modules.php?name=SNPtargets

11 TFSearch identifies transcription factor binding sites using methods described by Akiyama, Y. in TFSEARCH: Searching Transcription Factor Binding Sites (1998). The TFSearch database can be accessed at http://www.cbrc.jp/research/db/TFSEARCH.html

12 The MARIE-GENICA Consortium on Genetic Susceptibility for Menopausal Hormone Therapy Related Breast Cancer Risk Polymorphisms in the BRCA1 and ABCB1 genes modulate menopausal hormone therapy associated breast cancer risk in postmenopausal women, Breast Cancer Res Treat (2010) 120:727–736

Table 2.15 Assessment of potential functionality of single nucleotide polymorphisms (SNPs) on candidate gene *AURKA* genotyped in the CBCS

SNP	Status	FS Score[1]	Polyphen prediction[2]	SIFT prediction[3]	Description
rs1047972	Probably not functional	0.273	Benign	Tolerated	Predicted to impact splicing regulation by ESEfinder[5] and ESRSearch[6].
rs34987347	Possibly functional	0.5	N/A	N/A	Predicted to impact splicing regulation by ESEfinder[5] and ESRSearch[6]; predicted to impact pos-translation by OGPET[7]. Not significantly researched in the literature.
rs1468056	Not functional	0	N/A	N/A	
rs16979826	Probably not functional	0.208	N/A	N/A	Predicted to impact transcriptional regulation by TFSearch[11].
rs16979829	Probably not functional	0.208	N/A	N/A	Predicted to impact transcriptional regulation by TFSearch[11].
rs16979865	Not functional	0.05	N/A	N/A	
rs2064863	Probably not functional	0.176	N/A	N/A	Predicted to impact transcriptional regulation by TFSearch[11].
rs2180691	Not functional	0.05	N/A	N/A	
rs2236207	Probably not functional	0.109	N/A	N/A	
rs2273535	Possibly functional	0.5	Benign	Tolerated	Predicted to impact splicing regulation by ESEfinder[5], ESRSearch[6], PESX[8] and RESCUE_EXE[9]. Has been associated with risk of breast cancer in several studies.
rs2298016	Probably not functional	0.176	N/A	N/A	Predicted to impact transcriptional regulation by TFSearch[11]. Has been associated with a decreased odds of breast cancer in a case control study of *AURKA* in a Han Chinese population (OR=0.38, 95%CI=0.18-0.82) PMID= 21598251
rs6014711	Not functional	0.05	N/A	N/A	
rs6014712	Probably not functional	0.208	N/A	N/A	Predicted to impact transcriptional regulation by TFSearch[11].
rs6024840	Probably not functional	0.158	N/A	N/A	Predicted to impact transcriptional regulation by TFSearch[11].

rs6092309	Probably not functional	0.176	N/A	N/A	Predicted to impact transcriptional regulation by TFSearch[11].
rs6099120	Probably not functional	0.208	N/A	N/A	Predicted to impact transcriptional regulation by TFSearch[11].
rs6099122	Probably not functional	0.158	N/A	N/A	Predicted to impact transcriptional regulation by TFSearch[11].
rs6099126	Probably not functional	0.208	N/A	N/A	Predicted to impact transcriptional regulation by TFSearch[11].
rs6099127	Not functional	0.05	N/A	N/A	
rs6099128	Not functional	0.05	N/A	N/A	
rs1468055	Not functional	0	N/A	N/A	
rs6024836	Probably not functional	0.208	N/A	N/A	Predicted to impact transcriptional regulation by TFSearch[11]. Has been associated with an increased odds of breast cancer in a case control study of *AURKA* in a Han Chinese population (OR=1.54, 95%CI=1.18-2.00) PMID= 21598251
rs33923703	Probably not functional	0.103	N/A	Damaging	Predicted to be deleterious by SIFT[3]; predicted to impact splicing regulation by ESRSearch[6].
rs6099119	Possibly functional	0.39	N/A	N/A	Predicted to impact transcriptional regulation by TFSearch[11]. Ensembl-NS predicts frameshift coding.
rs911162	Probably not functional	0.176	N/A	N/A	Predicted to impact transcriptional regulation by TFSearch[11].

1 F-SNP database and algorithms are the work of Phil H. Lee and Hagit Shatkay of Queen's University; the database provides information about the functional effects of SNPs by integrating results from 16 bioinformatics tools and databases. SNPs are evaluated on the basis of four integral functions: protein coding, splicing regulation, transcriptional regulation and post-translation. SNPs then receive an FS Score which ranges from 0-1, with scores above 0.5 indicating high probability of functionality. A detailed explanation of how the algorithm scores SNPs can be found at http://compbio.cs.queensu.ca/F-SNP/

2 PolyPhen is a website and method for predicting SNP variant impact on protein structure and function developed by Ramensky V., Bork P., and Sunyaev S. and described in Human non-synonymous SNPs: server and survey. Nucleic Acids Res. (2002) 30(17): 3894-900 (PMID: 12202775)

3 SIFT predicts whether an amino acid substitution affects protein function and is based on the conservation of amino acid residues in the genome. SIFT was developed by Ng, P. and Henikoff, S. and described in Predicting deleterious amino acid substitutions. Genome Research, (2001) 11, 863-874. The SIFT database can be accessed at http://blocks.fhcrc.org/sift/SIFT.html

4 SNPeffect predicts deleterious missense SNPs using methods described by Reumers, J., Schymkowitz, J., Ferkinghoff-Borg, J., Stricher, F., Serrano, L., and Rousseau, F. in SNPeffect: a database mapping molecular phenotypic effects of human non-synonymous coding SNPs. Nucleic Acid Research, (2005) 33 (Database issue), D527–532. The SNPeffect database can be accessed at http://snpeffect.switchlab.org/index.php

5 ESEfinder identifies exonic splice sites using methods described by Cartegni, L., Wang, J., Zhu, Z., Zhang, M. Q., and Krainer, A. R. in ESEfinder: A web resource to identify exonic splicing enhancers. Nucleic Acids Research, (2003), 31(13), 3568–3571. The ESEfinder database can be accessed at http://rulai.cshl.edu/cgi-bin/tools/ESE3/esefinder.cgi

6 ESRSearch identifies exonic splice sites using methods described by Fairbrother, W. G., Yeh, R. F., Sharp, P. A., and Burge, C. B. in Predictive identification of exonic splicing enhancers in human genes.. Science, (2002), 297, 1007–1013. The ESRSearch database can be accessed at

http://www.tau.ac.il/lifesci/

7 OGPET identifies O-glycosylation sites, which are associated with post-translational functionality. Methods employed by OGPET are described by Gerken, T., Tep, C., and Rarick, J. in The role of peptide sequence and neighboring residue glycosylation on the substrate specificity of the uridine 5 diphosphate-alpha-n-acetylgalactosamine:polypeptide n-acetylgalactosaminyl transferases t1 and t2: kinetic modeling of the porcine and canine submaxillary gland mucin tandem repeats. (2004) Biochemistry, 43, 9888–9900. The OGPET database may be accessed at http://ogpet.utep.edu/main.php

8 PESX identifies exonic splice sites using methods described by Zhang et al. in Exon inclusion is dependent on predictable exonic splicing enhancers. Molecular and Cellular Biology, (2005), 25(16), 7323–7332. The PESX database may be accessed at http://cubweb.biology.columbia.edu/pesx/

9 RESCUE_EXE identifies exonic splice sites using methods described by Yeo, G. and Burge, C. B. in Variation in sequence and organization of splicing regulatory elements in vertebrate genes. In the Proceeding of Proc. Natl. Acad. Sci., (2004), 101(44), 15700–15705. 5. The RESCUE_EXE database can be accessed at http://genes.mit.edu/burgelab/rescue-ese/

10 SNPs3D predicts deleterious missense mutations using methods described by Yue, P., Melamud, E., and Moult, J. in SNPs3D: candidate gene and SNP selection for association studies. BMC Bioinformatics, (2006), 7, 166. The SNPs3D database can be accessed at http://www.snps3d.org/modules.php?name=SNPtargets

11 TFSearch identifies transcription factor binding sites using methods described by Akiyama, Y. in TFSEARCH: Searching Transcription Factor Binding Sites (1998). The TFSearch database can be accessed at http://www.cbrc.jp/research/db/TFSEARCH.html

12 The MARIE-GENICA Consortium on Genetic Susceptibility for Menopausal Hormone Therapy Related Breast Cancer Risk Polymorphisms in the BRCA1 and ABCB1 genes modulate menopausal hormone therapy associated breast cancer risk in postmenopausal women, Breast Cancer Res Treat (2010) 120:727–736

Table 2.16 Race-specific allele and genotype frequencies for *AURKA, BRCA1,* and *BRCA1*-interacting genes genotyped in CBCS participants enrolled 1993-2001.

Gene	dbSNP rs	SNP	Race	Allele	Allele Count and Frequency				Genotype	Genotype Count and Frequency				HWE[a]
					Cases		Controls			Cases		Controls		P-value
					Count	AF	Count	AF		Count	GF	Count	GF	
AURKA	rs1047972	G/A	African American	G	1269	0.86	1121	0.85	GG	539	0.73	475	0.72	0.53
				A	215	0.14	195	0.15	GA	191	0.26	171	0.26	
									AA	12	0.02	12	0.02	
			Caucasian	G	2022	0.84	1810	0.83	GG	838	0.70	757	0.70	0.28
				A	386	0.16	368	0.17	GA	346	0.29	296	0.27	
									AA	20	0.02	36	0.03	
	rs34987347	C/T	African American	C	1477	1.00	1308	1.00	CC	735	0.99	651	0.99	1.00
				T	7	0.00	6	0.00	CT	7	0.01	6	0.01	
									TT	0	0.00	0	0.00	
			Caucasian	C	2408	1.00	2178	1.00	CC	1204	1.00	1089	1.00	--
				T	0	0.00	0	0.00	CT	0	0.00	0	0.00	
									TT	0	0.00	0	0.00	
	rs1468056	G/C	African American	G	526	0.35	453	0.34	GG	88	0.12	80	0.12	0.72
				C	958	0.65	863	0.66	GC	350	0.47	293	0.45	
									CC	304	0.41	285	0.43	
			Caucasian	G	1621	0.67	1494	0.69	GG	544	0.45	515	0.47	0.72
				C	787	0.33	684	0.31	GC	533	0.44	464	0.43	
									CC	127	0.11	110	0.10	
	rs16979826	T/C	African American	T	1331	0.90	1180	0.90	TT	596	0.80	528	0.80	0.53
				C	153	0.10	134	0.10	TC	139	0.19	124	0.19	
									CC	7	0.01	5	0.01	
			Caucasian	T	2404	1.00	2175	1.00	TT	1200	1.00	1086	1.00	1.00
				C	4	0.00	3	0.00	TC	4	0.00	3	0.00	
									CC	0	0.00	0	0.00	
	rs16979829	T/G	African American	T	1236	0.83	1096	0.83	TT	515	0.69	451	0.69	0.15
				G	248	0.17	220	0.17	TG	206	0.28	194	0.29	
									GG	21	0.03	13	0.02	
			Caucasian	T	2325	0.97	2112	0.97	TT	1123	0.93	1025	0.94	0.26
				G	83	0.03	66	0.03	TG	79	0.07	62	0.06	
									GG	2	0.00	2	0.00	
	rs16979865	A/C	African American	A	1351	0.91	1182	0.90	AA	612	0.82	531	0.81	1.00
				C	133	0.09	132	0.10	AC	127	0.17	120	0.18	
									CC	3	0.00	6	0.01	
			Caucasian	A	2222	0.92	2027	0.93	AA	1031	0.86	941	0.86	0.47
				C	184	0.08	151	0.07	AC	160	0.13	145	0.13	
									CC	12	0.01	3	0.00	
	rs2180691	G/A	African American	G	359	0.24	304	0.23	GG	42	0.06	37	0.06	0.67
				A	1121	0.76	1008	0.77	GA	275	0.37	230	0.35	
									AA	423	0.57	389	0.59	
			Caucasian	G	1743	0.72	1594	0.73	GG	641	0.53	590	0.54	0.31
				A	663	0.28	584	0.27	GA	461	0.38	414	0.38	
									AA	101	0.08	85	0.08	
	rs2273535	A/T	African American	A	1242	0.84	1113	0.85	AA	517	0.70	477	0.72	0.07
				T	240	0.16	203	0.15	AT	208	0.28	159	0.24	
									TT	16	0.02	22	0.03	
			Caucasian	A	1889	0.78	1701	0.78	AA	740	0.61	673	0.62	0.13
				T	519	0.22	477	0.22	AT	409	0.34	355	0.33	
									TT	55	0.05	61	0.06	
	rs2298016	G/C	African American	G	1129	0.76	1011	0.77	GG	433	0.58	393	0.60	0.31
				C	355	0.24	305	0.23	GC	263	0.35	225	0.34	
									CC	46	0.06	40	0.06	
			Caucasian	G	1873	0.78	1693	0.78	GG	729	0.61	665	0.61	0.22
				C	535	0.22	485	0.22	GC	415	0.34	363	0.33	
									CC	60	0.05	61	0.06	
	rs6014711	G/A	African American	G	1270	0.86	1121	0.85	GG	540	0.73	475	0.72	0.53
				A	214	0.14	195	0.15	GA	190	0.26	171	0.26	
									AA	12	0.02	12	0.02	
			Caucasian	G	2021	0.84	1806	0.83	GG	838	0.70	753	0.69	0.39
				A	387	0.16	372	0.17	GA	345	0.29	300	0.28	
									AA	21	0.02	36	0.03	

Table 2.16 (cont.) Race-specific allele and genotype frequencies for *AURKA, BRCA1,* and *BRCA1*-interacting genes genotyped in CBCS participants enrolled 1993-2001.

Gene	dbSNP rs	SNP	Race	Allele	Allele Count and Frequency				Genotype	Genotype Count and Frequency				HWE[a]
					Cases		Controls			Cases		Controls		P-value
					Count	AF	Count	AF		Count	GF	Count	GF	
AURKA	rs6014712	C/G	African American	C	1484	1.00	1316	1.00	CC	742	1.00	658	1.00	--
				G	0	0.00	0	0.00	CG	0	0.00	0	0.00	
									GG	0	0.00	0	0.00	
			Caucasian	C	2408	1.00	2178	1.00	CC	1204	1.00	1089	1.00	--
				G	0	0.00	0	0.00	CG	0	0.00	0	0.00	
									GG	0	0.00	0	0.00	
	rs6024840	A/G	African American	A	741	0.50	628	0.48	AA	183	0.25	143	0.22	0.31
				G	743	0.50	688	0.52	AG	375	0.51	342	0.52	
									GG	184	0.25	173	0.26	
			Caucasian	A	1808	0.75	1648	0.76	AA	684	0.57	628	0.58	0.45
				G	600	0.25	530	0.24	AG	440	0.37	392	0.36	
									GG	80	0.07	69	0.06	
	rs6092309	G/A	African American	G	1340	0.90	1141	0.87	GG	605	0.82	494	0.75	1.00
				A	144	0.10	175	0.13	GA	130	0.18	153	0.23	
									AA	7	0.01	11	0.02	
			Caucasian	G	2403	1.00	2174	1.00	GG	1199	1.00	1085	1.00	1.00
				A	5	0.00	4	0.00	GA	5	0.00	4	0.00	
									AA	0	0.00	0	0.00	
	rs6099120	C/T	African American	C	1484	1.00	1316	1.00	CC	742	1.00	658	1.00	--
				T	0	0.00	0	0.00	CT	0	0.00	0	0.00	
									TT	0	0.00	0	0.00	
			Caucasian	C	2408	1.00	2178	1.00	CC	1204	1.00	1089	1.00	--
				T	0	0.00	0	0.00	CT	0	0.00	0	0.00	
									TT	0	0.00	0	0.00	
	rs6099122	T/G	African American	T	1038	0.70	891	0.68	TT	368	0.50	302	0.46	0.93
				G	446	0.30	425	0.32	TG	302	0.41	287	0.44	
									GG	72	0.10	69	0.10	
			Caucasian	T	2344	0.97	2130	0.98	TT	1142	0.95	1041	0.96	1.00
				G	64	0.03	48	0.02	TG	60	0.05	48	0.04	
									GG	2	0.00	0	0.00	
	rs6099126	C/T	African American	C	902	0.61	779	0.59	CC	280	0.38	235	0.36	0.46
				T	582	0.39	537	0.41	CT	342	0.46	309	0.47	
									TT	120	0.16	114	0.17	
			Caucasian	C	2340	0.97	2124	0.98	CC	1137	0.94	1035	0.95	1.00
				T	68	0.03	54	0.02	CT	66	0.05	54	0.05	
									TT	1	0.00	0	0.00	
	rs6099128	T/G	African American	T	1248	0.84	1079	0.82	TT	523	0.70	442	0.67	1.00
				G	236	0.16	237	0.18	TG	202	0.27	195	0.30	
									GG	17	0.02	21	0.03	
			Caucasian	T	2174	0.90	1942	0.89	TT	989	0.82	871	0.80	0.12
				G	232	0.10	236	0.11	TG	196	0.16	200	0.18	
									GG	18	0.01	18	0.02	
	rs1468055	C/A	African American	C	1428	0.96	1260	0.96	CC	687	0.93	602	0.91	0.62
				A	56	0.04	56	0.04	CA	54	0.07	56	0.09	
									AA	1	0.00	0	0.00	
			Caucasian	C	1885	0.78	1714	0.79	CC	730	0.61	674	0.62	1.00
				A	523	0.22	464	0.21	CA	425	0.35	366	0.34	
									AA	49	0.04	49	0.04	
	rs6024836	G/A	African American	G	493	0.33	437	0.33	GG	80	0.11	68	0.10	0.48
				A	991	0.67	879	0.67	GA	333	0.45	301	0.46	
									AA	329	0.44	289	0.44	
			Caucasian	G	1783	0.74	1620	0.74	GG	660	0.55	611	0.56	0.18
				A	625	0.26	558	0.26	GA	463	0.38	398	0.37	
									AA	81	0.07	80	0.07	
	rs2064863	A/C	African American	C	321	0.22	255	0.19	CC	27	0.04	23	0.03	0.80
				A	1163	0.78	1061	0.81	CA	267	0.36	209	0.32	
									AA	448	0.60	426	0.65	
			Caucasian	C	1418	0.59	1272	0.58	CC	419	0.35	376	0.35	0.56
				A	988	0.41	906	0.42	CA	580	0.48	520	0.48	
									AA	204	0.17	193	0.18	

Table 2.16 (cont.) Race-specific allele and genotype frequencies for *AURKA, BRCA1,* and *BRCA1*-interacting genes genotyped in CBCS participants enrolled 1993-2001.

Gene	dbSNP rs	SNP	Race	Allele	Allele Count and Frequency				Genotype	Genotype Count and Frequency				HWE[a]
					Cases		Controls			Cases		Controls		P-value
					Count	AF	Count	AF		Count	GF	Count	GF	
AURKA	rs6099119	A/G	African American	A	1408	0.95	1255	0.96	AA	668	0.90	599	0.91	0.63
				G	72	0.05	57	0.04	AG	72	0.10	57	0.09	
									GG	0	0.00	0	0.00	
			Caucasian	A	2388	1.00	2158	1.00	AA	1193	1.00	1078	1.00	1.00
				G	2	0.00	2	0.00	AG	2	0.00	2	0.00	
									GG	0	0.00	0	0.00	
	rs911162	G/A	African American	G	1409	0.96	1258	0.96	GG	673	0.91	606	0.93	0.24
				A	65	0.04	50	0.04	GA	63	0.09	46	0.07	
									AA	1	0.00	2	0.00	
			Caucasian	G	2373	0.99	2148	0.99	GG	1178	0.99	1065	0.98	1.00
				A	17	0.01	18	0.01	GA	17	0.01	18	0.02	
									AA	0	0.00	0	0.00	
	rs2236207	G/A	African American	G	1105	0.74	962	0.73	GG	407	0.55	339	0.52	0.01
				A	379	0.26	354	0.27	GA	291	0.39	284	0.43	
									AA	44	0.06	35	0.05	
			Caucasian	G	2376	0.99	2150	0.99	GG	1172	0.97	1061	0.97	1.00
				A	32	0.01	28	0.01	GA	32	0.03	28	0.03	
									AA	0	0.00	0	0.00	
	rs6099127	C/T	African American	C	736	0.50	629	0.48	CC	187	0.25	159	0.24	0.24
				T	748	0.50	679	0.52	CT	362	0.49	311	0.48	
									TT	193	0.26	184	0.28	
			Caucasian	C	2249	0.95	2048	0.94	CC	1060	0.89	971	0.89	0.005
				T	129	0.05	124	0.06	CT	129	0.11	106	0.10	
									TT	0	0.00	9	0.01	
	rs33923703	T/C	African American	T	1472	1.00	1309	1.00	TT	733	0.99	653	0.99	0.01
				C	6	0.00	5	0.00	TC	6	0.01	3	0.00	
									CC	0	0.00	1	0.00	
			Caucasian	T	2345	0.98	2114	0.98	TT	1147	0.96	1032	0.95	0.14
				C	55	0.02	54	0.02	TC	51	0.04	50	0.05	
									CC	2	0.00	2	0.00	
BARD1	rs10932568	A/C	African American	A	1211	0.82	1063	0.81	AA	487	0.66	430	0.65	0.90
				C	273	0.18	253	0.19	AC	237	0.32	203	0.31	
									CC	18	0.02	25	0.04	
			Caucasian	A	1891	0.79	1663	0.76	AA	747	0.62	642	0.59	0.31
				C	517	0.21	511	0.24	AC	397	0.33	379	0.35	
									CC	60	0.05	66	0.06	
	rs10221582	C/T	African American	C	761	0.51	644	0.49	CC	201	0.27	157	0.24	1.00
				T	723	0.49	672	0.51	CT	359	0.48	330	0.50	
									TT	182	0.25	171	0.26	
			Caucasian	C	1421	0.59	1263	0.58	CC	420	0.35	372	0.34	0.50
				T	985	0.41	915	0.42	CT	581	0.48	519	0.48	
									TT	202	0.17	198	0.18	
	rs10932573	T/C	African American	T	908	0.61	838	0.64	TT	282	0.38	262	0.40	0.44
				C	576	0.39	478	0.36	TC	344	0.46	314	0.48	
									CC	116	0.16	82	0.12	
			Caucasian	T	1203	0.50	1107	0.51	TT	297	0.25	284	0.26	0.77
				C	1203	0.50	1069	0.49	TC	609	0.51	539	0.50	
									CC	297	0.25	265	0.24	
	rs12474696	A/G	African American	A	1047	0.71	926	0.70	AA	371	0.50	329	0.50	0.63
				G	437	0.29	388	0.30	AG	305	0.41	268	0.41	
									GG	66	0.09	60	0.09	
			Caucasian	A	1436	0.60	1320	0.61	AA	443	0.37	393	0.36	0.37
				G	968	0.40	856	0.39	AG	550	0.46	534	0.49	
									GG	209	0.17	161	0.15	
	rs12477063	C/T	African American	C	956	0.65	862	0.66	CC	306	0.41	287	0.44	0.44
				T	522	0.35	454	0.34	CT	344	0.47	288	0.44	
									TT	89	0.12	83	0.13	
			Caucasian	C	988	0.41	900	0.41	CC	195	0.16	183	0.17	0.66
				T	1414	0.59	1270	0.59	CT	598	0.50	534	0.49	
									TT	408	0.34	368	0.34	

Table 2.16 (cont.) Race-specific allele and genotype frequencies for *AURKA, BRCA1,* and *BRCA1*-interacting genes genotyped in CBCS participants enrolled 1993-2001.

Gene	dbSNP rs	SNP	Race	Allele	Allele Count and Frequency				Genotype	Genotype Count and Frequency				HWE[a]
					Cases		Controls			Cases		Controls		P-value
					Count	AF	Count	AF		Count	GF	Count	GF	
BARD1	rs1542173	A/G	African American	A	998	0.67	912	0.69	AA	342	0.46	321	0.49	0.36
				G	486	0.33	404	0.31	AG	314	0.42	270	0.41	
									GG	86	0.12	67	0.10	
			Caucasian	A	1806	0.75	1692	0.78	AA	671	0.56	658	0.60	0.87
				G	602	0.25	486	0.22	AG	464	0.39	376	0.35	
									GG	69	0.06	55	0.05	
	rs16852761	G/A	African American	G	1341	0.90	1184	0.90	GG	604	0.81	534	0.81	0.52
				A	143	0.10	132	0.10	GA	133	0.18	116	0.18	
									AA	5	0.01	8	0.01	
			Caucasian	G	2372	0.99	2137	0.98	GG	1168	0.97	1050	0.96	0.05
				A	36	0.01	41	0.02	GA	36	0.03	37	0.03	
									AA	0	0.00	2	0.00	
	rs16852798	C/T	African American	C	1290	0.87	1148	0.87	CC	562	0.76	501	0.76	0.87
				T	194	0.13	168	0.13	CT	166	0.22	146	0.22	
									TT	14	0.02	11	0.02	
			Caucasian	C	2364	0.98	2129	0.98	CC	1161	0.96	1040	0.96	1.00
				T	44	0.02	49	0.02	CT	42	0.03	49	0.04	
									TT	1	0.00	0	0.00	
	rs16852799	A/G	African American	A	1295	0.87	1139	0.87	AA	568	0.77	490	0.74	0.40
				G	189	0.13	177	0.13	AG	159	0.21	159	0.24	
									GG	15	0.02	9	0.01	
			Caucasian	A	2262	0.94	2023	0.93	AA	1065	0.89	942	0.87	0.25
				G	144	0.06	155	0.07	AG	132	0.11	139	0.13	
									GG	6	0.00	8	0.01	
	rs17487827	C/G	African American	C	1250	0.84	1105	0.84	CC	527	0.71	464	0.71	1.00
				G	234	0.16	211	0.16	CG	196	0.26	177	0.27	
									GG	19	0.03	17	0.03	
			Caucasian	C	1852	0.77	1746	0.80	CC	708	0.59	697	0.64	0.63
				G	556	0.23	432	0.20	CG	436	0.36	352	0.32	
									GG	60	0.05	40	0.04	
	rs1979028	T/A	African American	T	919	0.62	836	0.64	TT	282	0.38	261	0.40	0.50
				A	563	0.38	480	0.36	TA	355	0.48	314	0.48	
									AA	104	0.14	83	0.13	
			Caucasian	T	1691	0.70	1532	0.70	TT	604	0.50	549	0.51	0.19
				A	715	0.30	642	0.30	TA	483	0.40	434	0.40	
									AA	116	0.10	104	0.10	
	rs2075622	C/A	African American	C	1140	0.77	1011	0.77	CC	445	0.60	392	0.60	0.51
				A	344	0.23	303	0.23	CA	250	0.34	227	0.35	
									AA	47	0.06	38	0.06	
			Caucasian	C	1442	0.60	1348	0.62	CC	442	0.37	414	0.38	0.71
				A	964	0.40	830	0.38	CA	558	0.46	520	0.48	
									AA	203	0.17	155	0.14	
	rs2888294	C/G	African American	C	734	0.50	641	0.49	CC	177	0.24	152	0.23	0.52
				G	744	0.50	671	0.51	CG	380	0.51	337	0.51	
									GG	182	0.25	167	0.25	
			Caucasian	C	1225	0.51	1132	0.52	CC	322	0.27	290	0.27	0.63
				G	1183	0.49	1046	0.48	CG	581	0.48	552	0.51	
									GG	301	0.25	247	0.23	
	rs3768704	G/A	African American	G	1282	0.86	1135	0.86	GG	552	0.74	490	0.75	1.00
				A	202	0.14	179	0.14	GA	178	0.24	155	0.24	
									AA	12	0.02	12	0.02	
			Caucasian	G	2092	0.87	1900	0.87	GG	902	0.75	829	0.76	0.89
				A	316	0.13	278	0.13	GA	288	0.24	242	0.22	
									AA	14	0.01	18	0.02	
	rs3768707	C/T	African American	C	991	0.67	906	0.69	CC	337	0.45	314	0.48	0.72
				T	493	0.33	410	0.31	CT	317	0.43	278	0.42	
									TT	88	0.12	66	0.10	
			Caucasian	C	1807	0.75	1690	0.78	CC	672	0.56	656	0.60	0.93
				T	601	0.25	488	0.22	CT	463	0.38	378	0.35	
									TT	69	0.06	55	0.05	

Table 2.16 (cont.) Race-specific allele and genotype frequencies for *AURKA, BRCA1,* and *BRCA1*-interacting genes genotyped in CBCS participants enrolled 1993-2001.

Gene	dbSNP rs	SNP	Race	Allele	Allele Count and Frequency				Genotype	Genotype Count and Frequency				HWE[a]
					Cases		Controls			Cases		Controls		P-value
					Count	AF	Count	AF		Count	GF	Count	GF	
BARD1	rs3768708	A/G	African American	A	885	0.60	802	0.61	AA	263	0.35	238	0.36	0.32
				G	597	0.40	514	0.39	AG	359	0.48	326	0.50	
									GG	119	0.16	94	0.14	
			Caucasian	A	1801	0.75	1686	0.77	AA	667	0.55	653	0.60	0.93
				G	607	0.25	492	0.23	AG	467	0.39	380	0.35	
									GG	70	0.06	56	0.05	
	rs3820727	T/G	African American	T	1180	0.80	1036	0.79	TT	477	0.64	413	0.63	0.24
				G	304	0.20	280	0.21	TG	226	0.30	210	0.32	
									GG	39	0.05	35	0.05	
			Caucasian	T	1321	0.55	1197	0.55	TT	356	0.30	334	0.31	0.58
				G	1087	0.45	979	0.45	TG	609	0.51	529	0.49	
									GG	239	0.20	225	0.21	
	rs4672729	G/A	African American	G	939	0.63	860	0.65	GG	307	0.41	282	0.43	0.86
				A	543	0.37	456	0.35	GA	325	0.44	296	0.45	
									AA	109	0.15	80	0.12	
			Caucasian	G	1609	0.67	1504	0.69	GG	539	0.45	519	0.48	1.00
				A	797	0.33	674	0.31	GA	531	0.44	466	0.43	
									AA	133	0.11	104	0.10	
	rs6706777	G/C	African American	G	873	0.59	771	0.59	GG	268	0.36	213	0.32	0.04
				C	611	0.41	545	0.41	GC	337	0.45	345	0.52	
									CC	137	0.18	100	0.15	
			Caucasian	G	1404	0.58	1245	0.57	GG	412	0.34	360	0.33	0.62
				C	1004	0.42	933	0.43	GC	580	0.48	525	0.48	
									CC	212	0.18	204	0.19	
	rs6712055	T/C	African American	T	1220	0.82	1080	0.82	TT	500	0.67	445	0.68	0.60
				C	264	0.18	236	0.18	TC	220	0.30	190	0.29	
									CC	22	0.03	23	0.03	
			Caucasian	T	1677	0.70	1568	0.72	TT	585	0.49	550	0.51	0.03
				C	731	0.30	610	0.28	TC	507	0.42	468	0.43	
									CC	112	0.09	71	0.07	
	rs6749828	C/G	African American	C	899	0.61	797	0.61	CC	280	0.38	255	0.39	0.10
				G	571	0.39	501	0.39	CG	339	0.46	287	0.44	
									GG	116	0.16	107	0.16	
			Caucasian	C	1071	0.45	929	0.43	CC	237	0.20	199	0.18	0.90
				G	1335	0.55	1249	0.57	CG	597	0.50	531	0.49	
									GG	369	0.31	359	0.33	
	rs6751923	T/C	African American	T	986	0.66	894	0.68	TT	324	0.44	308	0.47	0.47
				C	498	0.34	422	0.32	TC	338	0.46	278	0.42	
									CC	80	0.11	72	0.11	
			Caucasian	T	2088	0.87	1842	0.85	TT	898	0.75	784	0.72	0.25
				C	320	0.13	336	0.15	TC	292	0.24	274	0.25	
									CC	14	0.01	31	0.03	
	rs6753417	C/T	African American	C	1164	0.78	1030	0.78	CC	455	0.61	407	0.62	0.36
				T	320	0.22	286	0.22	CT	254	0.34	216	0.33	
									TT	33	0.04	35	0.05	
			Caucasian	C	1564	0.65	1432	0.66	CC	508	0.42	468	0.43	0.69
				T	842	0.35	744	0.34	CT	548	0.46	496	0.46	
									TT	147	0.12	124	0.11	
	rs6756902	C/T	African American	C	960	0.65	865	0.66	CC	306	0.41	288	0.44	0.54
				T	524	0.35	451	0.34	CT	348	0.47	289	0.44	
									TT	88	0.12	81	0.12	
			Caucasian	C	986	0.41	899	0.41	CC	194	0.16	183	0.17	0.80
				T	1420	0.59	1279	0.59	CT	598	0.50	533	0.49	
									TT	411	0.34	373	0.34	
	rs7557557	C/T	African American	C	990	0.67	904	0.69	CC	336	0.45	308	0.47	0.71
				T	492	0.33	412	0.31	CT	318	0.43	288	0.44	
									TT	87	0.12	62	0.09	
			Caucasian	C	1205	0.50	1111	0.51	CC	300	0.25	286	0.26	0.76
				T	1201	0.50	1067	0.49	CT	605	0.50	539	0.49	
									TT	298	0.25	264	0.24	

Table 2.16 (cont.) Race-specific allele and genotype frequencies for *AURKA, BRCA1,* and *BRCA1*-interacting genes genotyped in CBCS participants enrolled 1993-2001.

Gene	dbSNP rs	SNP	Race	Allele	Allele Count and Frequency				Genotype	Genotype Count and Frequency				HWE[a]
					Cases		Controls			Cases		Controls		P-value
					Count	AF	Count	AF		Count	GF	Count	GF	
BARD1	rs7566806	G/C	African American	G	1240	0.84	1094	0.83	GG	513	0.69	458	0.70	0.40
				C	244	0.16	222	0.17	GC	214	0.29	178	0.27	
									CC	15	0.02	22	0.03	
			Caucasian	G	2039	0.85	1818	0.83	GG	857	0.71	763	0.70	0.37
				C	369	0.15	360	0.17	GC	325	0.27	292	0.27	
									CC	22	0.02	34	0.03	
	rs7585356	G/A	African American	G	1231	0.83	1081	0.82	GG	515	0.69	446	0.68	0.59
				A	253	0.17	235	0.18	GA	201	0.27	189	0.29	
									AA	26	0.04	23	0.03	
			Caucasian	G	1699	0.71	1542	0.71	GG	610	0.51	556	0.51	0.16
				A	709	0.29	634	0.29	GA	479	0.40	430	0.40	
									AA	115	0.10	102	0.09	
	rs1048108	G/A	African American	G	1151	0.78	1017	0.78	GG	450	0.61	401	0.61	0.26
				A	323	0.22	289	0.22	GA	251	0.34	215	0.33	
									AA	36	0.05	37	0.06	
			Caucasian	G	1543	0.65	1422	0.66	GG	503	0.42	462	0.43	0.51
				A	827	0.35	740	0.34	GA	537	0.45	498	0.46	
									AA	145	0.12	121	0.11	
	rs3738888	G/A	African American	G	1471	1.00	1310	1.00	GG	734	1.00	655	1.00	--
				A	3	0.00	0	0.00	GA	3	0.00	0	0.00	
									AA	0	0.00	0	0.00	
			Caucasian	G	2372	0.99	2152	0.99	GG	1176	0.98	1069	0.99	0.06
				A	20	0.01	16	0.01	GA	20	0.02	14	0.01	
									AA	0	0.00	1	0.00	
	rs28997576	G/C	African American	G	1476	0.99	1313	1.00	GG	734	0.99	655	1.00	1.00
				C	8	0.01	3	0.00	GC	8	0.01	3	0.00	
									CC	0	0.00	0	0.00	
			Caucasian	G	2315	0.96	2118	0.97	GG	1114	0.93	1029	0.94	1.00
				C	93	0.04	60	0.03	GC	87	0.07	60	0.06	
									CC	3	0.00	0	0.00	
	rs2229571	C/G	African American	C	611	0.41	540	0.41	CC	131	0.18	111	0.17	1.00
				G	871	0.59	776	0.59	CG	349	0.47	318	0.48	
									GG	261	0.35	229	0.35	
			Caucasian	C	1426	0.59	1301	0.60	CC	424	0.35	391	0.36	0.85
				G	982	0.41	873	0.40	CG	578	0.48	519	0.48	
									GG	202	0.17	177	0.16	
BRCA1	rs4986850	G/A	African American	G	1449	0.98	1277	0.97	GG	707	0.95	619	0.94	1.00
				A	35	0.02	39	0.03	GA	35	0.05	39	0.06	
									AA	0	0.00	0	0.00	
			Caucasian	G	2226	0.92	2005	0.92	GG	1029	0.85	929	0.85	0.02
				A	182	0.08	173	0.08	GA	168	0.14	147	0.13	
									AA	7	0.01	13	0.01	
	rs1799950	A/G	African American	A	1464	0.99	1297	0.99	AA	722	0.97	639	0.97	1.00
				G	20	0.01	19	0.01	AG	20	0.03	19	0.03	
									GG	0	0.00	0	0.00	
			Caucasian	A	2272	0.94	2046	0.94	AA	1072	0.89	963	0.88	0.28
				G	136	0.06	132	0.06	AG	128	0.11	120	0.11	
									GG	4	0.00	6	0.01	
	rs16941	A/G	African American	A	1214	0.82	1072	0.81	AA	491	0.66	439	0.67	0.52
				G	270	0.18	244	0.19	AG	232	0.31	194	0.29	
									GG	19	0.03	25	0.04	
			Caucasian	A	1626	0.68	1501	0.69	AA	541	0.45	529	0.49	0.11
				G	782	0.32	677	0.31	AG	544	0.45	443	0.41	
									GG	119	0.10	117	0.11	
	rs16942	A/G	African American	A	1131	0.76	1012	0.77	AA	427	0.58	390	0.59	0.82
				G	351	0.24	304	0.23	AG	277	0.37	232	0.35	
									GG	37	0.05	36	0.05	
			Caucasian	A	1625	0.67	1496	0.69	AA	540	0.45	526	0.48	0.10
				G	783	0.33	680	0.31	AG	545	0.45	444	0.41	
									GG	119	0.10	118	0.11	

Table 2.16 (cont.) Race-specific allele and genotype frequencies for *AURKA, BRCA1,* and *BRCA1*-interacting genes genotyped in CBCS participants enrolled 1993-2001.

Gene	dbSNP rs	SNP	Race	Allele	Allele Count and Frequency				Genotype	Genotype Count and Frequency				HWE[a]
					Cases		Controls			Cases		Controls		P-value
					Count	AF	Count	AF		Count	GF	Count	GF	
BRCA1	rs1799966	A/G	African American	A	1120	0.76	1004	0.76	AA	419	0.57	385	0.59	0.75
				G	360	0.24	310	0.24	AG	282	0.38	234	0.36	
									GG	39	0.05	38	0.06	
			Caucasian	A	1622	0.67	1496	0.69	AA	538	0.45	526	0.48	0.10
				G	784	0.33	680	0.31	AG	546	0.45	444	0.41	
									GG	119	0.10	118	0.11	
	rs799917	C/T	African American	C	293	0.20	243	0.18	CC	34	0.05	23	0.04	0.89
				T	1187	0.80	1071	0.82	CT	225	0.30	197	0.30	
									TT	481	0.65	437	0.67	
			Caucasian	C	1590	0.66	1466	0.67	CC	517	0.43	506	0.47	0.10
				T	816	0.34	710	0.33	CT	556	0.46	454	0.42	
									TT	130	0.11	128	0.12	
	rs4986852	C/T	African American	C	1469	0.99	1303	0.99	CC	730	0.99	648	0.99	1.00
				T	9	0.01	7	0.01	CT	9	0.01	7	0.01	
									TT	0	0.00	0	0.00	
			Caucasian	C	2341	0.98	2110	0.97	CC	1141	0.95	1028	0.95	0.17
				T	59	0.02	58	0.03	CT	59	0.05	54	0.05	
									TT	0	0.00	2	0.00	
	rs3737559	G/A	African American	G	1430	0.96	1274	0.97	GG	688	0.93	616	0.94	1.00
				A	54	0.04	42	0.03	GA	54	0.07	42	0.06	
									AA	0	0.00	0	0.00	
			Caucasian	G	2203	0.91	2015	0.93	GG	1012	0.84	938	0.86	0.02
				A	205	0.09	163	0.07	GA	179	0.15	139	0.13	
									AA	13	0.01	12	0.01	
	rs799923	G/A	African American	G	1420	0.96	1259	0.96	GG	684	0.92	601	0.91	0.63
				A	64	0.04	57	0.04	GA	52	0.07	57	0.09	
									AA	6	0.01	0	0.00	
			Caucasian	G	1809	0.75	1655	0.76	GG	677	0.56	629	0.58	1.00
				A	597	0.25	523	0.24	GA	455	0.38	397	0.36	
									AA	71	0.06	63	0.06	
BRIP1	rs4986764	C/T	African American	C	994	0.67	861	0.65	CC	333	0.45	274	0.42	0.20
				T	490	0.33	455	0.35	CT	328	0.44	313	0.48	
									TT	81	0.11	71	0.11	
			Caucasian	C	1442	0.60	1320	0.61	CC	438	0.36	406	0.37	0.44
				T	966	0.40	858	0.39	CT	566	0.47	508	0.47	
									TT	200	0.17	175	0.16	
	rs7213430	A/G	African American	A	992	0.67	857	0.65	AA	331	0.45	271	0.41	0.20
				G	492	0.33	459	0.35	AG	330	0.44	315	0.48	
									GG	81	0.11	72	0.11	
			Caucasian	A	1437	0.60	1317	0.60	AA	433	0.36	404	0.37	0.49
				G	971	0.40	861	0.40	AG	571	0.47	509	0.47	
									GG	200	0.17	176	0.16	
	rs4988350	T/G	African American	T	1484	1.00	1316	1.00	TT	742	1.00	658	1.00	--
				G	0	0.00	0	0.00	TG	0	0.00	0	0.00	
									GG	0	0.00	0	0.00	
			Caucasian	T	2408	1.00	2178	1.00	TT	1204	1.00	1089	1.00	--
				G	0	0.00	0	0.00	TG	0	0.00	0	0.00	
									GG	0	0.00	0	0.00	
	rs4988346	G/A	African American	G	1484	1.00	1315	1.00	GG	742	1.00	657	1.00	1.00
				A	0	0.00	1	0.00	GA	0	0.00	1	0.00	
									AA	0	0.00	0	0.00	
			Caucasian	G	2391	0.99	2169	1.00	GG	1187	0.99	1081	0.99	0.01
				A	17	0.01	9	0.00	GA	17	0.01	7	0.01	
									AA	0	0.00	1	0.00	
	rs4988351	G/C	African American	G	1254	0.85	1126	0.86	GG	531	0.72	485	0.74	0.87
				C	224	0.15	182	0.14	GC	192	0.26	156	0.24	
									CC	16	0.02	13	0.02	
			Caucasian	G	1745	0.73	1622	0.75	GG	637	0.53	611	0.56	0.57
				C	649	0.27	544	0.25	GC	471	0.39	400	0.37	
									CC	89	0.07	72	0.07	

Table 2.16 (cont.) Race-specific allele and genotype frequencies for *AURKA, BRCA1,* and *BRCA1*-interacting genes genotyped in CBCS participants enrolled 1993-2001.

Gene	dbSNP rs	SNP	Race	Allele	Allele Count and Frequency				Genotype	Genotype Count and Frequency				HWE[a]
					Cases		Controls			Cases		Controls		P-value
					Count	AF	Count	AF		Count	GF	Count	GF	
BRIP1	rs2048718	C/T	African American	C	1143	0.78	996	0.78	CC	446	0.61	399	0.62	0.01
				T	325	0.22	286	0.22	CT	251	0.34	198	0.31	
									TT	37	0.05	44	0.07	
			Caucasian	C	1313	0.55	1232	0.57	CC	359	0.30	356	0.33	0.53
				T	1077	0.45	930	0.43	CT	595	0.50	520	0.48	
									TT	241	0.20	205	0.19	
	rs1978111	C/T	African American	C	961	0.66	812	0.64	CC	321	0.44	253	0.40	0.35
				T	505	0.34	462	0.36	CT	319	0.44	306	0.48	
									TT	93	0.13	78	0.12	
			Caucasian	C	1429	0.60	1309	0.60	CC	433	0.36	401	0.37	0.53
				T	959	0.40	855	0.40	CT	563	0.47	507	0.47	
									TT	198	0.17	174	0.16	
ZNF350	rs4986773	T/C	African American	T	530	0.36	508	0.39	TT	105	0.14	103	0.16	0.51
				C	924	0.64	794	0.61	TC	320	0.44	302	0.46	
									CC	302	0.42	246	0.38	
			Caucasian	T	1731	0.72	1603	0.74	TT	634	0.53	596	0.55	0.63
				C	661	0.28	563	0.26	TC	463	0.39	411	0.38	
									CC	99	0.08	76	0.07	
	rs2278420	A/G	African American	A	927	0.62	819	0.62	AA	290	0.39	249	0.38	0.36
				G	557	0.38	497	0.38	AG	347	0.47	321	0.49	
									GG	105	0.14	88	0.13	
			Caucasian	A	2019	0.84	1839	0.84	AA	851	0.71	784	0.72	0.08
				G	389	0.16	339	0.16	AG	317	0.26	271	0.25	
									GG	36	0.03	34	0.03	
	rs3764538	G/T	African American	G	1293	0.87	1138	0.87	GG	565	0.76	484	0.74	0.00
				T	187	0.13	174	0.13	GT	163	0.22	170	0.26	
									TT	12	0.02	2	0.00	
			Caucasian	G	2067	0.87	1894	0.87	GG	902	0.76	835	0.77	0.07
				T	321	0.13	272	0.13	GT	263	0.22	224	0.21	
									TT	29	0.02	24	0.02	
	rs4986771	T/C	African American	T	1475	0.99	1307	0.99	TT	734	0.99	649	0.99	1.00
				C	9	0.01	9	0.01	TC	7	0.01	9	0.01	
									CC	1	0.00	0	0.00	
			Caucasian	T	2302	0.96	2101	0.97	TT	1103	0.92	1017	0.93	0.04
				C	106	0.04	75	0.03	TC	96	0.08	67	0.06	
									CC	5	0.00	4	0.00	
	rs2278415	T/A	African American	T	1298	0.87	1142	0.87	TT	566	0.76	486	0.74	0.00
				A	186	0.13	174	0.13	TA	166	0.22	170	0.26	
									AA	10	0.01	2	0.00	
			Caucasian	T	2083	0.87	1903	0.87	TT	908	0.75	838	0.77	0.07
				A	325	0.13	275	0.13	TA	267	0.22	227	0.21	
									AA	29	0.02	24	0.02	
	rs11879758	G/C	African American	G	1291	0.87	1153	0.88	GG	561	0.76	502	0.77	0.10
				C	193	0.13	159	0.12	GC	169	0.23	149	0.23	
									CC	12	0.02	5	0.01	
			Caucasian	G	2067	0.86	1882	0.86	GG	887	0.74	809	0.74	0.36
				C	339	0.14	296	0.14	GC	293	0.24	264	0.24	
									CC	23	0.02	16	0.01	
	rs2278417	C/T	African American	C	535	0.36	502	0.38	CC	104	0.14	99	0.15	0.62
				T	949	0.64	814	0.62	CT	327	0.44	304	0.46	
									TT	311	0.42	255	0.39	
			Caucasian	C	1739	0.72	1609	0.74	CC	638	0.53	599	0.55	0.48
				T	669	0.28	569	0.26	CT	463	0.38	411	0.38	
									TT	103	0.09	79	0.07	
	rs4986770	C/T	African American	C	1341	0.91	1205	0.92	CC	609	0.82	549	0.83	0.30
				T	139	0.09	111	0.08	CT	123	0.17	107	0.16	
									TT	8	0.01	2	0.00	
			Caucasian	C	2239	0.93	2025	0.93	CC	1043	0.87	942	0.87	0.64
				T	167	0.07	153	0.07	CT	153	0.13	141	0.13	
									TT	7	0.01	6	0.01	

Table 2.16 (cont.) Race-specific allele and genotype frequencies for *AURKA, BRCA1,* and *BRCA1*-interacting genes genotyped in CBCS participants enrolled 1993-2001.

Gene	dbSNP rs	SNP	Race	Allele	Allele Count and Frequency				Genotype	Genotype Count and Frequency				HWE[a]
					Cases		Controls			Cases		Controls		P-value
					Count	AF	Count	AF		Count	GF	Count	GF	
ZNF350	rs4988334	T/C	African American	T	1051	0.71	930	0.71	TT	375	0.51	319	0.48	0.07
				C	433	0.29	386	0.29	TC	301	0.41	292	0.44	
									CC	66	0.09	47	0.07	
			Caucasian	T	2057	0.85	1883	0.86	TT	885	0.74	820	0.75	0.12
				C	351	0.15	295	0.14	TC	287	0.24	243	0.22	
									CC	32	0.03	26	0.02	
	rs8102072	T/C	African American	T	1187	0.81	1081	0.82	TT	481	0.65	437	0.67	0.03
				C	287	0.19	231	0.18	TC	225	0.31	207	0.32	
									CC	31	0.04	12	0.02	
			Caucasian	T	1870	0.78	1725	0.79	TT	739	0.61	687	0.63	0.46
				C	538	0.22	453	0.21	TC	392	0.33	351	0.32	
									CC	73	0.06	51	0.05	

[a] HWE assesed in controls only; exact p-value corresponding to a 1 df chi-square test

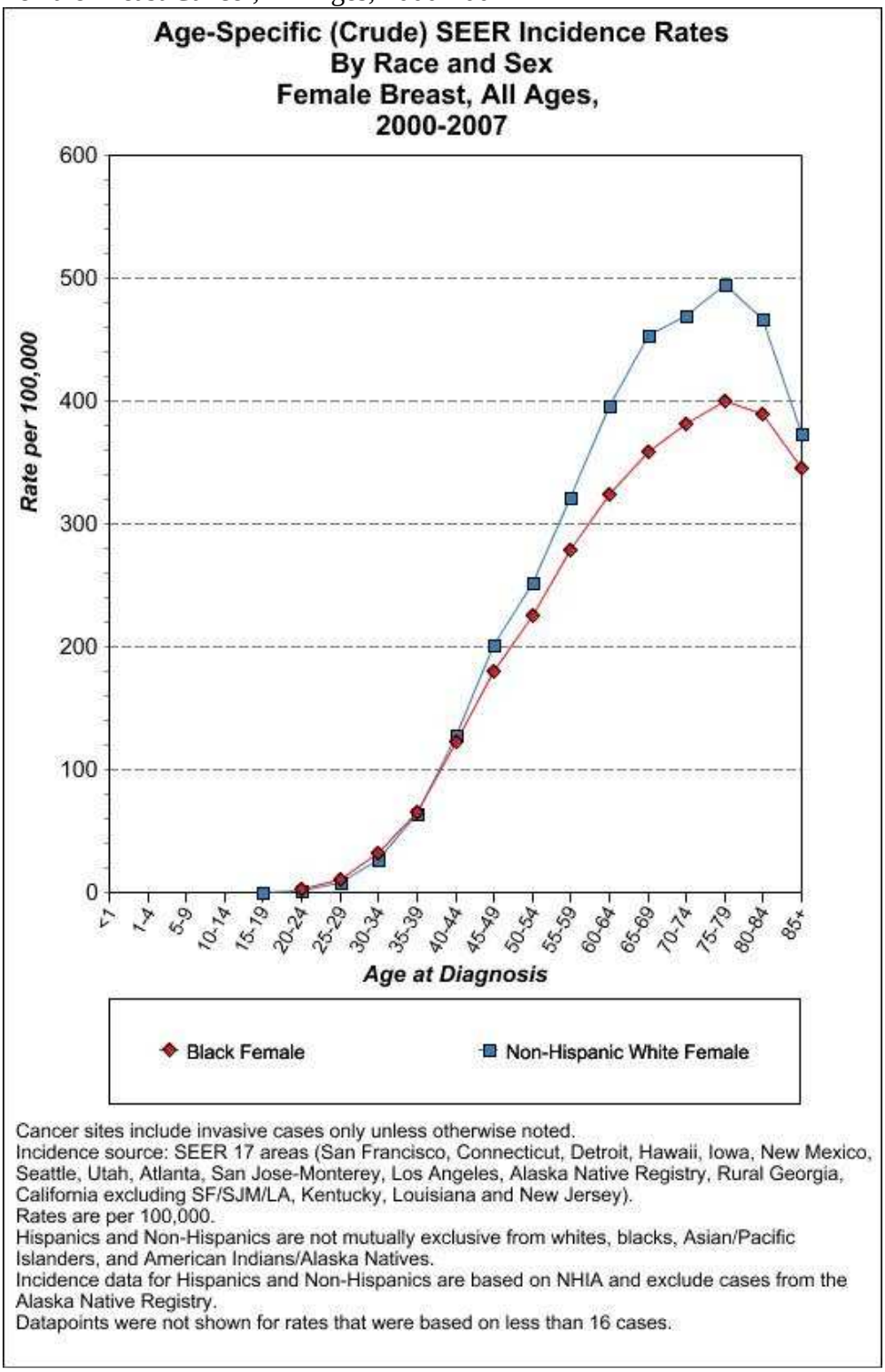
Table 2.17 Characteristics of CBCS case participants with genotype data (*N*=1,946), case participants missing genotype data (*N*=331), controls with genotype data (*N*=1,747), and controls missing genotype data (*N*=238)

	Cases with genotype data *(85%)*	Cases missing genotype data *(15%)*	Controls with genotype data *(88%)*	Controls missing genotype data *(12%)*
Race				
African American	742 *(38)*	151 *(46)*	658 *(38)*	128 *(54)*
Caucasian	1204 *(62)*	180 *(54)*	1089 *(62)*	110 *(46)*
Age				
20-24	6 *(0)*	0 *(0)*	1 *(0)*	0 *(0)*
25-29	21 *(1)*	4 *(1)*	10 *(0)*	1 *(0)*
30-34	85 *(4)*	10 *(3)*	60 *(3)*	7 *(3)*
35-39	172 *(9)*	24 *(7)*	133 *(8)*	11 *(5)*
40-44	276 *(14)*	55 *(17)*	242 *(14)*	39 *(16)*
45-49	387 *(20)*	82 *(25)*	359 *(21)*	56 *(24)*
50-54	208 *(11)*	29 *(9)*	237 *(14)*	28 *(12)*
55-59	216 *(11)*	37 *(11)*	191 *(13)*	27 *(11)*
60-64	201 *(10)*	23 *(7)*	166 *(10)*	22 *(9)*
65-69	200 *(10)*	43 *(13)*	185 *(11)*	17 *(7)*
70-74	174 *(9)*	25 *(8)*	163 *(9)*	30 *(13)*
Menopausal Status				
Premenopausal	864 *(44)*	149 *(45)*	746 *(43)*	105 *(44)*
Postmenopausal	1082 *(56)*	182 *(55)*	1001 *(57)*	133 *(56)*
Stage				
1	609 *(31)*	94 *(28)*		
2	627 *(32)*	129 *(39)*		
3	144 *(7)*	21 *(6)*		
4	42 *(2)*	8 *(2)*		
CIS	437 *(22)*	59 *(18)*		
Missing	87 *(4)*	20 *(6)*		
Tumor Size[a]				
≤2cm	769 *(40)*	125 *(38)*		
>2cm - 5cm	502 *(26)*	107 *(32)*		
>5cm	146 *(8)*	23 *(7)*		
Missing	529 *(27)*	76 *(23)*		
Subtype				
Luminal A	674 *(35)*	116 *(35)*		
Luminal B	114 *(6)*	21 *(6)*		
HER2+/ER-	94 *(5)*	22 *(7)*		
Basal-like	199 *(10)*	25 *(8)*		
Unclassified	129 *(7)*	18 *(5)*		
Missing	736 *(38)*	129 *(39)*		

[a] Not available for carcinoma *in situ* (CIS) cases

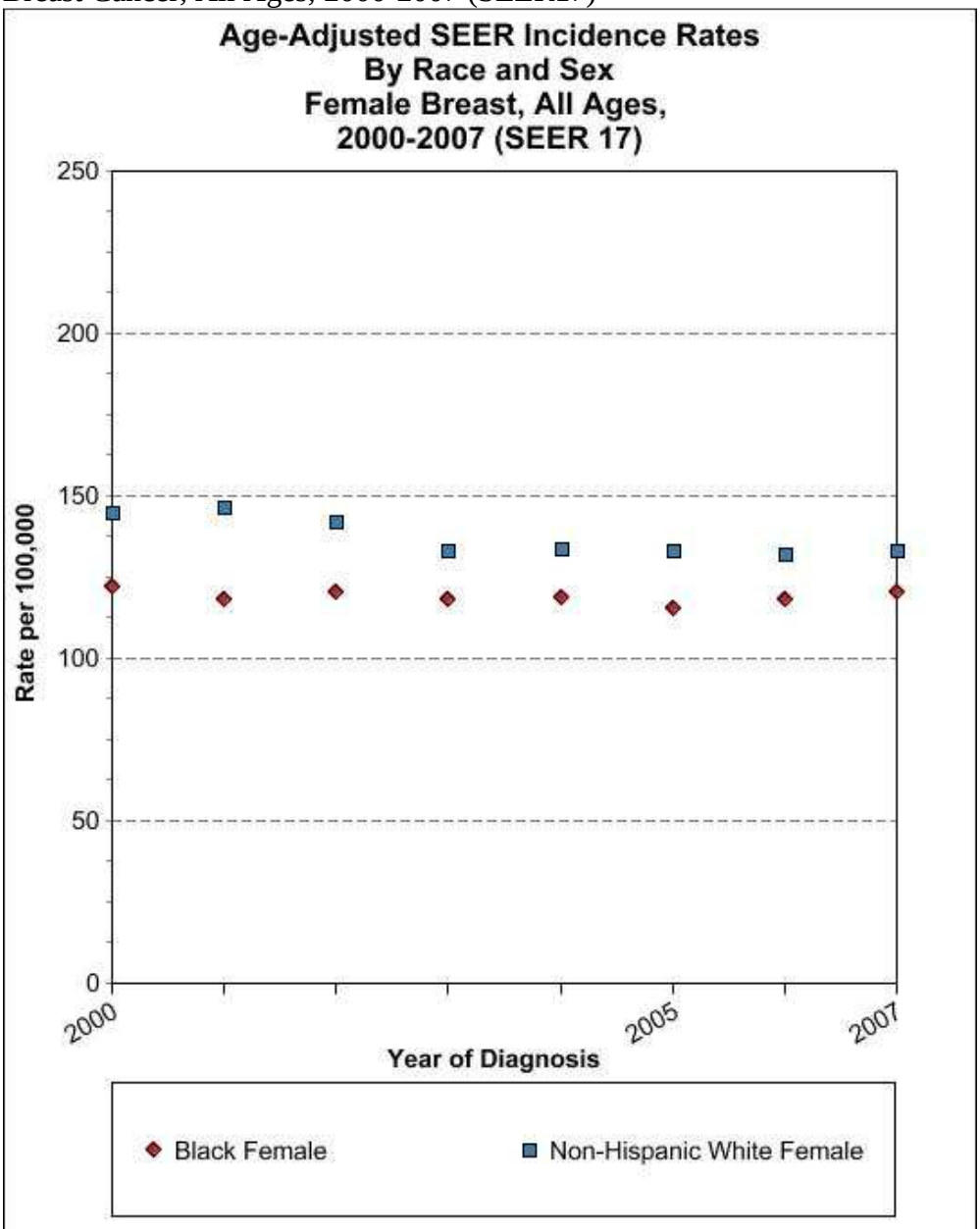
2.17 Figures

Figure 2.1 Age-specific (Crude) SEER Incidence Rates by Race and Sex, Female Breast Cancer, All Ages, 2000-2007



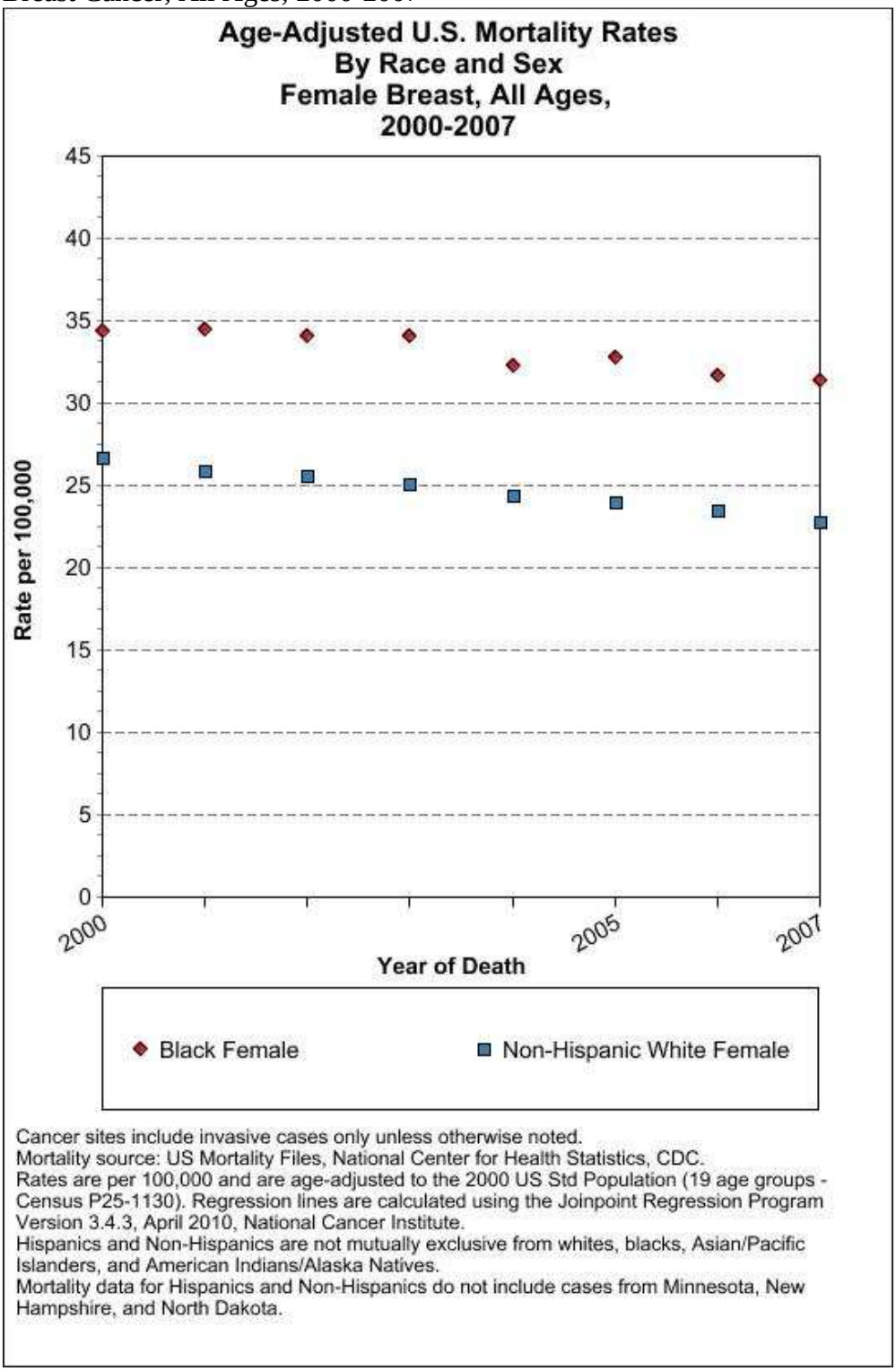
Cancer sites include invasive cases only unless otherwise noted.
Incidence source: SEER 17 areas (San Francisco, Connecticut, Detroit, Hawaii, Iowa, New Mexico, Seattle, Utah, Atlanta, San Jose-Monterey, Los Angeles, Alaska Native Registry, Rural Georgia, California excluding SF/SJM/LA, Kentucky, Louisiana and New Jersey).
Rates are per 100,000.
Hispanics and Non-Hispanics are not mutually exclusive from whites, blacks, Asian/Pacific Islanders, and American Indians/Alaska Natives.
Incidence data for Hispanics and Non-Hispanics are based on NHIA and exclude cases from the Alaska Native Registry.
Datapoints were not shown for rates that were based on less than 16 cases.

Figure 2.2 Age Adjusted SEER Incidence Rates by Race and Sex, Female Breast Cancer, All Ages, 2000-2007 (SEER17)



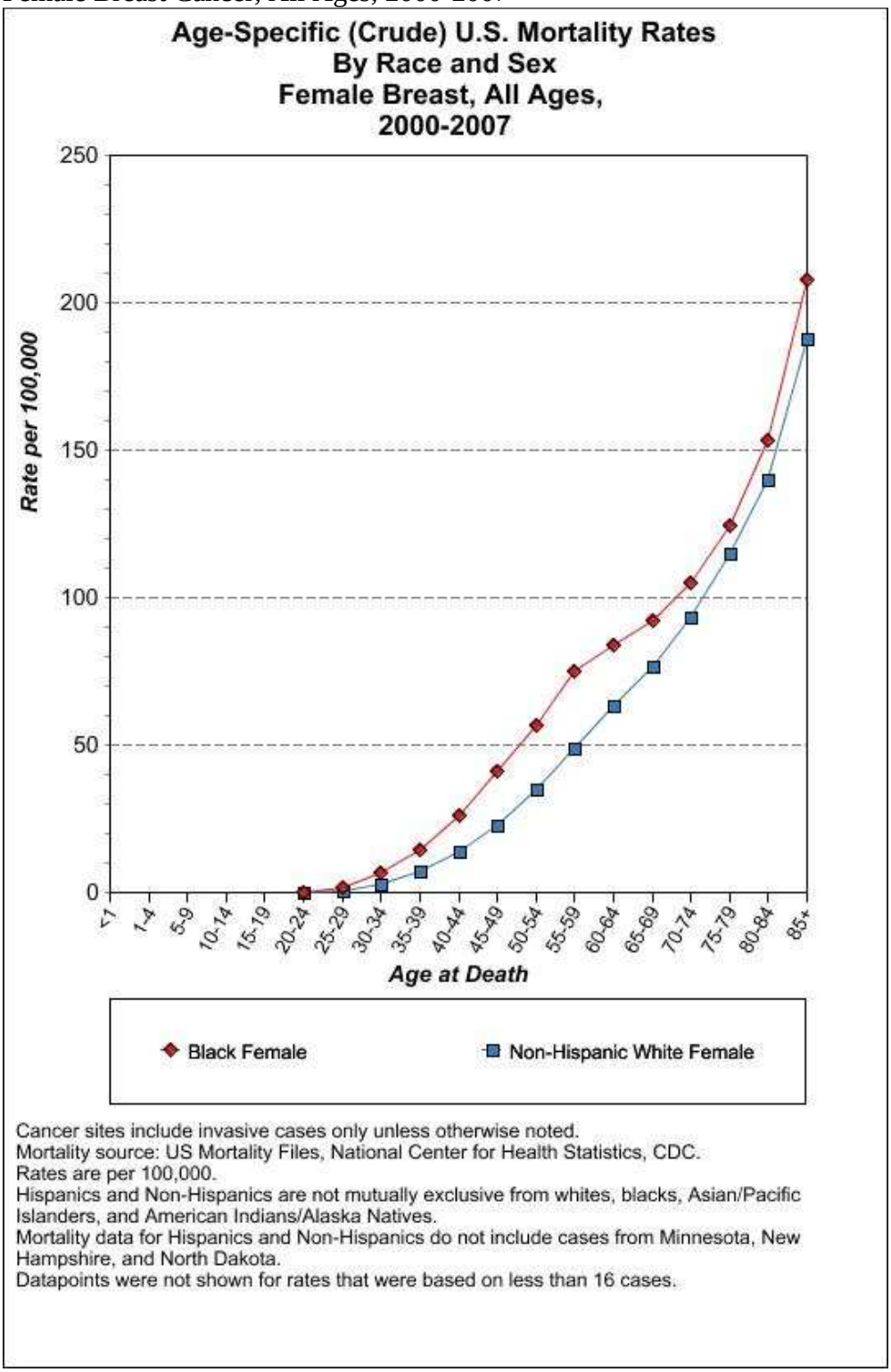
Cancer sites include invasive cases only unless otherwise noted.
Incidence source: SEER 17 areas (San Francisco, Connecticut, Detroit, Hawaii, Iowa, New Mexico, Seattle, Utah, Atlanta, San Jose-Monterey, Los Angeles, Alaska Native Registry, Rural Georgia, California excluding SF/SJM/LA, Kentucky, Louisiana and New Jersey).
Rates are per 100,000 and are age-adjusted to the 2000 US Std Population (19 age groups - Census P25-1130). Regression lines are calculated using the Joinpoint Regression Program Version 3.4.3, April 2010, National Cancer Institute.
Hispanics and Non-Hispanics are not mutually exclusive from whites, blacks, Asian/Pacific Islanders, and American Indians/Alaska Natives.
Incidence data for Hispanics and Non-Hispanics are based on NHIA and exclude cases from the Alaska Native Registry.

Figure 2.3 Age-Adjusted U.S. Mortality Rates by Race and Sex, Female Breast Cancer, All Ages, 2000-2007



Cancer sites include invasive cases only unless otherwise noted.
Mortality source: US Mortality Files, National Center for Health Statistics, CDC.
Rates are per 100,000 and are age-adjusted to the 2000 US Std Population (19 age groups - Census P25-1130). Regression lines are calculated using the Joinpoint Regression Program Version 3.4.3, April 2010, National Cancer Institute.
Hispanics and Non-Hispanics are not mutually exclusive from whites, blacks, Asian/Pacific Islanders, and American Indians/Alaska Natives.
Mortality data for Hispanics and Non-Hispanics do not include cases from Minnesota, New Hampshire, and North Dakota.

Figure 2.4 Age-Specific (Crude) U.S. Mortality Rates by Race and Sex, Female Breast Cancer, All Ages, 2000-2007



Cancer sites include invasive cases only unless otherwise noted.
Mortality source: US Mortality Files, National Center for Health Statistics, CDC.
Rates are per 100,000.
Hispanics and Non-Hispanics are not mutually exclusive from whites, blacks, Asian/Pacific Islanders, and American Indians/Alaska Natives.
Mortality data for Hispanics and Non-Hispanics do not include cases from Minnesota, New Hampshire, and North Dakota.
Datapoints were not shown for rates that were based on less than 16 cases.

Figure 2.5 Immunohistochemical Identification of Breast Tumor Intrinsic Subtypes

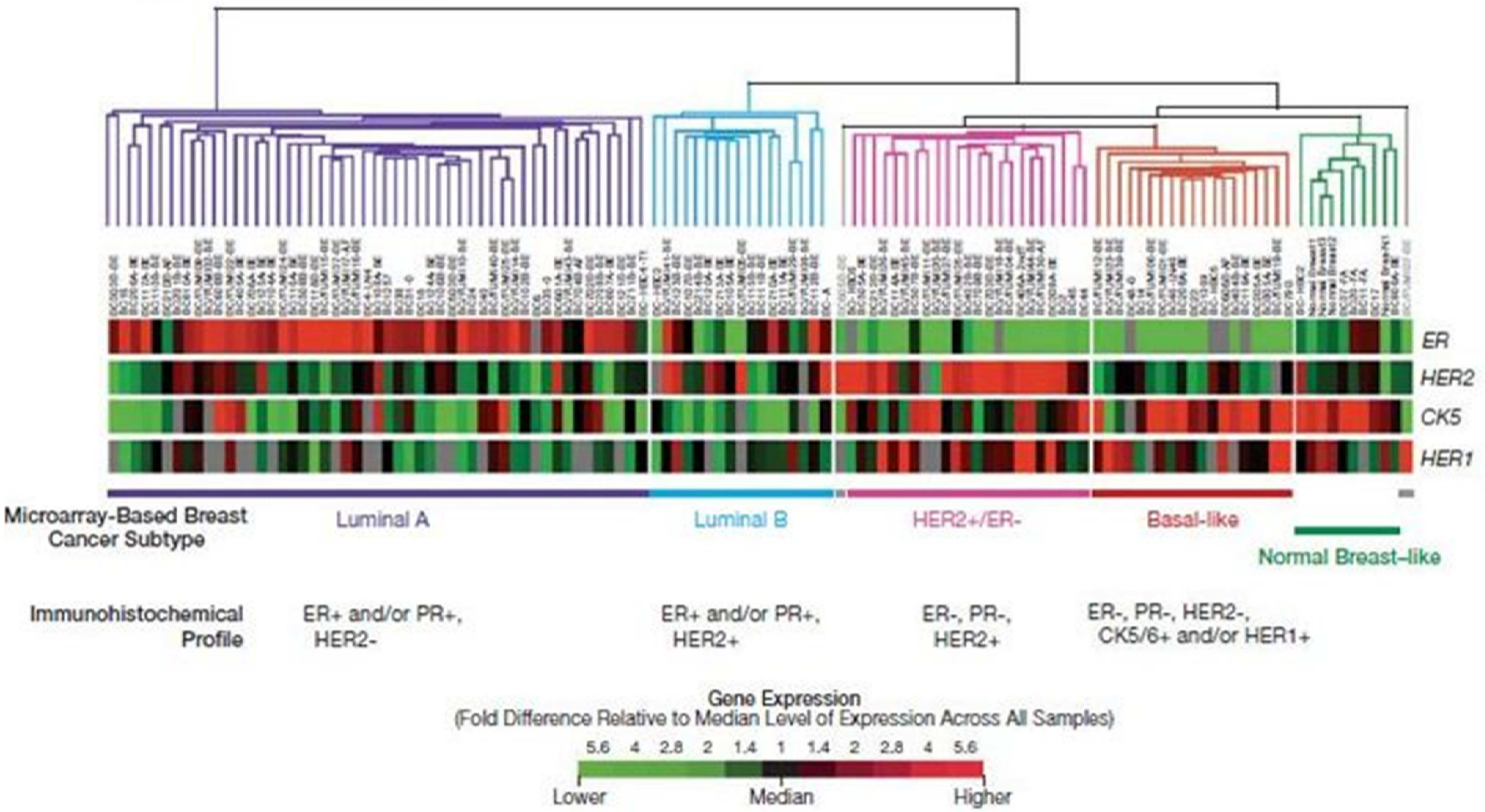


Figure 2.6 Carolina Breast Cancer Study Geographic Study Area

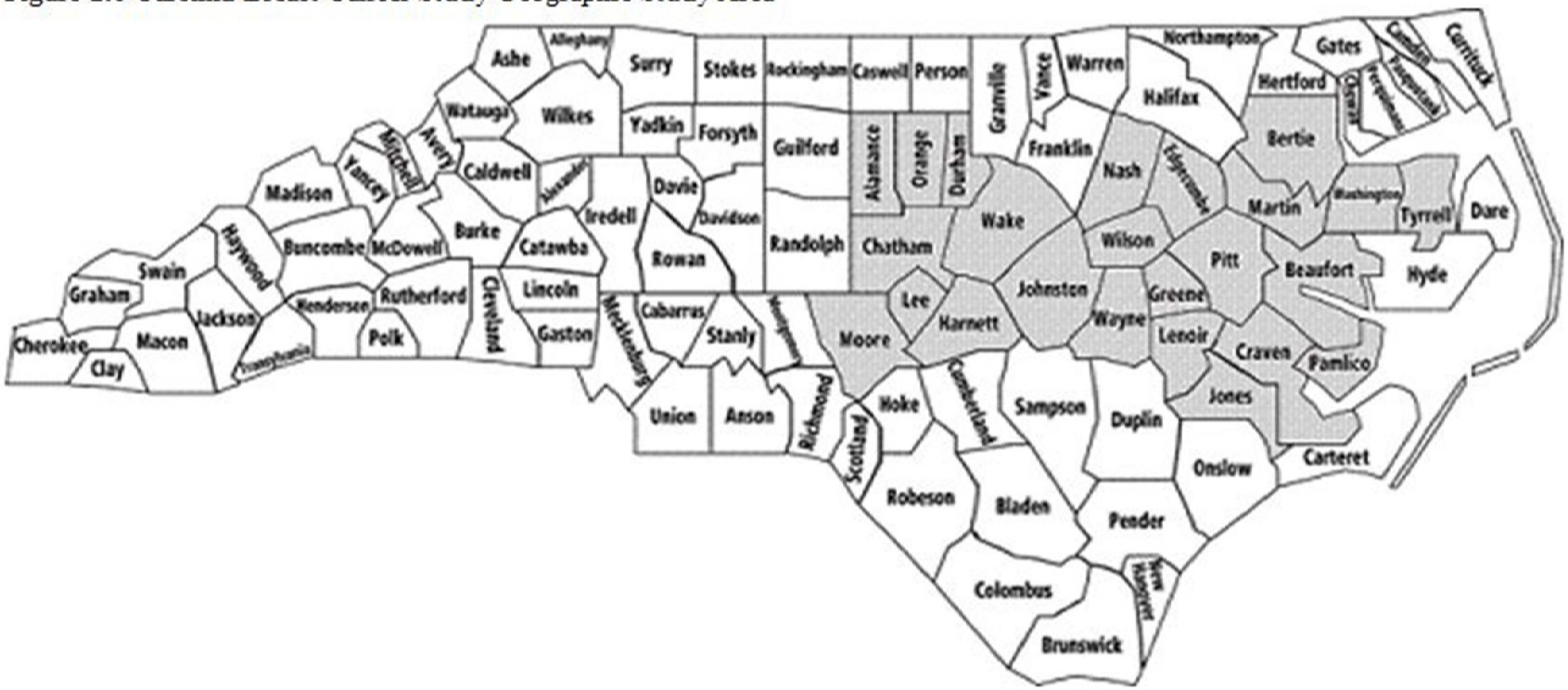


Figure 2.7 Decision Tree for Inclusion of Single Nucleotide Polymorphisms (SNPs) into Interaction Study

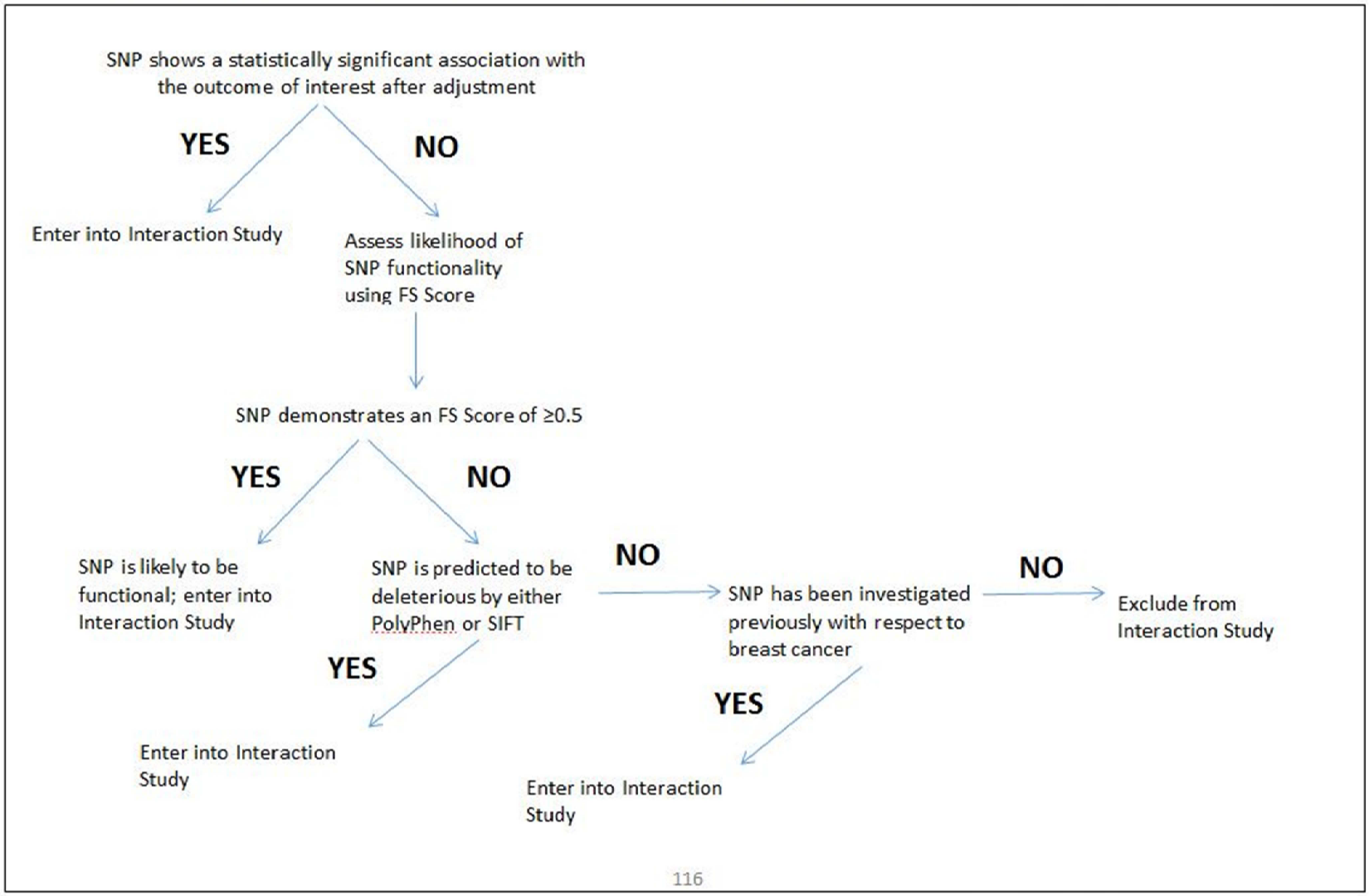


Figure 2.8 Power in Caucasian participants (Cases=1,204, Controls=1,089) given a genotype prevalence of 5%

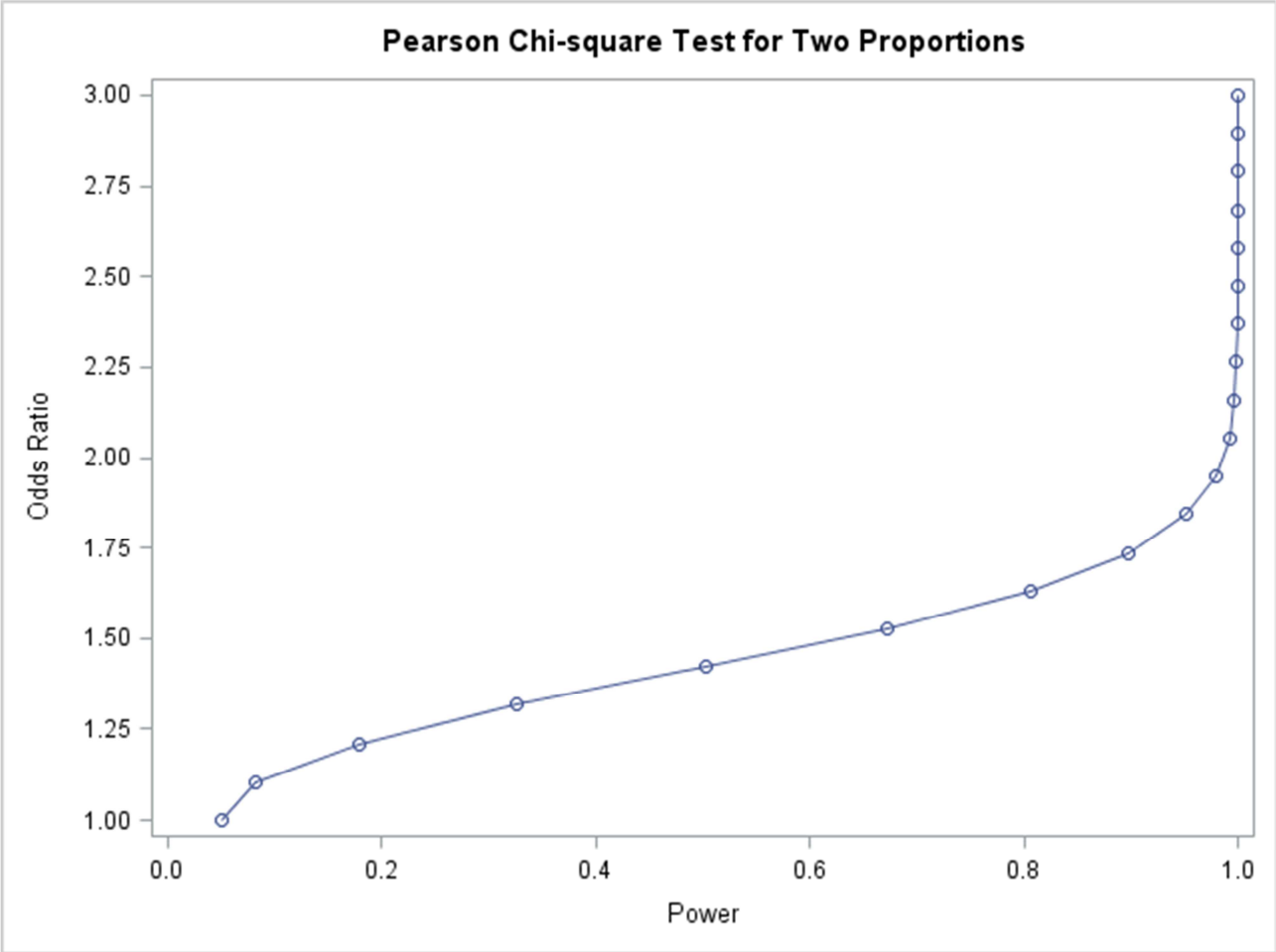


Figure 2.9 Power in Caucasian participants (Cases=1,204, Controls=1,089) given a genotype prevalence of 10%

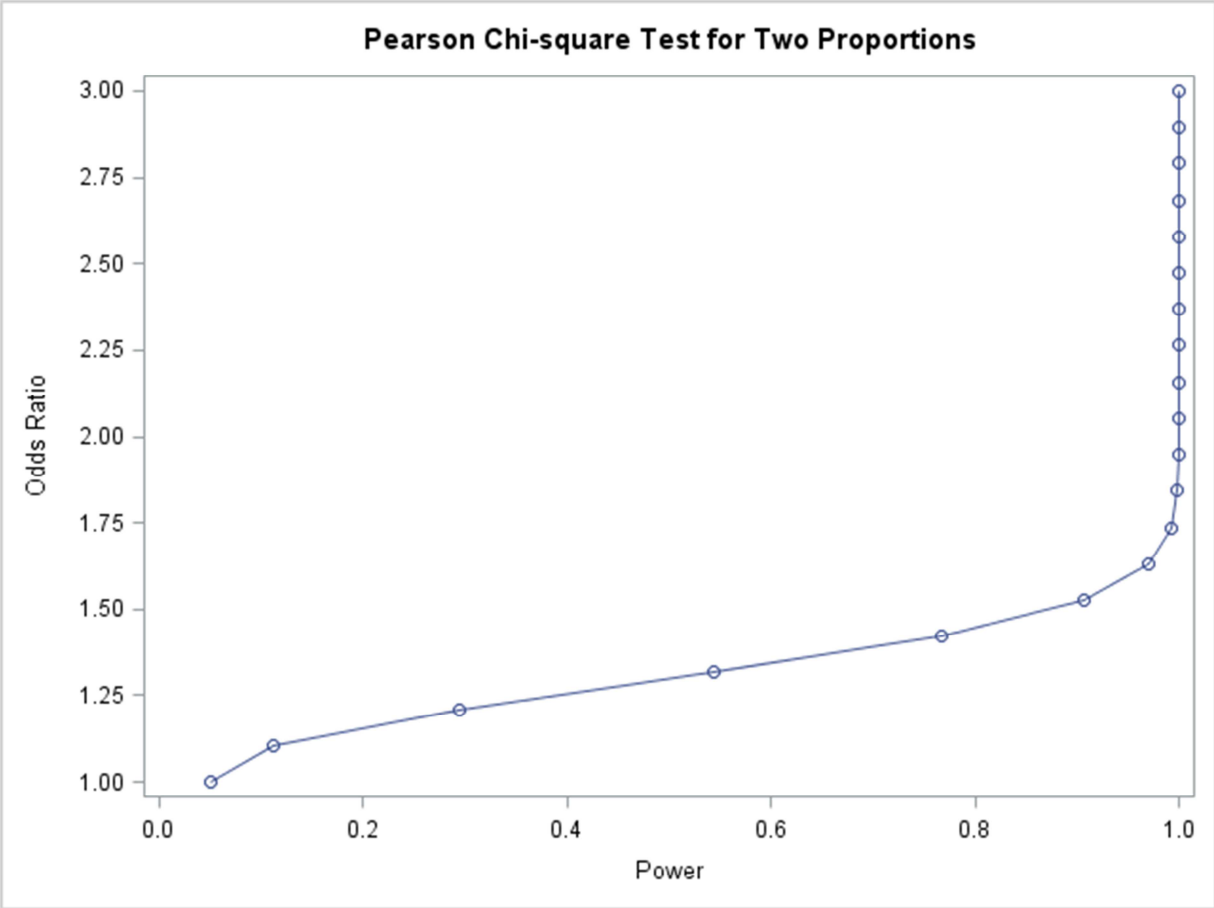


Figure 2.10 Power in Caucasian participants (Cases=1,204, Controls=1,089) given a genotype prevalence of 20%

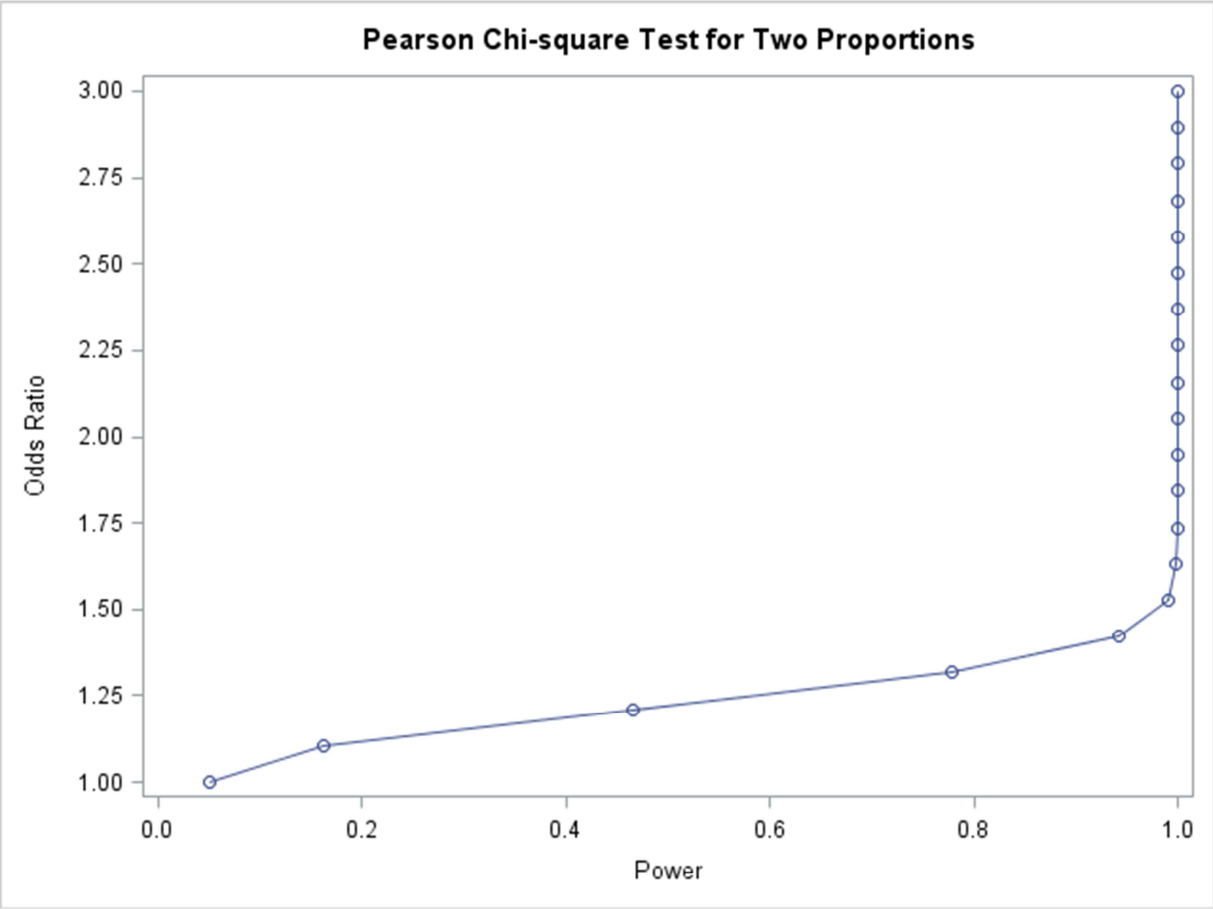


Figure 2.11 Power in African American participants (Cases=742, Controls=658) given a genotype prevalence of 5%

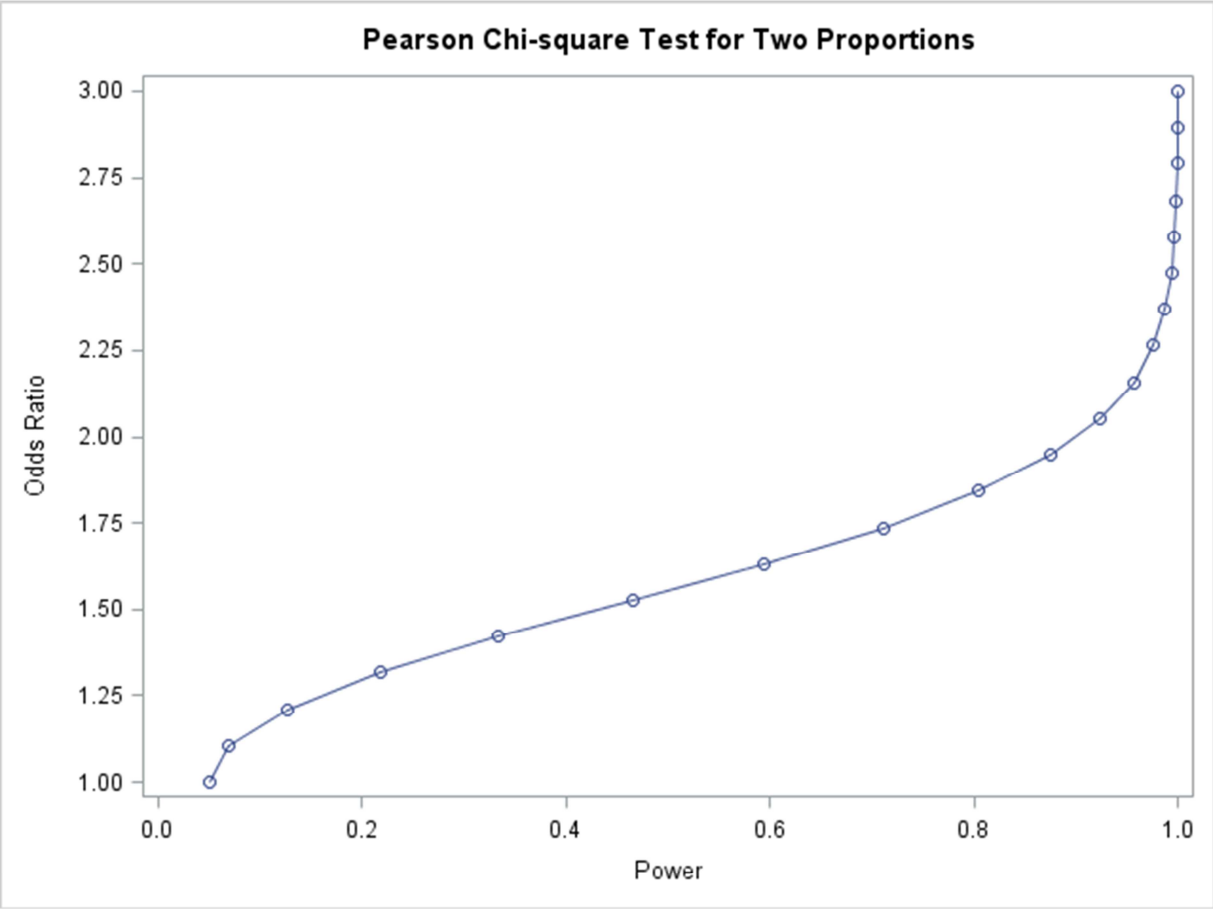


Figure 2.12 Power in African American participants (Cases=742, Controls=658) given a genotype prevalence of 10%

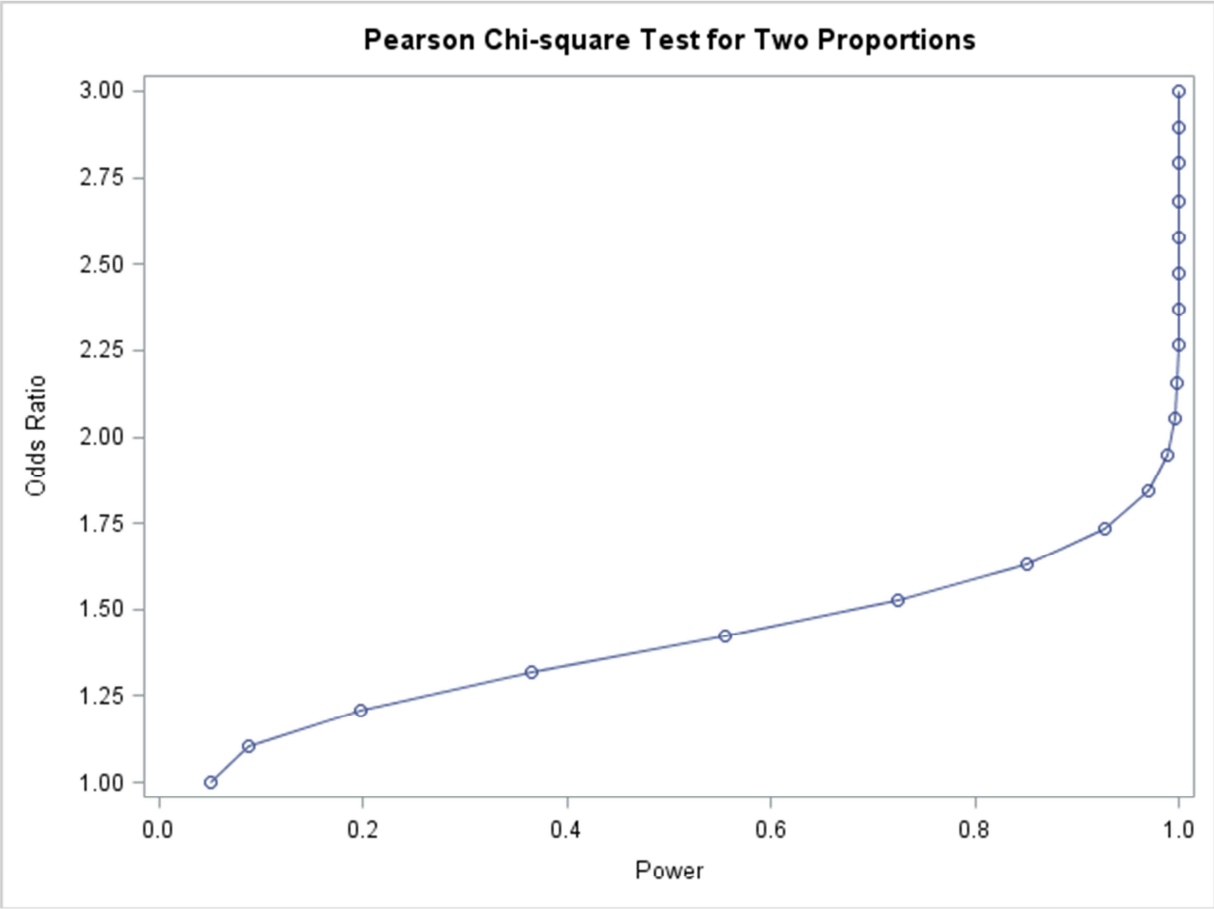


Figure 2.13 Power in African American participants (Cases=742, Controls=658) given a genotype prevalence of 20%

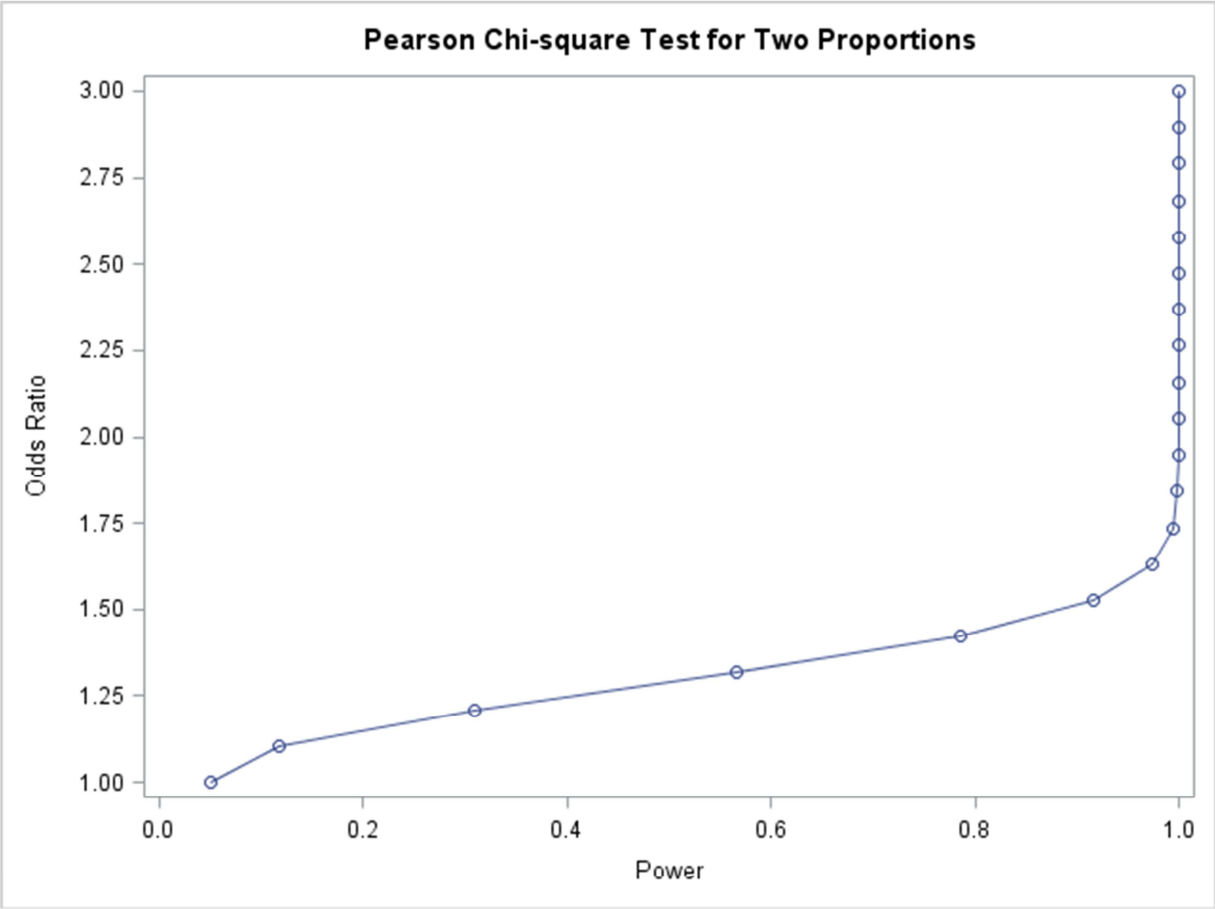
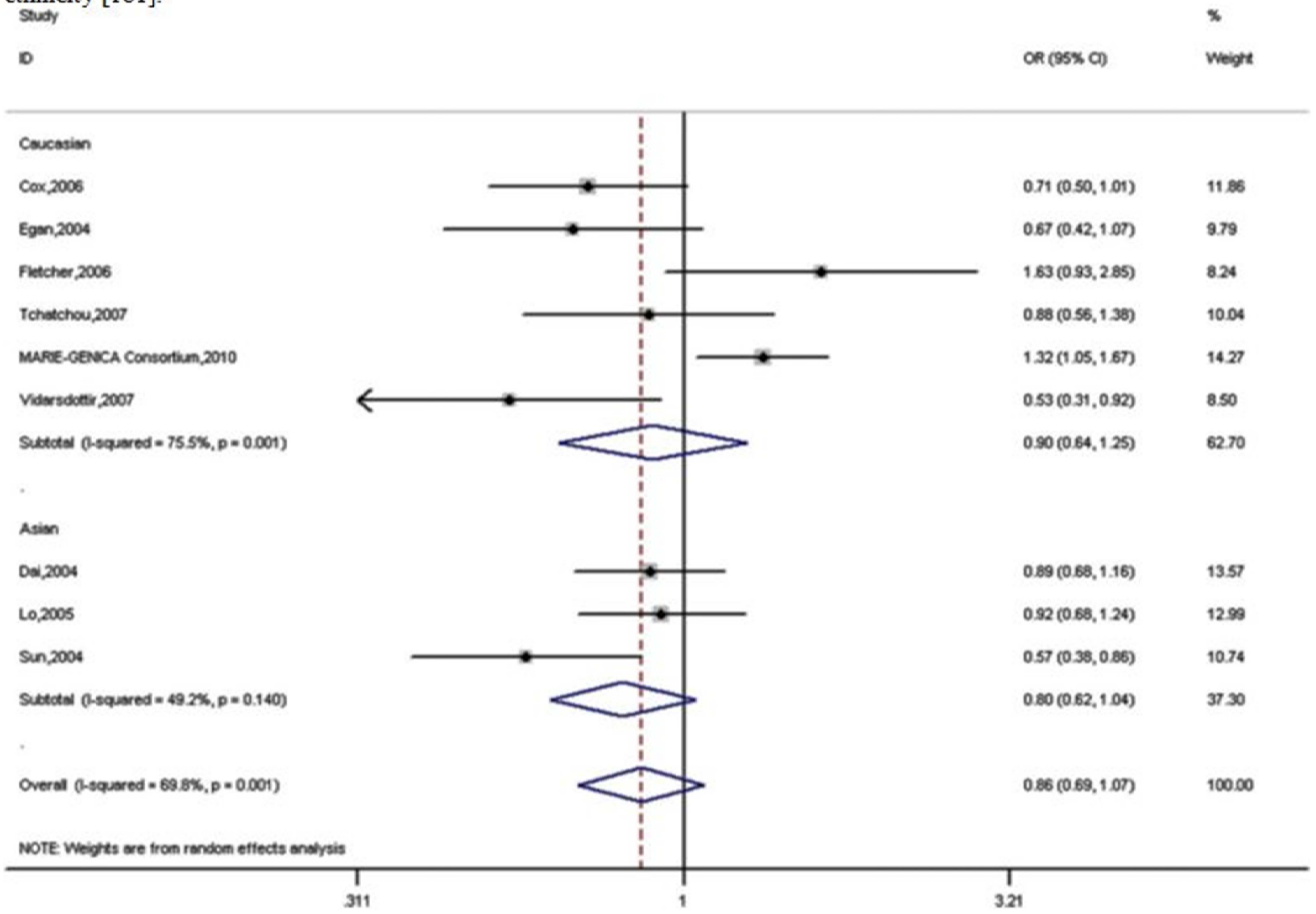


Figure 2.14 Forest plot of the association between *AURKA* functional polymorphism rs2273535 and breast cancer risk stratified by ethnicity [181].



REFERENCES

1. American Cancer Society. *Cancer Facts and Figures 2011*. In. Atlanta: American Cancer Society.

2. Carey LA, Perou CM, Livasy CA, Dressler LG, Cowan D, Conway K, Karaca G, Troester MA, Tse CK, Edmiston S *et al*: Race, breast cancer subtypes, and survival in the Carolina Breast Cancer Study. *JAMA* 2006, 295(21):2492-2502.

3. Bradbury AR, Olopade OI: Genetic susceptibility to breast cancer. *Rev Endocr Metab Disord* 2007, 8(3):255-267.

4. Lichtenstein P, Holm NV, Verkasalo PK, Iliadou A, Kaprio J, Koskenvuo M, Pukkala E, Skytthe A, Hemminki K: 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.

5. Antoniou AC, Pharoah PD, McMullan G, Day NE, Stratton MR, Peto J, Ponder BJ, Easton DF: A comprehensive model for familial breast cancer incorporating BRCA1, BRCA2 and other genes. *Br J Cancer* 2002, 86(1):76-83.

6. 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):33-36.

7. 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.

8. Antoniou AC, Easton DF: Polygenic inheritance of breast cancer: Implications for design of association studies. *Genet Epidemiol* 2003, 25(3):190-202.

9. Perou CM, Sorlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, Pollack JR, Ross DT, Johnsen H, Akslen LA *et al*: Molecular portraits of human breast tumours. *Nature* 2000, 406(6797):747-752.

10. Nielsen TO, Hsu FD, Jensen K, Cheang M, Karaca G, Hu Z, Hernandez-Boussard T, Livasy C, Cowan D, Dressler L *et al*: Immunohistochemical and clinical characterization of the basal-like subtype of invasive breast carcinoma. *Clin Cancer Res* 2004, 10(16):5367-5374.

11. Sorlie T, Tibshirani R, Parker J, Hastie T, Marron JS, Nobel A, Deng S, Johnsen H, Pesich R, Geisler S *et al*: Repeated observation of breast tumor subtypes in independent gene expression data sets. *Proc Natl Acad Sci U S A* 2003, 100(14):8418-8423.

12. Sorlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, Hastie T, Eisen MB, van de Rijn M, Jeffrey SS *et al*: Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci U S A* 2001, 98(19):10869-10874.

13. Millikan RC, Newman B, Tse CK, Moorman PG, Conway K, Dressler LG, Smith LV, Labbok MH, Geradts J, Bensen JT *et al*: Epidemiology of basal-like breast cancer. *Breast Cancer Res Treat* 2008, 109(1):123-139.

14. Calza S, Hall P, Auer G, Bjohle J, Klaar S, Kronenwett U, Liu ET, Miller L, Ploner A, Smeds J *et al*: Intrinsic molecular signature of breast cancer in a population-based cohort of 412 patients. *Breast Cancer Res* 2006, 8(4):R34.

15. Livasy CA, Karaca G, Nanda R, Tretiakova MS, Olopade OI, Moore DT, Perou CM: Phenotypic evaluation of the basal-like subtype of invasive breast carcinoma. *Mod Pathol* 2006, 19(2):264-271.

16. Foulkes WD, Smith IE, Reis-Filho JS: Triple-negative breast cancer. *N Engl J Med* 2010, 363(20):1938-1948.

17. Anders CK, Winer EP, Ford JM, Dent R, Silver DP, Sledge GW, Carey LA: Poly(ADP-Ribose) polymerase inhibition: "Targeted" therapy for triple-negative breast cancer. *Clin Cancer Res* 2010, 16(19):4702-4710.

18. Maegawa RO, Tang SC: Triple-negative breast cancer: unique biology and its management. *Cancer Invest* 2010, 28(8):878-883.

19. Kennecke H, Yerushalmi R, Woods R, Cheang MC, Voduc D, Speers CH, Nielsen TO, Gelmon K: Metastatic behavior of breast cancer subtypes. *J Clin Oncol* 2010, 28(20):3271-3277.

20. Honrado E, Benitez J, Palacios J: The molecular pathology of hereditary breast cancer: genetic testing and therapeutic implications. *Mod Pathol* 2005, 18(10):1305-1320.

21. Foulkes WD, Reis-Filho JS, Narod SA: Tumor size and survival in breast cancer--a reappraisal. *Nat Rev Clin Oncol* 2010, 7(6):348-353.

22. Foulkes WD, Stefansson IM, Chappuis PO, Begin LR, Goffin JR, Wong N, Trudel M, Akslen LA: Germline BRCA1 mutations and a basal epithelial phenotype in breast cancer. *J Natl Cancer Inst* 2003, 95(19):1482-1485.

23. Yang XR, Sherman ME, Rimm DL, Lissowska J, Brinton LA, Peplonska B, Hewitt SM, Anderson WF, Szeszenia-Dabrowska N, Bardin-Mikolajczak A *et al*: Differences in risk factors for breast cancer molecular subtypes in a population-based study. *Cancer Epidemiol Biomarkers Prev* 2007, 16(3):439-443.

24. Lukasiewicz KB, Lingle WL: Aurora A, centrosome structure, and the centrosome cycle. *Environ Mol Mutagen* 2009, 50(8):602-619.

25. Glover DM, Leibowitz MH, McLean DA, Parry H: Mutations in aurora prevent centrosome separation leading to the formation of monopolar spindles. *Cell* 1995, 81(1):95-105.

26. Polymorphisms in the BRCA1 and ABCB1 genes modulate menopausal hormone therapy associated breast cancer risk in postmenopausal women. *Breast Cancer Res Treat* 2010, 120(3):727-736.

27. Dai Q, Cai QY, Shu XO, Ewart-Toland A, Wen WQ, Balmain A, Gao YT, Zheng W: Synergistic effects of STK15 gene polymorphisms and endogenous estrogen exposure in the risk of breast cancer. *Cancer Epidemiol Biomarkers Prev* 2004, 13(12):2065-2070.

28. Sun T, Miao X, Wang J, Tan W, Zhou Y, Yu C, Lin D: Functional Phe31Ile polymorphism in Aurora A and risk of breast carcinoma. *Carcinogenesis* 2004, 25(11):2225-2230.

29. Lo YL, Yu JC, Chen ST, Yang HC, Fann CS, Mau YC, Shen CY: Breast cancer risk associated with genotypic polymorphism of the mitosis-regulating gene Aurora-A/STK15/BTAK. *Int J Cancer* 2005, 115(2):276-283.

30. Egan KM, Newcomb PA, Ambrosone CB, Trentham-Dietz A, Titus-Ernstoff L, Hampton JM, Kimura MT, Nagase H: STK15 polymorphism and breast cancer risk in a population-based study. *Carcinogenesis* 2004, 25(11):2149-2153.

31. Perou CM, Jeffrey SS, van de Rijn M, Rees CA, Eisen MB, Ross DT, Pergamenschikov A, Williams CF, Zhu SX, Lee JC *et al*: Distinctive gene expression patterns in human mammary epithelial cells and breast cancers. *Proc Natl Acad Sci U S A* 1999, 96(16):9212-9217.

32. National Cancer Institute SEER Fact Stat Sheets [http://seer.cancer.gov/statfacts/html/breast.html]

33. Carey LA: Through a glass darkly: advances in understanding breast cancer biology, 2000-2010. *Clin Breast Cancer* 2010, 10(3):188-195.

34. Guenard F, Labrie Y, Ouellette G, Beauparlant CJ, Durocher F: Genetic sequence variations of BRCA1-interacting genes AURKA, BAP1, BARD1 and DHX9 in French Canadian families with high risk of breast cancer. *J Hum Genet* 2009, 54(3):152-161.

35. Hannemann J, Kristel P, van Tinteren H, Bontenbal M, van Hoesel QG, Smit WM, Nooij MA, Voest EE, van der Wall E, Hupperets P *et al*: Molecular subtypes of breast cancer and amplification of topoisomerase II alpha: predictive role in dose intensive adjuvant chemotherapy. *Br J Cancer* 2006, 95(10):1334-1341.

36. Fan C, Oh DS, Wessels L, Weigelt B, Nuyten DS, Nobel AB, van't Veer LJ, Perou CM: Concordance among gene-expression-based predictors for breast cancer. *N Engl J Med* 2006, 355(6):560-569.

37. 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):478-485.

38. Stead LA, Lash TL, Sobieraj JE, Chi DD, Westrup JL, Charlot M, Blanchard RA, Lee JC, King TC, Rosenberg CL: Triple-negative breast cancers are increased in black women regardless of age or body mass index. *Breast Cancer Res* 2009, 11(2):R18.

39. Lund MJ, Butler EN, Hair BY, Ward KC, Andrews JH, Oprea-Ilies G, Bayakly AR, O'Regan RM, Vertino PM, Eley JW: Age/race differences in HER2 testing and in incidence rates for breast cancer triple subtypes: a population-based study and first report. *Cancer* 2010, 116(11):2549-2559.

40. Yang XR, Chang-Claude J, Goode EL, Couch FJ, Nevanlinna H, Milne RL, Gaudet M, Schmidt MK, Broeks A, Cox A *et al*: Associations of breast cancer risk factors with tumor subtypes: a pooled analysis from the Breast Cancer Association Consortium studies. *J Natl Cancer Inst* 2011, 103(3):250-263.

41. Easton DF: How many more breast cancer predisposition genes are there? *Breast Cancer Res* 1999, 1(1):14-17.

42. Prevalence and penetrance of BRCA1 and BRCA2 mutations in a population-based series of breast cancer cases. Anglian Breast Cancer Study Group. *Br J Cancer* 2000, 83(10):1301-1308.

43. Serova OM, Mazoyer S, Puget N, Dubois V, Tonin P, Shugart YY, Goldgar D, Narod SA, Lynch HT, Lenoir GM: Mutations in BRCA1 and BRCA2 in breast cancer families: are there more breast cancer-susceptibility genes? *Am J Hum Genet* 1997, 60(3):486-495.

44. Bischoff JR, Anderson L, Zhu Y, Mossie K, Ng L, Souza B, Schryver B, Flanagan P, Clairvoyant F, Ginther C *et al*: A homologue of Drosophila aurora kinase is oncogenic and amplified in human colorectal cancers. *EMBO J* 1998, 17(11):3052-3065.

45. Staff S, Isola J, Jumppanen M, Tanner M: Aurora-A gene is frequently amplified in basal-like breast cancer. *Oncol Rep* 2010, 23(2):307-312.

46. Cox DG, Hankinson SE, Hunter DJ: Polymorphisms of the AURKA (STK15/Aurora Kinase) Gene and Breast Cancer Risk (United States). *Cancer Causes Control* 2006, 17(1):81-83.

47. Fletcher O, Johnson N, Palles C, dos Santos Silva I, McCormack V, Whittaker J, Ashworth A, Peto J: Inconsistent association between the STK15 F31I genetic polymorphism and breast cancer risk. *J Natl Cancer Inst* 2006, 98(14):1014-1018.

48. Tchatchou S, Wirtenberger M, Hemminki K, Sutter C, Meindl A, Wappenschmidt B, Kiechle M, Bugert P, Schmutzler RK, Bartram CR *et al*: Aurora kinases A and B and familial breast cancer risk. *Cancer Lett* 2007, 247(2):266-272.

49. Vidarsdottir L, Bodvarsdottir SK, Hilmarsdottir H, Tryggvadottir L, Eyfjord JE: Breast cancer risk associated with AURKA 91T -->A polymorphism in relation to BRCA mutations. *Cancer Lett* 2007, 250(2):206-212.

50. Couch FJ, Sinilnikova O, Vierkant RA, Pankratz VS, Fredericksen ZS, Stoppa-Lyonnet D, Coupier I, Hughes D, Hardouin A, Berthet P *et al*: AURKA F31I polymorphism and breast cancer risk in BRCA1 and BRCA2 mutation carriers: a consortium of investigators of modifiers of BRCA1/2 study. *Cancer Epidemiol Biomarkers Prev* 2007, 16(7):1416-1421.

51. Newman B, Moorman PG, Millikan R, Qaqish BF, Geradts J, Aldrich TE, Liu ET: The Carolina Breast Cancer Study: integrating population-based epidemiology and molecular biology. *Breast Cancer Res Treat* 1995, 35(1):51-60.

52. Weinberg CR, Wacholder S: The design and analysis of case-control studies with biased sampling. *Biometrics* 1990, 46(4):963-975.

53. Weinberg CR, Sandler DP: Randomized recruitment in case-control studies. *Am J Epidemiol* 1991, 134(4):421-432.

54. Millikan R, Eaton A, Worley K, Biscocho L, Hodgson E, Huang WY, Geradts J, Iacocca M, Cowan D, Conway K *et al*: HER2 codon 655 polymorphism and risk of breast cancer in African Americans and whites. *Breast Cancer Res Treat* 2003, 79(3):355-364.

55. 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):703-714.

56. Age-Adjusted U.S. Mortality Rates By Race/Ethnicity Female Breast, All Ages, 2000-2010 [http://www.seer.cancer.gov/faststats/]

57. van de Rijn M, Perou CM, Tibshirani R, Haas P, Kallioniemi O, Kononen J, Torhorst J, Sauter G, Zuber M, Kochli OR *et al*: Expression of cytokeratins 17 and 5 identifies a

group of breast carcinomas with poor clinical outcome. *Am J Pathol* 2002, 161(6):1991-1996.

58. Livasy CA, Perou CM, Karaca G, Cowan DW, Maia D, Jackson S, Tse CK, Nyante S, Millikan RC: Identification of a basal-like subtype of breast ductal carcinoma in situ. *Hum Pathol* 2007, 38(2):197-204.

59. 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):188-195.

60. [www.hapmap.org]

61. 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):1217-1223.

62. Barnholtz-Sloan JS, Shetty PB, Guan X, Nyante SJ, Luo J, Brennan DJ, Millikan RC: FGFR2 and other loci identified in genome-wide association studies are associated with breast cancer in African-American and younger women. *Carcinogenesis* 2010, 31(8):1417-1423.

63. Nyante SJ, Gammon MD, Kaufman JS, Bensen JT, Lin DY, Barnholtz-Sloan JS, Hu Y, He Q, Luo J, Millikan RC: Common genetic variation in adiponectin, leptin, and leptin receptor and association with breast cancer subtypes. *Breast Cancer Res Treat* 2011.

64. Thomas DC, Witte JS: Point: population stratification: a problem for case-control studies of candidate-gene associations? *Cancer Epidemiol Biomarkers Prev* 2002, 11(6):505-512.

65. Freedman ML, Reich D, Penney KL, McDonald GJ, Mignault AA, Patterson N, Gabriel SB, Topol EJ, Smoller JW, Pato CN *et al*: Assessing the impact of population stratification on genetic association studies. *Nat Genet* 2004, 36(4):388-393.

66. Devlin B, Roeder K, Wasserman L: Genomic control, a new approach to genetic-based association studies. *Theor Popul Biol* 2001, 60(3):155-166.

67. 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):904-909.

68. Jackson JE: A User's guide to Principal Components. New York: John Wiley & Sons; 2003.

69. Pritchard JK, Stephens M, Donnelly P: Inference of population structure using multilocus genotype data. *Genetics* 2000, 155(2):945-959.

70. Pfaff CL, Barnholtz-Sloan J, Wagner JK, Long JC: Information on ancestry from genetic markers. *Genet Epidemiol* 2004, 26(4):305-315.

71. 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):640-649.

72. 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):1545-1551.

73. SAS Institute Inc. SAS OnlineDoc© 9.2

74. Rothman KG, Sander: Modern Epidemiology. Philadelphia: Maple Press; 1998.

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

76. Li B, Leal SM: Deviations from hardy-weinberg equilibrium in parental and unaffected sibling genotype data. *Hum Hered* 2009, 67(2):104-115.

77. Stern C: The Hardy-Weinberg Law. *Science* 1943, 97(2510):137-138.

78. Benjamini YH, Yosef: Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statisical Society* 1995, 57(1):289-300.

79. Selection Bias in Epidemiological Studies [http://www.teachepi.org/documents/courses/fundamentals/Pai_Lecture6_Selection%20bias.pdf]

80. Rothman K: Episheet. In.; 2002, 2004.

81. Ewart-Toland A, Dai Q, Gao YT, Nagase H, Dunlop MG, Farrington SM, Barnetson RA, Anton-Culver H, Peel D, Ziogas A *et al*: Aurora-A/STK15 T+91A is a general low penetrance cancer susceptibility gene: a meta-analysis of multiple cancer types. *Carcinogenesis* 2005, 26(8):1368-1373.

82. Lubin JH, Gail MH: On power and sample size for studying features of the relative odds of disease. *Am J Epidemiol* 1990, 131(3):552-566.

83. O'Brien KM, Cole SR, Tse CK, Perou CM, Carey LA, Foulkes WD, Dressler LG, Geradts J, Millikan RC: Intrinsic breast tumor subtypes, race, and long-term survival in the Carolina Breast Cancer Study. *Clin Cancer Res* 2010, 16(24):6100-6110.

84. 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):12-15.

85. 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.

86. Age-Adjusted SEER Incidence Rates By Race/Ethnicity Female Breast, Ages<50, 2000-2010 (SEER18) [http://www.seer.cancer.gov/faststats/]

87. Zheng W, Cai Q, Signorello LB, Long J, Hargreaves MK, Deming SL, Li G, Li C, Cui Y, Blot WJ: Evaluation of 11 breast cancer susceptibility loci in African-American women. *Cancer Epidemiol Biomarkers Prev* 2009, 18(10):2761-2764.

88. Bhargava R, Striebel J, Beriwal S, Flickinger JC, Onisko A, Ahrendt G, Dabbs DJ: Prevalence, morphologic features and proliferation indices of breast carcinoma molecular classes using immunohistochemical surrogate markers. *Int J Clin Exp Pathol* 2009, 2(5):444-455.

89. Lee PH, Shatkay H: An integrative scoring system for ranking SNPs by their potential deleterious effects. Bioinformatics 2009, 25(8

Chapter 3. Results Manuscript 1: Genetic variation in cell cycle regulatory gene *AURKA* and association with intrinsic breast cancer subtype

3.1 Background

Previous research has established at least five distinct breast cancer subtypes that vary in their gene expression profiles and in their responsiveness to endocrine therapies [1-4]. Furthermore, risk factors for breast cancer have been shown to differ by intrinsic subtype [5], suggesting distinct etiologic and molecular pathways of carcinogenesis. Common low-penetrant susceptibility single nucleotide polymorphisms (SNPs) may play an important role in the etiology of breast cancer, individually conferring small increases in risk [6-10]. In aggregate, these increases in risk may become substantial [6-10]. *AURKA*, encoding a serine/threonine kinase (Aurora-A), is a putative oncogene that plays a role in cell cycle regulation [11]. Overexpression of *AURKA* has been associated with centrosomal duplication abnormalities, chromosomal instability and aneuploidy in mammalian cells, common characteristics of cancer cells [12,13]. *AURKA* overexpression has been demonstrated in several types of cancer and has been correlated with poor prognosis [14-16]. Previous studies of genetic variation in *AURKA* and risk of breast cancer have been largely limited to investigations of a single polymorphism (rs2273535) in Asian and Caucasian (Cau) populations, and none have focused on African Americans (AA). Some effect estimates among Asian and Cau populations were increased [17-21], some decreased [22], and some suggested no association [23,24]. These inconsistent results could be due to tumor heterogeneity and/or differences in population substructure. Importantly, these associations have not been previously investigated by breast cancer subtype, and this

approach could elucidate important subtype-specific associations, as has been shown in previous studies of other breast cancer risk factors [5,25-27].

We evaluated SNPs on *AURKA* in association with breast cancer rate in the Carolina Breast Cancer Study (CBCS), a large population-based case-control study of breast cancer in AA and Cau women in North Carolina. The CBCS allowed us to examine genetic risk factors given the increased incidence of breast cancer in younger AA women [28], as well as increased mortality and a preponderance of the basal-like subtype among AA women [25,29]. Capitalizing on the CBCS study design which oversampled African American women, we examined main effects of *AURKA* SNPs on breast cancer rate stratified by race. We also utilized the carefully characterized intrinsic subtype information in this study to evaluate *AURKA* genetic variation in association with specific intrinsic subtypes. This subtype-specific analysis is important because *AURKA* overexpression has been associated with aneuploidy and basal-like tumors have been shown to demonstrate a high degree of aneuploidy [30,31].

3.2 Methods

Study Population

The CBCS is a population-based, case-control study of genetic and environmental risk factors for breast cancer among AA and Cau women residing in North Carolina [32]. CBCS study design and methods have been previously described by Newman *et al.* [32]. Study participants were recruited and selected from 24 contiguous counties in central and eastern North Carolina [32]. CBCS recruitment was conducted in two phases—from 1993 through 1995 (Phase 1) and from 1996 through 2001 (Phase 2). Women living in the study area between the ages of 20 and 74 and diagnosed with invasive breast cancer for the first time were eligible cases

in Phase 1. CBCS Phase 2 included women diagnosed with *in situ* breast cancer (CIS) as well as those diagnosed with invasive breast cancer. Cases were identified using a rapid case ascertainment system via the North Carolina Central Cancer Registry (NCCCR). After eligibility criteria were met, randomized recruitment case sampling was undertaken to ensure adequate representation of AA and younger women [33]. Phase 2 CIS cases did not undergo random recruitment sampling; all eligible CIS cases were enrolled.

Controls were selected from two sources: women younger than 65 were selected from a list maintained by the North Carolina Division of Motor Vehicles; women between the ages of 65 and 74 were selected from Health Care Financing Administration records. Controls were sampled from these lists using modified randomized recruitment, and sampling fractions were designed to ensure frequency-matching of cases to controls by race and five-year age interval [33,34].

Potential cases and controls were contacted first by letter and then by telephone, if available. Women agreeing to participate were scheduled for an in-home visit by a registered nurse interviewer. The nurse interviewer collected anthropometric measurements, questionnaires, permission/consent to obtain tumor tissue, and a 30cc blood sample. Germline DNA was extracted from peripheral blood lymphocytes and stored at -80°C for future analysis [32]. The CBCS pathologist performed a standardized review of all breast tissue received to confirm the diagnosis of breast cancer and to characterize histology [32]. Slides were cut from paraffin blocks for molecular and immunohistochemical (IHC) assays, procedures for which have been described previously [29,35,36]. The study procedures for recruitment and enrollment into the CBCS were approved by the Institutional Review Board of the University of North Carolina (UNC), and all study participants gave written informed consent.

Subtyping of Cases by Immunohistochemistry (IHC)

For invasive cases, estrogen receptor (ER) and progesterone receptor (PR) status were primarily obtained from medical records (80%). Clinical laboratories determined ER/PR results on these cases. Approximately half of the clinical laboratories used IHC on paraffin-embedded tissue, and employed cutoffs for receptor positivity from more than 0% to more than 20%. The other half performed biochemical assays on frozen tissue with cutoffs for receptor positivity of 10-15 fmol/mg [36]. For approximately 11% of invasive cases, ER/PR status was not available in the medical record; however, paraffin-embedded tissue was available and ER/PR status was ascertained by the UNC IHC Core laboratory. For these cases, IHC scoring was based on UNC Hospitals Department of Pathology standards, using a cutoff of 5% positive nuclei staining in invasive breast cancer cells [29]. A random sample of ER+ and ER- cases based on medical record abstraction was drawn to compare with IHC performed by the UNC IHC Core laboratory. A kappa statistic of 0.62 and concordance of 81% resulted from the comparison, indicating good agreement [37]. Nine percent of invasive cases had missing data for ER/PR status [3].

CBCS intrinsic breast cancer subtypes were based on expression of ER, PR, human epidermal growth factor receptor 2 (HER2), cytokines (CK) 5/6, and human epidermal growth factor receptor (HER1) according to previously published definitions [29]. Tumors that were negative for expression of all five markers were unclassified. Negative staining for all markers is not necessarily indicative of receptor negativity in the tumor, and can result from poor tumor block quality or inadequate tissue present in the tumor block [29]. Tissue subtype analysis was performed in the following manner: HER2 status in invasive cases was determined using the CB11 monoclonal antibody as previously described [35]. HER2 positivity was defined by weak to strong staining of membrane or membrane plus cytoplasm in at least 10% of tumor cells [29].

Interscorer agreement of the HER2 IHC assay was evaluated on a subset of cases, yielding overall concordance of 82% [29]. HER1 and cytokeratin (CK) 5/6 characterization have been previously described [38,39], and invasive cases demonstrating any staining were classified as positive [29]. All assays for HER1, HER2, and CK5/6 were performed by the UNC IHC Core laboratory. ER, HER2, CK5/6, and HER1 classification and determination for CIS cases were described in detail previously [40]. PR status was not determined for CIS cases due to its high correlation with ER expression and to preserve tissue [25].

SNP Selection

SNPs in this study were genotyped as part of a larger panel of 1,536 SNPs by the UNC Mammalian Genotyping Core using the Illumina Golden Gate Assay (Illumina, San Diego, CA). Detailed genotyping procedures and quality control measures for the entire 1,536 SNP panel were described previously [41,42]. Assay intensity data and genotype cluster images for all SNPs were reviewed individually. To ensure quality control of genetic data, SNPs with low signal intensity or SNPs that were unable to be distinguished by genotype cluster were excluded. For each SNP, Hardy-Weinberg equilibrium (HWE) was evaluated in SAS v9.3 (SAS, Cary, NC) using a one-degree-of-freedom chi square exact test among race-stratified controls to determine if genotype frequencies were distributed as expected given the allele frequencies. Specifically for the evaluation of *AURKA*, a combination of tag and candidate SNPs were selected for genotyping. Tag SNPs were identified for Cau and AA from Utah residents with ancestry from northern and western Europe (CEU) and individuals of Yoruban descent from Idaban, Nigeria (YRI) HapMap populations respectively [43], and selected using the Tagger program developed by de Bakker *et al.* [44]. Tag SNPs were selected based on a linkage

disequilibrium (LD) $r^2 \geq 0.80$ and a minor allele frequency (MAF) of ≥0.10 in either CEU or YRI populations. Tag SNPs in each population were then combined and CBCS participants were genotyped for the pooled list. Candidate SNPs were chosen based on a literature review or previous GWAS association [41]. Five SNPs in *AURKA* were excluded from the overall analysis due to HWE *P*-values <0.05 in either AA or Cau (*N=3*) or because they were not polymorphic in the CBCS population (*N=2*); one SNP was excluded from the combined race subtype analysis because it was not polymorphic in Cau (rs34987347). Detailed genotyping procedures and quality control measures were described previously [41,42]. The software package *Structure* and a set of 144 ancestry informative markers (AIMs) were used to determine the proportion of African and European ancestry for each participant [41,45].

Statistical Analysis

Odds ratios (ORs) and 95% confidence intervals (CIs) were calculated, as estimates of the rate ratios [46], for genotype associations with breast cancer overall and by immunohistochemical (IHC) subtype using unconditional binary logistic regression in SAS v9.3 (SAS, Cary, NC). SNPs were coded using a dominant model, with the most common allele in Cau as the reference allele in both race groups for SNPs that were tags in both CEU and YRI HapMap populations to facilitate race comparisons. In race-stratified analyses, YRI tag SNPs that were not tag SNPs in the CEU population were analyzed in AA only, using the major allele in AA as the reference allele; likewise, CEU tag SNPs that were not tag SNPs in the YRI population were analyzed in Cau only, using the major allele in Cau as the reference allele. Candidate SNPs were analyzed in both race groups, using the major allele in Cau as the reference allele to facilitate race comparisons. Genotype associations were adjusted for age,

potential population stratification using the AIMs variable, and an offset term (defined as the natural log of recruitment probability of cases/recruitment probability of controls) to adjust for differing randomized recruitment sampling probabilities between phases of CBCS [41,47]. Subtype-specific analyses were performed in the combined race group rather than by race due to small sample numbers within strata of subtype, and were adjusted for self-identified race, age, the AIMs variable, and the offset term. Subtype-specific analyses included all tag and candidate SNPs, and assigned the major allele in Cau as the reference allele.

3.3 Results

Participant Characteristics

Among self-reported AA, the median proportion of African ancestry was 81%. The median proportion of African ancestry among self-reported Cau was 6%. Immunohistochemical subtype data was available for 1,412 of 2,277 (62%) cases, and successful genotyping data was collected for 1,946 of 2,277 (85%) cases. Of the 2,277 cases, 1,210 (53%) were successfully genotyped and subtyped (742 AA/1,204 Cau) (Table 1). The distribution of tumor subtype in cases with genotype data was as follows: 199 basal-like, 674 luminal A, 114 luminal B, 94 HER2+/ER-, and 129 unclassified (Table 1). Cases with missing subtype data were more likely to be Cau and have an earlier stage at diagnosis [25]. Of 1,985 controls, 1,747 (88%) were successfully genotyped (658 AA/1,089 Cau) (Table 1). Participants were excluded from analysis because of genotype calls for <95% of SNPs (*N=569*), gender mismatch (*N=5*), and suspected contamination of DNA specimen (*N=1*) [41]. Participants missing genotype data were more likely to be AA cases.

Genotype Associations

Here we focus on patterns to identify those SNPs for which the effect estimates were pronounced; and, we highlight estimates that were least influenced by chance (*i.e.* those estimates with the lowest confidence limit ratios (CLRs); the ratio of the upper to lower 95% confidence limits-a measure of precision [48]). Odds ratios for *AURKA* SNPs in the race-stratified analysis with breast cancer, not divided by subtype, were all close to 1.00 (Table 2). Among AA, rs6092309 showed a decreased odds ratio and rs911162 had a slightly elevated odds ratio with breast cancer. Table 3 presents the subtype-specific (race-combined) results. Rs6092309 and rs6099128 had decreased ORs for all subtypes, except the luminal B subtype which had imprecise effect measure estimates close to 1.00. Three *AURKA* SNPs (rs6014711, rs911162, rs1047972) had elevated ORs for basal-like breast cancer, and ORs reduced or close to 1.00 for all other subtypes. One SNP (rs16979826) showed a two-fold elevated odds ratio for HER2+/ER- breast cancer.

3.4 Discussion

Compared to previous studies, this study represents a more comprehensive investigation of *AURKA* related to breast cancer in a population of AA and Cau women. Previous studies of *AURKA* have focused largely on a few functional SNPs (rs2273535—Phe31Ile, rs1047972—Val57Ile) in Cau and Asian populations and have not investigated the influence of subtype. Our main finding was a decreased association between rs6092309 and breast cancer among AA women. Among Cau women this SNP led to an elevated but very imprecise odds ratio estimate because of a minor allele frequency of less than 1% in both Cau cases and controls. In the combined race group subtype-specific analysis, rs6092309 showed odds ratios less than one

across all subtypes. These results suggest that the association of *AURKA* genetic variation with subtype-specific breast cancer may differ by race. Rs6092309 is located within an intronic region of *AURKA*, is not predicted to be deleterious by SIFT or PolyPhen, and has not been previously studied with respect to breast cancer. Rs6092309 is in weak LD with other SNPs on *AURKA* in the HapMap YRI population (Release #27), demonstrates weak residual LD among SNPs genotyped in CBCS AA controls, and may be a marker for an ungenotyped genetic factor.

The importance of population stratification and race also emerged in subtype specific analyses, where there was evidence of heterogeneity in the relationships between *AURKA* SNPs and luminal A and basal-like breast cancer. Intronic SNPs rs2298016 and rs6099128 both demonstrated decreased odds ratios for basal-like breast cancer (Table 3). A population-based case-control study of breast cancer in Han Chinese women found rs2298016 to be inversely associated with breast cancer (OR = 0.52, 95% CI = 0.32-0.87, p = 0.01) [49]. However, the minor/test allele in the Han Chinese population was opposite that in the CBCS population and subtype-specific results were not reported in that study. Furthermore, rs2298016 was positively associated with both HER2+/ER- and unclassified breast cancer subtypes in CBCS cases. The instability of ORs for these SNPs across populations suggests significant differences in LD structure and/or different subtype distributions among the study populations. Allele and genotype frequencies for rs2298016 among AA cases and controls were comparable to those in Cau (Table 2.16), however LD structure was considerably different between races. This study was not powered to examine associations by race and breast cancer subtype, but exploratory subtype analysis of rs2298016 showed a decreased association between rs2298016 and basal-like breast cancer in AA (OR = 0.55, 95 % CI = 0.35-0.88), with weaker effects among Cau (OR = 0.81, 95 % CI = 0.51-1.28). Allele and genotype frequencies for rs6099128 among AA cases

and controls were also similar to those in Cau (Table 2.16), and LD structure was similar between races. Exploratory subtype analysis by race showed an odds ratio less than one for the association between rs6099128 and basal-like breast cancer among AA (OR = 0.45, 95 % CI = 0.27-0.75), with weaker effects in Cau (OR = 0.71, 95 % CI = 0.39-1.28). Rs6099128 was negatively associated with luminal A breast cancer; upon exploratory race-specific subtype analysis, a stronger negative association (OR = 0.68, 95% CI = 0.49-0.93) among Cau women compared to AA (OR = 0.86, 95 % CI = 0.61-1.20). These results should be considered in the context of small sample sizes and imprecise effect estimates, but may suggest race-specific differences by breast cancer subtype.

Several published studies have investigated the effects of missense SNP rs2273535 (Phe31Ile) and rs1047972 (Val57Ile) in association with breast cancer overall. Sun T. *et al.* found increased risk for breast carcinoma associated with the Ile/Ile genotype of rs2273535 (OR=1.66, 95% CI = 1.29-2.12) in a case-control study of unrelated Han Chinese women [17]. Additional studies of rs2273535 in both Chinese [18,23] and Cau [20] populations failed to replicate the finding. A 2011 meta-analysis of rs2273535, which included 11 case-control studies, reported a slight inverse association between the Ile/Ile genotype and risk of breast cancer (OR=0.86, 95% CI = 0.74-0.99), but only in Asian populations [50]. Our study found no association between rs2273535 and breast cancer overall in Cau or among AA women. We also found no association for rs2273535 among luminal A cases, and a slightly negative association with basal-like breast cancer. The coding region polymorphism rs1047972 on *AURKA* resulting in a valine to isoleucine substitution has also been heavily investigated for association with risk of breast cancer. Egan *et al.* reported no association with breast cancer risk among Cau women with the Ile/Ile genotype (OR = 0.92, 95 % CI = 0.50-1.71) in a population-based case-control

study [20]. Our study found no association between rs1047972 and breast cancer overall or luminal A breast cancer. However, an elevated odds ratio for rs1047972 and basal-like breast cancer was found (OR=1.34, 95%CI=0.97-1.85).

Limitations of this study include diminished statistical power to detect subtype-specific effects of *AURKA* due to small numbers of cases within strata of breast cancer subtype. Furthermore, whereas this study employed IHC to classify breast cancer subtypes, gene expression profiling using mRNA-based assays containing thousands of genes was originally used to characterize intrinsic breast cancer subtypes [2,3]. IHC assays do not provide as much information about tumor biology as mRNA-based expression assays do, and could result in misclassification of subtype [29]. However, IHC-based subtyping has been shown to identify common tumor subtypes with similar biologic characteristics, does not require fresh tissue, and has been widely used in population-based studies as a surrogate for gene expression profiling methods [29,51]. Although our study population was large, the effect sizes of *AURKA* SNP associations with breast cancer risk are likely small and thus more subtle main or subtype effects will require a much larger study sample to determine more accurate estimates. Additionally, sample sizes were not sufficient to reliably conduct subtype-specific race stratified analyses of *AURKA*. A third phase of the CBCS is underway to augment the number of AA cases with characterized tumor subtype, which will allow for further genetic evaluation to address this limitation. There was potential for selection bias to influence study results since 38% of cases were unable to be subtyped. However, genotyping distributions were similar between cases with and without subtype data (data not shown). Likewise, subtype distributions were similar between cases with and without genotyping data (Table 2.17). This suggests that the genotype distribution in cases with subtype data is likely representative of the genotype distribution in all

cases. Similarly, the subtype distribution in cases with genotype data is likely representative of the subtype distribution in all cases.

This study applied a candidate gene approach that was based on a plausible biological mechanism involving the cell-cycle regulatory gene *AURKA*, which is implicated in oncogensis [12,13,52]. Strengths of this study include (1) the availability of a comprehensive set of tag and candidate SNPs in *AURKA*, which improves our survey and coverage of this important oncogene, (2) inclusion of a relatively large number of AA women, (3) inclusion of 5-marker intrinsic subtype data based on the most current understanding of breast tumor heterogeneity, and (4) use of AIMS to adjust for population stratification, a factor which has been shown to impact effect estimates significantly if not controlled for [41].

In summary, these results represent the first comprehensive examination of *AURKA* SNPs in a population-based study with a large group of African American participants. Odds ratios for associations between *AURKA* SNPs and breast cancer overall were modest and consistent by race. Associations by intrinsic breast cancer subtype were relatively imprecise compared to overall estimates, but results were suggestive of decreased associations between a few *AURKA* SNPs and breast cancer subtype. Exploratory results also suggested race-specific effects within subtype. Given the likelihood of small effect sizes of *AURKA* SNPs on rate of breast cancer, evaluating subtype-specific effects in larger groups of AA and Cau women may better estimate the effect of *AURKA* on the rate of distinct breast cancer subtypes.

3.5 Tables

Table 3.1 Characteristics of CBCS participants with genotype data.

	Cases *(%)*	Controls *(%)*
N	1,946 *(100)*	1,747 *(100)*
Self-identified race		
African American	742 *(38.1)*	658 *(37.7)*
Caucasian	1,204 *(61.9)*	1,089 *(62.3)*
Age		
20-24	6 *(0.3)*	1 *(0.0)*
25-29	21 *(1.1)*	10 *(0.6)*
30-34	85 *(4.4)*	60 *(3.4)*
35-39	172 *(8.8)*	133 *(7.6)*
40-44	276 *(14.2)*	242 *(13.9)*
45-49	387 *(19.9)*	359 *(20.5)*
50-54	208 *(10.7)*	237 *(13.6)*
55-59	216 *(11.1)*	191 *(10.9)*
60-64	201 *(10.3)*	166 *(9.5)*
65-69	200 *(10.3)*	185 *(10.6)*
70-74	174 *(8.9)*	163 *(9.3)*
Menopausal Status		
Premenopausal	864 *(44.4)*	746 *(42.7)*
Postmenopausal	1,082 *(55.6)*	1,001 *(57.3)*
Stage		
1	609 *(31.3)*	
2	627 *(32.3)*	
3	144 *(7.4)*	
4	42 *(2.2)*	
CIS	437 *(22.5)*	
Missing[a]	87 *(4.5)*	
Tumor size[b]		
≤2 cm	769 *(51.0)*	
>2 – 5 cm	502 *(33.3)*	
>5 cm	146 *(9.7)*	
Missing	92 *(6.1)*	
Subtype		
Luminal A	674 *(34.6)*	
Luminal B	114 *(5.9)*	
HER2+/ER-	94 *(4.8)*	
Basal-like	199 *(10.2)*	
Unclassified	129 *(6.6)*	
Missing	736 *(37.8)*	

[a] Invasive breast cancer cases

[b] Not available for CIS (carcinoma *in situ*) cases

Table 3.2 Odds ratios (Ors) and 95% confidence intervals (CIs) for the association between single nucleotide polymorphisms (SNPs) on AURKA and all incident cases of breast cancer by race.

	Caucasian cases & controls	African American cases & controls
SNP	ORa (95% CI)	ORa (95% CI)
rs1468055[b]		
AC + AA	1.06 (0.88, 1.26)	0.95 (0.63, 1.41)
CC	Referent	Referent
rs1468056[b]		
CG + CC	1.09 (0.91, 1.30)	1.05 (0.75, 1.47)
GG	Referent	Referent
rs16979829[b]		
GT + GG	1.10 (0.77, 1.58)	0.97 (0.77, 1.23)
TT	Referent	Referent
rs2064863[b]		
AC + AA	1.00 (0.83, 1.20)	0.91 (0.50, 1.64)
CC	Referent	Referent
rs2180691[b]		
AG + AA	0.99 (0.84, 1.18)	0.90 (0.56, 1.45)
GG	Referent	Referent
rs2273535[b]		
TA + TT	1.00 (0.84, 1.20)	1.07 (0.84, 1.36)
AA	Referent	Referent
rs6099122[b]		
GT + GG	1.20 (0.79, 1.81)	0.86 (0.69, 1.07)
TT	Referent	Referent
rs6099128[b]		
GT + GG	0.85 (0.98, 1.06)	0.81 (0.64, 1.02)
TT	Referent	Referent
rs911162[b]		
AG + AA	0.82 (0.41, 1.67)	1.23 (0.82, 1.84)
GG	Referent	Referent
rs6014711[b]		
AG + AA	1.04 (0.86, 1.25)	0.97 (0.76, 1.23)
GG	Referent	Referent

Table 3.2 (cont.) Odds ratios (Ors) and 95% confidence intervals (CIs) for the association between single nucleotide polymorphisms (SNPs) on AURKA and all incident cases of breast cancer by race.

	Caucasian cases & controls	African American cases & controls
SNP	ORa (95% CI)	ORa (95% CI)
rs1047972[c]		
AG + AA	1.05 (0.87, 1.27)	0.97 (0.76, 1.24)
GG	Referent	Referent
rs6024836[c]		
AG + AA	1.02 (0.86, 1.22)	0.95 (0.67, 1.35)
GG	Referent	Referent
rs34987347[c,d]		
TC + TT		1.04 (0.34, 1.27)
CC		Referent
rs16979865[e]		
CA + CC		0.85 (0.64, 1.13)
AA		Referent
rs2298016[e]		
CG + CC		0.97 (0.78, 1.21)
GG		Referent
rs16979826[e]		
CT + CC		0.96 (0.73, 1.27)
TT		Referent
rs6092309[e]		
AG + AA		0.69 (0.53, 0.90)
GG		Referent
rs6099119[e]		
GA + GG		1.11 (0.77, 1.62)
AA		Referent
rs6099126[e]		
TC + TT		0.89 (0.71, 1.12)
CC		Referent
rs6024840[e]		
GA + GG		1.13 (0.88, 1.45)
AA		Referent

[a] Case-control odds ratio and 95% confidence interval adjusted for age, African ancestry

[b] Tag SNP in both CEU and YRI HapMap populations.

[c] Candidate SNP.

[d] Too few heterozygotes and homozygotes for the minor allele in

[e] Tag SNP in YRI HapMap population only.

Table 3.3 Odds ratios (ORs) and 95% confidence intervals (CIs) for the association between single nucleotide polymorphisms (SNPs) on *AURKA* and intrinsic subtype of breast cancer

	Luminal A (N_{cases}=674)	Luminal B (N_{cases}=114)	HER2+/ER- (N_{cases}=94)	Basal-like (N_{cases}=199)	Unclassified (N_{cases}=129)
SNP	OR[a] (95% CI)	OR[a] (95% CI)	OR[a] (95% CI)	OR[a] (95% CI)	OR[a] (95% CI)
rs1047972					
AG + AA	0.96 (0.78, 1.18)	0.75 (0.48, 1.18)	0.84 (0.52, 1.36)	1.34 (0.97, 1.85)	0.93 (0.62, 1.40)
GG	Referent	Referent	Referent	Referent	Referent
rs1468056					
CG + CC	1.28 (1.03, 1.59)	1.06 (0.68, 1.64)	1.04 (0.64, 1.70)	1.13 (0.78, 1.63)	0.90 (0.58, 4.41)
GG	Referent	Referent	Referent	Referent	Referent
rs16979826					
CT + CC	0.81 (0.54, 1.21)	0.69 (0.28, 1.72)	2.14 (1.06, 4.29)	0.88 (0.51, 1.50)	1.18 (0.65, 2.15)
TT	Referent	Referent	Referent	Referent	Referent
rs16979829					
GT + GG	0.92 (0.69, 1.23)	0.97 (0.54, 1.77)	1.05 (0.57, 1.92)	1.22 (0.82, 1.82)	1.03 (0.63, 1.69)
TT	Referent	Referent	Referent	Referent	Referent
rs16979865					
CA + CC	1.20 (0.93, 1.55)	0.71 (0.39, 1.31)	1.01 (0.56, 1.82)	0.91 (0.59, 1.41)	1.05 (0.64, 1.73)
AA	Referent	Referent	Referent	Referent	Referent
rs2180691					
AG + AA	0.96 (0.77, 1.20)	1.02 (0.65, 1.61)	1.25 (0.74, 2.11)	0.96 (0.64, 1.43)	1.01 (0.62, 1.65)
GG	Referent	Referent	Referent	Referent	Referent
rs2273535					
TA + TT	0.96 (0.78, 1.17)	1.30 (0.87, 1.94)	1.11 (0.71, 1.74)	0.81 (0.58, 1.13)	1.36 (0.92, 2.00)
AA	Referent	Referent	Referent	Referent	Referent
rs2298016					
CG + CC	0.96 (0.79, 1.16)	1.14 (0.77, 1.70)	1.38 (0.90, 2.11)	0.67 (0.48, 0.93)	1.34 (0.93, 1.95)
GG	Referent	Referent	Referent	Referent	Referent
rs6014711					
AG + AA	0.95 (0.77, 1.17)	0.74 (0.47, 1.17)	0.79 (0.48, 1.29)	1.33 (0.97, 1.84)	0.93 (0.61, 1.40)
GG	Referent	Referent	Referent	Referent	Referent
rs6024840					
GA + GG	0.95 (0.78, 1.16)	1.21 (0.80, 1.84)	1.28 (0.80, 2.05)	0.76 (0.54, 1.06)	1.00 (0.66, 1.50)
AA	Referent	Referent	Referent	Referent	Referent

Table 3.3 (cont.) Odds ratios (ORs) and 95% confidence intervals (CIs) for the association between single nucleotide polymorphisms (SNPs) on *AURKA* and intrinsic subtype of breast cancer

	Luminal A (N_{cases}=674)	Luminal B (N_{cases}=114)	HER2+/ER- (N_{cases}=94)	Basal-like (N_{cases}=199)	Unclassified (N_{cases}=129)
SNP	OR[a] (95% CI)	OR[a] (95% CI)	OR[a] (95% CI)	OR[a] (95% CI)	OR[a] (95% CI)
rs6092309					
AG + AA	0.61 (0.41, 0.92)	0.95 (0.43, 2.09)	0.69 (0.29, 1.61)	0.75 (0.44, 1.25)	0.62 (0.32, 1.20)
GG	Referent	Referent	Referent	Referent	Referent
rs6099122					
GT + GG	1.08 (0.81, 1.42)	1.06 (0.58, 1.92)	1.56 (0.86, 2.83)	1.15 (0.77, 1.71)	0.71 (0.44, 1.17)
TT	Referent	Referent	Referent	Referent	Referent
rs6099126					
TC + TT	1.05 (0.79, 1.39)	0.83 (0.45, 1.52)	1.35 (0.73, 2.49)	1.10 (0.73, 1.66)	0.86 (0.53, 1.41)
CC	Referent	Referent	Referent	Referent	Referent
rs6099128					
GT + GG	0.76 (0.60, 0.95)	1.09 (0.70, 1.69)	0.86 (0.52, 1.41)	0.54 (0.37, 0.80)	0.61 (0.38, 0.97)
TT	Referent	Referent	Referent	Referent	Referent
rs1468055					
AC + AA	1.01 (0.81, 1.27)	1.28 (0.82, 2.00)	0.80 (0.46, 1.37)	1.10 (0.75, 1.63)	0.79 (0.48, 1.30)
CC	Referent	Referent	Referent	Referent	Referent
rs6024836					
AG + AA	1.02 (0.82, 1.27)	0.90 (0.58, 1.39)	0.85 (0.52, 1.39)	0.85 (0.59, 1.24)	0.63 (0.41, 0.98)
GG	Referent	Referent	Referent	Referent	Referent
rs2064863					
AC + AA	0.87 (0.69, 1.11)	1.25 (0.76, 2.06)	1.12 (0.64, 1.97)	0.93 (0.61, 1.43)	0.94 (0.56, 1.58)
CC	Referent	Referent	Referent	Referent	Referent
rs6099119					
GA + GG	1.15 (0.68, 1.93)	0.28 (0.04, 2.11)	0.58 (0.13, 2.49)	1.61 (0.85, 3.06)	1.37 (0.62, 3.05)
AA	Referent	Referent	Referent	Referent	Referent
rs911162					
AG + AA	0.83 (0.50, 1.40)	0.85 (0.29, 2.47)	0.87 (0.30, 2.55)	1.32 (0.70, 2.51)	0.68 (0.26, 1.79)
GG	Referent	Referent	Referent	Referent	Referent

[a] Case-control odds ratio and 95% confidence interval adjusted for age, self-reported race, African ancestry and offset term

REFERENCES

1. Sorlie T, Tibshirani R, Parker J, Hastie T, Marron JS, Nobel A, Deng S, Johnsen H, Pesich R, Geisler S, Demeter J, Perou CM, Lonning PE, Brown PO, Borresen-Dale AL, Botstein D (2003) Repeated observation of breast tumor subtypes in independent gene expression data sets. Proc Natl Acad Sci U S A 100 (14):8418-8423. doi:10.1073/pnas.0932692100 0932692100 [pii]

2. Perou CM, Sorlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, Pollack JR, Ross DT, Johnsen H, Akslen LA, Fluge O, Pergamenschikov A, Williams C, Zhu SX, Lonning PE, Borresen-Dale AL, Brown PO, Botstein D (2000) Molecular portraits of human breast tumours. Nature 406 (6797):747-752. doi:10.1038/35021093

3. Sorlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, Hastie T, Eisen MB, van de Rijn M, Jeffrey SS, Thorsen T, Quist H, Matese JC, Brown PO, Botstein D, Eystein Lonning P, Borresen-Dale AL (2001) Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proc Natl Acad Sci U S A 98 (19):10869-10874. doi:10.1073/pnas.191367098 98/19/10869 [pii]

4. Perou CM, Jeffrey SS, van de Rijn M, Rees CA, Eisen MB, Ross DT, Pergamenschikov A, Williams CF, Zhu SX, Lee JC, Lashkari D, Shalon D, Brown PO, Botstein D (1999) Distinctive gene expression patterns in human mammary epithelial cells and breast cancers. Proc Natl Acad Sci U S A 96 (16):9212-9217

5. Yang XR, Sherman ME, Rimm DL, Lissowska J, Brinton LA, Peplonska B, Hewitt SM, Anderson WF, Szeszenia-Dabrowska N, Bardin-Mikolajczak A, Zatonski W, Cartun R, Mandich D, Rymkiewicz G, Ligaj M, Lukaszek S, Kordek R, Garcia-Closas M (2007) Differences in risk factors for breast cancer molecular subtypes in a population-based study. Cancer Epidemiol Biomarkers Prev 16 (3):439-443. doi:16/3/439 [pii]10.1158/1055-9965.EPI-06-0806

6. Lichtenstein P, Holm NV, Verkasalo PK, Iliadou A, Kaprio J, Koskenvuo M, Pukkala E, Skytthe A, Hemminki K (2000) Environmental and heritable factors in the causation of cancer--analyses of cohorts of twins from Sweden, Denmark, and Finland. N Engl J Med 343 (2):78-85. doi:10.1056/NEJM200007133430201

7. Antoniou AC, Pharoah PD, McMullan G, Day NE, Stratton MR, Peto J, Ponder BJ, Easton DF (2002) A comprehensive model for familial breast cancer incorporating BRCA1, BRCA2 and other genes. Br J Cancer 86 (1):76-83. doi:10.1038/sj.bjc.6600008

8. Pharoah PD, Antoniou A, Bobrow M, Zimmern RL, Easton DF, Ponder BA (2002) Polygenic susceptibility to breast cancer and implications for prevention. Nat Genet 31 (1):33-36. doi:10.1038/ng853ng853 [pii]

9. Antoniou AC, Pharoah PD, McMullan G, Day NE, Ponder BA, Easton D (2001) Evidence for further breast cancer susceptibility genes in addition to BRCA1 and BRCA2 in a population-based study. Genet Epidemiol 21 (1):1-18. doi:10.1002/gepi.1014 [pii]10.1002/gepi.1014

10. Antoniou AC, Easton DF (2003) Polygenic inheritance of breast cancer: Implications for design of association studies. Genet Epidemiol 25 (3):190-202. doi:10.1002/gepi.10261

11. Glover DM, Leibowitz MH, McLean DA, Parry H (1995) Mutations in aurora prevent centrosome separation leading to the formation of monopolar spindles. Cell 81 (1):95-105. doi:0092-8674(95)90374-7 [pii]

12. Zhou H, Kuang J, Zhong L, Kuo WL, Gray JW, Sahin A, Brinkley BR, Sen S (1998) Tumour amplified kinase STK15/BTAK induces centrosome amplification, aneuploidy and transformation. Nat Genet 20 (2):189-193. doi:10.1038/2496

13. Hoque A, Carter J, Xia W, Hung MC, Sahin AA, Sen S, Lippman SM (2003) Loss of aurora A/STK15/BTAK overexpression correlates with transition of in situ to invasive ductal carcinoma of the breast. Cancer Epidemiol Biomarkers Prev 12 (12):1518-1522

14. Bischoff JR, Anderson L, Zhu Y, Mossie K, Ng L, Souza B, Schryver B, Flanagan P, Clairvoyant F, Ginther C, Chan CS, Novotny M, Slamon DJ, Plowman GD (1998) A homologue of Drosophila aurora kinase is oncogenic and amplified in human colorectal cancers. EMBO J 17 (11):3052-3065. doi:10.1093/emboj/17.11.3052

15. Staff S, Isola J, Jumppanen M, Tanner M (2010) Aurora-A gene is frequently amplified in basal-like breast cancer. Oncol Rep 23 (2):307-312

16. Lukasiewicz KB, Lingle WL (2009) Aurora A, centrosome structure, and the centrosome cycle. Environ Mol Mutagen 50 (8):602-619. doi:10.1002/em.20533

17. Sun T, Miao X, Wang J, Tan W, Zhou Y, Yu C, Lin D (2004) Functional Phe31Ile polymorphism in Aurora A and risk of breast carcinoma. Carcinogenesis 25 (11):2225-2230. doi:10.1093/carcin/bgh244bgh244 [pii]

18. Dai Q, Cai QY, Shu XO, Ewart-Toland A, Wen WQ, Balmain A, Gao YT, Zheng W (2004) Synergistic effects of STK15 gene polymorphisms and endogenous estrogen exposure in the risk of breast cancer. Cancer Epidemiol Biomarkers Prev 13 (12):2065-2070. doi:13/12/2065 [pii]

19. Cox DG, Hankinson SE, Hunter DJ (2006) Polymorphisms of the AURKA (STK15/Aurora Kinase) Gene and Breast Cancer Risk (United States). Cancer Causes Control 17 (1):81-83. doi:10.1007/s10552-005-0429-9

20. Egan KM, Newcomb PA, Ambrosone CB, Trentham-Dietz A, Titus-Ernstoff L, Hampton JM, Kimura MT, Nagase H (2004) STK15 polymorphism and breast cancer risk in a population-based study. Carcinogenesis 25 (11):2149-2153. doi:10.1093/carcin/bgh231bgh231 [pii]

21. Vidarsdottir L, Bodvarsdottir SK, Hilmarsdottir H, Tryggvadottir L, Eyfjord JE (2007) Breast cancer risk associated with AURKA 91T -->A polymorphism in relation to BRCA mutations. Cancer Lett 250 (2):206-212. doi:S0304-3835(06)00556-8 [pii]10.1016/j.canlet.2006.10.003

22. Fletcher O, Johnson N, Palles C, dos Santos Silva I, McCormack V, Whittaker J, Ashworth A, Peto J (2006) Inconsistent association between the STK15 F31I genetic polymorphism and breast cancer risk. J Natl Cancer Inst 98 (14):1014-1018. doi:98/14/1014 [pii]10.1093/jnci/djj268

23. Lo YL, Yu JC, Chen ST, Yang HC, Fann CS, Mau YC, Shen CY (2005) Breast cancer risk associated with genotypic polymorphism of the mitosis-regulating gene Aurora-A/STK15/BTAK. Int J Cancer 115 (2):276-283. doi:10.1002/ijc.20855

24. Tchatchou S, Wirtenberger M, Hemminki K, Sutter C, Meindl A, Wappenschmidt B, Kiechle M, Bugert P, Schmutzler RK, Bartram CR, Burwinkel B (2007) Aurora kinases A and B and familial breast cancer risk. Cancer Lett 247 (2):266-272. doi:S0304-3835(06)00376-4 [pii]10.1016/j.canlet.2006.05.002

25. Millikan RC, Newman B, Tse CK, Moorman PG, Conway K, Dressler LG, Smith LV, Labbok MH, Geradts J, Bensen JT, Jackson S, Nyante S, Livasy C, Carey L, Earp HS, Perou CM (2008) Epidemiology of basal-like breast cancer. Breast Cancer Res Treat 109 (1):123-139. doi:10.1007/s10549-007-9632-6

26. Kristensen VN, Borresen-Dale AL (2008) 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 2 (1):12-15. doi:10.1016/j.molonc.2008.02.003S1574-7891(08)00027-6 [pii]

27. Nordgard SH, Johansen FE, Alnaes GI, Naume B, Borresen-Dale AL, Kristensen VN (2007) Genes harbouring susceptibility SNPs are differentially expressed in the breast cancer subtypes. Breast Cancer Res 9 (6):113. doi:bcr1784 [pii]10.1186/bcr1784

28. Altekruse SF KC, Krapcho M, Neyman N, Aminou R, Waldron W, Ruhl J, Howlader N, Tatalovich Z, Cho H, Mariotto A, Eisner MP, Lewis DR, Cronin K, Chen HS, Feuer EJ, Stinchcomb DG, Edwards BK (2010) SEER Stat Fact Sheets: Breast. National Cancer Institute. http://seer.cancer.gov/statfacts/html/breast.html. Accessed 11/16/2010 2010

29. Carey LA, Perou CM, Livasy CA, Dressler LG, Cowan D, Conway K, Karaca G, Troester MA, Tse CK, Edmiston S, Deming SL, Geradts J, Cheang MC, Nielsen TO, Moorman PG, Earp HS, Millikan RC (2006) Race, breast cancer subtypes, and survival in the Carolina Breast Cancer Study. JAMA 295 (21):2492-2502. doi:295/21/2492 [pii]10.1001/jama.295.21.2492

30. Foulkes WD, Smith IE, Reis-Filho JS (2010) Triple-negative breast cancer. N Engl J Med 363 (20):1938-1948. doi:10.1056/NEJMra1001389

31. Anders CK, Winer EP, Ford JM, Dent R, Silver DP, Sledge GW, Carey LA (2010) Poly(ADP-Ribose) polymerase inhibition: "Targeted" therapy for triple-negative breast cancer. Clin Cancer Res 16 (19):4702-4710. doi:1078-0432.CCR-10-0939 [pii]10.1158/1078-0432.CCR-10-0939

32. Newman B, Moorman PG, Millikan R, Qaqish BF, Geradts J, Aldrich TE, Liu ET (1995) The Carolina Breast Cancer Study: integrating population-based epidemiology and molecular biology. Breast Cancer Res Treat 35 (1):51-60

33. Weinberg CR, Sandler DP (1991) Randomized recruitment in case-control studies. Am J Epidemiol 134 (4):421-432

34. Weinberg CR, Wacholder S (1990) The design and analysis of case-control studies with biased sampling. Biometrics 46 (4):963-975

35. Millikan R, Eaton A, Worley K, Biscocho L, Hodgson E, Huang WY, Geradts J, Iacocca M, Cowan D, Conway K, Dressler L (2003) HER2 codon 655 polymorphism and risk of breast cancer in African Americans and whites. Breast Cancer Res Treat 79 (3):355-364

36. Huang WY, Newman B, Millikan RC, Schell MJ, Hulka BS, Moorman PG (2000) Hormone-related factors and risk of breast cancer in relation to estrogen receptor and progesterone receptor status. Am J Epidemiol 151 (7):703-714

37. Landis JR, Koch GG (1977) The measurement of observer agreement for categorical data. Biometrics 33 (1):159-174

38. Nielsen TO, Hsu FD, Jensen K, Cheang M, Karaca G, Hu Z, Hernandez-Boussard T, Livasy C, Cowan D, Dressler L, Akslen LA, Ragaz J, Gown AM, Gilks CB, van de Rijn M, Perou CM (2004) Immunohistochemical and clinical characterization of the basal-like subtype of invasive breast carcinoma. Clin Cancer Res 10 (16):5367-5374. doi:10.1158/1078-0432.CCR-04-022010/16/5367 [pii]

39. van de Rijn M, Perou CM, Tibshirani R, Haas P, Kallioniemi O, Kononen J, Torhorst J, Sauter G, Zuber M, Kochli OR, Mross F, Dieterich H, Seitz R, Ross D, Botstein D, Brown P (2002) Expression of cytokeratins 17 and 5 identifies a group of breast carcinomas with poor clinical outcome. Am J Pathol 161 (6):1991-1996. doi:S0002-9440(10)64476-8 [pii]10.1016/S0002-9440(10)64476-8

40. Livasy CA, Perou CM, Karaca G, Cowan DW, Maia D, Jackson S, Tse CK, Nyante S, Millikan RC (2007) Identification of a basal-like subtype of breast ductal carcinoma in situ. Hum Pathol 38 (2):197-204. doi:S0046-8177(06)00534-X [pii]10.1016/j.humpath.2006.08.017

41. Barnholtz-Sloan JS, Shetty PB, Guan X, Nyante SJ, Luo J, Brennan DJ, Millikan RC (2010) FGFR2 and other loci identified in genome-wide association studies are associated with breast cancer in African-American and younger women. Carcinogenesis 31 (8):1417-1423. doi:bgq128 [pii]10.1093/carcin/bgq128

42. Nyante SJ, Gammon MD, Kaufman JS, Bensen JT, Lin DY, Barnholtz-Sloan JS, Hu Y, He Q, Luo J, Millikan RC (2011) Common genetic variation in adiponectin, leptin, and leptin receptor and association with breast cancer subtypes. Breast Cancer Res Treat. doi:10.1007/s10549-011-1517-z

43. www.hapmap.org.

44. de Bakker PI, Yelensky R, Pe'er I, Gabriel SB, Daly MJ, Altshuler D (2005) Efficiency and power in genetic association studies. Nat Genet 37 (11):1217-1223. doi:ng1669 [pii]10.1038/ng1669

45. Barnholtz-Sloan JS, Chakraborty R, Sellers TA, Schwartz AG (2005) Examining population stratification via individual ancestry estimates versus self-reported race. Cancer Epidemiol Biomarkers Prev 14 (6):1545-1551. doi:14/6/1545 [pii]10.1158/1055-9965.EPI-04-0832

46. Pearce N (1993) What does the odds ratio estimate in a case-control study? Int J Epidemiol 22 (6):1189-1192

47. Pfaff CL, Barnholtz-Sloan J, Wagner JK, Long JC (2004) Information on ancestry from genetic markers. Genet Epidemiol 26 (4):305-315. doi:10.1002/gepi.10319

48. Poole C (2001) Low P-values or narrow confidence intervals: which are more durable? Epidemiology 12 (3):291-294

49. Ruan Y, Song AP, Wang H, Xie YT, Han JY, Sajdik C, Tian XX, Fang WG (2011) Genetic polymorphisms in AURKA and BRCA1 are associated with breast cancer susceptibility in a Chinese Han population. J Pathol. doi:10.1002/path.2902

50. Sun H, Bai J, Chen F, Jin Y, Yu Y, Fu S (2011) Lack of an association between AURKA T91A polymorphisms and breast cancer: a meta-analysis involving 32,141 subjects. Breast Cancer Res Treat 125 (1):175-179. doi:10.1007/s10549-010-0936-6

51. Selvin S (1996) A note on the power to detect interaction effects. In: Kesley JM, M.; Stolley, P.; Vessey, M. (ed) Statistical Analysis of Epidemiologic Data. Oxford University Press, New York

52. Balczon R, Bao L, Zimmer WE, Brown K, Zinkowski RP, Brinkley BR (1995) Dissociation of centrosome replication events from cycles of DNA synthesis and mitotic division in hydroxyurea-arrested Chinese hamster ovary cells. J Cell Biol 130 (1):105-115

Chapter 4. Results Manuscript 2: Genetic variation in *BRCA1* and *BRCA1*-interacting genes in association with intrinsic breast cancer subtype

4.1 Background

Mutations in *BRCA1* are likely to account for fewer than 10% of hereditary cases of breast cancer and between 1-2% of all breast cancers [1]. The large proportion of unexplained risk may depend on unidentified genetic traits, environmental risk factors, or a combination of both. There is considerable debate as to which of these factors predominates, and the magnitude of the genetic contribution to the causation of breast cancer remains unclear [2, 3]. Twin-studies and studies of familial inheritance have suggested that common, low penetrance genetic factors may account for residual familial risk [2, 4]. The polygenic model proposes that genetic susceptibility to breast cancer is not entirely predicted by rare, highly penetrant genes but more often stems from several common loci that each confer smaller increases in risk [4-7]. Under this model it would be rare to observe multiple-affected case families (as is the case for those demonstrating highly penetrant mutations in genes such as *BRCA1*) since multiple individuals in a family would each have to inherit several different less penetrant susceptibility variants.

Studies have shown that the tumor suppressor activity of *BRCA1* influences several pathways, including DNA damage repair and cell cycle regulation [8]. Through these pathways, BRCA1 interacts with numerous other proteins that are important for cell cycle progression [8]. Due to the prominent role *BRCA1* plays throughout the cell cycle and the relatively high risk for breast cancer conferred by mutations in it, *BRCA1* and lesser penetrant genes encoding *BRCA1*-interacting proteins are logical targets for further investigation [9]. *BARD1*, *BRIP1* and *ZNF350*

are three putative low penetrance breast cancer susceptibility genes whose protein products are known to interact with BRCA1. The *BARD1* protein markedly enhances the tumor suppression activity of *BRCA1* by forming a heterodimer *BRCA1/BARD1* complex [10]. Mutations in *BRCA1* are known to deactivate this heterodimer complex [10, 11], suggesting a role for *BARD1* in DNA repair processes. *BRIP1* encodes a helicase that binds to the C-terminus of *BRCA1*, contributing to its double-strand break repair function [12], and has previously been identified as a potential breast cancer susceptibility gene [13]. *ZNF350* and *BRCA1* are corepressors of *GADD45*, which is involved in cell cycle arrest at the G2/M checkpoint subsequent to DNA damage [14, 15]. *ZNF350* has been associated with breast cancer risk in previous DNA repair pathway-based studies of breast cancer [16, 17].

Based on the plausible etiologic role for genetic variants of three *BRCA1*-interacting genes, we investigated tag and candidate single nucleotide polymorphisms (SNPs) on these genes and their associations with breast cancer and intrinsic breast cancer subtype. We used data from the Carolina Breast Cancer Study (CBCS), a large population-based case-control study of breast cancer in African American (AA) and Caucasian (Cau) women in North Carolina. We also examined associations between candidate SNPs on *BRCA1* and breast cancer. In addition, and building on our previous investigation of another *BRCA1*-interacting gene, *AURKA*, we explored gene—gene interactions between candidate SNPs on *AURKA* and candidate SNPs on each of *BRCA1*, *BARD1*, *BRIP1*, and *ZNF350*. Several epidemiologic studies have examined common genetic variation in *BRCA1* and *BRCA1*-interacting genes in association with breast cancer risk [16-18], but none of them were conducted in large groups of African Americans or by breast cancer subtype. Taking advantage of the CBCS study design, which oversampled AA women and classified samples as to intrinsic subtype, we estimated the association of *BRCA1*

and *BRCA1*-interacting SNPs on overall breast cancer and with stratification by race (all cases and controls) and by subtype (race-combined).

4.2 Methods

Study Population

The CBCS is a population-based, case-control study of genetic and environmental risk factors for breast cancer among AA and Cau women residing in North Carolina [19]. CBCS study design and methods have been previously described by Newman *et al.* [19]. Study participants were recruited and selected from 24 contiguous counties in central and eastern North Carolina [19]. CBCS recruitment was conducted in two phases—from 1993 through 1995 (Phase 1) and from 1996 through 2001 (Phase 2). Women living in the study area between the ages of 20 and 74 and diagnosed with invasive breast cancer for the first time were eligible cases in Phase 1. CBCS Phase 2 included women diagnosed with *in situ* breast cancer (CIS) as well as those diagnosed with invasive breast cancer. Cases were identified using a rapid case ascertainment system via the North Carolina Central Cancer Registry (NCCCR). After eligibility criteria were met, randomized recruitment case sampling was undertaken to ensure adequate representation of AA and younger women [20]. Phase 2 CIS cases were not included in random recruitment sampling and all eligible CIS cases were enrolled.

Controls were selected from two sources: women younger than 65 were selected from a list maintained by the North Carolina Division of Motor Vehicles; women between the ages of 65 and 74 were selected from Health Care Financing Administration records. Controls were sampled from these lists using modified randomized recruitment, and sampling fractions were

designed to ensure frequency-matching of cases to controls by race and five-year age interval [20, 21].

Potential cases and controls were contacted first by letter and then by telephone, if available. Women agreeing to participate were scheduled for an in-home visit by a registered nurse interviewer. The nurse interviewer collected anthropometric measurements, questionnaires, permission/consent to obtain tumor tissue, and a 30cc blood sample. Germline DNA was extracted from peripheral blood lymphocytes and stored at -80°C for future analysis [19]. The CBCS pathologist performed a standardized review of all breast tissue received to confirm the diagnosis of breast cancer and to characterize histology [19]. Slides were cut from paraffin blocks for molecular and immunohistochemical (IHC) assays, procedures for which have been described previously [22-24]. The study procedures for recruitment and enrollment into the CBCS were approved by the Institutional Review Board of the University of North Carolina, Chapel Hill (UNC), and all study participants gave written informed consent.

Subtyping of Cases by Immunohistochemistry (IHC)

For invasive cases, estrogen receptor (ER) and progesterone receptor (PR) status were primarily obtained from medical records (80%). Clinical laboratories determined ER/PR results on these cases. Approximately half of the clinical laboratories used IHC on paraffin-embedded tissue, and employed cutoffs for receptor positivity from more than 0% to more than 20%. The other half performed biochemical assays on frozen tissue with cutoffs for receptor positivity of 10-15 fmol/mg [24]. For approximately 11% of invasive cases, ER/PR status was not available in the medical record; however, paraffin-embedded tissue was available and ER/PR status was ascertained by the UNC IHC Core laboratory. For these cases, IHC scoring was based on UNC

Hospitals Department of Pathology standards, using a cutoff of 5% positive nuclei staining in invasive breast cancer cells [22]. A random sample of ER+ and ER- cases based on medical record abstraction was drawn to compare with IHC performed by the UNC IHC Core laboratory. A kappa statistic of 0.62 and concordance of 81% resulted from the comparison, indicating good agreement [25]. Nine percent of invasive cases had missing data for ER/PR status [22].

CBCS intrinsic breast cancer subtypes were based on expression of ER, PR, human epidermal growth factor receptor 2 (HER2), cytokines (CK) 5/6, and human epidermal growth factor receptor 1 (HER1) according to previously published definitions [22]. Tumors that were negative for expression of all five markers were unclassified. Negative staining for all markers is not necessarily indicative of receptor negativity in the tumor, and can result from poor tumor block quality or inadequate tissue present in the tumor block [22]. HER2, CK5/6 and HER1 assays were performed by the UNC IHC Core laboratory (IC). Tissue subtype analysis was performed in the following manner: HER2 status in invasive cases was determined using the CB11 monoclonal antibody as previously described [23]. HER2 positivity was defined by weak to strong staining of membrane or membrane plus cytoplasm in at least 10% of tumor cells [22]. Interscorer agreement of the HER2 IHC assay was evaluated on a subset of cases, yielding overall concordance of 82% [22]. HER1 and cytokeratin (CK) 5/6 characterization have been previously described [26, 27], and invasive cases demonstrating any staining were classified as positive [22]. ER, HER2, CK5/6, and HER1 classification and determination for CIS cases were described in detail previously [28]. PR status was not determined for CIS cases due to its high correlation with ER expression and to preserve tissue [29].

SNP Selection

SNPs in this study were genotyped as part of a larger panel of 1,536 SNPs by the UNC Mammalian Genotyping Core using the Illumina Golden Gate Assay (Illumina, San Diego, CA). Detailed genotyping procedures and quality control measures for the entire 1,536 SNP panel were described previously [30, 31]. Assay intensity data and genotype cluster images for all SNPs were reviewed individually. To ensure quality control of genetic data, SNPs with low signal intensity or SNPs that were unable to be distinguished by genotype cluster were excluded. For each SNP, Hardy-Weinberg equilibrium (HWE) was evaluated in SAS v9.3 (SAS, Cary, NC) using a one-degree-of-freedom chi square exact test among race-stratified controls to determine if genotype frequencies were distributed as expected given the allele frequencies. For the evaluation of *AURKA, BRCA1* and *BRCA1*-interacting genes, a combination of tag and candidate SNPs were selected for genotyping. Tag SNPs were identified for Cau and AA from CEU (Utah residents with ancestry from northern and western Europe) and YRI (individuals of Yoruban descent from Idaban, Nigeria) HapMap populations respectively [32], and selected using the Tagger program developed by de Bakker *et al.* [33]. Tag SNPs were selected based on a linkage disequilibrium (LD) $r^2 \geq 0.80$ and a minor allele frequency (MAF) of ≥0.10 in either CEU or YRI populations. Tag SNPs in each population were then combined and CBCS participants were genotyped for the pooled list. Candidate SNPs were chosen based on a literature review and previous GWAS hits [30]. Twelve SNPs on *BRCA1* and the *BRCA1*-interacting genes were excluded due to HWE *P*-values <0.05 in either AA or Cau (*N=10*) or because they were not polymorphic in the CBCS population (*N=2*) (Table 2.16). Six SNPs on *AURKA* were excluded from consideration for the interaction analysis due to HWE *P*-values <0.05 in either AA or Cau (*N=3*) or because they were not polymorphic in the CBCS population

(*N=3*). One SNP on *BARD1* was excluded from the combined race subtype analysis because it was not polymorphic in African Americans (rs28997576). Detailed genotyping procedures and quality control measures were described previously [30, 31]. The software package *Structure* and a set of 144 ancestry informative markers (AIMs) were used to determine the proportion of African ancestry for each participant [30, 34].

SNPs were chosen for inclusion into the interaction analysis based on a decision tree (Figure 2.7). The primary criterion for inclusion was based on the likelihood that a SNP was functional. Likelihood of SNP functionality was determined using the FS Score, an integrative *in silico* scoring system for assessing potential SNP functionality based on protein coding, splicing regulation, transcriptional regulation, and post-translation [35]. SNPs demonstrating FS Scores of ≥0.50 were included in the interaction study.

Statistical Analysis

We used multivariable logistic regression to identify patterns among SNPs having effect estimates that were most different from the null; and, were least influenced by chance (*i.e.* those estimates with the lowest confidence limit ratios (CLRs); the ratio of the upper to lower 95% confidence limits-a measure of precision [41]). We defined relatively good estimate precision to correspond to a CLR of ≤3.0. Odds ratios (ORs), as estimates of rate ratios [36], and 95% confidence intervals (CIs) were calculated for genotype associations with breast cancer overall, by race, and by immunohistochemical (IHC) subtype using unconditional binary logistic regression in SAS v9.3 (SAS, Cary, NC). SNPs were coded using a dominant model, with the most common allele in Cau as the reference allele in both race groups for SNPs that were tags in both CEU and YRI HapMap populations to facilitate race comparisons. In race-stratified

analyses, YRI tag SNPs that were not tag SNPs in the CEU population were analyzed in AA only, using the major allele in AA as the reference allele; likewise, CEU tag SNPs that were not tag SNPs in the YRI population were analyzed in Cau only, using the major allele in Cau as the reference allele. Candidate SNPs were analyzed in both race groups, using the major allele in Cau as the reference allele to facilitate race comparisons. Genotype associations were adjusted for age, potential population stratification using the AIMs variable, and an offset term (defined as the natural log of recruitment probability of cases/recruitment probability of controls) to adjust for differing randomized recruitment sampling probabilities between phases of CBCS [41,47]. Subtype-specific analyses were performed in the combined race group rather than by race due to small sample numbers within strata of subtype, and were adjusted for self-identified race, age, the AIMs variable, and the offset term. Subtype-specific analyses included all tag and candidate SNPs and assigned the major allele in Cau as the reference allele. Additive interaction between selected SNPs on *AURKA*, *BRCA1*, and *BRCA1*-interacting genes was assessed using the relative excess risk due to interaction (RERI) based on the formula $RERI=OR_{11} - OR_{01} - OR_{10} + 1$ [39], with 95% confidence intervals calculated based on the method proposed by Hosmer and Lemeshow [40].

4.3 Results

Participation

Among self-reported AA, the median proportion of African ancestry was 81%. The median proportion of African ancestry among self-reported Cau was 6%. Immunohistochemical subtype data was available for 1,412 of 2,277 (62%) cases, and successful genotyping data was collected for 1,946 of 2,277 (85%) cases. Of the 2,277 cases, 1,210 (53%) were successfully genotyped

and assigned a breast cancer intrinsic subtype (742 AA/1,204 Cau) (Table 3.1). The distribution of tumor subtype in cases with genotype data was as follows: 199 basal-like, 674 luminal A, 114 luminal B, 94 HER2+/ER-, and 129 unclassified (Table 3.1). Cases with missing subtype data were more likely to be Cau and have an earlier stage at diagnosis [29]. Of 1,985 controls, 1,747 (88%) were successfully genotyped (658 AA/1,089 Cau) (Table 3.1). Participants were excluded from analysis if genotype calls were missing for ≥95% of SNPs (*N=569*), gender was mismatched (*N=5*), or due to suspected contamination of DNA specimen (*N=1*) [30]. Participants missing genotype data were more likely to be AA cases.

Genotype Associations

Odds ratios for SNPs on *BRCA1* and breast cancer were all close to 1.00 among AA (Table 4.1). Among Cau, three SNPs on *BRCA1* (rs16941, rs16942, and rs1799966) showed positive associations with breast cancer and had relatively good estimate precision. The majority of SNPs on *BARD1* had ORs close to 1.00 in AA and Cau. Rs16852799 on *BARD1* had an inverse association with breast cancer among AA (OR=0.87, 95% CI: 0.68-1.13) and Cau (OR=0.75, 95% CI: 0.58-0.98). One *BARD1* SNP (rs28997576: OR=1.42, 95% CI: 1.00-2.03) showed an elevated OR among Cau but was not polymorphic among African American CBCS participants (not shown in Table 4.1). Among AA, three *BRIP1* SNPs (rs4986764, rs7213430, and rs1978111) had inverse associations with breast cancer, with relatively good estimate precision. Results for *BRIP1* SNPs among Cau were consistent with little or no association. Similarly, ORs for SNPs on *ZNF350* were all close to 1.00 for both AA and Cau.

Table 4.2 presents intrinsic breast cancer subtype-specific (race-combined) results. Three SNPs on *BRCA1* (rs16941, rs16942, and rs1799966) had inverse associations with HER2+/ER-

breast cancer and positive associations with luminal A and basal-like subtypes, with relatively good estimate precision. Two *BRCA1* SNPs (rs1799950 and rs799923) had reduced ORs for luminal A breast cancer and increased ORs for luminal B breast cancer. Rs1799950 also showed inverse associations with HER2+/ER- and basal-like breast cancer. Exploratory race-stratified analysis of rs1777950 showed an inverse association with basal-like breast cancer among Cau (OR=0.15, 95%CI: 0.04-0.61), with results in AA imprecise, but consistent with no association (OR=1.02, 95%CI: 0.30-3.46). Exploratory analysis of rs799923 by race and subtype showed an inverse association with luminal A breast cancer and a positive association with the basal-like subtype among Cau (luminal A OR=0.76, 95%CI: 0.60-0.97; luminal B OR=1.76, 95%CI: 1.10-2.84).

Two SNPs on *BARD1* (rs16852761 and rs3768704) had positive associations with luminal B breast cancer and negative associations with basal-like breast cancer, with ORs closer to 1.00 for other subtypes. Three *BARD1* SNPs (rs12474696, rs2075622, and rs2888294) had decreased ORs for luminal A breast cancer and elevated ORs for basal-like breast cancer. One *BARD1* SNP (rs16852799) showed decreased ORs for both luminal A and basal-like subtypes.

Three SNPs on *BRIP1* (rs4986764, rs7213430, and rs1978111) were positively associated with HER2+/ER- breast cancer and negatively associated with basal-like breast cancer. Two SNPs on *ZNF350* (rs2278420 and rs4988334) showed elevated ORs for luminal B and HER2+/ER- breast cancer. Assessment of additive interactions between candidate SNPs on *AURKA* and *BRCA1* and *BRCA1*-interacting genes using RERI yielded results that were consistent with little or no departure from additivity (Tables 4.3-4.6).

4.4 Discussion

We estimated associations between tag and candidate SNPs on *BRCA1* and *BRCA1*-interacting genes and rate of breast cancer overall and by intrinsic subtype using data from a large population based case-control study. Our main findings were positive associations between three candidate SNPs on *BRCA1* (rs16941, rs16942, and rs1799966) and breast cancer overall that demonstrated relatively good estimate precision among Cau women. All three nonsynonymous missense SNPs demonstrated estimates of similar magnitude and precision, which is likely due to the high degree of LD between them ($r^2 \geq 0.90$). Furthermore, among the HapMap CEU population (Release #27), all three SNPs are also in high LD with 36 other SNPs on *BRCA1* that were not genotyped in the CBCS population. It is possible that these three coding SNPs along with other SNPs in LD alter function of *BRCA1* and together are responsible for the associations we observed. Both rs16941 and rs1799966 are predicted to be deleterious by SIFT and showed FS scores of ≥0.5, indicting a strong probability of functionality. Rs16942 was also predicted to be functional, with an FS score of 0.9, but was predicted to be tolerated by SIFT. Using data on *BRCA1* mutation carriers from the Consortium of Investigators of Modifiers of *BRCA1/2* (CIMBA), Cox *et al.* reported a decreased risk of breast cancer among women carrying the minor allele of rs16942 on the wild-type copy of *BRCA1* (hazard ratio=0.86, 95%CI: 0.77-0.95), contrary to our results. The proportion of CBCS case participants with disease-related *BRCA1* mutations is largely unknown, but its population-based design and a previous study by Newman *et al.* suggests it may be small [42]. This difference may account for contrasting results reported by Cox *et al.* Furthermore, the study group evaluated by Cox *et al.* (CIMBA) is a hospital-based study comprised of study participants from 18 different countries that may not be comparable to the CBCS study population.

Subtype-specific analyses suggested heterogeneity in the relationship between genotyped SNPs and intrinsic breast cancer subtype. *BRCA1* missense SNP rs1799950, resulting in a glutamine to arginine amino acid change, has been examined in large population-based studies of breast cancer. Baynes *et al.* reported an inverse association between rs1799950 and overall breast cancer risk in a large case-control study of European women (OR=0.63, 95%CI: 0.23-1.23) [43]. Similar results were reported by Dunning *et al.* in a case-control study of Caucasian women from the United Kingdom [44]. These prior results are consistent with our overall findings for rs1799950 among Caucasians. Subtype-specific results also indicated inverse associations between rs1799950 and luminal A, HER2+/ER-, and basal-like breast cancers in CBCS participants. Since genotype and allele frequencies for rs1777950 were similar between races, we conducted exploratory subtype analyses stratified by race (data not shown). Race-stratified estimates for luminal A breast cancer were similar in magnitude and direction to those reported in our combined race subtype-specific analysis. An inverse association was also noted between rs1799950 and HER2+/ER- breast cancer among Cau, with inadequate cell sizes precluding calculation of an estimate in AA. Intronic *BRCA1* SNP rs799923 demonstrated an inverse association with luminal A breast cancer and a positive association with luminal B breast cancer. Exploratory subtype analysis by race showed this pattern repeated among Cau, with highly imprecise estimates among AA. Rs799923 is not in high LD with any other SNP on *BRCA1* in Cau. Three intronic SNPs on *BARD1* (rs12474696, rs2075622, and rs6749828) demonstrated positive associations with basal-like breast cancer. Exploratory subtype analysis by race demonstrated the same pattern of association in both Cau and AA. All three SNPs are predicted to be nonfunctional by FS score and are in high LD with other SNPS that were not

genotyped by CBCS. These exploratory results should be considered in the context of the imprecise effect estimates, but may suggest race-specific differences by breast cancer subtype.

Limitations of this study include limited statistical power to detect intrinsic breast cancer subtype-specific associations of candidate genes due to small numbers of cases within these strata. Although our study population was large, effect sizes of selected candidate gene SNP associations with breast cancer risk are likely small, and thus more subtle main or subtype effects will require a much larger study sample to estimate accurately. Additionally, samples sizes were not sufficient to reliably conduct subtype-specific race-stratified analyses. A third phase of the CBCS is underway to augment the number of AA cases with tumor subtype data, which will allow for better powered genetic analyses. There was potential for selection bias to influence study results since 38% of cases were unable to be subtyped. However, genotyping distributions were similar between cases with and without subtype data. Likewise, subtype distributions were similar between cases with and without genotyping data (data not shown). This suggests that the genotype distribution in cases with subtype data is likely representative of the genotype distribution in all cases. Similarly, the subtype distribution in cases with genotype data is likely representative of the subtype distribution in all cases.

This study applied a candidate gene approach that was based on plausible biological oncogenic mechanisms involving candidate genes *BRCA1, BARD1, BRIP1,* and *ZNF350*. Additional strengths of this study include (1) inclusion of a relatively large number of AA women, which are drawn from the largest case-control study of breast cancer among AA women available to date; (2) inclusion of 5-marker intrinsic subtype data, and (3) use of AIMS to adjust for population stratification [30].

In summary, we observed positive associations between breast cancer and three candidate SNPs on *BRCA1* (rs16941, rs16942, and rs1799966). These results represent the first candidate gene study of genetic variation in *BARD1*, *BRIP1*, and *ZNF350* in a population-based study with a large group of African American participants. Odds ratios for associations between SNPs on these candidate genes and breast cancer overall were close to 1.00 and consistent by race. Associations by intrinsic breast cancer subtype were relatively imprecise compared to overall estimates, but results were suggestive of differential associations between candidate genes and intrinsic breast cancer subtype. Exploratory results also suggested race-specific effects within subtype. Given the likelihood of small effect sizes of candidate gene SNPs on rate of breast cancer, evaluating subtype-specific effects in larger groups of AA and Cau women may better estimate the effects of genetic variation in *BRCA1* and *BRCA1*-interacting genes on the rate of distinct breast cancer subtypes.

4.5 Tables

Table 4.1 Odds ratios (ORs) and 95% confidence intervals (CIs) for the association between single nucleotide polymorphisms (SNPs) on BRCA1 and BRCA1-interacting genes and all incident cases of breast cancer by race

	Caucasian cases & controls	African American cases & controls
SNP	ORa (95% CI)	ORa (95% CI)
BRCA1		
rs1799950[c]		
AG+GG	0.89 (0.68, 1.17)	0.94 (0.48, 1.83)
AA	Referent	Referent
rs16941[c]		
AG+GG	1.24 (1.04, 1.48)	1.04 (0.83, 1.31)
AA	Referent	Referent
rs16942[c]		
AG+GG	1.24 (1.04, 1.48)	1.10 (0.88, 1.37)
AA	Referent	Referent
rs1799966[c]		
AG+GG	1.25 (1.05, 1.48)	1.11 (0.89, 1.38)
AA	Referent	Referent
rs799917[c]		
CT+TT	1.06 (0.80, 1.40)	1.06 (0.83, 1.34)
CC	Referent	Referent
rs4986852[c]		
CT+TT	0.98 (0.66, 1.45)	1.07 (0.38, 3.02)
CC	Referent	Referent
rs799923[f]		
GA+AA	1.05 (0.88, 1.26)	
GG	Referent	
BARD1		
rs12477063[b]		
TC+CC	1.10 (0.87, 1.40)	1.10 (0.88, 1.37)
TT	Referent	Referent
rs6751923[b]		
TC+CC	0.83 (0.68, 1.01)	1.13 (0.91, 1.41)
TT	Referent	Referent
rs1542173[b]		
AG+GG	1.16 (0.97, 1.39)	1.11 (0.89, 1.38)
AA	Referent	Referent

Table 4.1 (cont.) Odds ratios (ORs) and 95% confidence intervals (CIs) for the association between single nucleotide polymorphisms (SNPs) on BRCA1 and BRCA1-interacting genes and all incident cases of breast cancer by race

	Caucasian cases & controls	African American cases & controls
SNP	ORa (95% CI)	ORa (95% CI)
rs16852761[b]		
GA+AA	0.92 (0.57, 1.49)	0.99 (0.75, 1.31)
GG	Referent	Referent
rs16852798[b]		
CT+TT	0.85 (0.55, 1.32)	1.00 (0.77, 1.28)
CC	Referent	Referent
rs16852799[b]		
AG+GG	0.75 (0.58, 0.98)	0.87 (0.68, 1.13)
AA	Referent	Referent
rs2075622[b]		
CA+AA	1.00 (0.84, 1.20)	0.95 (0.76, 1.19)
CC	Referent	Referent
rs2888294[b]		
CG+GG	0.93 (0.77, 1.13)	0.96 (0.74, 1.24)
CC	Referent	Referent
rs3768704[b]		
GA+AA	1.07 (0.88, 1.31)	1.02 (0.79, 1.30)
GG	Referent	Referent
rs4672729[b]		
GA+AA	1.08 (0.91, 1.29)	1.06 (0.85, 1.32)
GG	Referent	Referent
rs6749828[b]		
GC+CC	0.99, (0.79, 1.24)	1.09 (0.87, 1.36)
GG	Referent	Referent
rs1979028[c]		
TA+AA	1.07 (0.90, 1.28)	1.05 (0.84, 1.31)
TT	Referent	Referent
rs7585356[c]		
GA+AA	1.07 (0.90, 1.28)	0.92 (0.73, 1.16)
GG	Referent	Referent

Table 4.1 (cont.) Odds ratios (ORs) and 95% confidence intervals (CIs) for the association between single nucleotide polymorphisms (SNPs) on BRCA1 and BRCA1-interacting genes and all incident cases of breast cancer by race

	Caucasian cases & controls	African American cases & controls
SNP	ORa (95% CI)	ORa (95% CI)
rs1048108[c]		
GA+AA	1.03 (0.86, 1.23)	1.02 (0.81, 1.27)
GG	Referent	Referent
rs28997576[c,d]		
GC+CC	1.42 (1.00, 2.03)	
GG	Referent	
rs2229571[c]		
CG+GG	0.99 (0.82, 1.18)	0.93 (0.70, 1.24)
CC	Referent	Referent
rs10221582[e]		
CT+TT		0.81 (0.63, 1.04)
CC		Referent
rs10932573[e]		
TC+CC		1.06 (0.85, 1.33)
TT		Referent
rs12474696[e]		
AG+GG		1.00 (0.80, 1.24)
AA		Referent
rs3768707[e]		
CT+TT		1.09 (0.87, 1.35)
CC		Referent
rs3768708[e]		
AG+GG		1.02 (0.81, 1.28)
AA		Referent
rs3820727[e]		
TG+GG		0.94 (0.75, 1.18)
TT		Referent
rs6756902[e]		
TC+CC		1.11 (0.89, 1.39)
TT		Referent

Table 4.1 (cont.) Odds ratios (ORs) and 95% confidence intervals (CIs) for the association between single nucleotide polymorphisms (SNPs) on BRCA1 and BRCA1-interacting genes and all incident cases of breast cancer by race

	Caucasian cases & controls	African American cases & controls
SNP	ORa (95% CI)	ORa (95% CI)
rs7557557[e]		
CT+TT		1.04 (0.84, 1.29)
CC		Referent
rs7566806[e]		
GC+CC		1.04 (0.82, 1.32)
GG		Referent
rs6753417[e]		
CT+TT		1.02 (0.82, 1.28)
CC		Referent
rs10932568[f]		
AC+CC	0.91 (0.76, 1.08)	
AA	Referent	
rs17487827[f]		
CG+GG	1.18 (0.99, 1.42)	
CC	Referent	
BRIP1		
rs4986764[c]		
CT+TT	1.07 (0.89, 1.28)	0.83 (0.67, 1.03)
CC	Referent	Referent
rs7213430[c]		
AG+GG	1.09 (0.91, 1.30)	0.82 (0.66, 1.03)
AA	Referent	Referent
rs4988351[c]		
GC+CC	1.17 (0.98, 1.40)	1.15 (0.90, 1.47)
GG	Referent	Referent
rs1978111[c]		
CT+TT	1.06 (0.89, 1.27)	0.81 (0.65, 1.01)
CC	Referent	Referent

Table 4.1 (cont.) Odds ratios (ORs) and 95% confidence intervals (CIs) for the association between single nucleotide polymorphisms (SNPs) on BRCA1 and BRCA1-interacting genes and all incident cases of breast cancer by race

	Caucasian cases & controls	African American cases & controls
SNP	ORa (95% CI)	ORa (95% CI)
ZNF350		
rs4986773[c]		
TC+CC	1.08 (0.91, 1.29)	1.18 (0.86, 1.61)
TT	Referent	Referent
rs2278420[c]		
AG+GG	1.10 (0.91, 1.33)	0.98 (0.78, 1.22)
AA	Referent	Referent
rs11879758[c]		
GC+CC	1.02 (0.84, 1.25)	1.06 (0.82, 1.37)
GG	Referent	Referent
rs2278417[c]		
CT+TT	1.08 (0.91, 1.29)	1.16 (0.85, 1.60)
CC	Referent	Referent
rs4986770[c]		
CT+TT	0.96 (0.74, 1.24)	1.09 (0.82, 1.46)
CC	Referent	Referent
rs4988334[c]		
TC+CC	1.15 (0.84, 1.40)	0.93 (0.75, 1.15)
TT	Referent	Referent

[a] Case-control odds ratio and 95% confidence interval adjusted for age, African ancestry, and offset term

[b] Tag SNP in both CEU and YRI HapMap populations

[c] Candidate SNP

[d] Too few heterozygotes and homozygotes for the minor allele in African Americans

[e] Tag SNP in YRI HapMap population only

[f] Tag SNP in CEU HapMap population only

Table 4.2 Odds ratios (ORs) and 95% confidence intervals (CIs) for the association between single nucleotide polymorphisms (SNPs) on BRCA1 and BRCA1-interacting genes and breast cancer subtype

	Luminal A (N_{CASES}= 674)	Luminal B (N_{CASES}= 114)	HER2+/ER- (N_{CASES}= 94)	Basal-like (N_{CASES}= 199)	Unclassified (N_{CASES}= 129)
SNP	OR[a] (95% CI)	OR[a] (95% CI)	OR[a] (95% CI)	OR[a] (95% CI)	OR[a] (95% CI)
BRCA1					
rs1799950					
AG+GG	0.62 (0.42, 0.90)	1.36 (0.74, 2.51)	0.55 (0.51, 1.41)	0.33 (0.14, 0.79)	1.43 (0.76, 2.72)
AA	Referent	Referent	Referent	Referent	Referent
rs16941					
AG+GG	1.17 (0.97, 1.42)	1.02 (0.69, 1.51)	0.64 (0.41, 1.01)	1.21 (0.88, 1.65)	0.86 (0.58, 1.27)
AA	Referent	Referent	Referent	Referent	Referent
rs16942					
AG+GG	1.15 (0.95, 1.39)	1.01 (0.68, 1.49)	0.69 (0.44, 1.06)	1.19 (0.88, 1.63)	0.96 (0.66, 1.40)
AA	Referent	Referent	Referent	Referent	Referent
rs1799966					
AG+GG	1.17 (0.97, 1.42)	1.00 (0.68, 1.48)	0.71 (0.46, 1.10)	1.20 (0.88, 1.64)	0.95 (0.65, 1.38)
AA	Referent	Referent	Referent	Referent	Referent
rs799917					
CT+TT	1.10 (0.86, 1.42)	1.12 (0.65, 1.90)	1.32 (0.74, 2.35)	0.88 (0.60, 1.30)	0.78 (0.49, 1.24)
CC	Referent	Referent	Referent	Referent	Referent
rs4986852					
CT+TT	1.01 (0.61, 1.67)	1.67 (0.69, 4.04)	0.33 (0.05, 2.47)	1.13 (0.46, 2.73)	1.25 (0.43, 3.61)
CC	Referent	Referent	Referent	Referent	Referent
rs799923					
GA+AA	0.79 (0.63, 0.99)	1.58 (1.02, 2.45)	1.08 (0.65, 1.79)	0.95 (0.64, 1.41)	1.43 (0.91, 2.25)
GG	Referent	Referent	Referent	Referent	Referent
BARD1					
rs10932568					
AC+CC	1.05 (0.87, 1.28)	0.88 (0.58, 1.32)	1.10 (0.71, 1.71)	0.86 (0.63, 1.20)	0.86 (0.58, 1.27)
AA	Referent	Referent	Referent	Referent	Referent
rs10221582					
CT+TT	1.10 (0.89, 1.35)	0.86 (0.57, 1.30)	0.71 (0.46, 1.11)	1.02 (0.73, 1.43)	0.82 (0.55, 1.22)
CC	Referent	Referent	Referent	Referent	Referent

Table 4.2 (cont.) Odds ratios (ORs) and 95% confidence intervals (CIs) for the association between single nucleotide polymorphisms (SNPs) on BRCA1 and BRCA1-interacting genes and breast cancer subtype

	Luminal A (N_{CASES}= 674)	Luminal B (N_{CASES}= 114)	HER2+/ER- (N_{CASES}= 94)	Basal-like (N_{CASES}= 199)	Unclassified (N_{CASES}= 129)
SNP	OR[a] (95% CI)	OR[a] (95% CI)	OR[a] (95% CI)	OR[a] (95% CI)	OR[a] (95% CI)
rs10932573					
TC+CC	1.05 (0.85, 1.29)	1.10 (0.71, 1.69)	1.17 (0.73, 1.89)	1.29 (0.62, 1.82)	1.08 (0.72, 1.62)
TT	Referent	Referent	Referent	Referent	Referent
rs12474696					
AG+GG	0.78 (0.64, 0.94)	1.00 (0.67, 1.49)	0.95 (0.61, 1.46)	1.40 (1.01, 1.93)	1.03 (0.70, 1.51)
AA	Referent	Referent	Referent	Referent	Referent
rs12477063					
TC+CC	1.21 (0.96, 1.52)	0.93 (0.59, 1.47)	1.29 (0.77, 2.16)	1.06 (0.75, 1.51)	1.32 (0.85, 2.03)
TT	Referent	Referent	Referent	Referent	Referent
rs1542173					
AG+GG	1.04 (0.86, 1.26)	1.00 (0.68, 1.49)	1.20 (0.78, 1.84)	1.24 (0.91, 1.69)	1.25 (0.86, 1.82)
AA	Referent	Referent	Referent	Referent	Referent
rs16852761					
GA+AA	1.14 (0.81, 1.59)	1.39 (0.70, 2.77)	0.87 (0.38, 1.97)	0.75 (0.44, 1.30)	0.87 (0.46, 1.63)
GG	Referent	Referent	Referent	Referent	Referent
rs16852798					
CT+TT	0.95 (0.69, 1.30)	0.82 (0.41, 1.65)	0.78 (0.37, 1.60)	0.76 (0.47, 1.23)	1.27 (0.76, 2.14)
CC	Referent	Referent	Referent	Referent	Referent
rs16852799					
AG+GG	0.71 (0.55, 0.93)	0.84 (0.49, 1.44)	1.08 (0.63, 1.84)	0.74 (0.49, 1.13)	1.30 (0.84, 2.01)
AA	Referent	Referent	Referent	Referent	Referent
rs17487827					
CG+GG	0.98 (0.81, 1.20)	0.89 (0.59, 1.36)	1.09 (0.70, 1.69)	1.32 (0.96, 1.81)	0.95 (0.64, 1.42)
CC	Referent	Referent	Referent	Referent	Referent
rs1979028					
TA+AA	1.18 (0.98, 1.43)	0.79 (0.54, 1.18)	1.02 (0.66, 1.56)	1.06 (0.77, 1.44)	1.17 (0.80, 1.71)
TT	Referent	Referent	Referent	Referent	Referent
rs2075622					
CA+AA	0.86 (0.71, 1.05)	1.06 (0.71, 1.59)	0.89 (0.58, 1.38)	1.38 (1.00, 1.90)	0.94 (0.64, 1.38)
CC	Referent	Referent	Referent	Referent	Referent

Table 4.2 (cont.) Odds ratios (ORs) and 95% confidence intervals (CIs) for the association between single nucleotide polymorphisms (SNPs) on BRCA1 and BRCA1-interacting genes and breast cancer subtype

	Luminal A (N_{CASES}= 674)	Luminal B (N_{CASES}= 114)	HER2+/ER- (N_{CASES}= 94)	Basal-like (N_{CASES}= 199)	Unclassified (N_{CASES}= 129)
SNP	OR[a] (95% CI)	OR[a] (95% CI)	OR[a] (95% CI)	OR[a] (95% CI)	OR[a] (95% CI)
rs2888294					
CG+GG	0.79 (0.64, 0.98)	0.86 (0.56, 1.34)	0.87 (0.53, 1.40)	1.18 (0.82, 1.71)	1.18 (0.75, 1.84)
CC	Referent	Referent	Referent	Referent	Referent
rs3768704					
GA+AA	1.10 (0.88, 1.37)	1.51 (0.99, 2.30)	1.02 (0.62, 1.68)	0.72 (0.49, 1.06)	1.19 (0.79, 1.81)
GG	Referent	Referent	Referent	Referent	Referent
rs3768707					
CT+TT	1.01 (0.84, 1.22)	0.99 (0.67, 1.47)	1.17 (0.76, 1.80)	1.23 (0.90, 1.67)	1.22 (0.84, 1.77)
CC	Referent	Referent	Referent	Referent	Referent
rs3768708					
AG+GG	0.99 (0.81, 1.20)	1.02 (0.68, 1.52)	1.00 (0.65, 1.55)	1.21 (0.88, 1.66)	1.19 (0.81, 1.75)
AA	Referent	Referent	Referent	Referent	Referent
rs3820727					
TG+GG	1.14 (0.93, 1.39)	0.90 (0.60, 1.37)	1.25 (0.79, 1.98)	1.30 (0.94, 1.81)	0.81 (0.55, 1.21)
TT	Referent	Referent	Referent	Referent	Referent
rs4672729					
GA+AA	0.95 (0.79, 1.15)	1.10 (0.75, 1.64)	1.08 (0.71, 1.66)	1.12 (0.82, 1.53)	1.22 (0.83, 1.78)
GG	Referent	Referent	Referent	Referent	Referent
rs6749828					
GC+CC	1.03 (0.83, 1.29)	0.88 (0.56, 1.37)	1.27 (0.76, 2.11)	1.35 (0.94, 1.93)	1.04 (0.69, 1.58)
GG	Referent	Referent	Referent	Referent	Referent
rs6751923					
TC+CC	0.94 (0.77, 1.15)	1.04 (0.69, 1.58)	0.98 (0.62, 1.54)	0.84 (0.61, 1.17)	1.09 (0.74, 1.60)
TT	Referent	Referent	Referent	Referent	Referent
rs6753417					
CT+TT	1.15 (0.95, 1.39)	0.97 (0.65, 1.44)	1.07 (0.69, 1.64)	1.21 (0.89, 1.66)	1.03 (0.70, 1.50)
CC	Referent	Referent	Referent	Referent	Referent
rs6756902					
TC+CC	1.21 (0.96, 1.52)	0.93 (0.59, 1.47)	1.29 (0.77, 2.16)	1.07 (0.75, 1.52)	1.32 (0.86, 2.03)
TT	Referent	Referent	Referent	Referent	Referent

Table 4.2 (cont.) Odds ratios (ORs) and 95% confidence intervals (CIs) for the association between single nucleotide polymorphisms (SNPs) on BRCA1 and BRCA1-interacting genes and breast cancer subtype

	Luminal A (N_{CASES}= 674)	Luminal B (N_{CASES}= 114)	HER2+/ER- (N_{CASES}= 94)	Basal-like (N_{CASES}= 199)	Unclassified (N_{CASES}= 129)
SNP	OR[a] (95% CI)	OR[a] (95% CI)	OR[a] (95% CI)	OR[a] (95% CI)	OR[a] (95% CI)
rs7557557					
CT+TT	1.12 (0.91, 1.38)	1.00 (0.65, 1.52)	1.21 (0.76, 1.95)	1.23 (0.88, 1.71)	0.95 (0.64, 1.40)
CC	Referent	Referent	Referent	Referent	Referent
rs7566806					
GC+CC	1.07 (0.87, 1.31)	1.09 (0.71, 1.67)	0.97 (0.60, 1.57)	0.73 (0.51, 1.05)	1.13 (0.76, 1.69)
GG	Referent	Referent	Referent	Referent	Referent
rs7585356					
GA+AA	1.09 (0.90, 1.32)	0.77 (0.51, 1.16)	1.15 (0.75, 1.77)	1.01 (0.74, 1.39)	1.07 (0.73, 1.58)
GG	Referent	Referent	Referent	Referent	Referent
rs1048108					
GA+AA	1.11 (0.91, 1.34)	0.94 (0.63, 1.39)	1.03 (0.67, 1.59)	1.13 (0.82, 1.54)	1.02 (0.69, 1.49)
GG	Referent	Referent	Referent	Referent	Referent
rs3738888					
GA+AA	0.96 (0.36, 2.53)	2.08 (0.45, 9.69)	2.65 (0.57, 12.33)	1.39 (0.30, 6.50)	----
GG	Referent	Referent	Referent	Referent	Referent
rs2229571					
CG+GG	0.86 (0.70, 1.06)	1.16 (0.74, 1.80)	0.88 (0.55, 1.40)	1.24 (0.85, 1.80)	1.42 (0.89, 2.25)
CC	Referent	Referent	Referent	Referent	Referent
BRIP1					
rs4986764					
CT+TT	1.01 (0.83, 1.22)	0.91 (0.61, 1.35)	1.30 (0.83, 2.05)	0.85 (0.63, 1.17)	0.76 (0.52, 1.10)
CC	Referent	Referent	Referent	Referent	Referent
rs7213430					
AG+GG	1.01 (0.84, 1.23)	0.94 (0.63, 1.39)	1.36 (0.87, 2.14)	0.86 (0.63, 1.17)	0.75 (0.52, 1.09)
AA	Referent	Referent	Referent	Referent	Referent
rs4988351					
GC+CC	1.09 (0.90, 1.33)	0.98 (0.65, 1.48)	0.91 (0.57, 1.43)	0.99 (0.71, 1.38)	1.73 (1.18, 2.55)
GG	Referent	Referent	Referent	Referent	Referent
rs1978111					
CT+TT	0.96 (0.79, 1.17)	0.93 (0.63, 1.39)	1.32 (0.84, 2.08)	0.84 (0.61, 1.15)	0.74 (0.51, 1.08)
CC	Referent	Referent	Referent	Referent	Referent

Table 4.2 (cont.) Odds ratios (ORs) and 95% confidence intervals (CIs) for the association between single nucleotide polymorphisms (SNPs) on BRCA1 and BRCA1-interacting genes and breast cancer subtype

	Luminal A (N_{CASES}= 674)	Luminal B (N_{CASES}= 114)	HER2+/ER- (N_{CASES}= 94)	Basal-like (N_{CASES}= 199)	Unclassified (N_{CASES}= 129)
SNP	OR[a] (95% CI)	OR[a] (95% CI)	OR[a] (95% CI)	OR[a] (95% CI)	OR[a] (95% CI)
ZNF350					
rs4986773					
TC+CC	1.09 (0.89, 1.35)	1.04 (0.67, 1.59)	1.04 (0.64, 1.67)	1.08 (0.75, 1.54)	1.28 (0.82, 2.01)
TT	Referent	Referent	Referent	Referent	Referent
rs2278420					
AG+GG	0.99 (0.81, 1.21)	1.31 (0.86, 1.98)	1.47 (0.94, 2.31)	0.93 (0.67, 1.29)	1.23 (0.83, 1.81)
AA	Referent	Referent	Referent	Referent	Referent
rs11879758					
GC+CC	1.04 (0.84, 1.30)	1.24 (0.81, 1.91)	0.99 (0.61, 1.63)	1.16 (0.82, 1.64)	0.72 (0.45, 1.14)
GG	Referent	Referent	Referent	Referent	Referent
rs2278417					
CT+TT	1.09 (0.88, 1.34)	1.08 (0.70, 1.65)	1.03 (0.64, 1.65)	1.03 (0.72, 1.47)	1.37 (0.87, 2.13)
CC	Referent	Referent	Referent	Referent	Referent
rs4986770					
CT+TT	0.87 (0.66, 1.14)	0.83 (0.46, 1.47)	1.21 (0.69, 2.12)	0.93 (0.60, 1.43)	0.93 (0.55, 1.55)
CC	Referent	Referent	Referent	Referent	Referent
rs4988334					
TC+CC	0.95 (0.77, 1.16)	1.44 (0.96, 2.18)	1.32 (0.84, 2.06)	0.97 (0.70, 1.35)	1.27 (0.86, 1.87)
TT	Referent	Referent	Referent	Referent	Referent

[a] Case-control odds ratio and 95% confidence interval adjusted for age, self-identified race, African ancestry and offset term

Table 4.3 Additive interaction analysis between select SNPs on *AURKA* and *BRCA1*

AURKA-BRCA1

1.	SNP1	SNP2				
	rs2273535	rs1799950	OR	RD	RERI	95% CI
00	AA	AA	1.00	0.00		
01	AA	GA+GG	1.09	**0.09**	0.01	(-0.57-0.60)
10	TA+TT	AA	1.03	0.00		
11	TA+TT	GA+GG	1.14	**0.11**		
	SNP1	SNP2				
	rs1799950	rs2273535	OR	RD	RERI	95% CI
00	AA	AA	1.00	0.00		
01	AA	TA+TT	1.03	**0.03**	0.01	(-0.57-0.60)
10	GA+GG	AA	1.09	0.00		
11	GA+GG	TA+TT	1.14	**0.05**		

2.	SNP1	SNP2				
	rs2273535	rs16941	OR	RD	RERI	95% CI
00	AA	AA	1.00	0.00		
01	AA	GA+GG	1.27	**0.27**	-0.28	(-0.62-0.06)
10	TA+TT	AA	1.15	0.00		
11	TA+TT	GA+GG	1.14	**-0.01**		
	SNP1	SNP2				
	rs16941	rs2273535	OR	RD	RERI	95% CI
00	AA	AA	1.00	0.00		
01	AA	TA+TT	1.15	**0.15**	-0.28	(-0.62-0.06)
10	GA+GG	AA	1.27	0.00		
11	GA+GG	TA+TT	1.14	**-0.13**		

3.	SNP1	SNP2				
	rs2273535	rs16942	OR	RD	RERI	95% CI
00	AA	AA	1.00	0.00		
01	AA	GA+GG	1.27	**0.27**	-0.25	(-0.60-0.09)
10	TA+TT	AA	1.15	0.00		
11	TA+TT	GA+GG	1.17	**0.02**		
	SNP1	SNP2				
	rs16942	rs2273535	OR	RD	RERI	95% CI
00	AA	AA	1.00	0.00		
01	GA+GG	TA+TT	1.15	**0.15**	-0.25	(-0.60-0.09)
10	AA	AA	1.27	0.00		
11	GA+GG	TA+TT	1.17	**-0.10**		

4.	SNP1	SNP2				
	rs2273535	rs1799966	OR	RD	RERI	95% CI
00	AA	AA	1.00	0.00		
01	AA	GA+GG	1.28	**0.28**	-0.25	(-0.60-0.09)
10	TA+TT	AA	1.15	0.00		
11	TA+TT	GA+GG	1.18	**0.03**		
	SNP1	SNP2				
	rs1799966	rs2273535	OR	RD	RERI	95% CI
00	AA	AA	1.00	0.00		
01	GA+GG	TA+TT	1.15	**0.15**	-0.25	(-0.60-0.09)
10	AA	AA	1.28	0.00		
11	GA+GG	TA+TT	1.18	**-0.10**		

5.	SNP1	SNP2				
	rs2273535	rs799917	OR	RD	RERI	95% CI
00	AA	TT	1.00	0.00		
01	AA	CT+CC	1.11	**0.11**	-0.15	(-0.50-0.21)
10	TA+TT	TT	1.13	0.00		
11	TA+TT	CT+CC	1.09	**-0.04**		
	SNP1	SNP2				
	rs799917	rs2273535	OR	RD	RERI	95% CI
00	TT	AA	1.00	0.00		
01	TT	TA+TT	1.13	**0.13**	-0.15	(-0.50-0.21)
10	CT+CC	AA	1.11	0.00		
11	CT+CC	TA+TT	1.09	**-0.01**		

6.	SNP1	SNP2				
	rs2298016	rs1799950	OR	RD	RERI	95% CI
00	GG	AA	1.00	0.00		
01	GG	GA+GG	1.07	**0.07**	0.10	(-0.49-0.69)
10	GC+CC	AA	1.01	0.00		
11	GC+CC	GA+GG	1.18	**0.17**		
	SNP1	SNP2				
	rs1799950	rs2298016	OR	RD	RERI	95% CI
00	AA	GG	1.00	0.00		
01	AA	GC+CC	1.01	**0.01**	0.10	(-0.49-0.69)
10	GA+GG	GG	1.07	0.00		
11	GA+GG	GC+CC	1.18	**0.11**		

7.	SNP1	SNP2				
	rs2298016	rs16941	OR	RD	RERI	95% CI
00	GG	AA	1.00	0.00		
01	GG	GA+GG	1.29	**0.29**	-0.30	(-0.63-0.04)
10	GC+GG	AA	1.13	0.00		
11	GC+GG	GA+GG	1.12	**-0.01**		
	SNP1	SNP2				
	rs16941	rs2298016	OR	RD	RERI	95% CI
00	AA	GG	1.00	0.00		
01	AA	GC+GG	1.13	**0.13**	-0.30	(-0.63-0.04)
10	GA+GG	GG	1.29	0.00		
11	GA+GG	GC+GG	1.12	**-0.17**		

8.	SNP1	SNP2				
	rs2298016	rs16942	OR	RD	RERI	95% CI
00	GG	AA	1.00	0.00		
01	GG	GA+GG	1.28	**0.28**	-0.23	(-0.56-0.10)
10	GC+CC	AA	1.11	0.00		
11	GC+CC	GA+GG	1.15	**0.05**		
	SNP1	SNP2				
	rs16942	rs2298016	OR	RD	RERI	95% CI
00	AA	GG	1.00	0.00		
01	AA	GC+CC	1.11	**0.11**	-0.23	(-0.56-0.10)
10	GA+GG	GG	1.28	0.00		
11	GA+GG	GC+CC	1.15	**-0.13**		

9.	SNP1	SNP2				
	rs2298016	rs1799966	OR	RD	RERI	95% CI
00	GG	AA	1.00	0.00		
01	GG	GA+GG	1.30	**0.30**	-0.25	(-0.58-0.08)
10	GC+CC	AA	1.11	0.00		
11	GC+CC	GA+GG	1.16	**0.05**		
	SNP1	SNP2				
	rs1799966	rs2298016	OR	RD	RERI	95% CI
00	AA	GG	1.00	0.00		
01	AA	GC+CC	1.11	**0.11**	-0.25	(-0.58-0.08)
10	GA+GG	GG	1.30	0.00		
11	GA+GG	GC+CC	1.16	**-0.14**		

10.	SNP1	SNP2				
	rs2298016	rs799917	OR	RD	RERI	95% CI
00	GG	TT	1.00	0.00		
01	GG	TC+CC	1.05	**0.05**	-0.01	(-0.31-0.30)
10	CG+CC	TT	1.02	0.00		
11	CG+CC	TC+CC	1.06	**0.04**		
	SNP1	SNP2				
	rs799917	rs2298016	OR	RD	RERI	95% CI
00	TT	GG	1.00	0.00		
01	TT	CG+CC	1.02	**0.02**	-0.01	(-0.31-0.30)
10	TC+CC	GG	1.05	0.00		
11	TC+CC	CG+CC	1.06	**0.01**		

11.	SNP1	SNP2				
	rs6024836	rs1799950	OR	RD	RERI	95% CI
00	GG	AA	1.00	0.00		
01	GG	AG+GG	1.07	**0.07**	0.10	(-0.49-0.69)
10	GA+AA	AA	1.01	0.00		
11	GA+AA	AG+GG	1.18	**0.17**		
	SNP1	SNP2				
	rs1799950	rs6024836	OR	RD	RERI	95% CI
00	AA	GG	1.00	0.00		
01	AA	GA+AA	1.01	**0.01**	0.10	(-0.49-0.69)
10	AG+GG	GG	1.07	0.00		
11	AG+GG	GA+AA	1.18	**0.11**		

12.	SNP1	SNP2				
	rs6024836	rs16941	OR	RD	RERI	95% CI
00	GG	AA	1.00	0.00		
01	GG	GA+GG	1.11	**0.11**	0.01	(-0.33-0.36)
10	GA+AA	AA	1.04	0.00		
11	GA+AA	GA+GG	1.16	**0.12**		
	SNP1	SNP2				
	rs16941	rs6024836	OR	RD	RERI	95% CI
00	AA	GG	1.00	0.00		
01	AA	GA+AA	1.04	**0.04**	0.01	(-0.33-0.36)
10	GA+GG	GG	1.11	0.00		
11	GA+GG	GA+AA	1.16	**0.05**		

13.	SNP1	SNP2				
	rs6024836	rs16942	OR	RD	RERI	95% CI
00	GG	AA	1.00	0.00		
01	GG	GA+GG	1.08	**0.08**	0.02	(-0.34-0.37)
10	GA+AA	AA	1.07	0.00		
11	GA+AA	GA+GG	1.17	**0.09**		
	SNP1	SNP2				
	rs16942	rs6024836	OR	RD	RERI	95% CI
00	AA	GG	1.00	0.00		
01	AA	GA+AA	1.07	**0.07**	0.02	(-0.34-0.37)
10	GA+GG	GG	1.08	0.00		
11	GA+GG	GA+AA	1.17	**0.09**		

14.	SNP1	SNP2				
	rs6024836	rs1799966	OR	RD	RERI	95% CI
00	GG	AA	1.00	0.00		
01	GG	GA+GG	1.08	**0.08**	0.01	(-0.34-0.37)
10	GA+AA	AA	1.08	0.00		
11	GA+AA	GA+GG	1.17	**0.09**		
	SNP1	SNP2				
	rs1799966	rs6024836	OR	RD	RERI	95% CI
00	AA	GG	1.00	0.00		
01	AA	GA+AA	1.08	**0.08**	0.01	(-0.34-0.37)
10	GA+GG	GG	1.08	0.00		
11	GA+GG	GA+AA	1.17	**0.09**		

15.	SNP1	SNP2				
	rs6024836	rs799917	OR	RD	RERI	95% CI
00	GG	TT	1.00	0.00		
01	GG	TC+CC	1.01	**0.01**	-0.03	(-0.42-0.35)
10	GA+AA	TT	1.05	0.00		
11	GA+AA	TC+CC	1.03	**0.02**		
	SNP1	SNP2				
	rs799917	rs6024836	OR	RD	RERI	95% CI
00	TT	GG	1.00	0.00		
01	TT	GA+AA	1.05	**0.05**	-0.03	(-0.42-0.35)
10	TC+CC	GG	1.01	0.00		
11	TC+CC	GA+AA	1.03	**0.02**		

Table 4.4 Additive interaction analysis between select SNPs on *AURKA* and *BARD1*

AURKA-BARD1

16.	SNP1	SNP2				
	rs2273535	rs1048108	OR	RD	RERI	95% CI
00	AA	GG	1.00	0.00		
01	AA	GA+AA	1.07	**0.07**	-0.13	(-0.44-0.18)
10	AT+TT	GG	1.10	0.00		
11	AT+TT	GA+AA	1.04	**-0.06**		
	SNP1	SNP2				
	rs1048108	rs2273535	OR	RD	RERI	95% CI
00	GG	AA	1.00	0.00		
01	GG	AT+TT	1.10	**0.10**	-0.13	(-0.44-0.18)
10	GA+AA	AA	1.07	0.00		
11	GA+AA	AT+TT	1.04	**-0.04**		

17.	SNP1	SNP2				
	rs2273535	rs2229571	OR	RD	RERI	95% CI
00	AA	CC	1.00	0.00		
01	AA	CG+GG	1.16	**0.16**	-0.35	(-0.74-0.03)
10	AT+TT	CC	1.22	0.00		
11	AT+TT	CG+GG	1.03	**-0.19**		
	SNP1	SNP2				
	rs2229571	rs2273535	OR	RD	RERI	95% CI
00	CC	AA	1.00	0.00		
01	CC	AT+TT	1.22	**0.22**	-0.35	(-0.74-0.03)
10	CG+GG	AA	1.16	0.00		
11	CG+GG	AT+TT	1.03	**-0.13**		

18.	SNP1	SNP2				
	rs2298016	rs1048108	OR	RD	RERI	95% CI
00	GG	CC	1.00	0.00		
01	GG	CG+GG	1.04	**0.04**	-0.03	(-0.32-0.25)
10	GC+CC	CC	1.02	0.00		
11	GC+CC	CG+GG	1.02	**0.00**		
	SNP1	SNP2				
	rs1048108	rs2298016	OR	RD	RERI	95% CI
00	CC	GG	1.00	0.00		
01	CC	GC+CC	1.02	**0.02**	-0.03	(-0.32-0.25)
10	CG+GG	GG	1.04	0.00		
11	CG+GG	GC+CC	1.02	**-0.02**		

19.	SNP1	SNP2				
	rs2298016	rs2229571	OR	RD	RERI	95% CI
00	GG	GG	1.00	0.00		
01	GG	GA+AA	1.13	**0.13**	-0.23	(-0.59-0.13)
10	GC+CC	GG	1.16	0.00		
11	GC+CC	GA+AA	1.06	**-0.10**		
	SNP1	SNP2				
	rs2229571	rs2298016	OR	RD	RERI	95% CI
00	GG	GG	1.00	0.00		
01	GG	GC+CC	1.16	**0.16**	-0.23	(-0.59-0.13)
10	GA+AA	GG	1.13	0.00		
11	GA+AA	GC+CC	1.06	**-0.06**		

20.	SNP1	SNP2				
	rs6024836	rs1048108	OR	RD	RERI	95% CI
00	GG	GG	1.00	0.00		
01	GG	GA+AA	1.01	**0.01**	0.02	(-0.27-0.31)
10	GA+AA	GG	1.01	0.00		
11	GA+AA	GA+AA	1.04	**0.03**		
	SNP1	SNP2				
	rs1048108	rs6024836	OR	RD	RERI	95% CI
00	GG	GG	1.00	0.00		
01	GG	GA+AA	1.01	**0.01**	0.02	(-0.27-0.31)
10	GA+AA	GG	1.01	0.00		
11	GA+AA	GA+AA	1.04	**0.03**		

21.	SNP1	SNP2				
	rs6024836	rs2229571	OR	RD	RERI	95% CI
00	GG	CC	1.00	0.00		
01	GG	CG+GG	1.00	**0.00**	-0.03	(-0.37-0.32)
10	GA+AA	CC	1.04	0.00		
11	GA+AA	CG+GG	1.02	**-0.02**		
	SNP1	SNP2				
	rs2229571	rs6024836	OR	RD	RERI	95% CI
00	CC	GG	1.00	0.00		
01	CC	GA+AA	1.04	**0.04**	-0.03	(-0.37-0.32)
10	CG+GG	GG	1.00	0.00		
11	CG+GG	GA+AA	1.02	**0.01**		

Table 4.5 Additive interaction analysis between select SNPs on *AURKA* and *BRIP1*

AURKA-BRIP1

22.	SNP1	SNP2				
	rs2273535	rs4986764	OR	RD	RERI	95% CI
00	AA	CC	1.00	0.00		
01	AA	CT+TT	1.10	**0.10**	-0.17	(-0.49-0.15)
10	AT+TT	CC	1.08	0.00		
11	AT+TT	CT+TT	1.00	**-0.07**		
	SNP1	SNP2				
	rs4986764	rs2273535	OR	RD	RERI	95% CI
00	CC	AA	1.00	0.00		
01	CC	AT+TT	1.08	**0.08**	-0.17	(-0.49-0.15)
10	CT+TT	AA	1.10	0.00		
11	CT+TT	AT+TT	1.00	**-0.10**		

23.	SNP1	SNP2				
	rs2298016	rs4986764	OR	RD	RERI	95% CI
00	GG	CC	1.00	0.00		
01	GG	CT+TT	1.12	**0.12**	-0.21	(-0.53-0.11)
10	GC+CC	CC	1.13	0.00		
11	GC+CC	CT+TT	1.04	**-0.09**		
	SNP1	SNP2				
	rs4986764	rs2298016	OR	RD	RERI	95% CI
00	CC	GG	1.00	0.00		
01	CC	GC+CC	1.13	**0.13**	-0.21	(-0.53-0.11)
10	CT+TT	GG	1.12	0.00		
11	CT+TT	GC+CC	1.04	**-0.08**		

24.	SNP1	SNP2				
	rs6024836	rs4986764	OR	RD	RERI	95% CI
00	GG	CC	1.00	0.00		
01	GG	CT+TT	1.10	**0.10**	-0.22	(-0.55-0.11)
10	GA+AA	CC	1.14	0.00		
11	GA+AA	CT+TT	1.02	**-0.12**		
	SNP1	SNP2				
	rs4986764	rs6024836	OR	RD	RERI	95% CI
00	CC	GG	1.00	0.00		
01	CC	GA+AA	1.14	**0.14**	-0.22	(-0.55-0.11)
10	CT+TT	GG	1.10	0.00		
11	CT+TT	GA+AA	1.02	**-0.08**		

Table 4.6 Additive interaction analysis between select SNPs on *AURKA* and *ZNF350*

AURKA-ZNF350

25.	SNP1	SNP2				
	rs2273535	rs4986773	OR	RD	RERI	95% CI
00	AA	TT	1.00	0.00		
01	AA	TC+CC	1.20	**0.20**	-0.25	(-0.59-0.09)
10	AT+TT	TT	1.16	0.00		
11	AT+TT	TC+CC	1.10	**-0.05**		
	SNP1	SNP2				
	rs4986773	rs2273535	OR	RD	RERI	95% CI
00	TT	AA	1.00	0.00		
01	TT	AT+TT	1.16	**0.16**	-0.25	(-0.59-0.09)
10	TC+CC	AA	1.20	0.00		
11	TC+CC	AT+TT	1.10	**-0.09**		

26.	SNP1	SNP2				
	rs2273535	rs2278420	OR	RD	RERI	95% CI
00	AA	AA	1.00	0.00		
01	AA	AG+GG	1.04	**0.04**	0.03	(-0.45-0.50)
10	AT+TT	AA	1.07	0.00		
11	AT+TT	AG+GG	1.14	**0.07**		
	SNP1	SNP2				
	rs2278420	rs2273535	OR	RD	RERI	95% CI
00	AA	AA	1.00	0.00		
01	AA	AT+TT	1.07	**0.07**	0.03	(-0.45-0.50)
10	AG+GG	AA	1.04	0.00		
11	AG+GG	AT+TT	1.14	**0.10**		

27.	SNP1	SNP2				
	rs2273535	rs11879758	OR	RD	RERI	95% CI
00	AA	GG	1.00	0.00		
01	AA	GC+CC	1.15	**0.15**	-0.19	(-0.66-0.29)
10	AT+TT	GG	1.10	0.00		
11	AT+TT	GC+CC	1.07	**-0.04**		
	SNP1	SNP2				
	rs1187975	rs2273535	OR	RD	RERI	95% CI
00	GG	AA	1.00	0.00		
01	GG	AT+TT	1.10	**0.10**	-0.19	(-0.66-0.29)
10	GC+CC	AA	1.15	0.00		
11	GC+CC	AT+TT	1.07	**-0.08**		

28.	SNP1	SNP2				
	rs2273535	rs4986770	OR	RD	RERI	95% CI
00	AA	CC	1.00	0.00		
01	AA	CT+TT	1.18	**0.18**	-0.07	(-0.63-0.50)
10	AT+TT	CC	1.09	0.00		
11	AT+TT	CT+TT	1.20	**0.11**		
	SNP1	SNP2				
	rs4986770	rs2273535	OR	RD	RERI	95% CI
00	CC	AA	1.00	0.00		
01	CC	AT+TT	1.09	**0.09**	-0.07	(-0.63-0.50)
10	CT+TT	AA	1.18	0.00		
11	CT+TT	AT+TT	1.20	**0.02**		

29.	SNP1	SNP2				
	rs2298016	rs4986773	OR	RD	RERI	95% CI
00	GG	TT	1.00	0.00		
01	GG	TC+CC	1.18	**0.18**	-0.17	(-0.49-0.15)
10	GC+CC	TT	1.08	0.00		
11	GC+CC	TC+CC	1.08	**0.00**		
	SNP1	SNP2				
	rs4986773	rs2298016	OR	RD	RERI	95% CI
00	TT	GG	1.00	0.00		
01	TT	GC+CC	1.08	**0.08**	-0.17	(-0.49-0.15)
10	TC+CC	GG	1.18	0.00		
11	TC+CC	GC+CC	1.08	**-0.09**		

30.	SNP1	SNP2				
	rs2298016	rs2278420	OR	RD	RERI	95% CI
00	GG	AA	1.00	0.00		
01	GG	AG+GG	1.04	**0.04**	-0.04	(-0.51-0.43)
10	GC+CC	AA	1.09	0.00		
11	GC+CC	AG+GG	1.09	**0.00**		
	SNP1	SNP2				
	rs2278420	rs2298016	OR	RD	RERI	95% CI
00	AA	GG	1.00	0.00		
01	AA	GC+CC	1.09	**0.09**	-0.04	(-0.51-0.43)
10	AG+GG	GG	1.04	0.00		
11	AG+GG	GC+CC	1.09	**0.05**		

31.	SNP1	SNP2				
	rs2298016	rs11879758	OR	RD	RERI	95% CI
00	GG	GG	1.00	0.00		
01	GG	GC+CC	1.13	**0.13**	-0.13	(-0.60-0.34)
10	GC+CC	GG	1.05	0.00		
11	GC+CC	GC+CC	1.05	**0.00**		
	SNP1	SNP2				
	rs1187975{	rs2298016	OR	RD	RERI	95% CI
00	GG	GG	1.00	0.00		
01	GG	GC+CC	1.05	**0.05**	-0.13	(-0.60-0.34)
10	GC+CC	GG	1.13	0.00		
11	GC+CC	GC+CC	1.05	**-0.08**		

32.	SNP1	SNP2				
	rs2298016	rs4986770	OR	RD	RERI	95% CI
00	GG	CC	1.00	0.00		
01	GG	CT+TT	1.20	**0.20**	-0.04	(-0.61-0.52)
10	GC+CC	CC	1.06	0.00		
11	GC+CC	CT+TT	1.21	**0.15**		
	SNP1	SNP2				
	rs4986770	rs2298016	OR	RD	RERI	95% CI
00	CC	GG	1.00	0.00		
01	CC	GC+CC	1.06	**0.06**	-0.04	(-0.61-0.52)
10	CT+TT	GG	1.20	0.00		
11	CT+TT	GC+CC	1.21	**0.02**		

33.	SNP1	SNP2				
	rs6024836	rs4986773	OR	RD	RERI	95% CI
00	GG	TT	1.00	0.00		
01	GG	TC+CC	1.09	**0.09**	0.00	(-0.33-0.33)
10	GA+AA	TT	1.01	0.00		
11	GA+AA	TC+CC	1.10	**0.09**		
	SNP1	SNP2				
	rs4986773	rs6024836	OR	RD	RERI	95% CI
00	TT	GG	1.00	0.00		
01	TT	GA+AA	1.01	**0.01**	0.00	(-0.33-0.33)
10	TC+CC	GG	1.09	0.00		
11	TC+CC	GA+AA	1.10	**0.02**		

34.	SNP1	SNP2				
	rs6024836	rs2278420	OR	RD	RERI	95% CI
00	GG	AA	1.00	0.00		
01	GG	AG+GG	1.08	**0.08**	-0.05	(-0.36-0.27)
10	GA+AA	AA	1.02	0.00		
11	GA+AA	AG+GG	1.05	**0.03**		
	SNP1	SNP2				
	rs2278420	rs6024836	OR	RD	RERI	95% CI
00	AA	GG	1.00	0.00		
01	AA	GA+AA	1.02	**0.02**	-0.05	(-0.36-0.27)
10	AG+GG	GG	1.08	0.00		
11	AG+GG	GA+AA	1.05	**-0.03**		

35.	SNP1	SNP2				
	rs6024836	rs11879758	OR	RD	RERI	95% CI
00	GG	GG	1.00	0.00		
01	GG	GC+CC	1.09	**0.09**	-0.08	(-0.42-0.26)
10	GA+AA	GG	1.02	0.00		
11	GA+AA	GC+CC	1.03	**0.01**		
	SNP1	SNP2				
	rs11879758	rs6024836	OR	RD	RERI	95% CI
00	GG	GG	1.00	0.00		
01	GG	GA+AA	1.02	**0.02**	-0.08	(-0.42-0.26)
10	GC+CC	GG	1.09	0.00		
11	GC+CC	GA+AA	1.03	**-0.06**		

36.	SNP1	SNP2				
	rs6024836	rs4986770	OR	RD	RERI	95% CI
00	GG	CC	1.00	0.00		
01	GG	CT+TT	1.04	**0.04**	-0.05	(-0.46-0.36)
10	GA+AA	CC	1.01	0.00		
11	GA+AA	CT+TT	1.00	**0.00**		
	SNP1	SNP2				
	rs4986770	rs6024836	OR	RD	RERI	95% CI
00	CC	GG	1.00	0.00		
01	CC	GA+AA	1.01	**0.01**	-0.05	(-0.46-0.36)
10	CT+TT	GG	1.04	0.00		
11	CT+TT	GA+AA	1.00	**-0.04**		

REFERENCES

1. Easton DF: How many more breast cancer predisposition genes are there? Breast Cancer Res 1999, 1(1):14-17.

2. Lichtenstein P, Holm NV, Verkasalo PK, Iliadou A, Kaprio J, Koskenvuo M, Pukkala E, Skytthe A, Hemminki K: 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.

3. Peto J: Cancer epidemiology in the last century and the next decade. Nature 2001, 411(6835):390-395.

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

5. Antoniou AC, Easton DF: Polygenic inheritance of breast cancer: Implications for design of association studies. Genet Epidemiol 2003, 25(3):190-202.

6. 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.

7. 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):33-36.

8. Deng CX: BRCA1: cell cycle checkpoint, genetic instability, DNA damage response and cancer evolution. Nucleic Acids Res 2006, 34(5):1416-1426.

9. Guenard F, Labrie Y, Ouellette G, Beauparlant CJ, Durocher F: Genetic sequence variations of BRCA1-interacting genes AURKA, BAP1, BARD1 and DHX9 in French Canadian families with high risk of breast cancer. J Hum Genet 2009, 54(3):152-161.10. Chen A, Kleiman FE, Manley JL, Ouchi T, Pan ZQ: Autoubiquitination of the BRCA1*BARD1 RING ubiquitin ligase. J Biol Chem 2002, 277(24):22085-22092.

11. Brzovic PS, Rajagopal P, Hoyt DW, King MC, Klevit RE: Structure of a BRCA1-BARD1 heterodimeric RING-RING complex. Nat Struct Biol 2001, 8(10):833-837.

12. Zhang F, Fan Q, Ren K, Auerbach AD, Andreassen PR: FANCJ/BRIP1 recruitment and regulation of FANCD2 in DNA damage responses. Chromosoma 2010, 119(6):637-649.

13. Seal S, Thompson D, Renwick A, Elliott A, Kelly P, Barfoot R, Chagtai T, Jayatilake H, Ahmed M, Spanova K et al: Truncating mutations in the Fanconi anemia J gene BRIP1 are low-penetrance breast cancer susceptibility alleles. Nat Genet 2006, 38(11):1239-1241.

14. Desjardins S, Belleau P, Labrie Y, Ouellette G, Bessette P, Chiquette J, Laframboise R, Lepine J, Lesperance B, Pichette R et al: Genetic variants and haplotype analyses of the ZBRK1/ZNF350 gene in high-risk non BRCA1/2 French Canadian breast and ovarian cancer families. Int J Cancer 2008, 122(1):108-116.

15. Harkin DP, Bean JM, Miklos D, Song YH, Truong VB, Englert C, Christians FC, Ellisen LW, Maheswaran S, Oliner JD et al: Induction of GADD45 and JNK/SAPK-dependent apoptosis following inducible expression of BRCA1. Cell 1999, 97(5):575-586.

16. Garcia-Closas M, Egan KM, Newcomb PA, Brinton LA, Titus-Ernstoff L, Chanock S, Welch R, Lissowska J, Peplonska B, Szeszenia-Dabrowska N et al: Polymorphisms in DNA double-strand break repair genes and risk of breast cancer: two population-based studies in USA and Poland, and meta-analyses. Hum Genet 2006, 119(4):376-388.

17. Huo X, Lu C, Huang X, Hu Z, Jin G, Ma H, Wang X, Qin J, Shen H, Tang J: Polymorphisms in BRCA1, BRCA1-interacting genes and susceptibility of breast cancer in Chinese women. J Cancer Res Clin Oncol 2009, 135(11):1569-1575.

18. Ricks-Santi LJ, Nie J, Marian C, Ochs-Balcom HM, Trevisan M, Edge SB, Freudenheim JL, Shields PG: BRCA1 polymorphisms and breast cancer epidemiology in the Western New York exposures and breast cancer (WEB) study. Genet Epidemiol 2013, 37(5):504-511.

19. Newman B, Moorman PG, Millikan R, Qaqish BF, Geradts J, Aldrich TE, Liu ET: The Carolina Breast Cancer Study: integrating population-based epidemiology and molecular biology. Breast Cancer Res Treat 1995, 35(1):51-60.

20. Weinberg CR, Sandler DP: Randomized recruitment in case-control studies. Am J Epidemiol 1991, 134(4):421-432.

21. Weinberg CR, Wacholder S: The design and analysis of case-control studies with biased sampling. Biometrics 1990, 46(4):963-975.

22. Carey LA, Perou CM, Livasy CA, Dressler LG, Cowan D, Conway K, Karaca G, Troester MA, Tse CK, Edmiston S et al: Race, breast cancer subtypes, and survival in the Carolina Breast Cancer Study. JAMA 2006, 295(21):2492-2502.

23. Millikan R, Eaton A, Worley K, Biscocho L, Hodgson E, Huang WY, Geradts J, Iacocca M, Cowan D, Conway K et al: HER2 codon 655 polymorphism and risk of breast cancer in African Americans and whites. Breast Cancer Res Treat 2003, 79(3):355-364.

24. 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):703-714.

25. Landis JR, Koch GG: The measurement of observer agreement for categorical data. Biometrics 1977, 33(1):159-174.

26. Nielsen TO, Hsu FD, Jensen K, Cheang M, Karaca G, Hu Z, Hernandez-Boussard T, Livasy C, Cowan D, Dressler L et al: Immunohistochemical and clinical characterization of the basal-like subtype of invasive breast carcinoma. Clin Cancer Res 2004, 10(16):5367-5374.

27. van de Rijn M, Perou CM, Tibshirani R, Haas P, Kallioniemi O, Kononen J, Torhorst J, Sauter G, Zuber M, Kochli OR et al: Expression of cytokeratins 17 and 5 identifies a group of breast carcinomas with poor clinical outcome. Am J Pathol 2002, 161(6):1991-1996.

28. Livasy CA, Perou CM, Karaca G, Cowan DW, Maia D, Jackson S, Tse CK, Nyante S, Millikan RC: Identification of a basal-like subtype of breast ductal carcinoma in situ. Hum Pathol 2007, 38(2):197-204.

29. Millikan RC, Newman B, Tse CK, Moorman PG, Conway K, Dressler LG, Smith LV, Labbok MH, Geradts J, Bensen JT et al: Epidemiology of basal-like breast cancer. Breast Cancer Res Treat 2008, 109(1):123-139.

30. Barnholtz-Sloan JS, Shetty PB, Guan X, Nyante SJ, Luo J, Brennan DJ, Millikan RC: FGFR2 and other loci identified in genome-wide association studies are associated with breast cancer in African-American and younger women. Carcinogenesis 2010, 31(8):1417-1423.

31. Nyante SJ, Gammon MD, Kaufman JS, Bensen JT, Lin DY, Barnholtz-Sloan JS, Hu Y, He Q, Luo J, Millikan RC: Common genetic variation in adiponectin, leptin, and leptin receptor and association with breast cancer subtypes. Breast Cancer Res Treat 2011.

32. [www.hapmap.org]

33. 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):1217-1223.

34. 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):1545-1551.

35. Lee PH, Shatkay H: An integrative scoring system for ranking SNPs by their potential deleterious effects. Bioinformatics 2009, 25(8):1048-1055.

36. Pearce N: What does the odds ratio estimate in a case-control study? Int J Epidemiol 1993, 22(6):1189-1192.

37. Pfaff CL, Barnholtz-Sloan J, Wagner JK, Long JC: Information on ancestry from genetic markers. Genet Epidemiol 2004, 26(4):305-315.

38. Benjamini YH, Yosef: Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statisical Society 1995, 57(1):289-300.

39. Rothman KG, Sander: Modern Epidemiology. Philadelphia: Maple Press; 1998.

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

41. Poole C: Low P-values or narrow confidence intervals: which are more durable? Epidemiology 2001, 12(3):291-294.

42. 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):915-921.

43. Baynes C, Healey CS, Pooley KA, Scollen S, Luben RN, Thompson DJ, Pharoah PD, Easton DF, Ponder BA, Dunning AM: Common variants in the ATM, BRCA1, BRCA2, CHEK2 and TP53 cancer susceptibility genes are unlikely to increase breast cancer risk. Breast Cancer Res 2007, 9(2):R27.

44. Dunning AM, Chiano M, Smith NR, Dearden J, Gore M, Oakes S, Wilson C, Stratton M, Peto J, Easton D et al: Common BRCA1 variants and susceptibility to breast and ovarian cancer in the general population. Hum Mol Genet 1997, 6(2):285-289.

Chapter 5. Summary and Conclusions

5.1 Main Findings

The primary purpose of this dissertation was to utilize data from the CBCS and a candidate gene approach to investigate associations between common genetic variation in the oncogene *AURKA* (in the form of SNPs) and breast cancer overall, while also exploring the hypothesis that associations may differ by intrinsic subtype of breast cancer. This hypothesis was considered in light of previous CBCS findings that showed differences for clinical outcomes and non-genetic risk factors between intrinsic subtypes [1,2]. Because *AURKA* is a key regulator of the cell cycle, and overexpression of its encoded protein product (Aurora-A) has been demonstrated in human cancers [3-5], other candidate genes whose protein products are known to interact with Aurora-A were also logical targets for this investigation. The highly penetrant *BRCA1* is such a critical gene, and we also evaluated SNPs on it for associations with overall breast cancer and intrinsic breast cancer subtype in the CBCS study population. Mutations in *BRCA1* are known to confer large increases in lifetime risk of breast cancer [6], and because the BRCA1 protein is known to interact with Aurora-A during the cell cycle, it seemed logical to investigate *BRCA1*-interacting genes for associations with breast cancer and possible gene-gene interactions with *AURKA*. Therefore, SNPs in *BARD1, BRIP1,* and *ZNF350* were also examined. Although the CBCS was not fully powered to investigate gene-gene interactions, the biological plausibility of such interactions between loci on *AURKA* and each of *BRCA1, BARD1, BRIP1,* and *ZNF350* warranted analysis. To focus this endeavor and minimize the possibility of chance findings, each SNP was evaluated for presumed functionality using an integrative *in silico* scoring system based

on protein coding, splicing regulation, transcriptional regulation, and post-translation [7]. Additive interaction between selected SNPs was then assessed using the RERI.

Chapter 3 details the investigation of *AURKA* in relation to overall rate of breast cancer and intrinsic breast cancer subtype. Analyses of *AURKA* SNPs in association with breast cancer among all cases and controls were stratified by race due to differences in LD between African Americans and Caucasians. The CBCS's coverage of SNPs genotyped on *AURKA* was comprehensive, and represents the largest genotyping of tag SNPs on *AURKA* to date among African Americans. The patterns noted in the race stratified analysis suggested little or no association between most *AURKA* SNPs and overall rate of breast cancer. However, one tag SNP among African Americans (rs6092309; OR=0.69) demonstrated a pronounced inverse association with breast cancer and relatively good estimate precision (CLR=1.69). Future studies of variation in *AURKA* among African Americans will be required to replicate this finding and examine the biological consequences of this SNP. There was diminished statistical power to estimate intrinsic subtype-specific associations, but the innovative feature of the CBCS is in the 5-marker subtyping of cases based on our most current understanding of breast tumor heterogeneity which justifies the exploration of such associations. Due to small numbers within strata of intrinsic subtype, it was necessary to combine African Americans and Caucasians for this analysis. Several SNPs on *AURKA* showed subtype-specific estimates that suggest differences in risk by subtype. Further race-stratified exploratory analyses within subtype were also suggestive of racial differences by subtype. It is important to note that these exploratory analyses were not powered to detect race-stratified subtype-specific associations, and caution is advised when interpreting these results.

Chapter 4 focused on germline genetic variation in *BRCA1* and the *BRCA1*-interacting genes: *BARD1, BRIP1,* and *ZNF350* in association with overall rate of breast cancer and intrinsic subtype of breast cancer. These candidate genes were chosen based on their biologically plausible influence on *AURKA*. The BRCA1 protein is known to interact with Aurora-A to regulate cell cycle progression, so other candidate genes interacting with *BRCA1* may also influence *AURKA*. CBCS genotyping coverage of *BRCA1* was limited to several candidate SNPs and a single tag SNP in Caucasians. The main finding of manuscript 2 was a modest positive association between three nonsynonymous missense *BRCA1* SNPs and overall rate of breast cancer among Caucasians. All three SNPs are predicted to be functional by FS Score and demonstrated associations of similar magnitude and precision, which could be due to the high degree of LD between them (r^2>0.90). It is possible that the observed associations were caused by an ungenotyped locus that is also in high LD with the three candidate SNPs, and future studies may endeavor to sequence the region of *BRCA1* bound by these loci. Tag SNP genotyping coverage of *BARD1* by CBCS was comprehensive, but yielded few patterns of association with breast cancer in the race-stratified analysis. Previous genetic studies of *BARD1* are limited, and focus on a few functional variants. One such variant, rs28997576, results in a cysteine to serine amino acid substitution at codon 557, a missense mutation that has been suspected of increasing risk for breast cancer in Icelandic women (OR=1.82, 95%CI: 1.11-3.01) [8]. Our study also noted a positive association between rs28997576 and overall rate of breast cancer among Caucasians that was similar in magnitude and more precise, however a recent meta-analysis of ~12,000 cases and ~7500 controls reported no association between the polymorphism and breast cancer risk [9]. Subtype-specific associations among *BARD1* SNPs were suggestive of differences by subtype, but estimates were less precise and patterns more

difficult to ascertain. *BRIP1* and *ZNF350* polymorphisms genotyped in the CBCS were limited to candidate SNPs. Among African Americans, there was some evidence for inverse associations between three SNPs on *BRIP1* and overall rate of breast cancer, with no such evidence among Caucasians. Results for *ZNF350* among both race groups were consistent with little or no association. There was limited evidence for subtype-specific effects of SNPs on *BRIP1* and *ZNF350*. Future studies may require a more exhaustive genotyping of these genes in larger groups of African Americans and Caucasians to improve coverage and accuracy of estimates, especially for subtype analyses.

Lastly, we calculated RERIs to investigate the potential for gene-gene interactions on the additive scale between select SNPs on *AURKA* and select SNPs on each of *BRCA1, BARD1, BRIP1,* and *ZNF350* in association with overall rate of breast cancer. Although the CBCS was not powered to investigate gene-gene interactions, there were several interactions that suggested one SNP allele antagonistically eliminated or reversed a rate-increasing effect of another SNP allele. Some of these reductions or reversals were relatively large in magnitude and occurred between SNPs on *AURKA* and each of the other investigated candidate genes.

5.2 Future Directions

The results of this dissertation provide evidence that some genotypes are associated with breast cancer, and those associations may vary by race and intrinsic subtype of breast cancer. Although common genetic variation in the main candidate gene of interest, *AURKA*, has been studied previously, this investigation is the first comprehensive evaluation of *AURKA* in African American women with intrinsic subtype data and results will need to be replicated in yet larger studies with similar outcome assessment. Future studies may also consider fine mapping regions

of *AURKA* that include SNPs identified in this study as showing the largest and most precise associations with breast cancer and intrinsic breast cancer subtype.

Although germline genetic variation in the form of SNPs may be associated with breast cancer risk, little is known about how genetic variation contributes to mRNA and protein expression levels [10]. Since mRNA overexpression of *AURKA* has been previously associated with tumor characteristics as well as poor clinical outcomes, it may be important to know how *AURKA* is being expressed within intrinsic subtype of breast cancer. Future studies may endeavor to fine map *AURKA* and measure expression levels to look for patterns among intrinsically subtyped cases of breast cancer.

Future assessment of *AURKA* and other candidate genes with respect to intrinsic breast cancer subtype could benefit from more refined intrinsic subtype definitions to further reduce the chances of subtype misclassification. The CBCS used definitions based on immunohistochemistry as surrogates for subtypes defined by gene expression profiling using mRNA-based assays containing thousands of genes. IHC assays do not provide as much information about tumor biology as mRNA-based expression assays do, and could result in subtype misclassification. In addition, efforts could be made to better describe heterogeneity within the unclassified subtype of breast cancer. Tumors showing no expression for any of the five markers used to classify intrinsic breast cancer subtype in the CBCS were labeled unclassified. It is possible that mRNA-based expression assays may be better suited to characterize these tumors.

Future investigations of *AURKA* might also consider a pathway-based approach to improve our chances of discovering important risk loci for breast cancer. *AURKA* is known to play a vital role in regulating the cell cycle via its control over centrosomal function. Other important genes

function in this pathway as well, and focusing our investigations on the whole pathway may elucidate important susceptibility patterns for breast cancer.

REFERENCES

1. Carey LA, Perou CM, Livasy CA, Dressler LG, Cowan D, Conway K, Karaca G, Troester MA, Tse CK, Edmiston S et al: Race, breast cancer subtypes, and survival in the Carolina Breast Cancer Study. JAMA 2006, 295(21):2492-2502.

2. Millikan RC, Newman B, Tse CK, Moorman PG, Conway K, Dressler LG, Smith LV, Labbok MH, Geradts J, Bensen JT et al: Epidemiology of basal-like breast cancer. Breast Cancer Res Treat 2008, 109(1):123-139.

3. Baba Y, Nosho K, Shima K, Irahara N, Kure S, Toyoda S, Kirkner GJ, Fuchs CS, Ogino S. Aurora-A expression is independently associated with chromosomal instability in colorectal cancer. Neoplasia 2009, 11(5): 418-25.

4. Zhang H, Chen X, Jin Y, Liu B, Zhou L Overexpression of Aurora-A promotes laryngeal cancer progression by enhancing invasive ability and chromosomal instability. Eur Arch Otorhinolaryngol. 2012, 269(2): 607-614.

5. Li D, Zhu J, Firozi PF, Abbruzzese JL, Evans DB, Cleary K, Friess H, Sen S Overexpression of oncogenic STK15/BTAK/Aurora A kinase in human pancreatic cancer. Clin Cancer Res. 2003, 9(3): 991-7.

6. Ford D, Easton F, Stratton M *et al.* Genetic heterogeneity and penetrance analysis of the BRCA1 and BRCA2 genes in breast cancer families. Am. J. Hum. Genet. 1998, 62: 676-689.

7. Lee PH, Shatkay H: An integrative scoring system for ranking SNPs by their potential deleterious effects. Bioinformatics 2009, 25(8):1048-1055.

8. Stacey SN, Sulem P, Johannsson OT, *et al.* The BARD1 Cys557Ser variant and breast cancer risk in Iceland. PLoS Med. 2006, 3(7): e217.

9. Ding DP, Zhang Y, Ma WL, He XF, Wang W, Yu HL, Guo YB, Zheng WL. Lack of association between BARD1 Cys557Ser variant and breast cancer risk: a meta-analysis of 11,870 cases and 7,687 controls. J Cancer Res Clin Oncol. 2011, 137(10): 1463-8.

10. Garge N, Pan H, Rowland MD, Cargile BJ, Zhang X, Cooley PC, Page GP, Bunger MK. Identification of quantitative trait loci underlying proteome variation in human lymphoblastoid cells. Mol Cell Proteomics. 2010, 9(7): 1383-99.