RACE AND SUBTYPE DIFFERENCES IN THE REPLICATION OF PREVIOUSLY IDENTIFIED BREAST CANCER SUSCEPTIBILITY LOCI: A BAYESIAN APPROACH

Katie M. O'Brien

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

Chapel Hill
2013

Approved by:
Steven R. Cole
Robert C. Millikan
Jeannette T. Bensen
Amy H. Herring
Charles Poole
Lawrence S. Engel

Abstract

KATIE M. O'BRIEN: Race and subtype differences in the replication of previously identified breast cancer susceptibility loci: A Bayesian approach
(Under the direction of Robert C. Millikan and Stephen R. Cole)

Over the last twenty-five years, researchers have identified several dozen genetic polymorphisms associated with breast cancer susceptibility. While many of these loci are now considered well-established risk factors for the disease, previous attempts to replicate variant-disease associations in African Americans or to identify subtype-specific risk variants have been imprecise and inconsistent.

I examined the association between breast cancer subtypes and previously established candidate gene and genome-wide association study "hits" among white and African American women in the Carolina Breast Cancer Study. Maximum likelihood and Bayesian methods were used to estimate race and subtype-specific odds ratios (ORs) for each of 83 candidate single nucleotide polymorphisms (SNPs). Selected SNPs included several previous GWAS hits (n=22), near-GWAS hits (n=19), otherwise well-established risk loci (n=5), or SNPs in the same gene as another selected variant (n=37). Subtypes were defined using 5 immunohistochemical markers: estrogen receptors (ER), progesterone receptors (PR), human epidermal growth factor receptors 1 and 2 (HER1/2) and cytokeratin (CK) 5/6.

Eighteen GWAS-identified SNPs successfully replicated in whites and ten GWAS-identified SNPs successfully replicated in African Americans. SNPs in *FGFR2* and *TNRC9/TOX3* were strongly associated with breast cancer in both races. Additionally, SNPs in *MRPS30*, *MAP3K1, CDKN2A/B, ZM1Z1, LSP1, H19,* and *TP53* were

associated with breast cancer in whites and SNPs in *TLR1*, *ESR1*, and *H19* were associated with breast cancer in African Americans. Several SNPs in *TNRC9/TOX3* were associated with luminal A (ER/PR+, HER2-) or basal-like disease (ER-, PR-, HER2-, HER1 or CK 5/6+), and one SNP (rs3104746) was associated with both. SNPs in *FGFR2* were associated with luminal A, luminal B (ER/PR+, HER2+), and HER2+/ER-, but not basal-like disease. There were also subtype differences in the effects of SNPs in 2q35, 4p, *TLR1, MAP3K1*, *ESR1*, *CDKN2A/B*, *ANKRD16*, and *ZM1Z1*.

These analyses provide precise, well-informed race and subtype-stratified ORs for several key breast cancer-related SNPs. These results also demonstrate the utility of Bayesian methods in genetic epidemiology and provide evidence of subtype-specific etiologies. This work may help to identify specific causal variants, locate targets for research on directed therapies, and identify high-risk individuals.

Dedicated to Robert C. Millikan,

an extraordinary teacher, mentor and friend

"Remember that when you leave this earth,
you can take with you nothing that you have received –
only what you have given: a full heart,
enriched by honest service, love, sacrifice and courage."
- St. Francis of Assisi

Acknowledgements

I would first like to thank my committee members, for their advice and encouragement throughout this dissertation process. In particular, I would like to thank my chair, Dr. Stephen Cole, for his steadfast support and mentorship, and Dr. Lawrence Engel, for his willingness to join my committee belatedly and for his continued guidance and backing. I would also like to thank the staff of the Carolina Breast Cancer Study, as well as Nancy Colvin, Carmen Woody, Andy Olshan, and other members of the Epidemiology department who have provided the emotional and academic support I needed to complete this dissertation. I am also grateful to the participants of the Carolina Breast Cancer Study, who gave us their time and personal information in hopes of making a difference for future generations.

Additionally, I would like to thank my husband, Alexander Keil, for his love, patience, and commiseration during the last 6 years of graduate school. His programming, editing, and methodology skills are also much appreciated. I am also indebted to my fellow epidemiology students and unofficial support group members, including Cassidy Henegar, Pamela Klein, Jennifer Lund, Peter Samai, Leila Family, Jess Edwards, Christina Ludema, and Annah Wyss. A final thank you goes to my parents, who have done everything in their power to ensure that their daughters are happy and successful women. I am truly grateful for everything they have given me.

The Carolina Breast Cancer Study and this dissertation were funded in part by the University Cancer Research Fund of North Carolina, the National Cancer Institute

Specialized Program of Research Excellence (SPORE) in Breast Cancer (NIH/ NCI P50-CA58223), the Lineberger Comprehensive Cancer Center Core Grant (NIH/NCI P30-CA16086), and two NIH training grants (Cancer Education and Career Development Program: NIH/NCI 5R25CA057726-20 and National Research Service Award Institutional Training Grant for Cancer Epidemiology: 5T32CA009330-30). The UNC BioSpecimen Processing Facility, the UNC Mammalian Genotyping Core, and Jessica Tse provided technical assistance for the study.

Table of Contents

List of Tables

List of Figures

List of Abbreviations and Symbols

AIMs = Ancestry Informative Markers

ASW= HapMap population of African Americans living in the Southwest USA

BMI = Body Mass Index

BRCA1 = Breast Cancer Gene 1

BRCA2= Breast Cancer Gene 2

CBCS= Carolina Breast Cancer Study

ccOR = Case-case Odds Ratio

CEU = HapMap population of Utah residents with ancestry from northern and western Europe

CI = Confidence Interval

CK 5/6 = Cytokeratin 5/6

DAG= Directed Acyclic Graph

Df = Degree of Freedom

DNA = Deoxyribonucleic Acid

ER= Estrogen Receptor

HER1= Human Epidermal Growth Factor Receptor-1

HER2= Human Epidermal Growth Factor Receptor-2

HR= Hazard Ratio

HRT = Hormone Replacement Therapy

HWE = Hardy-Weinberg Equilibrium

IHC= Immunohistochemical

LD = Linkage Disequilibrium

MAF= Minor Allele Frequency

MCMC = Markov-Chain Monte Carlo

MLE = Maximum Likelihood Estimate

mRNA = Messenger Ribonucleic Acid

NC= North Carolina

NHS= Nurses Health Study

NOS = Not Otherwise Specified

OC= Oral Contraceptive

OR = Odds Ratio

PBCS= Polish Breast Cancer Study

PCR = Polymerase Chain Reaction

PI = Posterior Interval

PR = Progesterone Receptor

RAF = Risk Allele Frequency

SEER = Surveillance Epidemiology and End Results

SNP= Single Nucleotide Polymorphism

UNC = University of North Carolina

US = United States

WHR = Waist to Hip Ratio

YRI = HapMap population of Yorubans from Ibadan, Nigeria

1. Specific Aims

Although advances in breast cancer detection, prevention and treatment have helped reduce incidence and mortality over the last decade, breast cancer still afflicts nearly one in every eight US women and kills one in 36 [1, 2]. A crucial next step in reducing the public health burden of breast cancer is to identify and better characterize genetic mutations affecting breast tumorigenesis. By examining these mutations we can better understand how the disease develops and progresses, identify individuals at higher risk of developing the disease, and locate targets for further research on directed therapies. We can gain additional insights by investigating these genetic mutations within strata of disease subtype and race, as there is strong evidence that breast cancer is a heterogeneous disease with several distinct etiologies and that subtypes and risk variants are not evenly distributed across race or age groups [3-12].

Aside from a few highly influential genetic variants identified through linkage or candidate gene analyses, most of the known genetic risk factors for breast cancer were first detected in genome-wide association studies (GWAS). With so much about breast cancer etiology and the human genome still unknown, GWAS have been invaluable discovery tools. Yet, the expansiveness of these studies also limits their precision and thus their ability to identify causal variants or discern subtype-specific risk profiles.

To better understand the role of genetic polymorphisms in breast carcinogenesis and more specifically, the carcinogenesis of individual breast cancer subtypes in specific racial groups, I studied the association between several previously identified genetic risk factors

and the risk of breast cancer. In this analysis, I used Bayesian statistical methods to incorporate information on the linkage disequilibrium of the variants within the gene and the likely magnitude of the association to better inform statistical models. These theoretically sound but computationally challenging methods are expected to produce more precise and informative effect estimates than traditional statistical techniques [13-17]. I conducted this research using data from the Carolina Breast Cancer Study (CBCS), a population-based, case-control study with large samples of both white (n= 1247 cases, 1105 controls) and African American women (n=766 cases, n=681 controls). The specific aims of this study are as follows.

Specific aim 1: Estimate associations between identified genetic risk variants and overall breast cancer in whites and African Americans using Bayesian and frequentist methods.

a) Describe the application of Bayesian methods, including hierarchical modeling and full Bayes regression analysis

b) Generate race-specific effect estimates for the association between invasive and *in situ* breast cancer and 83 candidate polymorphisms from 28 well-established breast cancer-related genes or gene regions [18-20]

   i. Estimate individual SNP effects using frequentist logistic regression models and full Bayes logistic regression models with informative priors

   ii. Estimate individual and group-level SNP effects for SNPs in the same linkage disequilibrium block using hierarchical regression analysis

Specific aim 2: Estimate effects between the candidate SNPs and each breast cancer subtype.

a) Estimate the effect of each candidate gene on each breast cancer subtype, relative to controls, using frequentist and Bayesian methods

b) Estimate race-stratified effects for each candidate gene on each breast cancer subtype, relative to controls, using frequentist and Bayesian methods

c) Conduct sensitivity analyses to demonstrate the robustness of the effect estimates under alternative model assumptions

2. Review of the Literature

2.1 Public health impact of breast cancer

Despite recent advances in detection and prevention, breast cancer remains the most common female cancer in the United States (US) and worldwide [21]. In 2012, an estimated 226,870 US women were diagnosed with the disease [22], which corresponds to an average lifetime risk of 12%, or one in eight, for US women [23].

With a 5-year relative survival rate of nearly 90% in the US [23], deaths due to breast cancer are fairly rare, with approximately 1 in 35 women dying annually from the disease [2]. However, breast cancer is the second most common cause of cancer death in the US, claiming the lives of nearly 40,000 US women in 2012 [23], and remains the leading cause of cancer death in females worldwide [21].

Age-adjusted incidence rates for the state of North Carolina (NC) are slightly above the US average, with 127.5 NC cases versus 123.1 US cases per 100,000 women per year in 2009 [24]. Breast cancer mortality rates in NC are slightly lower than the US average, with 21.4 NC deaths and 22.2 US deaths per 100,000 in 2009 [24].

2.2 Race and age disparities in incidence and mortality

Data from the National Cancer Institute's Surveillance Epidemiology and End Results (SEER) database indicate that although breast cancer incidence rates are very similar for White and African American women younger than 50 at diagnosis (44.7 and 44.4 cases per 100,000 women in 2009 for Whites and African Americans, respectively), the disease is

more common among White women aged 50 or older than African American women of the same age group (360.9 versus 344.8 cases per 100,000 women) [1]. In contrast, mortality statistics from 2009 indicate much higher rates for African Americans than Whites in both age groups (8.1 versus 4.5 deaths per 100,000 women aged <50; 89.1 versus 66.6 deaths per 100,000 women aged 50 or older) [2]. As shown in Figure 1, a plot of race and age-group specific mortality rates from 1969-2009, racial differences in mortality in the 50 or older age group have not shown a meaningful decrease since reaching a peak in 1997, at which point there were approximately 23 more deaths per 100,000 per year in African Americans than Whites [2]. All rates have been decreasing over time since the mid-1990s.

2.3 Breast cancer histology

The National Cancer Institute defines breast cancer as "cancer that forms in the tissues of the breast, usually the ducts and lobules" [22]. More specifically, invasive breast cancer develops when abnormal cells multiply without constraint, forming a tumor within one part of the breast that eventually penetrates the basement membrane and infiltrates adjacent tissues. These tumors can be classified based on their histological subtype, which describes the anatomical structure in which the malignant cells originate [25].

The majority of invasive breast cancers are ductal carcinomas not otherwise specified (NOS) that originate in women's milk ducts, then grow through the ductal walls to invade other stromal tissues [25]. Approximately 8-15% of all invasive breast cancers occurring in US women originate in a lobule, or milk gland [26, 27]. Several rare histologic subtypes make up the remaining 15-20% of invasive breast cancers, including tubular carcinoma, medullary carcinoma, inflammatory breast carcinoma, papillary carcinoma, mucinous

carcinoma and Paget's disease, with each accounting for less than 2% of all invasive cases. Ductal carcinoma *in situ* and lobular carcinoma *in situ* are non-invasive breast tumors that have not spread beyond their point of origin [25]. These malignant cells may grow into invasive tumors if left untreated.

2.4 Breast cancer and hormone receptors

The presence or absence of specific hormone receptors within the breast tumor can also be used to classify breast tumors. Hormone receptor expression is determined using immunohistochemical (IHC) analysis and may be an important indicator of where and how the malignant cells first arose [28-30].

Estrogen receptors (ER) and progesterone receptors (PR) are the most common types of hormone receptor in breast tumor cells, with approximately 80% and 60-70% of all US breast cancers arising from ER positive cells and PR positive cells, respectively [7, 26, 31-33]. ER and PR status are highly correlated, with concordance estimated at 85% [34-37]. A third commonly evaluated hormone receptor is human epidermal growth factor receptor-2 (HER2), which appears in 12-20% of breast tumors collected from US population-based samples [7, 33, 38].

2.5 Breast cancer intrinsic subtypes

The development of gene expression analysis methods, which quantify the activity levels of certain genes as they are transcribed from deoxyribonucleic acid (DNA) into messenger ribonucleic acid (mRNA) and sent for translation into proteins, allows for more thorough evaluations of breast tumor heterogeneity. In one of the first gene expression

analyses of breast tissue, Perou et al. [39] identified 496 genes that sufficiently captured the variation between tumor cells. When these 496 genes were analyzed using a hierarchical clustering method to group tumors with similar expression patterns, the investigators observed interesting parallels with the traditional IHC markers of ER and HER2. For example, tumor samples that over-expressed the estrogen receptor-α gene and several other transcription factors were also ER positive. Similarly, HER2+ tumor samples were the only ones that expressed high levels of genes from a small region on chromosome 17. The remaining gene expression-derived subgroups could be differentiated based on the presence or absence of expression of cytokeratin 5/6 (CK 5/6), a protein found in basal epithelial cells but not in the more differentiated luminal epithelial cells.

Based on these findings and more refined analyses [40-42], breast cancer researchers created a classification system for breast cancer tumors based on five IHC tumor markers (ER, PR, HER2, CK 5/6, and human epidermal growth factor receptor-1 [HER1]) that serve as adequate, inexpensive surrogates for more complex gene expression profiles. Based on these IHC markers, breast cancer can be classified into four subtypes with unique biological characteristics: luminal A (ER+ and/or PR+, HER2-), luminal B (ER+ and/or PR+, HER2+), HER2+/ER- (ER-, PR-, HER2+), and basal-like (ER-, PR-, HER2-, HER1+ and/or CK 5/6+). Despite evidence that the gene expression of tumors negative for these five IHC markers is still rather varied, claudin-low, HER2-enriched, apocrine expressing, normal-like, and other potentially unique subtypes are generally lumped together as ‘unclassified’ due to their rarity and complicated features [43-46].

While the 5 IHC marker categorization system does not classify the heterogeneity of breast tumors perfectly [47], its adoption has already furthered our understanding of breast

cancer etiology and uncovered topics in need of further research. In particular, this subtype classification system has lead to insights in racial differences in the incidence of each breast cancer subtype, furthered explorations of how genetic and behavioral risk factors vary by subtype, and motivated development of targeted therapies.

2.5.1 Intrinsic subtypes and race

Luminal A is the most common subtype, but the proportion of tumors of each subtype varies widely depending on the age and race of the population [3, 5, 6, 48-59]. This phenomenon was first observed by Carey et al. [3] in the Carolina Breast Cancer Study (CBCS), a population-based, case-control study of breast cancer in North Carolina with approximately equal proportions of white and African American women, and pre and postmenopausal women. In this population, luminal A was the most common subtype in postmenopausal African Americans (59%), premenopausal non-African Americans (51%) and postmenopausal non-African Americans (58%), but not premenopausal African Americans, who had a higher proportion of basal-like breast cancer (39% basal-like vs. 36% luminal A). In contrast, postmenopausal African Americans, premenopausal non-African Americans and postmenopausal non-African Americans had 14%, 16%, and 16% basal-like breast cancer, respectively. Premenopausal African Americans also had slightly higher proportions of HER2+/ER- tumors than the other subgroups (9% vs. 6-7%) and lower proportions of luminal B tumors (9% vs. 16-18%). These results were replicated in later analyses of an expanded CBCS population [4].

Although data on basal markers (CK 5/6 and HER1) has not been collected for any other largely African American study populations, other studies have confirmed that African Americans, particularly young African Americans, have a higher proportion of ‘triple

negative' tumors (ER-, PR- and HER2-) [7, 8, 60-67] relative to other racial groups. Additionally, several studies of breast cancer subtype distributions in African populations have observed proportions of basal-like breast cancer [50, 68, 69] or triple negative breast cancer [64] that are equal to, if not greater than those found in US African American populations.

Asians and Europeans seem to have similar subtype distributions as white Americans, with studies reporting approximately 50-60% luminal A tumors and 10-20% basal-like tumors [7, 33, 48, 52, 53, 55, 56, 58-60, 62, 63, 70-83]. The only study to analyze basal markers in Hispanic individuals found a low prevalence of basal-like breast cancer (5%) [49], but the prevalence of triple negative disease was consistent with that of white Americans, Europeans, and Asians [7, 60, 62, 65, 67, 84, 85].

2.5.2 Intrinsic subtypes and prognosis

Carey et al. [3] was also the first study to examine whether intrinsic subtypes affect survival, though an update from O'Brien et al. [86] in 2010 examined an expanded study population and longer follow-up period. Here, 26% of CBCS participants with the HER2+/ER- subtype and 24% of those with basal-like breast cancer died of their disease within 5 years of their diagnosis, compared to only 9% of those with luminal A tumors. These breast cancer specific mortality rates corresponded to hazard ratios of 2.3 (95% confidence interval [CI]: 1.5- 3.6) and 1.7 (95% CI: 1.2- 2.4) for women with HER2+/ER- and basal-like tumors, respectively, relative to women with luminal A tumors.

Other investigators have observed that HER2+/ER- tumors and basal-like tumors have worse prognoses than luminal A or B tumors in their study populations [5, 51-53, 56, 58, 67, 87], thereby lending support to the theory that these subtypes are biologically distinct

diseases with diverse prognoses. As breast cancer specific survival was worse for basal-like tumors than tumors negative for all 5 IHC markers in most of these studies [5, 51, 53, 56, 58, 86-88], these analyses provide further evidence that combining these two diseases into a single 'triple-negative' subtype limits our ability to study their divergent outcomes and possibly unique etiologic origins [28, 29, 41, 89-91].

Of note, while these prognostic discrepancies do provide evidence of distinct subtype etiologies, the availability and effectiveness of subtype-targeted treatments also plays a substantial role in subtype-specific mortality rates. Anti-estrogen drugs, such as tamoxifen or raloxifen, can bind to the estrogen receptors in ER positive tumors, thereby inhibiting growth-stimulating estrogen molecules from binding to the receptors [92]. Similarly, aromatase inhibitors such as letrozole (Femara®) decrease the amount of estrogen a woman produces [93]. Although the development and FDA approval of trastuzumab (brand name Herceptin®) is too recent to affect survival statistics in any long-term observational studies, the HER2 protein binding anti-body is now used to treat HER2 positive tumors [94, 95]. To date, there is no FDA approved targeted therapy for women with triple negative disease, though additional research on gene expression patterns among triple negative tumors has identified some potential therapeutic markers [44, 45, 96].

2.6 Non-genetic risk factors for breast cancer

Decades of research on breast cancer etiology have consistently shown associations between the disease and several non-genetic traits and lifestyle patterns. Female gender, older age, and African American race are three such well-established, positively associated risk factors [97-99]. Reproductive traits associated with an increased number of menstrual cycles,

including early age at menarche, older age at first birth, low parity (especially nulliparity), lack of breast-feeding, and late age at menopause have all been linked to higher disease rates as well [100]. Other well-established factors associated with increased disease risk are use of hormone replacement therapy (HRT), previous chest radiation, radiation treatment for lymphoma and other cancers, high alcohol consumption, being overweight or obese after menopause, lack of physical activity, having dense breasts, and a history of benign breast disease. Tobacco smoke, exposure to certain environmental chemicals, oral contraceptive (OC) use, and certain dietary patterns may also increase breast cancer risk, but the results from studies on these topics have been inconsistent [97-99].

2.6.1 Non-genetic risk factors by estrogen receptor (ER) status

Once researchers realized how strongly hormone receptors influence patients' responses to specific treatments and overall prognoses, they began to question whether breast cancer subtypes had unique etiologies and, accordingly, unique risk factors. In their earliest and most basic approach to this question, epidemiologists examined traditional risk factors in ER+ and ER- tumors separately, accounting for PR status if available. These studies revealed some noteworthy discrepancies.

Due to known associations between high-risk reproductive patterns and increased lifetime estrogen exposure [100], researchers originally hypothesized that ER+ tumors would be more strongly associated with the established reproductive risk factors than ER- tumors. Two systematic reviews [101, 102] and one pooled analysis [103], which together represent years of work and dozens of studies on the subject, partially corroborated this theory. These authors conclude that ER+ tumors were more strongly and consistently associated with early age at menarche, nulliparity, and later age at first full term pregnancy than ER- tumors,

though early menarche is also a risk factor for ER- disease. However, these authors also found that breastfeeding was associated with a reduced risk of both disease types, and that HRT and OC use did not meaningfully affect the risk of either. Studies of other established risk factors indicated that postmenopausal obesity is probably associated with ER+ but not ER- disease [101, 104], and that family history, smoking, and alcohol are associated with comparable increases in risk for both subgroups [101, 105, 106]. Consideration of PR status did not lead to any additional insights [103, 104, 106].

2.6.2 Non-genetic risk factors by subtype

Testing for HER2 status became more common after the development of Herceptin, and the identification of intrinsic subtypes prompted some researchers to assess basal-marker status as well. This allowed researchers to conduct more etiologically relevant risk factor analyses for each ER/PR/HER2 or intrinsic subtype separately. Although these studies were often small and not representative of all cases, they have contributed greatly to our understanding of what causes specific breast cancer subtypes and aided in the development of new treatment and prevention strategies.

To date, three research groups have published analyses of breast cancer risk factors by intrinsic subtype. This includes the Carolina Breast Cancer Study (CBCS) [4], the Nurses' Health Study (NHS) [83], and the Poland Breast Cancer Study (PBCS) [82]. Some effect estimates for basal-like breast cancer are also available from the pooled analysis by Yang and colleagues [103], which includes PBCS and 33 other studies.

Many other research groups assessed risk factors by combined ER, PR, and HER2 status, with no differentiation between basal-like breast cancer and other triple negative subtypes. While these less refined definitions may produce biased effect estimates for basal-

like breast cancer, the discrepancy is minimal if most triple-negative tumors are basal-like, as is usually the case [45], and relative risks for luminal A, luminal B, and HER2+/ER- breast cancers are the same for either classification system. In the following risk factor summary, conclusions are based on confounder-adjusted ORs comparing cases to non-cases, or confounder-adjusted case-case odds ratios (ccORs) comparing one subtype directly to another referent subtype, usually luminal A. The results are further summarized in Table 1 (case-control comparisons) and Table 2 (case-case comparisons).

2.6.2.1 Age

Millikan et al. [4] found that younger age was a stronger risk factor for luminal B, HER2+/ER-, basal-like, and unclassified breast cancers than for luminal A breast cancers. Studies without basal marker data found similar trends for all subtypes relative to luminal A [62, 66, 78, 107], for triple-negative relative to luminal A [108], or for triple-negative relative to non-triple negative [65].

Results from three case-control comparisons [66, 109, 110], were less consistent and poorly generalizable, as all three were conducted in relatively young populations (maximum ages of 45, 56, and 54, respectively). Dolle et al. [109] found that the risk of triple-negative breast cancer was highest for women in their 30s. Gaudet et al. [110] observed a positive correlation between age and the risk of luminal A or HER2+/ER- breast cancer, but not luminal B or triple-negative. Trivers et al. [66] reported a lower risk for luminal A breast cancer among younger women, but an increased risk for luminal B or triple-negative disease. In general, the combined case-case and case-control results suggest that younger women are more likely to be diagnosed with basal-like or other non-luminal A breast cancer than older women, who are more likely to get luminal A breast cancer.

2.6.2.2 Race

As expected based on previous comparisons of subtype prevalence by race, effect estimates from all studies reporting race-stratified analyses indicated that African Americans had a greater risk of basal-like, unclassified or triple-negative breast cancer than luminal A [4, 62, 66] or non triple-negative breast cancer [65, 108]. African Americans also had an elevated risk of HER2+/ER- breast cancer versus Luminal A breast cancer [4, 62, 66]. Hispanics had a higher risk of HER2+ than Luminal A disease when compared to non-Hispanic whites, but the relative risk of triple-negative disease was unclear [62, 65, 78]. Women of Asian descent also appeared to get HER2+ disease more often than non-Hispanic whites, but had a reduced risk of triple-negative disease [62, 78, 108]. In the only study to examine whether African American race was a risk factor for breast cancer subtypes relative to controls, Trivers et al. [66] reported an inverse association for luminal A, a null association for luminal B, and positive associations for HER2+/ER- and triple negative.

2.6.2.3 Family history

Most studies reported a positive association between family history and each subtype, relative to non-cases [79, 82, 83, 110-113], with a slightly weaker or null effect for luminal B, HER2+/ER-, basal-like, or triple-negative disease relative to luminal A [4, 62]. The only exception to this was the pooled analysis, which exhibited a positive OR for the effect of first degree family history on basal-like breast cancer relative to luminal A [103].

2.6.2.4 Menopause

While there were no consistent patterns for the effect of menopausal status on subtype, later age at menopause was a risk factor for luminal A breast cancer. In CBCS, being premenopausal was negatively associated with HER2+/ER- breast cancer, relative to

luminal A breast cancer, but did not affect the risk of luminal B, HER2+/ER-, basal-like or unclassified disease [4]. One other study identified a slight increase in luminal B versus luminal A breast cancer in premenopausal women [62], and a third study found that being premenopausal increased the risk of luminal A or B breast cancer, relative to controls [79].

The Nurses' Health Study [83], PBCS [82], a registry-based, US case-control study [114], the Women's Health Initiative [115] and a Chinese case-control study [79] all found evidence that increasing age at menopause was positively associated with risk of luminal A disease relative to non-cases. Both the Nurses' Health Study and the US case-control study also reported a non-null association between age at menopause and Luminal B or HER2+/ER- disease. As for basal-like and triple-negative breast cancers, effect estimates from the Nurses' Health Study, the US case-control study, the Women's Health Initiative and a Japanese case-control study supported a null association [83, 113-115], while a Chinese case-control studies reported slightly increased risk for triple-negative disease among women with late onset menopause [79], and PBCS reported an increased risk for unclassified, but not basal-like tumors [82]. No studies estimated ccORs.

2.6.2.5 Age at menarche

In nearly all of the populations examined, women with luminal A, basal-like, or triple negative breast cancer were more likely than non-cases to have started menstruating at an early age [4, 66, 79, 82, 83, 109, 110, 113-116]. Several studies also found a positive association between early age at menarche and luminal B or HER2+/ER- disease [66, 79, 110, 114], but a few others reported only null effects [66, 82, 83, 113, 116]. In two of the three studies with ccOR estimates, women who menstruated at a very young age were more likely to get basal-like or triple negative than luminal A breast cancer [4, 66, 103]. Trivers et

al. [66] was the only study to observe an increased risk for HER2+/ER- versus luminal A among women with early age at menarche, and all luminal B versus luminal A ccORs were near-null.

2.6.2.6 Parity and lactation

Giving birth to one or more children had an inverse association with luminal A breast cancer in all eleven case-control comparisons [4, 66, 79, 82, 83, 110, 114-118]. Most studies that evaluated the effect of parity on luminal B breast cancer also found a reduced risk [66, 79, 82, 110], and case-case analyses indicated that the effect was roughly equivalent for the luminal B versus luminal A subtype [4, 62, 66, 103]. Whether compared to controls or luminal A cases, most studies found no evidence of an association between parity and HER2+/ER- breast cancer. The two exceptions to this were Trivers et al. [66], who reported an increased risk of HER2+/ER- disease among women with higher parity, relative to either controls or luminal A cases, and Xing et al. [79], who identified a substantially reduced risk for HER2+/ER- disease among women with one child, relative to controls.

The effect of parity on basal-like or triple-negative breast cancer was less clear. While only two of the twelve case-control comparisons detected a positive association between any parity and risk of either basal-like or triple-negative breast cancer relative to controls [4, 115], parity was associated with an increased risk of triple-negative, basal-like or unclassified breast cancer relative to luminal A in several investigations [4, 62, 66, 103, 108]. Additionally, four studies examined the effect of increased parity (i.e. giving birth to more than one child versus only one child) on triple-negative disease, two of which estimated a positive association [108, 115]. The other two studies did not find a similar pattern [114, 116].

Breastfeeding, regardless of duration, had an inverse association with each breast cancer subtype in most case-control comparisons [4, 66, 79, 83, 109, 110, 113-116]. Estimated ccORs less than 1 for luminal B and triple-negative breast cancer suggest that lack of breast-feeding is a particularly important risk factor for these two subtypes [4, 62, 66, 108].

Interestingly, when Millikan et al. [4] examined the joint effects of parity and lactation in women with basal-like breast tumors, they found that parity was only associated with an increased risk of cancer in women who did not breastfeed. Similarly, Kwan et al. [62] found that higher parity was associated with greater odds of having luminal B, HER2+/ER- or triple-negative versus luminal A breast cancer in women who did not lactate, but not in women who did.

2.6.2.7 Age at first full term pregnancy

Subtype-specific effect estimates for late age at first full term pregnancy were inconsistent and difficult to separate out from the effect of parity in general. Lack of a common referent group made comparisons across studies especially difficult.

Four studies used nulliparous women as the referent group [4, 66, 109, 115] when comparing subtypes to non-cases. Accordingly, all effect estimates for luminal A cancers were less than 1, but were higher for women in older age categories than younger age categories. The same was true for luminal B cancers in the one study with estimated effects [66]. The reported findings for HER2+/ER- and basal-like or unclassified breast cancers were too inconsistent to draw conclusions.

Five other studies analyzed the effect of age at first full term pregnancy on breast cancer subtype as a continuous variable [82, 83, 103, 110, 116]. In four of these five studies,

every 1-5 year increase in age at first full term pregnancy was associated with an increased risk of luminal A breast cancer. The pattern held for luminal A in four of six studies [79, 113-117] comparing relative risks for older versus younger mothers.

The results from case-case comparison studies reported positive associations between parity and luminal B, HER2+/ER-, basal-like or unclassified breast cancer relative to luminal A in younger mothers, but noted attenuated effects as maternal age increased [4, 62, 66, 103].

2.6.2.8 Other reproductive risk factors

Subtype-stratified analyses of less established reproductive risk factors for breast cancer, such as OC use, HRT use, time since last full term pregnancy, and abortion also revealed some interesting results.

When OC use was limited to an ever versus never comparison, most subtype-specific estimates for non-luminal A breast cancer were null, though the luminal B estimates demonstrated some striking inconsistencies [4, 62, 110, 117]. Dolle et al. [109] identified statistically significant increases in the risk of triple-negative disease, relative to controls, with long term OC use, especially if women were under the age of 40 when they initiated use. Gaudet et al. [110] also found an elevated risk of triple-negative breast cancer for ever versus never OC use in a similar age group, as well as an inverse association between ever OC use and luminal A breast cancer.

Subtype-specific analyses of ever versus never HRT use revealed null associations between HRT and luminal A, basal-like or triple-negative disease, but contradictory findings for luminal B or HER2+/ER- disease [4, 62, 119]. More detailed exposure classifications indicated that current HRT users were at higher risk of luminal A, luminal B, and possibly

basal-like and unclassified breast cancer, relative to non-cases, particularly if progestin was also included [83, 114, 119].

Three studies examined the association between subtype and time since last pregnancy [66, 116, 120]. In two of these three studies, cancers diagnosed soon after giving birth were more likely to be HER2+ than hormone receptor positive cases, HER2- cases or controls. Trivers et al. [66] and Li et al. [116] both found that women who were more recently pregnant had reduced risk of luminal A or B disease, but only Trivers et al. observed a similar protective effect for recent pregnancy and triple-negative disease.

Lastly, in the only two studies to examine whether spontaneous or induced abortions were related to subtype-specific breast cancer risk, one found a non-significant, positive association between abortion and triple-negative breast cancer, relative to controls [109], and another found an increased risk of luminal A or HER2+/ER- breast cancer with induced abortions, but a decreased risk for all four subtypes for spontaneous abortions [79].

2.6.2.9 Body size

Because postmenopausal adiposity is a known risk factor for breast cancer, most subtype analyses of body size are stratified by menopausal status. Body mass index (BMI) is the most common measure of body size, and is usually categorized according to the World Health Organization criteria: underweight (<18.5), normal (18.5-24.99), overweight (25.0-29.99) or obese (≥30.00) [121]. Based on these criteria, Millikan et al. [4] found an inverse association between higher category of BMI and postmenopausal luminal A and postmenopausal basal-like breast cancer, relative to controls. Phipps et al. [122] found a positive association between higher category of BMI and risk of any breast cancer among postmenopausal women participating in the Women's Health Initiative. Three additional

studies of BMI in postmenopausal women, Phipps et al., Gaudet et al., and Yang et al. [82, 110, 123] reported only near-null findings. All case-case comparisons were also near-null [4, 103, 110].

In a pooled analysis, Pierobon et al. [124] found a positive association between obesity (BMI ≥30) and triple-negative breast cancer. The association was particularly strong among premenopausal women. Effect estimates for premenopausal BMI and breast cancer subtype revealed a possible negative association between high BMI and luminal A breast cancer, as found in three studies [4, 66, 82], but not two others [109, 110]. The relationship between high BMI and premenopausal luminal B cancer was inconsistent in these studies, with Trivers et al. [66] reporting an inverse association, Gaudet et al. [110] reporting a positive association and Yang et al. observing no effect [82]. Effect estimates from case-control comparisons of premenopausal HER2+/ER breast cancer were mostly null, though results from case-case comparisons indicated that women with larger body size had a higher risk of HER2+/ER- relative to luminal A [4, 62, 66, 103].

Waist-to-hip ratio (WHR), weight and hip circumference, and weight gain were also examined as potentially relevant body size indicators. Millikan et al. [4] and [122] both found positive associations between high WHR and postmenopausal luminal A breast cancer, relative to non-cases. This association held for premenopausal women in Millikan et al [4]. Millikan et al. also found an association between high WHR and basal-like breast cancer among pre- and postmenopausal women, but Phipps et al. saw no effect in their sample of postmenopausal women. In a case-case analysis, high WHR was a greater risk factor for postmenopausal luminal A breast cancer than either luminal B or HER2+/ER-, but a lesser risk factor than for basal-like breast cancer [4]. Among premenopausal women, basal-like

breast cancer again had the highest increase in risk associated with high WHR, while luminal B, HER2+/ER-, and unclassified tumors all had null effects relative to luminal A tumors [4].

Additionally, Tamimi et al. [83] found that higher BMI at age 18 was associated with a reduced risk in luminal A and basal-like breast cancer, and that weight gain since age 18 was positively associated with unclassified, but not basal-like breast cancer. Phipps et al. [122] reported positive effect estimates for the association of weight gain, waist circumference, and hip circumference with luminal A breast cancer, and negative effect estimates for the association of waist circumference and triple-negative breast cancer. These findings were inconsistent with an earlier paper by Phipps et al. [123], which reported predominantly null subtype-specific effect estimates for BMI at age 30 and weight gain or loss. Corresponding analyses for luminal B and HER2+/ER- subtypes were also mostly null, with the exception of a negative association between higher BMI at age 18 and HER2+/ER- breast cancer identified by Tamimi et al [83]. No case-case analyses of these alternative body size measures were conducted.

2.6.2.10 Physical activity

Increased physical activity was associated with a reduced risk of luminal A, HER2+/ER-, and triple-negative breast cancer in two studies [66, 122], but a third reported predominantly null findings [113]. The effect of increased physical activity on luminal B breast cancer was null in a case-control comparison, but was positively associated with risk relative to luminal A [66]. Increased physical activity was also associated with an increase in triple-negative breast cancer relative to luminal A.

2.6.2.11 Alcohol and smoking

The luminal A subtype has a strong positive association with high alcohol and cigarette consumption and the HER2+/ER- subtype may be associated with increased alcohol consumption and a history of smoking. Luminal B, basal-like, and triple-negative breast cancers are probably not associated with alcohol consumption, but former smoking may affect the risk of these subtypes.

Compared to never drinkers, women with increasingly higher alcohol consumption had greater relative risks of developing luminal A breast cancer than non-cases [66, 83, 113, 125]. Despite some evidence of a negative association between increased alcohol consumption and triple-negative breast cancer in the Women's Health Initiative [125] and an Atlanta-based case-control study [66], the Nurses' Health Study [83], and a SEER registry study [109] reported null effects, and a Japanese case-control study [113] reported a positive trend. Alcohol consumption was positively associated with HER2+/ER- breast cancer in both the Atlanta-based case-control study [66] and the Nurses' Health study [83], but not with luminal B breast cancer in either study. Relative to luminal A breast cancer, the effect of alcohol on luminal B, HER2+/ER- or triple-negative disease was either null or inconsistent [4, 62, 66].

Increased smoking intensity and duration were associated with an increased risk of luminal A breast cancer in one study [125] and with a decreased risk in another study [113]. Being a former smoker rather than a current smoker or non-smoker was associated with increased risk of luminal A disease [66, 125]. Former smokers also had an increased risk of luminal B and HER2+/ER- breast cancer [66], while current smokers had a reduced risk of luminal B, but an increased risk of HER2+/ER-. Estimates for the effect of smoking intensity

and duration on triple-negative tumors were mostly null and the effect of former or current smoking status was inconsistent [66, 109, 125]. In case-case comparisons, former smoking was associated with an increased risk of luminal B relative to luminal A, and current smoking was associated with an increased risk of HER2+/ER- relative to luminal A [66]. Subtype-specific ccORs for smoking intensity and duration were null for luminal B and HER2+/ER- and inconsistent for triple-negative tumors [4, 62, 113, 125].

2.6.2.12 Benign breast disease and breast density

Subtype-specific estimates for the effects of high breast density and previous benign breast disease were provided in one and two studies, respectively. Ma et al. [126] observed a positive association between high breast density and risk of both luminal A and basal-like breast tumors. The effect was roughly equivalent for both subtypes, as indicated by a null ccOR.

The Nurses' Health Study [66] and the Cancer and Steroid Hormone study [110] both reported a positive association between benign breast disease and luminal A breast cancer. The Nurses' Health Study also observed positive associations between benign breast disease and luminal B, basal-like and unclassified breast cancer, while the Cancer and Steroid Hormone study estimated a positive but imprecise association between benign breast disease and luminal B breast cancer and null effects for the remaining subtypes.

2.6.2.13 Summary of risk factors by subtype

Although the exact relationship between breast cancer subtypes and some of their possible risk factors is still poorly understood, many such associations are well characterized. The variability between each subtype-specific risk factor profile provides strong evidence that the subtypes represent etiologically distinct diseases.

Luminal A is the most common subtype and is associated with most previously established breast cancer risk factors, such as family history of breast cancer, later age at menopause, early age at menarche, nulliparity, lack of breastfeeding, later age at first full term pregnancy, decreased physical activity, increased alcohol consumption, history of benign breast disease and high breast density. The risk factors for luminal B breast cancer are fairly similar to those for luminal A, with the exception of older age at diagnosis, which is negatively associated with luminal B and positively associated with luminal A. Several other luminal A risk factors had null or inconsistent associations with luminal B disease, including race, age at menopause, age at menarche, lack of physical activity, and alcohol consumption. Only a few of the established risk factors are associated with HER2+/ER- disease, including younger age at diagnosis, African American race, family history of breast cancer, lack of breastfeeding, recent pregnancy, and high alcohol consumption. Lastly, triple-negative breast cancer is associated with several of the well-established risk factors, but its risk factor profile still varies quite a bit from that of luminal A. When compared directly to luminal A, triple negative tumors are more strongly associated with younger age at diagnosis, African American race, family history of breast cancer, early age at menarche, higher parity, lack of breastfeeding and high premenopausal BMI than luminal A tumors.

2.7 Genetic risk factors for breast cancer

The high correlation between family history and breast cancer risk has inspired countless investigations of the disease's genetic origins. Even though the currently identified susceptibility loci account for only a small proportion of the disease's measured heritability [127-129], their discoveries have helped elucidate some of the biological underpinnings of

the disease and offered new targets for screening or therapeutic intervention. Furthermore, the ever-growing body of literature on variability in genetic risk factors by breast cancer subtype provides additional evidence that these subtypes have unique etiologies.

2.7.1 *BRCA1* and *BRCA2*

The first high penetrance breast cancer susceptibility genes identified were aptly named Breast Cancer Gene 1 (*BRCA1*) and Breast Cancer Gene 2 (*BRCA2*) [130-132]. Both genes were discovered through linkage analysis, a technique that uses genetic markers to identify chromosomal regions disproportionately shared by diseased family members. Due to linkage disequilibrium (LD), proximal chromosomal regions are not randomly distributed during gametogenesis. Therefore, a marker strongly associated with disease incidence is likely physically near a causal locus, even if it does not tag the causal mutation. Family-based linkage analyses are particularly powerful for identifying susceptibility genes with high penetrance.

BRCA1 and *BRCA2* are both DNA repair genes and tumor suppressor genes. Although their specific mechanisms are slightly different, the proteins encoded by these genes help activate DNA double strand break repair mechanisms and initiate homologous recombination to replace damaged sequences [133, 134]. Individuals born with mutations in *BRCA1* or *BRCA2* may be less equipped to fix or suppress cells with damaged DNA, thereby increasing the likelihood that these damaged cells proliferate and form tumors.

BRCA mutations are very rare but highly penetrant. According to population-based estimates, approximately 0.04% and 0.4% of all individuals have *BRCA1* and *BRCA2* mutations, respectively [135], with a mutation prevalence of 2-3% for each gene among female breast cancer cases [135-137]. Mutation frequencies vary greatly by race and

ethnicity, with around 8% of all Ashkenazi Jewish cases exhibiting a mutation in *BRCA1*, but only 0.5% of Asian-Americans [137]. Compared to white women with breast cancer, African American cases are less likely to have a *BRCA1* mutation and more likely to have a *BRCA2* mutation [10, 135], although one study reported a *BRCA1* mutation prevalence of 17% (95% CI: 7%, 34%) in African Americans diagnosed before age 35 [137]. In terms of penetrance, around 60% of *BRCA1* mutation carriers will develop breast cancer by age 70, as will approximately 50% of *BRCA2* mutations carriers [127, 138].

Cases with *BRCA1* mutations are much more likely to have triple-negative disease than cases with wild type *BRCA1* [41, 139-147]. The association is particularly strong in young women [147]. *BRCA1* mutations are also associated with increased basal-marker expression [41, 148]. There is no consistent association between *BRCA2* and any breast cancer subtype.

2.7.2 Candidate genes

As noted previously, family-based linkage studies can effectively identify chromosomal regions associated with disease susceptibility in affected, related individuals, but they rarely pinpoint specific, causal mutations and cannot detect low-penetrant variants. The next step in determining which variants are causally related to the disease (or disease subtype) of interest is to conduct association analyses of single nucleotide polymorphisms (SNPs) or other genetic anomalies in unrelated individuals. Alternatively, investigators may select SNPs because of their location in a functionally-relevant gene or because a particular base-pair deletion, insertion, or mutation alters the amino acid sequence for a disease-related protein.

In a comprehensive review and meta-analysis published in 2011, Zhang et al. [20] summarized the findings of more than 1000 breast cancer candidate gene studies. Overall, they examined 521 candidate genes or chromosomal regions and conducted meta-analyses of 279 variants. Twenty-nine variants had statistically significant associations in meta-analyses, but only 10 variants in 6 genes met the authors' criteria for strong evidence of association, which required replication and protection from bias in addition to statistical significance. These 6 genes were: *ATM*, *CASP8*, *CHEK2*, *CTLA4*, *NBN*, and *TP53*. The authors considered three additional genes, *CYP19A*, *TERT* and *XRCC3,* to have moderate evidence of association. Though purposefully excluded from this review, *PTEN, BRIP1*, and *PALB2* were also mentioned as well-established, highly penetrant susceptibility genes. Variants identified in genome-wide association studies (GWAS) were also excluded.

Results from less comprehensive reviews of breast cancer candidate gene association studies are generally consistent with the findings of Zhang at el. For example, Antoniou and Easton [149] state that *BRCA1*, *BRCA2*, *TP53*, *PTEN*, *ATM*, and *CHEK2* are well-established breast cancer susceptibility genes, though they also mention *LKB1* (also known as *STK11*). Freisinger and Domchek [150] and Hirschfield et al. [151] consider *BRCA1*, *BRCA2*, *TP53*, *PTEN,* and *STK11* to be well-established, highly penetrant genes and *CHEK2, ATM*, *BRIP1* and *PALB2* to be uncommon, but well-established susceptibility genes of low to moderate penetrance.

Relatively few studies have addressed how these well-established breast cancer susceptibility genes relate to ER or PR status, and even fewer have included subtype-stratified analyses. Though limited, these reports provide evidence that some genes may be

differentially related to the expression of hormone receptors in female breast tumors. A summary of the existing evidence is provided in Table 3.

The best-studied example of subtype differentiation by candidate gene status is *CHEK2*, located on chromosome 22. Of the four specific mutations Zhang et al. [20] deemed strongly associated with breast cancer, three (1100delC, IVS2+1G>A, and I157T) have been assessed within strata of ER, PR or intrinsic subtype. A base pair deletion in exon 10 of the gene (1100delC) showed evidence of an association with ER+ breast cancer in four studies [152-155], while a fifth showed no association [156] and a sixth observed a stronger association with ER- than ER+ tumors [157]. The 1100delC mutation was also associated with ER- tumors in Cybulski et al. [153], though to a lesser degree than ER+ tumors. The same was true for a similar mutation involving deletions in exon 9 and 10 (del5395). In a gene-expression analysis, all CHEK2 1100delC tumors clustered with the luminal A and luminal B tumors [158].

The two other *CHEK2* mutations identified by Zhang et al. in their meta-analysis were not associated with ER status in Meyer et al. [156], though Cybulski et al. [153] found a strong association between ER+ breast cancer and IVS2+1G>A substitution, and Domagala et al. [159] observed a positive association between the I157T/ rs17879961 polymorphism and luminal A and B breast cancer, and an inverse association between I157T and basal-like and triple-negative breast cancer. The pattern held when Domagala et al. examined whether having any *CHEK2* mutation affected intrinsic subtype, but Cybulski observed statistically significant, positive associations between *CHEK2* mutations and both ER+ and ER- disease.

Subtype-stratified analyses for *ATM* (chromosome 11), *TERT* (chromosome 5), *CASP8* (chromosome 2) and *TP53* (chromosome 17) also revealed some possible

discrepancies in genetic risk factors by hormone receptor status. In a large consortium study, Cox et al. [160] identified a SNP in *ATM* (rs1800054) that was positively associated with PR+ breast tumors. Barroso et al. [161] replicated this association for nine SNPs in the same LD block (see Figures 2 and 3) [161]. Most SNPs in this LD block had no association with ER status, though three SNPs were positively associated with ER- disease [162, 163]. LD blocks were defined using Haploview (Haploview 4.2, Version 1.0, Broad Institute, Cambridge, MA, USA) [164], International HapMap version 3, release 2 for Utah residents with ancestry from northern and western Europe (CEU) and Yorubans from Ibadan, Nigeria (YRI) populations [9], and haplotype block criteria established in Gabriel et al. [165].

Although the association between breast cancer and *TERT* was first identified through a candidate gene approach, the only study with subtype-specific effect estimates is a genome-wide association study of African American and triple-negative cases [166]. Here, rs10069690 was a strong, positive genetic risk factor for triple-negative breast cancer. In a different pooled analysis of triple-negative cases, Stevens et al. [167] identified a rs17468277 in *CASP8* (Figures 4 and 5) that was inversely associated with triple-negative disease. Two other SNPs in *CASP8* (rs1861270 and rs1045485) were associated with all breast cancer subgroups with approximately equal magnitude and direction [160, 162, 168-170]. Lastly, multiple studies of rs1042522 in *TP53* (Figures 6 and 7) indicated a possible positive association between the SNP and ER+ disease [171-174].

Subtype-specific analyses were conducted for SNPs in *CTLA4*, *XRCC3*, *CYP19A1* (Figures 8 and 9), *PALB2* (Figures 10 and 11), and *BRIP1*, but the results were either too sparse or too conflicting to make meaningful inferences [161, 171, 172, 174-182]. To date, no studies have evaluated how *LKBI*, *NBN*, or *PTEN* affect hormone receptor status.

2.7.3 Genome-wide association studies

A genome-wide association study (GWAS) is "a study that compares the complete DNA of people with a disease or condition to the DNA of people without the disease or condition" [183]. In practice, GWAS examine the association between a disease of interest and at least 100,000 SNPs selected to represent the entirety of the genome via LD patterns. To be established as a true GWAS-hit, a SNP must have a p-value $<1 \times 10^{-5}$ in the overall study population, which usually includes the original sample and one or more replication samples [184]. In practice, this α-level is often set as high as 5×10^{-8}, as this corresponds to $\alpha=0.05$ corrected for 1,000,000 independent SNP tests.

As of January 2013, 23 breast cancer GWAS have been published, the first of which was published in June 2007. Basic descriptions of these 23 studies and the GWAS-significant SNPs identified by each can be found in Table 4. All information was downloaded from the National Human Genome Research Institute Catalog of Published Genome-Wide Association Studies [185].

In total, 58 unique GWAS hits have been identified in these 23 studies. These SNPs appear in the following genes or gene regions: 1 on chromosome 1 (1p11.2) [186], 3 on chromosome 2 (2p16.1, 2q35, and *ERBB4*) [186-192], 2 on chromosome 3 (*SLC4A7* and *SIAH2*) [189, 190, 193], 8 on chromosome 5 (*TERT*, *ROPN1L*, 2 on 5p12, *MRPS30*, 2 on *MAP3K1*, and 5q34) [166, 186, 188-191, 194, 195], 8 on chromosome 6 (*ECHDC1*/*RNF146*, 6q25, *FAM46A*, *TAB2* and 4 on *ESR1*/*C6orf97*) [189, 190, 196-199], 2 on chromosome 7 (7q11.22 and 7q32.3) [188, 200], 2 on chromosome 8 (both on 8q24.21) [189, 190, 194], 2 on chromosome 9 (*CDKN2A*/*CDKN2B* and *RAD23B*/*KLF4*/*ACTL7A*) [189, 190], 10 on chromosome 10 (*ANKRD16*/*FBX018*, 2 on *ZNF365*, *ZMIZ1*, and 6 on *FGFR2*) [186, 189-

191, 193, 194, 200-202], 4 on chromosome 11 (*MYEOV/CCND1*, *BARX2*, and 2 on *LSP1*) [189, 194, 199], 1 on chromosome 12 (12q21.1) [188], 1 on chromosome 13 (*ABCC4*) [188], 2 on chromosome 14 (*RAD51L1* and *GALC*) [186, 203], 1 on chromosome 15 (*FBN1*) [188], 4 on chromosome 16 (3 on *TOX3* and *GLG1*) [186, 187, 189-191, 194, 204, 205], 1 on chromosome 17 (*COL1A1*) [188], 1 on chromosome 18 (18q21.2) [188], 3 on chromosome 19 (19p13.11, *ABHD8* and *ZNF577*) [195, 198, 206], 1 on chromosome 20 (*RALY*) [198]and 1 on chromosome 21 (*GRIK1*) [188]. Li et al. [207] conducted a GWAS among women with ER- disease, but failed to identify any genome-wide significant associations. Of the remaining 22 GWAS studies, 13 provided effect estimates stratified by hormone-receptor status [186, 187, 189, 192, 193, 197-200, 203-206], though this was usually limited to those SNPs achieving GWAS-significance.

Additionally, both Easton et al. and Hunter et al. [194, 201] identified several strongly associated SNPs that were just shy of the GWAS inclusion criteria. Easton et al. obtained association p-values of 6×10^{-5}, 2×10^{-5} and 0.001 for rs4666451 on chromosome 2p, rs2107425 on the *H19* gene and rs30099 on chromosome 5q, respectively, while Hunter et al. identified 4 SNPs (rs12505080 on 4p, rs7696175 on *TLR1*/TLR6, rs17157903 on *RELN*, and rs10510126 on 10q) that had very strong associations with breast cancer in the initial study population, but failed to replicate in a secondary analyses. Two other SNPs of potential genome-wide significance were identified in Stacey et al. [208] and Ahmed et al. [209], both of which were follow-ups to previous GWAS (Stacey et al. [187] and Easton et al. [194], respectively). Although neither rs10941679 in *MRPS30* or rs6504950 in *COX11* reached GWAS-significance in the initial studies, further analyses in additional populations produced p-values less than 1×10^{-5}.

Of the 58 GWAS SNP hits and 9 nearly GWAS-significant SNP hits, 31 were successfully genotyped in Carolina Breast Cancer Study (CBCS) participants and are thus included in this review. The other GWAS hits were either not yet discovered at the time the genotyping was requested or failed preliminary quality control checks. This review also covers several additional SNPs from GWAS-identified genes that were genotyped by CBCS investigators to augment analyses of these genomic regions. A brief discussion of each of these regions is described below and in Table 5. Unless otherwise noted, this table includes the range of estimates for log-additive models with the major allele in the HapMap CEU population as the referent genotype [9]. Also included in this table are minor allele frequencies (MAFs) from the HapMap YRI population and African Americans living in the Southwest USA (ASW). Results from studies conducted in special populations (e.g. women with *BRCA1* or *BRCA2* mutations or women with contralateral disease) are not included in these summary tables. Previously reported effect estimates from the CBCS population are also excluded [210]. Lastly, as few studies collected data on HER2 status and even fewer collected data on intrinsic breast cancer subtypes, the term 'subtype' is used to indicate any type of differentiation by IHC marker status.

2.7.3.1 1p12: rs11249433

First identified in Thomas et al. [186], rs11249433 was later confirmed by both Turnbull and Li [189, 191]. The MAF was around 40% in most European samples, [169, 186, 189, 191, 211-213] but much lower in African Americans (13-16%) [11, 169, 214-216]. In women of European descent, the SNP's relationship with breast cancer appeared to follow a log-additive pattern with each copy of the risk allele corresponding to a 10-15% increase in the odds of developing breast cancer. There was no association among African Americans.

Subtype analyses indicated that the SNP was most strongly associated with ER+, PR+, HER2-, Luminal A and Luminal B breast cancer [167, 169, 186, 212, 214, 217].

2.7.3.2 2p: rs4666451

As mentioned previously, SNP rs4666451 just missed the criteria for GWAS significance in Easton et al. [194], with $p=6 \times 10^{-5}$ and OR=0.97 (95% CI: 0.94-1.00) for a log-additive model. Three of four subsequent studies confirmed the approximate magnitude and direction of the association [162, 186, 211, 218]. The effect was similar in both ER+ and ER- breast cancers, but there was no clear association with HER2 status [162, 186]. No studies have examined the effect of rs4666451 among African Americans.

2.7.3.3 2q35: rs13387042

Despite its location in a gene desert, rs13387042 is one of the most well studied breast cancer-related polymorphisms. Approximately half of all women of European descent carried the variant allele [162, 168, 169, 186, 187, 189-191, 213, 218-223], as did 25-30% of all African Americans [11, 12, 169, 187, 214-216]. Among women of European ancestry, each copy of the variant allele was associated with a 10-15% reduction in the odds of getting breast cancer. The association was weaker and less precise in African Americans, but seemed consistent for each subtype versus control comparison, including ER+, ER-, PR+, PR-, HER2+, HER2-, luminal A, luminal B, HER2+/ER-, triple-negative, and basal-like breast cancer [162, 167, 168, 169, 187, 198, 214, 216, 217, 219, 221, 222, 224].

2.7.3.4 SLC4A7: rs4973768

Located on chromosome 3 in the *SLC4A7* gene, rs4973768 was first identified in Turnbull et al [189]. With one exception [213], all studies found that the SNP was positively associated with the risk of breast cancer in women of European ancestry, with OR estimates

of 1.08-1.16 [168, 169, 189-191, 209, 211, 220, 223]. In contrast, the SNP had a much weaker and statistically null association with breast cancer among African Americans [11, 169, 214-216]. A recent meta-analysis found a strong overall association between the SNP and breast cancer, but null effects when limited to women of European or African ancestry [225]. In subtype-stratified analyses, rs4973768 was most strongly associated with luminal A breast cancer, with no apparent association with triple-negative disease [167-169, 209, 214, 216, 217, 226].

2.7.3.5 4p: rs12505080 and TLR1: rs7696175

These SNPs were identified in a GWAS investigation using participants of the Nurses' Health Study [201]. In both cases, effect estimates for log-additive model showed no association, but general model estimates indicated that compared to women homozygous for the wildtype allele, heterozygotes had an increased risk of breast cancer (OR=1.22, 95% CI: 1.02-1.45 for rs12505080 and OR=1.39, 95% CI: 1.15-1.68 for rs7696175), while women homozygous for the minor allele had a decreased risk of breast cancer (OR=0.51, 95% CI: 0.35-0.73 for rs12505080 and OR=0.86, 95% CI: 0.67-1.09 for rs7696175). These patterns were replicated in Stage II of Hunter et al., but no other studies have examined the association. To date, no one has examined subtype differences in the effect of rs7696175 in whites or African Americans, but one study of Chinese women observed statistically significant associations between one or more copies of the minor allele and ER- and HER2- disease [227].

2.7.3.6 MRPS30: rs4415084 and rs10941679

In a follow-up to their original GWAS paper [187], Stacey et al. [208] identified two correlated SNPs (r^2=0.51 in HapMap CEU) in the *MRPS30* gene that were strongly

associated with breast cancer risk. Fletcher et al. confirmed the association of rs4415084 and breast cancer in a later GWAS [190].

Among women of European ancestry, each copy of the rare variant at rs4415084 or rs10941679 conferred a 10-15% increase in the odds of breast cancer [169, 186, 190, 191, 208, 211, 220, 223, 228-230]. These SNPs were not correlated in HapMap YRI (r^2=0.12) and effect estimates among African American women were inconsistent, with many null findings [11, 12, 169, 208, 214-216, 231]. Both SNPs showed stronger associations with ER+, PR+, luminal A, and luminal B cancers, with null associations with ER-, PR- or triple-negative disease [167, 169, 186, 198, 208, 214, 216, 217, 230, 231].

2.7.3.7 5p12: rs981782

In 5 separate study populations, the presence of a rare variant at rs981782 was inversely associated with breast cancer risk in women of European ancestry [162, 194, 208, 218, 230]. In the only study to report subtype-stratified estimates, Reeves et al. [162] observed a similar reduction in risk in ER+, ER- and HER2+ disease. rs981782 was not polymorphic in women of African descent.

2.7.3.8 5q: rs30099

rs30099 was strongly associated with breast cancer risk in Easton et al. [194], but fell short of genome-wide significance (OR=1.05, 95% CI: 1.01-1.10, p=0.001). Harlid et al. [218] and Reeves et al. [162] confirmed the association in two later studies, with Reeves et al. reporting stronger effects for ER- and HER2+ disease subtypes. No studies have estimated the effect of rs30099 in African Americans.

2.7.3.9 MAP3K1: rs889312

SNP rs889312, located on the *MAP3K1* gene on chromosome 5, was equally common in women of European and African descent, with a MAF of approximately 30%. The 10-15% increase in risk was well-replicated among women of European descent [162, 168, 169, 189, 194, 213, 218, 220, 222, 229, 232-237], but was weak or null in African American women [11, 169, 214-216, 236, 238]. A meta-analysis including 6 separate studies estimated a pooled OR of 1.09 (95% CI: 1.07-1.12), assuming a log-additive model [239]. Subtype-stratified analyses indicated that the SNP was associated with approximately equal risk increases in all disease subtypes, including ER+, ER-, PR+, PR-, HER2+, HER2-, luminal A, luminal B, HER2+/ER-, triple-negative and basal-like breast cancer [162, 167-169, 198, 214, 216, 217, 222, 224, 226, 232, 234, 236, 238, 240].

2.7.3.10 ECHDC1: rs2180341

Gold et al. [196] discovered rs2180341 in a GWAS conducted among Ashkenazi Jewish women with strong family histories but no *BRCA1* or *2* mutations. Kirchhoff et al. [241] replicated these findings in both Jewish and non-Jewish women, but two other investigations yielded only null results [169, 213]. The SNP was not associated with breast cancer in African Americans [11, 12, 215, 216, 241] and subtype analyses generated predominantly null findings [169, 216, 241].

2.7.3.11 ESR1: rs2046210, rs851974, rs2077647, rs2234693, rs1801132, rs3020314 and rs3798577

Four SNPs near the *ESR1* gene were identified in GWASs, though only rs2046210 will be discussed here. Zheng et al. [197] discovered this SNP in a GWAS of Chinese women in the Shanghai Breast Cancer Study and replicated the finding in a sample of white women from the Nashville Breast Cancer Study. Three other large studies confirmed the positive

association among women of European ancestry [169, 242, 243], with a fourth smaller study reporting null results [213]. rs2046210 had no effect on breast cancer in a study of African Americans [12], but a nearby correlated SNP (r^2=0.38 in YRI), rs851974, had a positive association with the disease [12]. Most of the seven other studies of rs2046210 in African Americans estimated near-null associations [11, 169, 197, 214-216, 242, 243]. Subtype-stratified analyses suggested that rs2046210 had a positive association with all subtypes, though most of these analyses were conducted in Asian women [167, 169, 192, 197, 198, 214, 216, 217, 224, 226, 227, 242, 244].

As *ESR1* is the gene responsible for encoding the alpha form of estrogen receptors, it is a frequent candidate in studies of genetic risk factors for breast cancer. Among the most frequently studied SNPs are rs2234693 (also known as the *Pvull* T397C mutation), rs1801132 (Pro325Pro), rs2077647 (Ser10Ser), rs3798577 (UTR-3, T>C) and rs3020314 (C5029T), all of which were included in a meta-analysis by Zhang et al. [20]. Of these SNPs, rs3020314, rs1801132 and rs2234693 had statistically significant, but likely biased effect estimates (ORs for additive models were 1.12, 95% CI: 1.06-1.18; 0.95, 95% CI: 0.90-1.00; and 0.94, 95% CI: 0.89-1.00, respectively), with the effect of rs2234693 limited to women of Asian descent. The pooled effect estimates for rs2077647 and rs3798577 were both null (OR=0.98, 95% CI: 0.93-1.03 and OR=0.97, 95% CI: 0.90-1.04, respectively). An LD map of *ESR1* SNPs genotyped in HapMap is included as Figures 12 and 13. Though not picture here, the two other GWAS-significant SNPs, rs3734805 and rs3757318, are located even further away from the 5' end of *ESR1* than rs20406210.

2.7.3.12 RELN: rs17157903

SNP rs1715903 in *RELN* was strongly associated with breast cancer in Hunter et al. [201], but fell short of GWAS-defined significance levels in both the original and replication analyses. No other investigators have attempted to validate the association in whites or African Americans.

2.7.3.13 8q24: rs13281615 and rs1562430

A gene desert on 8q24 contains two SNPs that are strongly associated with breast cancer. rs13281615 was identified in the first breast cancer GWAS [194] and rs1562430 was identified a few years later [189]. These SNPs were in the same LD block in HapMap CEU (r^2=0.43) but not YRI (r^2=0.25). More than a dozen studies have validated the association between rs13281615 and breast cancer, with effect estimates ranging from 1.07 to 1.31 for the log-additive model [162, 168, 169, 189, 213, 218, 220, 229, 232-235, 237, 245]. The polymorphism was equally common in African Americans (MAF≅0.43) as in whites, though most studies observed only null associations [11, 12, 169, 214-216]. In subtype-specific analyses, rs13281615 was most strongly associated with luminal A and B tumors [162, 167-169, 198, 214, 216, 217, 232, 234, 246], with no observed association with triple-negative disease. rs1562430 was re-assessed in two other GWAS investigations and one replication study, all of which confirmed the inverse association between the rare variant and all breast cancer in women of European descent [186, 190, 223]. In a Chinese GWAS, rs1562430 was associated with ER+, ER-, PR+ and PR- disease [192].

2.7.3.14 CDKN2A/CDKN2B: rs1011970, rs3731257 and rs3731249

SNP rs1011970, located near *CDKN2A* and *CDKN2B*, was discovered by Turnbull et al. [189] and was later replicated in one study of women of European descent [247]. Four

studies of African Americans [11, 214-216] reported predominantly null associations. In both studies of whites, each additional copy of the rare allele was associated with an increased risk of 5-10%. In limited subtype analyses, rs1011970 was positively associated with ER+ breast cancer and triple-negative breast cancer, but not ER- breast cancer [167, 189, 198, 214, 216, 247]. Although poorly studied, analysis of rs3731257 and rs3731249 may provide additional information about the role of the *CDKN2A/CDKN2B* region, as both are located downstream of *CDKN2A*, rather than upstream of *CDKN2B*, as is the case with rs1011970 (Figures 14 and 15). In previous analyses, rs3731249 (A148T) was strongly associated with breast cancer in young Polish women (OR for dominant model =1.5, p=0.0002) [248], but rs3731257 had no effect in a population of British women (OR for homozygous variant versus homozygous wildtype= 0.95, 95% CI: 0.74-1.20) [249].

2.7.3.15 ANKRD16: rs2380205, ZNF365: rs10995190, and ZMIZ1: rs704010

Three SNPs in three separate genes on chromosome 10 were originally identified in Turnbull et al [189]. Both rs2380205 in *ANKRD16* and rs10995190 in *ZNF365* had inverse associations with breast cancer (OR=0.94, 95% CI: 0.91-0.98 and OR=0.86, 95% CI: 0.82-0.91 for rs2380205 and rs10995190, respectively), while rs704010 in *ZMIZ1* had a positive association with the disease (OR=1.07, 95% CI: 1.03-1.11). All three SNPs were confirmed in at least one other white population [247, 250]. All three were also examined in African American populations [11, 214-216], though rs704010 was the only SNP to demonstrate evidence of an association [216]. Both rs2380205 and rs10995190 were inversely associated with ER+ disease in women of European descent [189, 247], while rs704010 was positively correlated with all ER and PR subtypes [189, 198, 247]. Neither rs2380205 nor rs704010 were associated with triple-negative disease [167].

2.7.3.16 FGFR2: rs3750817, rs10736303, rs11200014, rs2981579, rs1708806, rs1219648, rs2912774, rs2936870, rs2420946, rs2981582, and rs3135718

Although the causal relationship between the *FGFR2* gene and breast cancer is not completely understood, multiple GWAS, GWAS-validation studies, and fine-mapping studies have confirmed that one or more SNPs in the gene are strongly associated with the disease. Easton et al. [194] were the first to detect a strong association between breast cancer and rs2981582. In Hunter et al. [201], rs1219648 had the strongest association and in Thomas et al. [186] it was rs2981579. In a later GWAS limited to Japanese hormone positive breast cancer cases [193], a fourth *FGFR2* SNP, rs3750817, showed the strongest association with disease. All four SNPs were strongly correlated in the HapMap CEU population, but not the YRI. Estimated effects for women of European descent were consistent for rs2981582, rs1219648 and rs2981579, with each copy of the rare allele conferring approximately a 20-30% increase in the odds of breast cancer [162, 168, 169, 186, 189-191, 194, 201, 213, 218-220, 222, 223, 229, 232-237, 251-256]. The rare allele of rs3750817 was inversely associated with disease in studies of whites [169, 252], but positively associated with disease in the Japanese GWAS [193].

In a recent meta-analyses, rs2981582 had an OR of 1.23 (95% CI: 1.20-1.26) among women of European descent [257]. In the same meta-analyses, the effect of rs1219648 was OR=1.26 (95% CI: 1.22-1.31). A meta-analysis of rs2981579 reported an OR of 1.34 (95% CI: 1.28, 1.38) in women of European descent [258].

Effect estimates for rs2981582 and rs2981579 were inconsistent or null in African American samples [11, 12, 169, 214-216, 236, 259], with meta-analysis ORs of 1.08 (95% CI: 0.97, 1.21) [257] and 1.09 (95% CI: 0.92, 1.30) [258], respectively. In contrast, rs1219648 was positively associated with breast cancer in African Americans (meta-analysis

OR= 1.13, 95% CI: 1.01, 1.26) [257]. To date, no studies have examined the relationship between rs3750817 and breast cancer in African Americans.

rs1219648 or rs2981582 were more strongly associated with ER+ and PR+ disease than ER- or PR- disease [257]. rs1219648 also demonstrated a strong association with HER2- disease in white US women [236]. In the three studies that examined the effect of rs2981582 by combined ER/PR/HER2 status, the SNP was associated with luminal A and B disease, but not HER2+/ER- or triple-negative disease [167, 168, 226]. rs2981579 was associated with ER+/PR+ disease in one study of African Americans [216] and with ER+, ER-, PR+, HER2+ and HER2- disease in a study of Chinese breast cancer patients [227]. The SNP was not associated with ER- disease in women of European descent [198]. rs3750817 was associated with ER+ and PR+ disease in whites, but effect was in the opposite direction of the Japanese GWAS [193].

Eight other SNPs in the same CEU LD block as the four GWAS hits were genotyped in CBCS (Figures 16 and 17). These include rs1076303, rs11200014, rs1078806, rs1219648, rs2912774, rs2936870, rs2420946, and rs3135718. Of these SNPs, rs2420946 and rs11200014 were the most extensively studied. Each had effect estimates comparable to those seen for the three GWAS SNPs. For rs2420946, the estimated meta-analysis ORs was 1.25 (95% CI: 1.19-1.32) [257]. Only one study looked at the SNP in an African American population, finding a small positive association [12]. No one has examined subtype-specific effects. For rs112000014, the log-additive meta-analysis ORs were 1.28 (95% CI: 1.21, 1.35) for whites and 1.03 (95% CI: 0.85, 1.24) for African Americans [258]. In a subtype analyses among Chinese women, the SNP was positively associated with ER+, PR+, HER2+ and HER2- disease [227].

rs2912774 also had well-replicated effect estimates of approximately 1.25 per copy of the rare allele among women of European descent [194, 252, 254-256] and the limited evidence on rs10736303, rs1778806, rs2981578, rs2936870, and rs3135718 suggested that these SNPs have similar associations [169, 194, 196, 252]. Two of the SNPs, rs10736303 and rs2981578, were also positively associated with breast cancer in African Americans, with some evidence that rs2981578 was a causal variant in this population [12, 214, 215, 258, 259]. When studied in subtype-specific analyses, most of these SNPs were associated with ER+ or PR+, but not ER- or PR- disease [169, 214, 259].

2.7.3.17 10q: rs10510126

Like rs12505080 on 4p, rs7696175 on *TLR1/TLR6*, and rs17157903 on *RELN*, rs10510126 had an extremely low p-value in Hunter et al.'s [201] initial analysis (OR=0.62, 95% CI: 0.51-0.75, $p=7.1 \times 10^{-7}$), but failed to reach genome-wide significance levels in replication analyses (OR=0.83, 95% CI: 0.74-0.93, p=0.001). The SNP was not assessed in any other study populations.

2.7.3.18 LSP: rs3817198 and rs909116

Easton et al. [194] discovered the association between rs3817198 and breast cancer in their 2007 GWAS. In this initial GWAS, the SNP was associated with a mere 7% increase in the odds of having breast cancer (OR=1.07, 95% CI: 1.04-1.11). Since then, several studies have replicated this small effect with reasonable precision [162, 168, 186, 189, 218, 220, 232, 233], while others reported null or even slightly contradictory findings [169, 211, 213, 229, 234, 235, 237]. Results from a meta-analysis of six of these studies indicated a null effect (OR=1.01, 95% CI: 0.94-1.09) [260]. Effect estimates among African Americans were imprecise and inconsistent [11, 12, 169, 214-216], as were subtype-specific estimates [162,

167-169, 198, 214, 216, 217, 232, 234, 246], though there was some indication that the SNP was associated with unclassified disease [168].

Of note, while attempting to replicate the association between rs3817198 and breast cancer, Turnbull et al. [189] found that rs909116, another SNP on *LSP1*, had a stronger effect and lower p-value in their population of white women (OR=1.17, 95% CI: 1.10-1.24, p=7.3 x 10^{-7}). The two SNPs were not correlated (CEU r^2=0.24, YRI r^2=0.03). No other studies have replicated this observation.

2.7.3.19 H19: rs2107425

In Easton et al. [194], rs2107425 on *H19* was inversely associated with breast cancer, but the p-value for the association was above the threshold for genome-wide significance. In the only other study to examine the effect of this SNP on breast cancer risk, Teraoka et al. [261] found that it was not related to the risk of contralateral versus unilateral disease.

2.7.3.20 MYEOV/CCND1: rs614367

A single SNP near the *MYEOV* and *CCND1* genes on chromosome 11 demonstrated a positive, genome-wide significant association with breast cancer in Turnbull et al. [189] (OR=1.15, 95% CI: 1.1-1.2, p=1.3 x 10^{-8}), which was later replicated by Lambrechts et al [247]. The effect seemed strongest in women with ER+ or PR+ disease [189, 198, 216, 247]. There was no clear association in African Americans [11, 214-216] or among those with triple-negative disease [167].

2.7.3.21 TNRC9/TOX3: rs8051542, rs12443621, rs3803662, rs4784227, rs3104746, and rs3112562

As with *FGFR2*, multiple GWAS pinpointed the *TNRC9* gene (also known as *TOX3*) as strongly associated with breast cancer. Of these GWAS, five identified rs3803662 as the SNP with the strongest signal [186, 187, 189, 194, 207], while a sixth [205] and a seventh

[190] detected stronger effects for rs4784227 and rs3112612, respectively. The latter SNP was not genotyped in CBCS, but rs3803662, rs4784227 and seven other *TNRC9* SNPs were.

Including the seven GWAS mentioned above, more than twenty papers have reported ORs for the effect of rs3803662 on breast cancer. In these papers, MAFs ranged from 0.20-0.39 for women of European descent [162, 163, 168, 169, 186, 187, 189, 194, 205, 207, 213, 218-220, 222, 229, 232-235, 237, 262] and 0.47-0.54 for women of African descent [11, 12, 169, 187, 214-216, 262, 263]. Other than one paper reporting an imprecise, inverse association [233], effect estimates for log-additive models in women of European descent ranged from 1.01-1.33, with most approximating a 20% increase in the odds of breast cancer per copy of the more rare 'A' allele. A meta-analysis of eight of these studies estimated an OR of 1.18 (95% CI: 1.09- 1.28) [264]. Interestingly, most studies among African Americans indicated that the 'A' allele had an inverse association with breast cancer (range 0.75-1.04), with most showing a 10% decrease in risk. In subtype-specific analyses, rs3803662 was positively associated with all subtypes [162, 167-169, 187, 192, 198, 214, 217, 222, 224, 226, 232, 234, 265, 266].

The other GWAS-identified SNP, rs4784227, was in LD with rs3803662 in the HapMap CEU population (r^2=0.81) but not the HapMap YRI population (r^2=0.03). Although far less studied that rs3803662, two studies in women of European ancestry [205, 262] and three studies in African American women [11, 214, 262] indicated that the SNP was positively associated with disease in both racial groups (OR range of 1.17-1.29 and 1.09-1.23, respectively). In the only study to assess subtype-specific effects, Kim et al. [192] found that all ER and PR-defined subtypes were positively associated with the minor allele at rs4784227.

None of the remaining *TNRC9* genotyped SNPs were correlated in CEU or YRI populations (Figures 18 and 19). Several investigators have estimated effects for rs8051542 and rs12443621, which are both located between the 3' end and rs3803662 [194, 205, 234, 237, 262]. Meta-analyses generated ambivalent results, with effect estimates on opposite sides of the null (OR=1.07, 95% CI: 0.93-1.23 for rs8051542 and OR=0.94, 95% CI: 0.85-1.05 for rs12443621) [264]. Both SNPs were analyzed in African Americans and within strata of ER and PR status, but only the effect of rs8051542 on ER+ breast cancer was non-null [12, 217, 224, 234, 263, 265, 266].

rs3104746 and rs3112562 are located on the other side of rs3803662 and rs4784227, closer to the 5' end of the gene. Ruiz-Narvaez et al. observed strong associations between breast cancer and rs3104746 (OR=1.23, 95% CI: 1.05-1.44) and rs3112562 (OR=1.17, 95% CI: 1.02-1.34) among participants in the Black Women's Health Study [263]. Chen et al. replicated these observations for rs3104746 and observed similar effects in ER+ and ER- breast cancer cases [214].

2.7.3.22 COX11- rs7222197 and rs6504950

In their follow-up to Easton et al. [194], Ahmed et al. [209] found that rs6504950, a SNP in the *COX11* gene, was strongly and inversely associated with breast cancer (OR=0.95, 95% CI: 0.92-0.97, with $p=1.4 \times 10^{-8}$ across all replication stages). These findings were replicated with varying degrees of precision in five other studies of Europeans or European-Americans (OR range 0.80- 0.95) [168, 169, 211, 213, 220] and a meta-analysis [267]. Two of five other studies in African Americans [11, 169, 214-216] reported effect estimates of similar magnitude and direction. Overall, these effects appeared to be the strongest in women with luminal A or B disease [167, 168]. The other *COX11* SNP, rs7222197, was in perfect

LD with rs6504950 in both HapMap CEU and YRI, and was inversely associated with breast cancer in Turnbull et al. (OR=0.89, 95% CI: 0.83-0.96, p=0.0009) [189].

2.7.4 Summary of genetic risk factors

As shown by this literature review, the search for genetic risk factors for breast cancer has identified dozens of associated variants at a variety of chromosomal loci. Some of these polymorphisms or mutations are extremely rare, but have large effects on disease risk. This includes *BRCA1*, *BRCA2*, *ATM*, *PTEN,* and some of the other genes identified in linkage analyses or candidate gene studies. More common variants, such as rs13387042, or SNPs in *FGFR2* or *TNRC9/TOX3*, tend to have weak associations with disease, especially because these polymorphisms are often proxies for nearby, correlated causal variants rather than contributory links themselves. Most of these variants have been replicated in several study populations, with each new analysis providing additional information on the form, magnitude, and direction of the association.

Despite heavy replication in women of European and Asian descent, far fewer studies include women of African descent. Studies that do include African Americans are often quite small, resulting in imprecise or inconsistent estimates. In many cases, it is not clear whether the failure to find and replicate associations is due to racial differences in genetic risk factors and linkage disequilibrium patterns [268-270], or if the study populations are simply too small to allow the detection of a weak association with sufficient precision. Subtype-specific analyses face the same methodological challenge. Differences in gene expression patterns, race and age distributions, prognoses, and risk factor profiles all imply that breast cancer subtypes have distinct etiologies and thus unique genetic origins, but results from subtype-stratified analyses are thus far scant and generally inconclusive.

If our ultimate goal is to improve breast cancer prevention, detection and treatment, we must first understand breast carcinogenesis, much of which is determined by individual variations in genetic make-up. Analyses within strata of race and subtype are of particular importance, as risk patterns vary according to these factors. Such analyses require either larger studies, which are expensive and logistically complicated, or the use of more powerful analysis methods within existing populations. The subsequent pages of this report describe the implementation, findings and conclusions of a Bayesian-based study of the effects of previously identified genetic risk factors on breast cancer and breast cancer subtypes. The use of Bayesian methods, which incorporate prior knowledge into effect estimation, generated more accurate and precise ORs than were possible with traditional frequentist techniques.

Figure 1: US breast cancer mortality rate by race, age group, and year

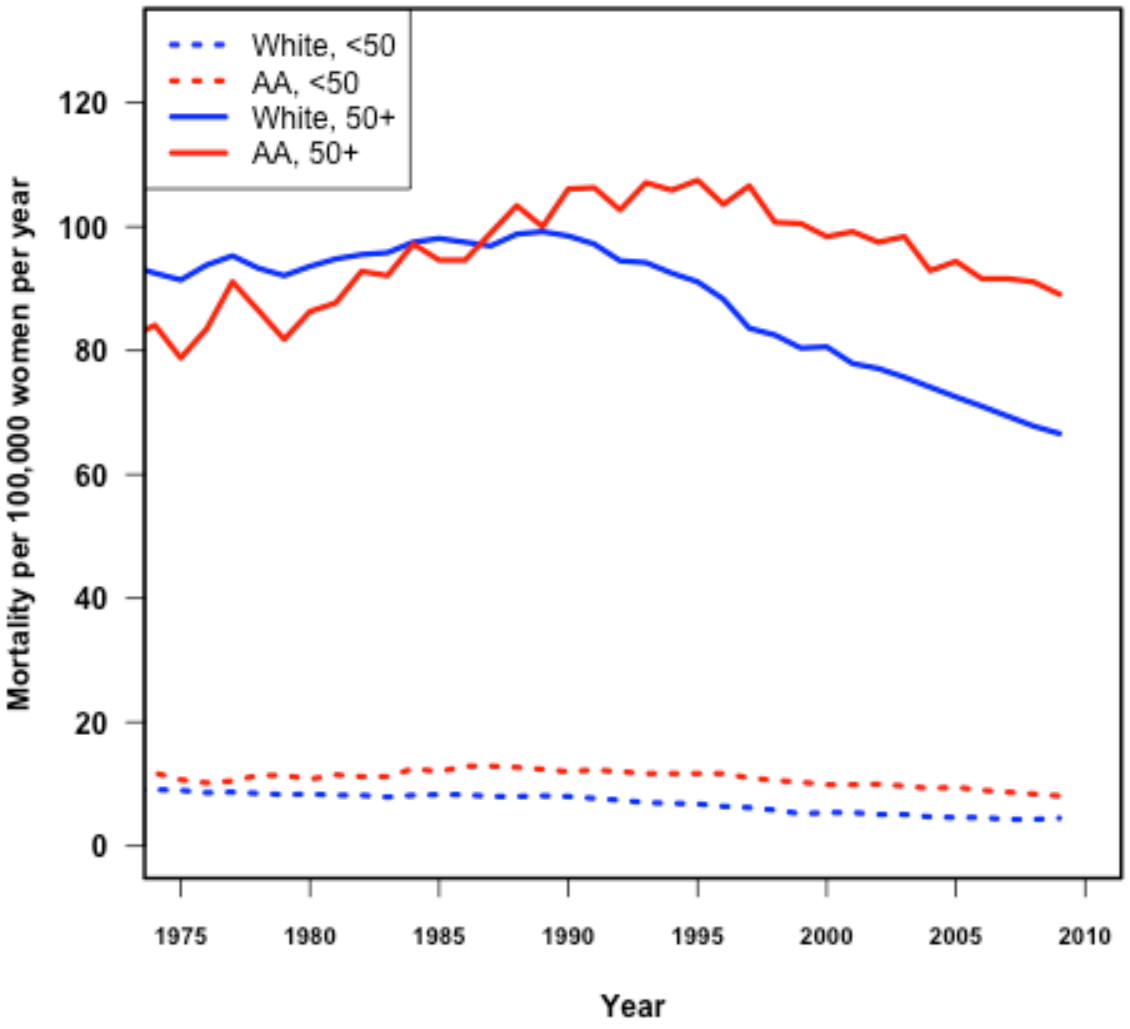


Figure 2: ATM linkage disequilibrium map for HapMap CEU

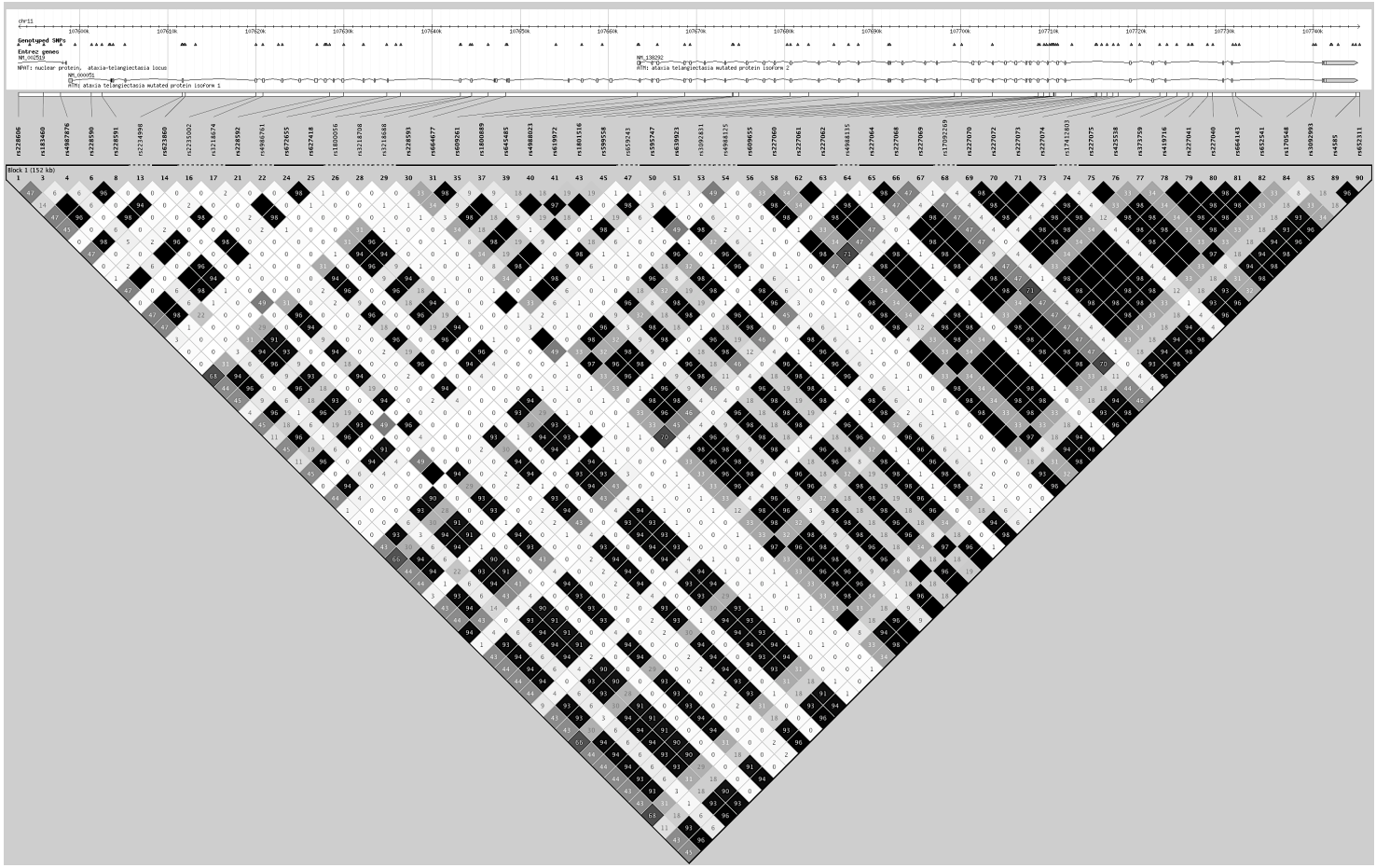


Figure 3: ATM linkage disequilibrium map for HapMap YRI

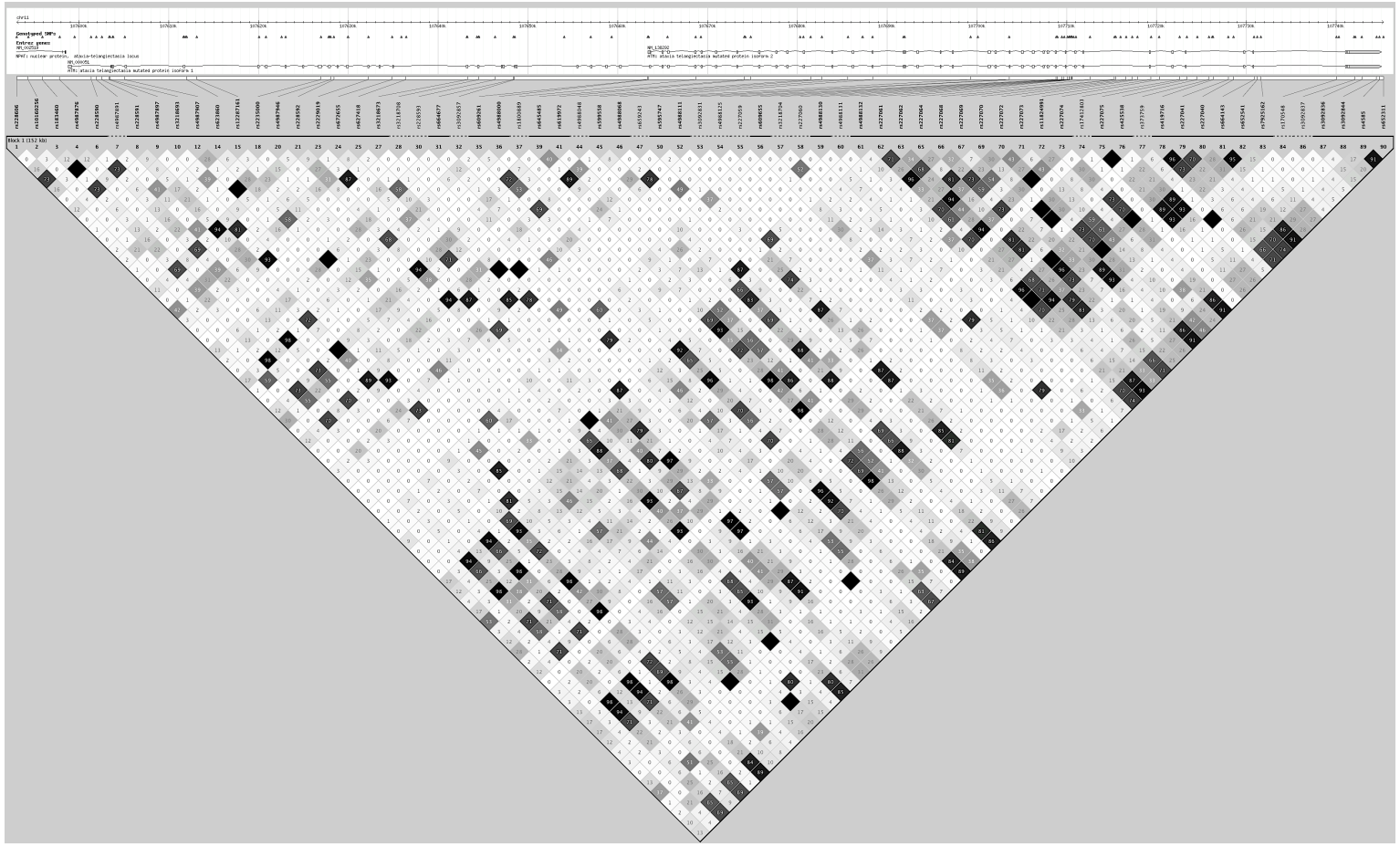


Figure 4: CASP8 linkage disequilibrium map for HapMap CEU

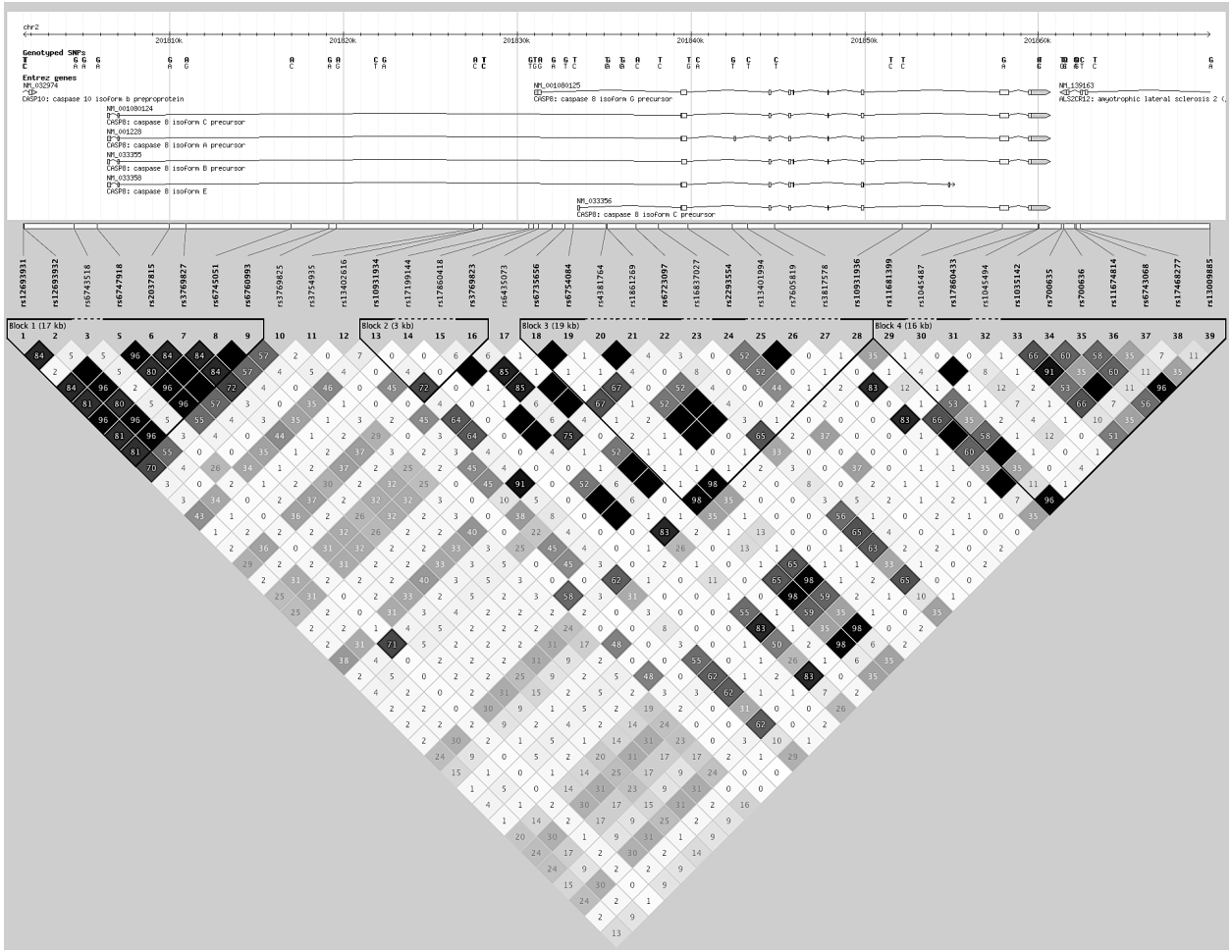


Figure 5: CASP8 linkage disequilibrium map for HapMap YRI

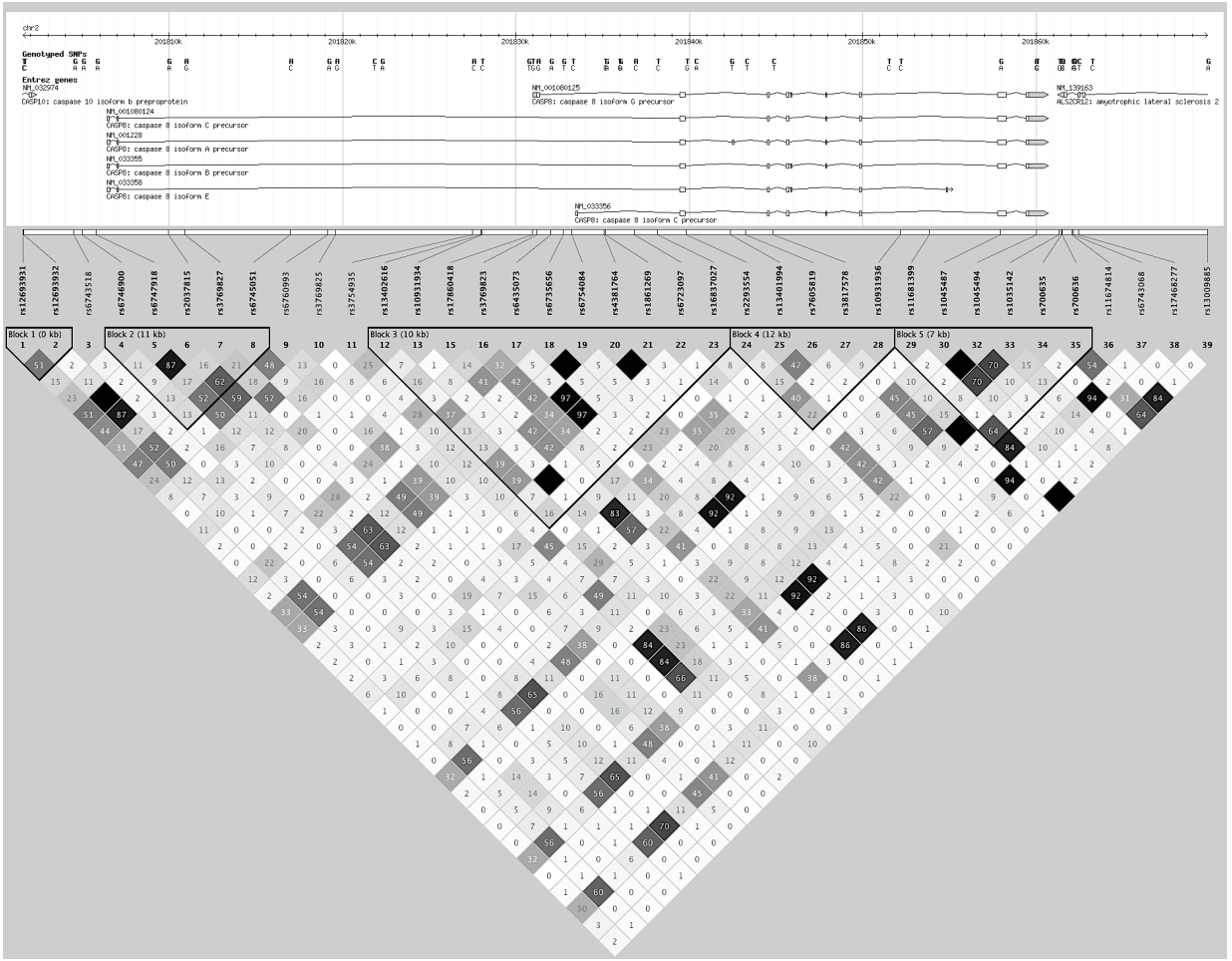


Figure 6: TP53 linkage disequilibrium map for HapMap CEU

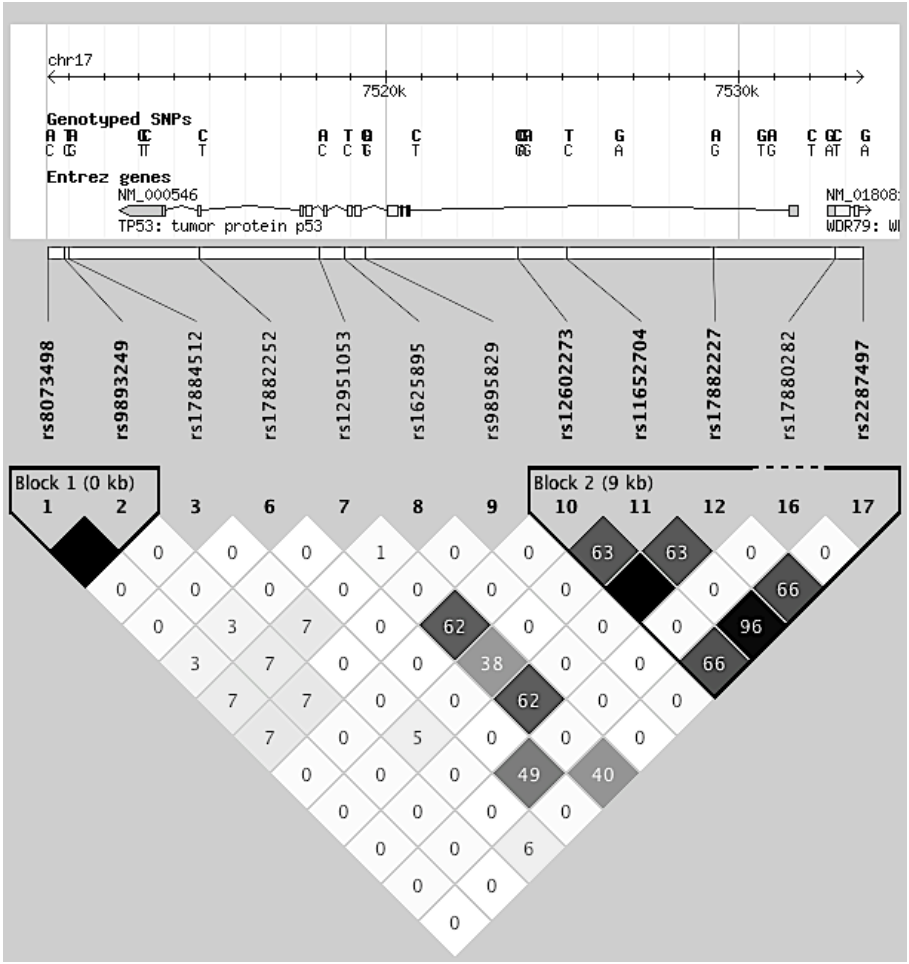


Figure 7: TP53 linkage disequilibrium map for HapMap YRI

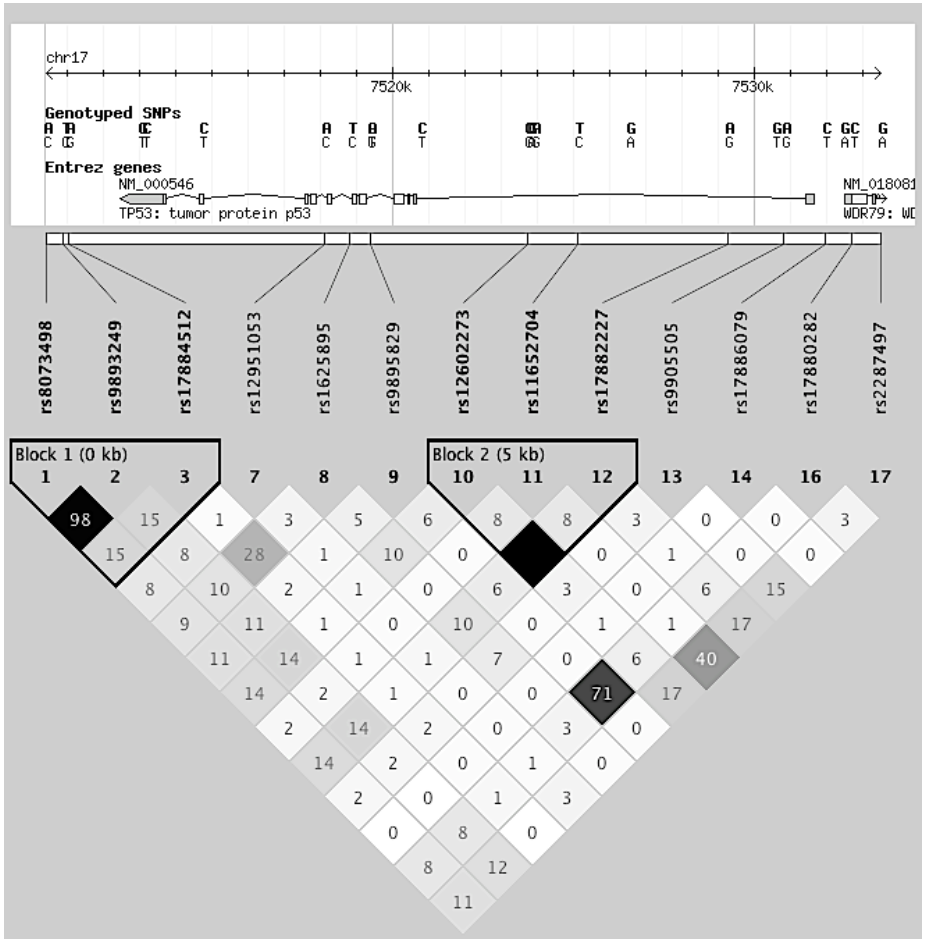


Figure 8: CYP19A1 linkage disequilibrium map for HapMap CEU

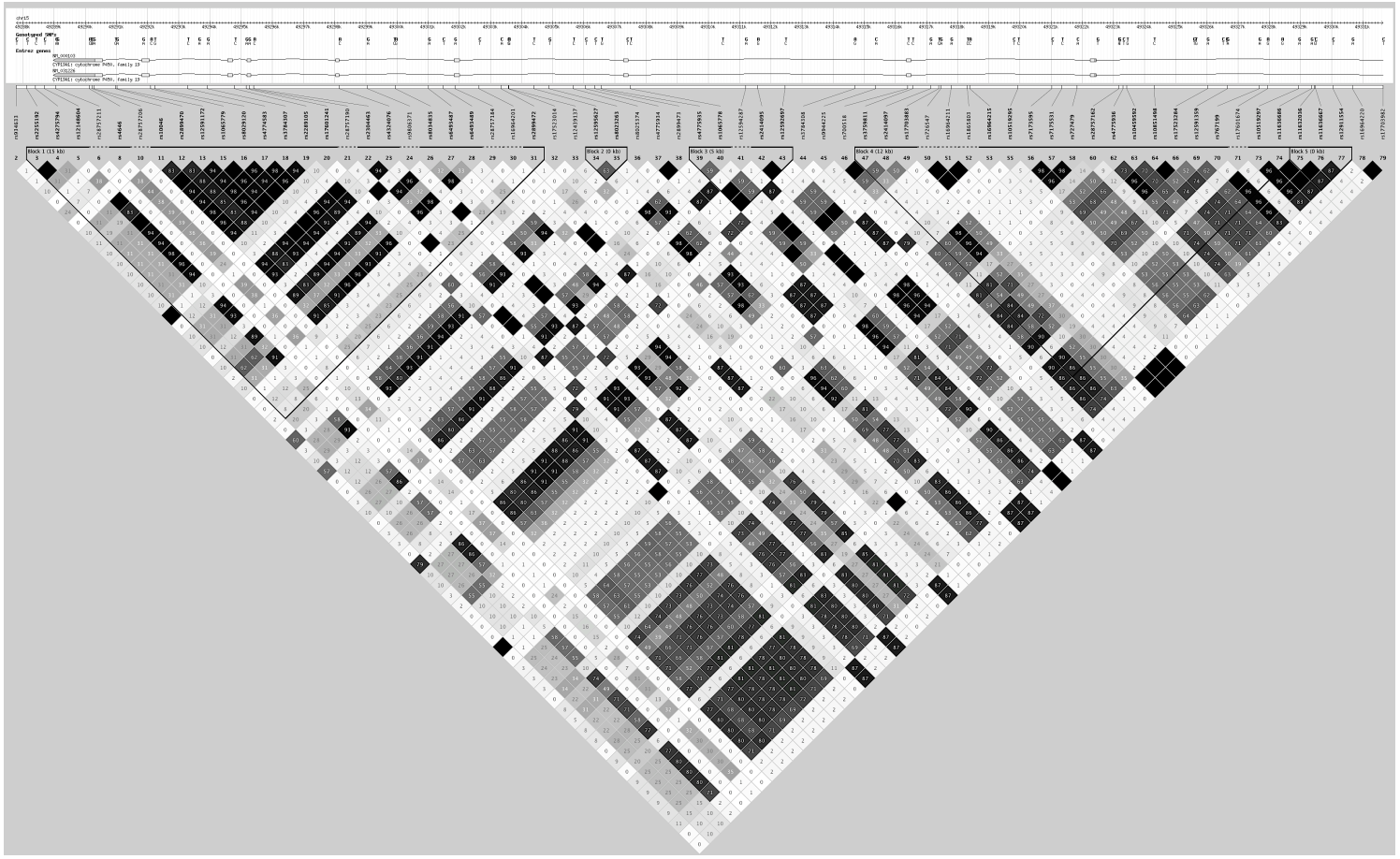


Figure 9: CYP19A1 linkage disequilibrium map for HapMap YRI

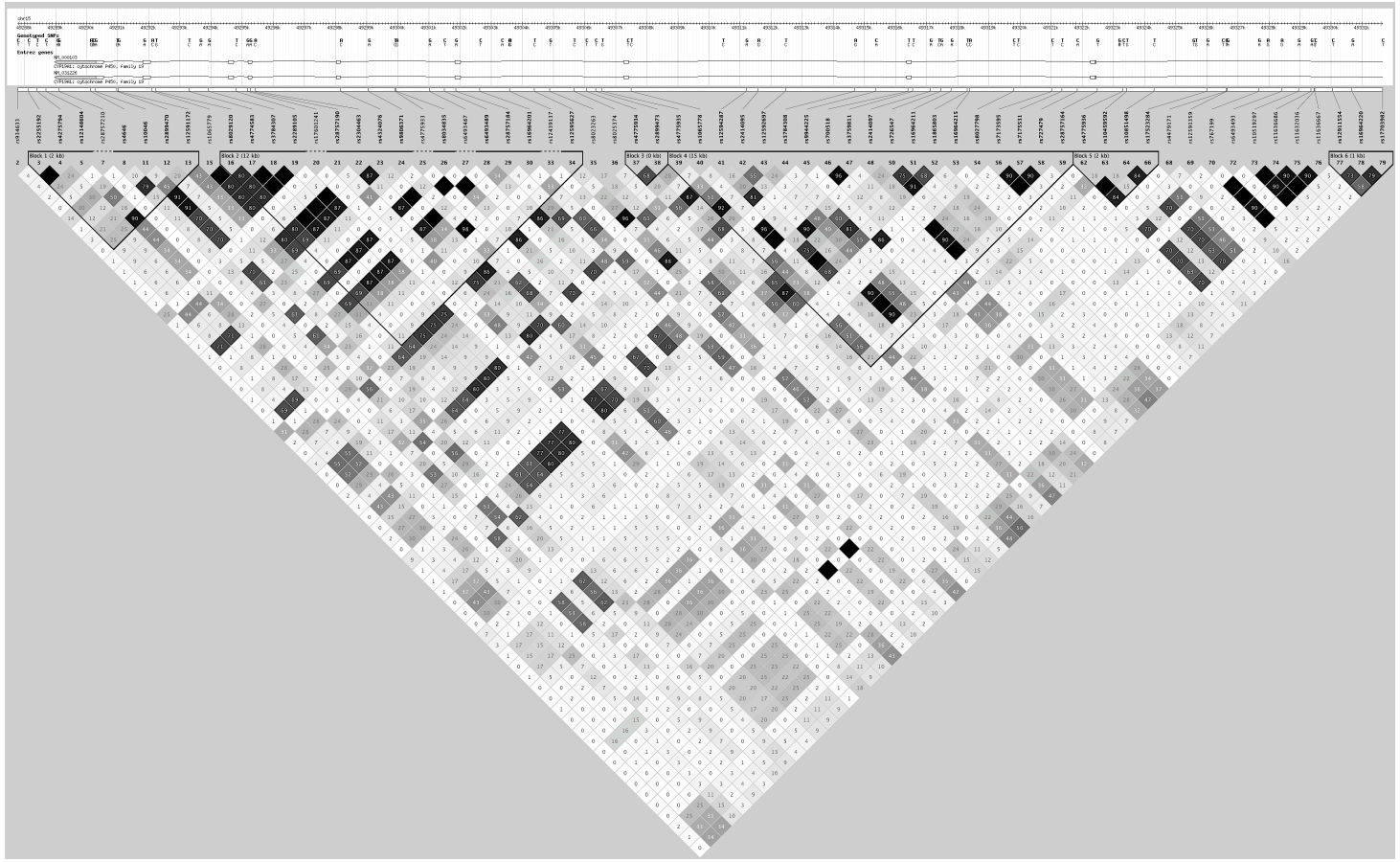


Figure 10: PALB2 linkage disequilibrium map for HapMap CEU

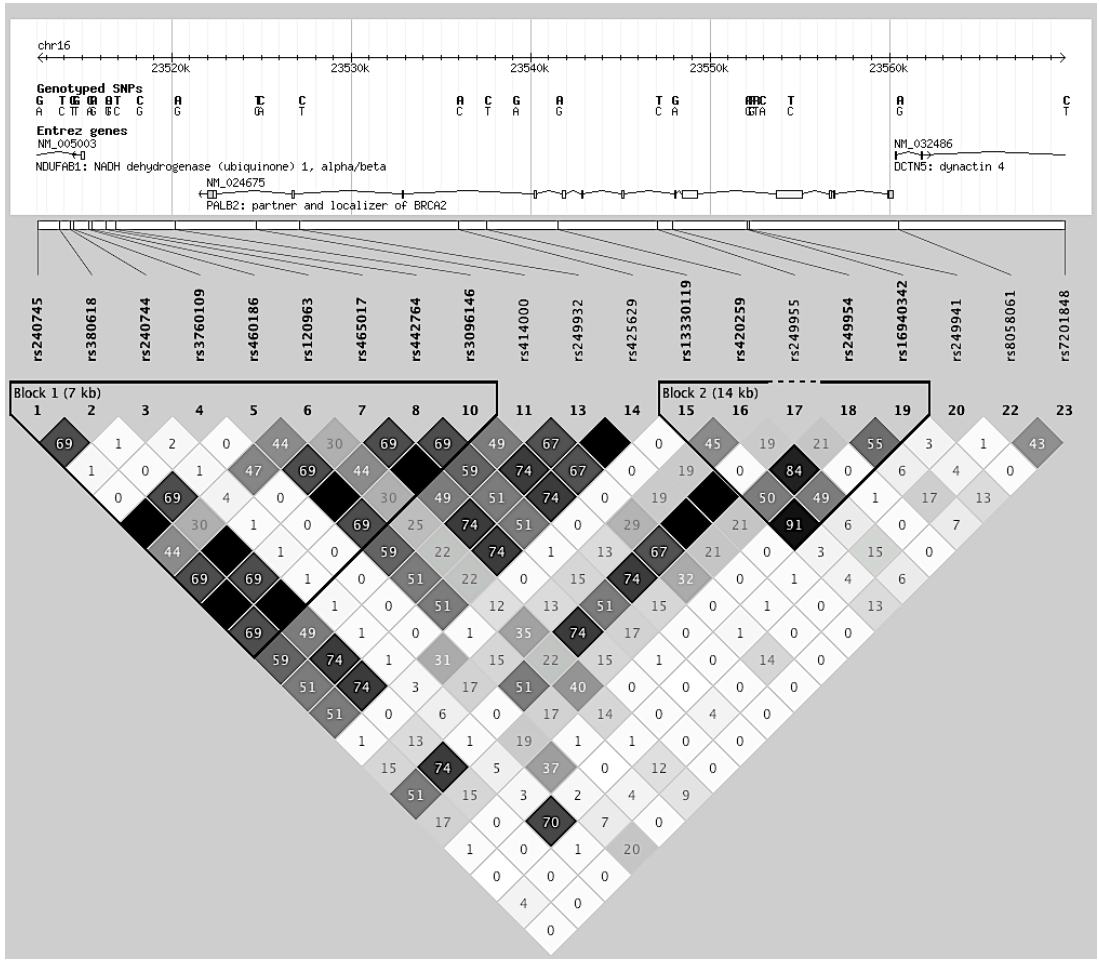


Figure 11: PALB2 linkage disequilibrium map for HapMap YRI

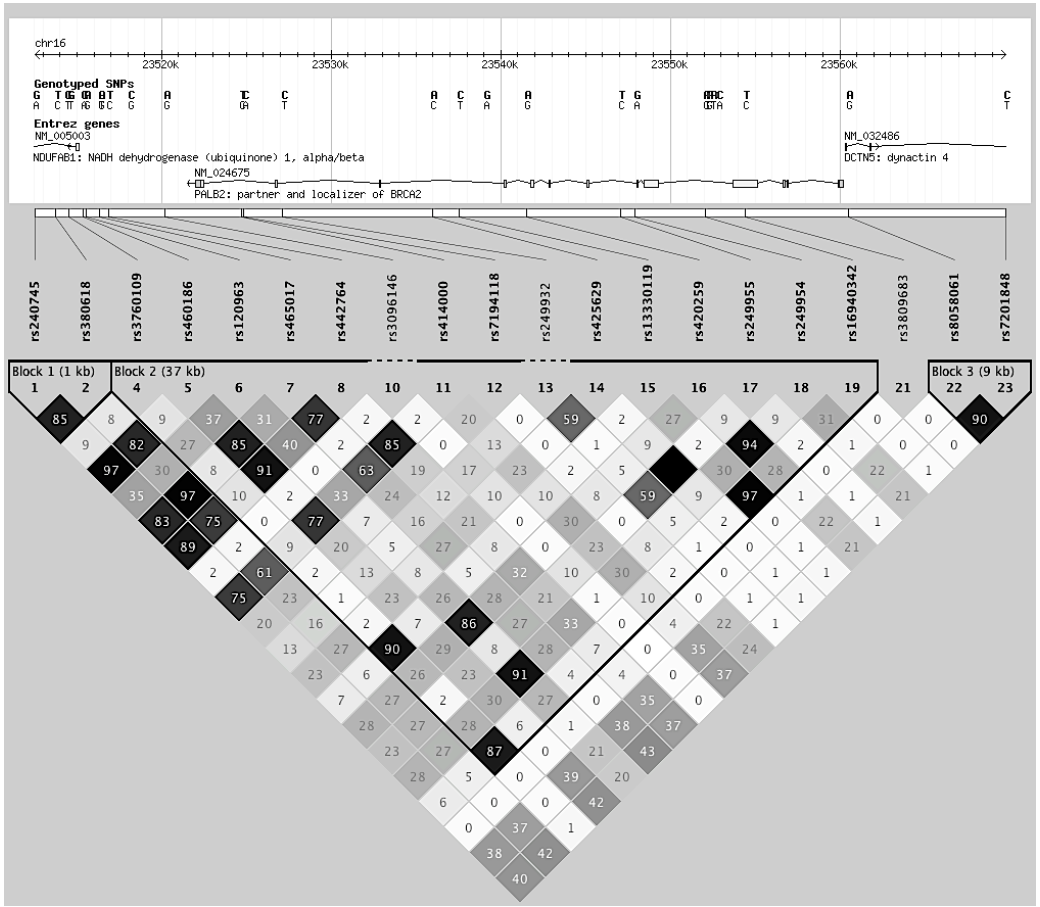


Figure 12: ESR1 linkage disequilibrium map for HapMap CEU

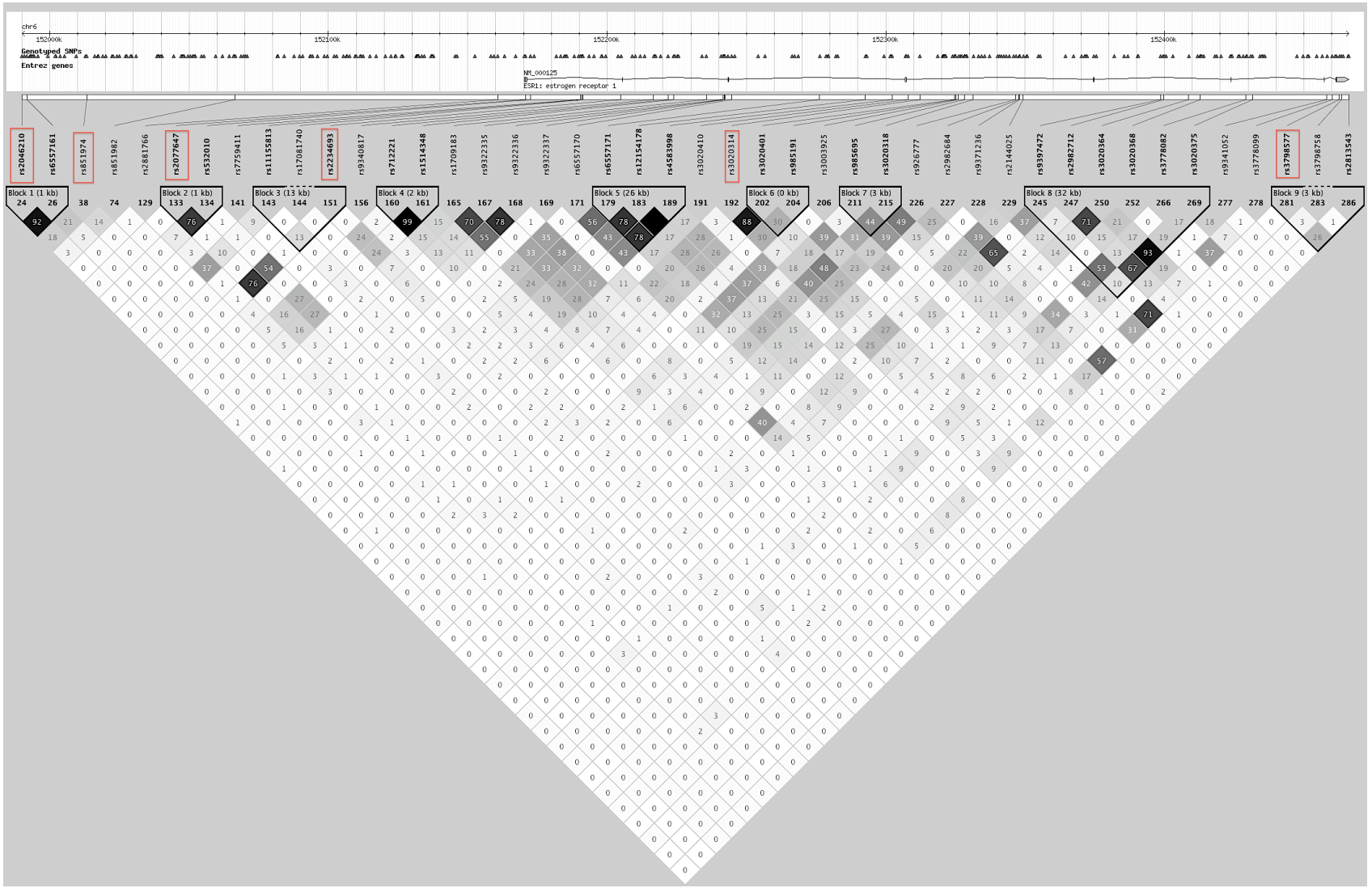


Figure 13: ESR1 linkage disequilibrium map for HapMap YRI

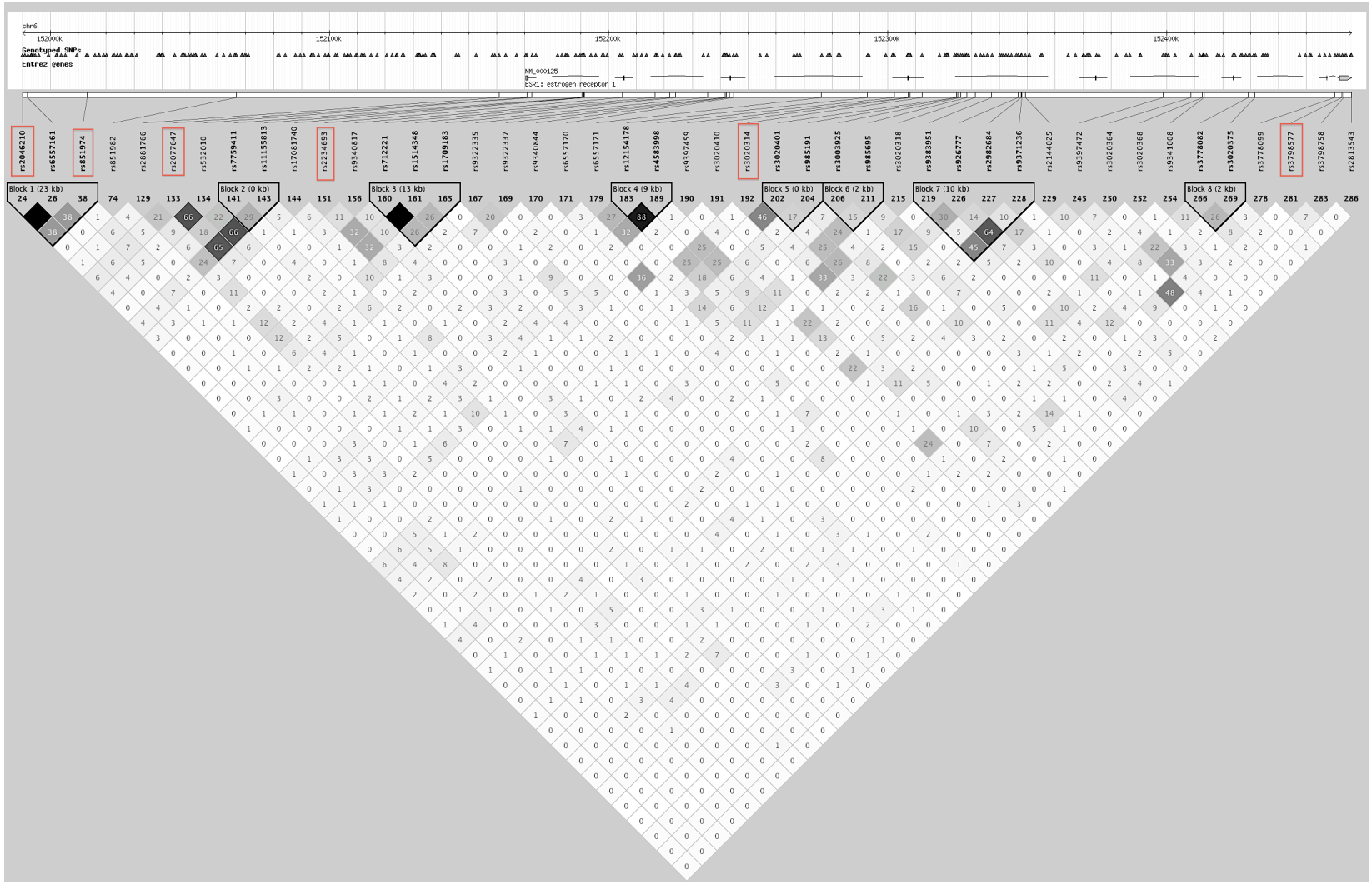


Figure 14: CDKN2A/CDKN2B linkage disequilibrium map for HapMap CEU

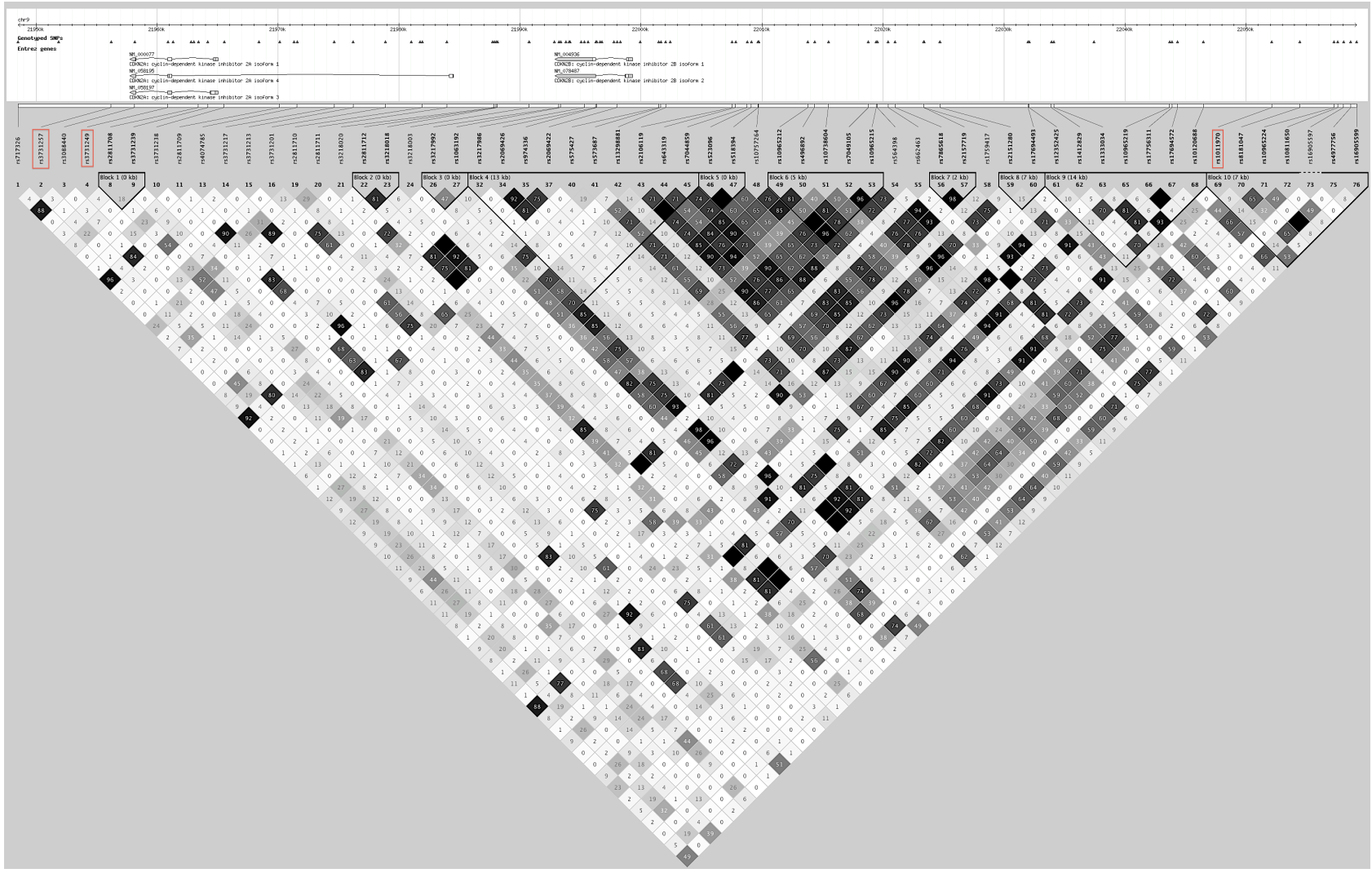


Figure 15: CDKN2A/CDKN2B linkage disequilibrium map for HapMap YRI

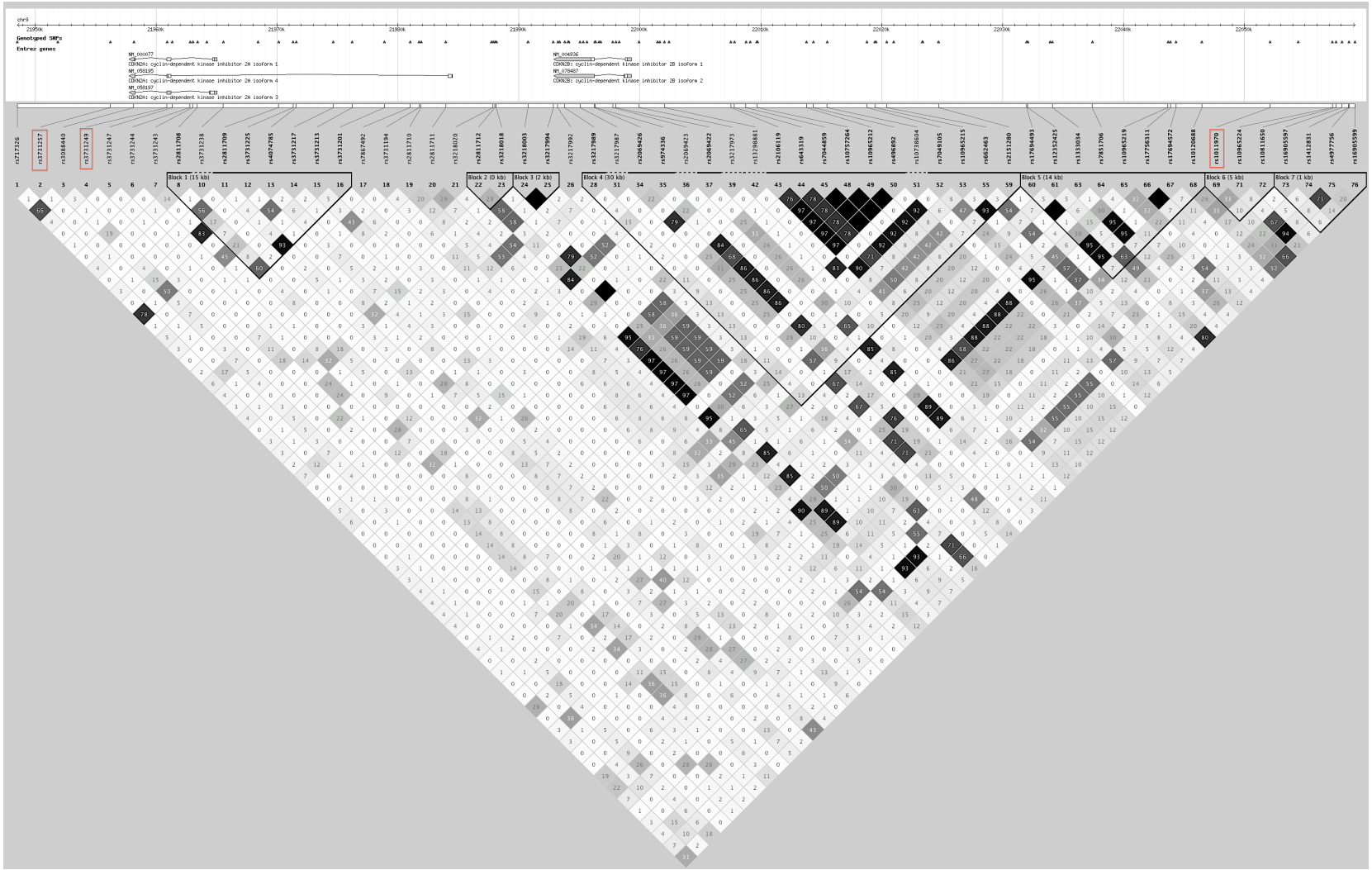


Figure 16: FGFR2 linkage disequilibrium map for HapMap CEU

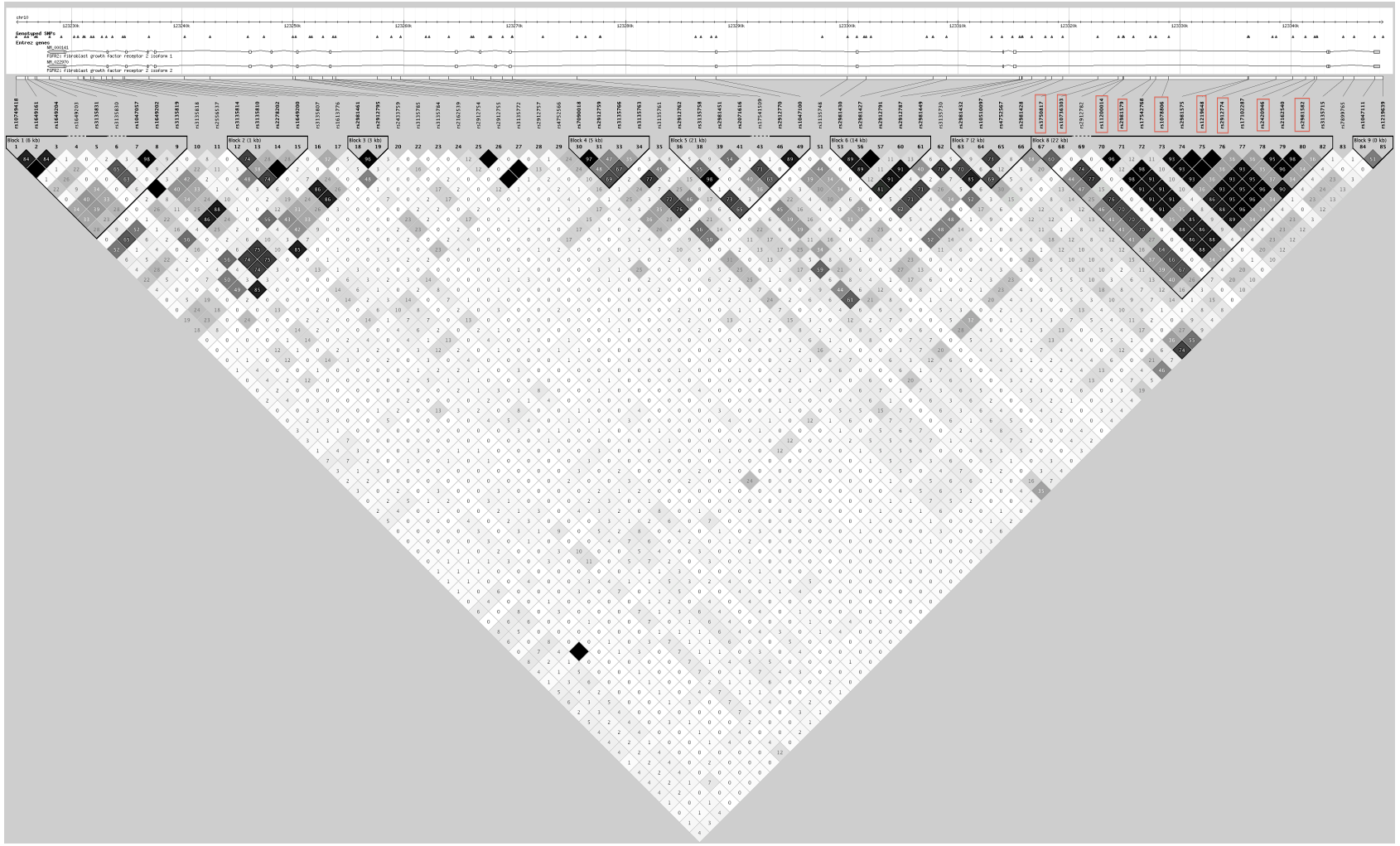


Figure 17: FGFR2 linkage disequilibrium map for HapMap YRI

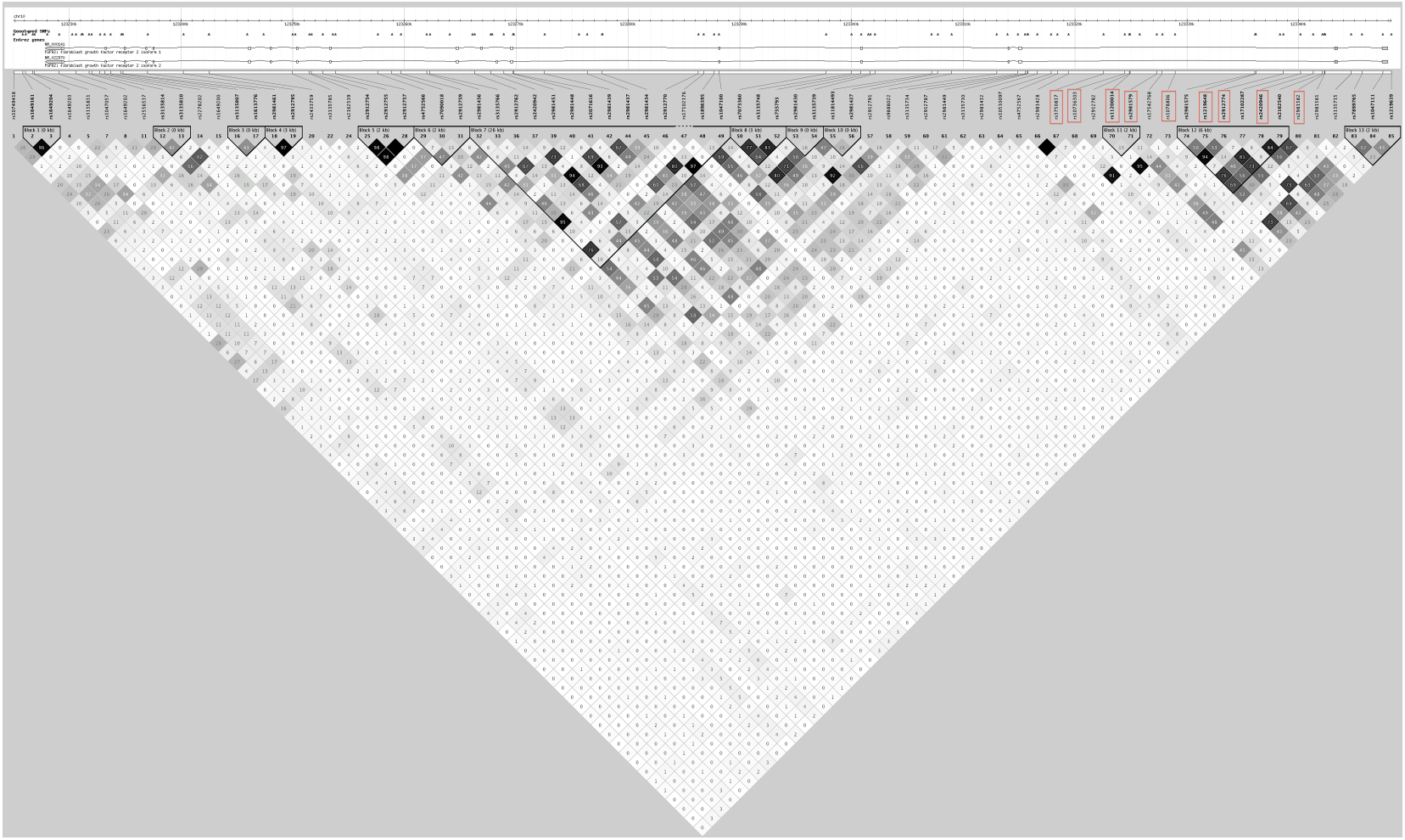


Figure 18: TNRC9/TOX3 linkage disequilibrium map for HapMap CEU

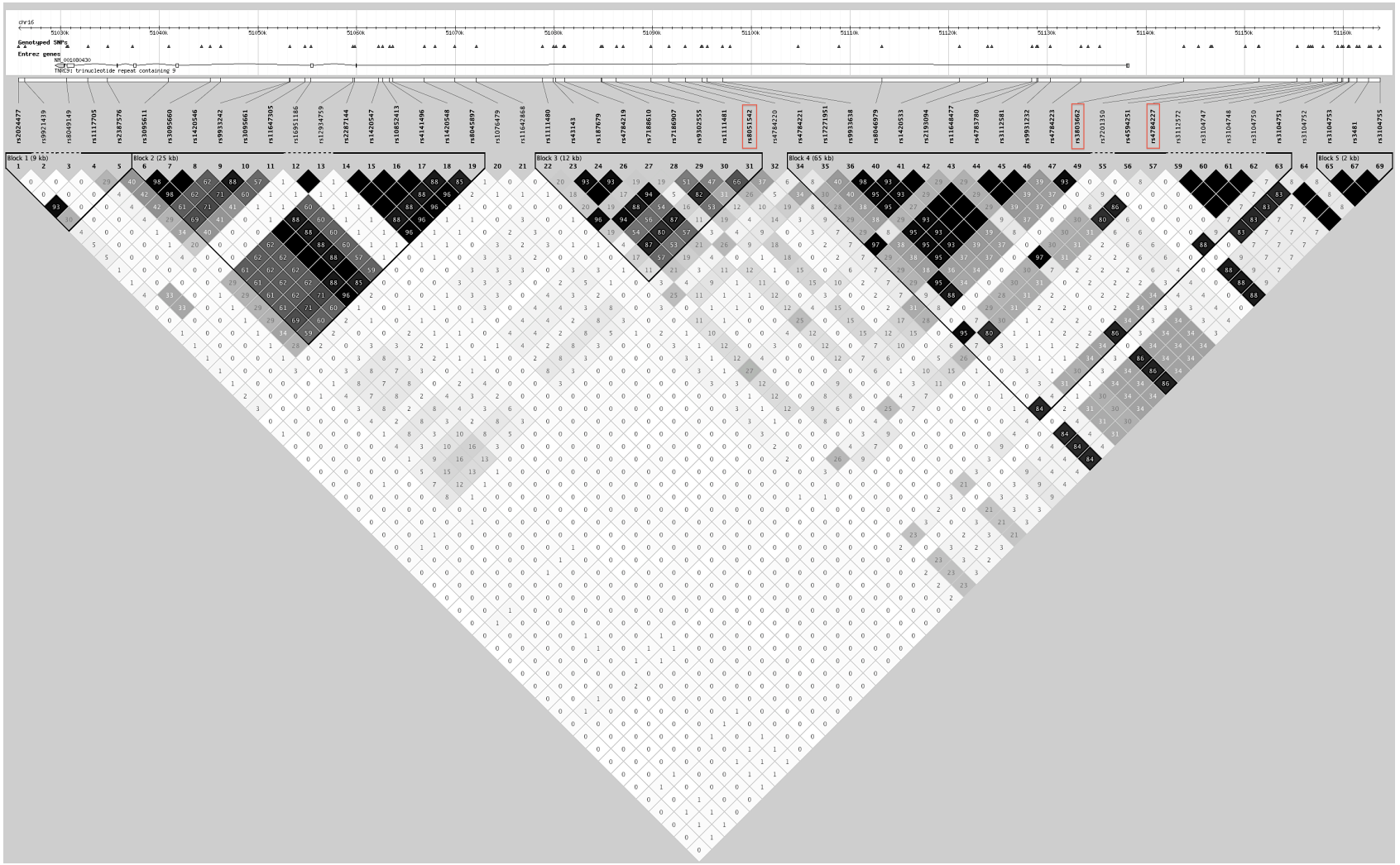


Figure 19: TNRC9/TOX3 linkage disequilibrium map for HapMap YRI

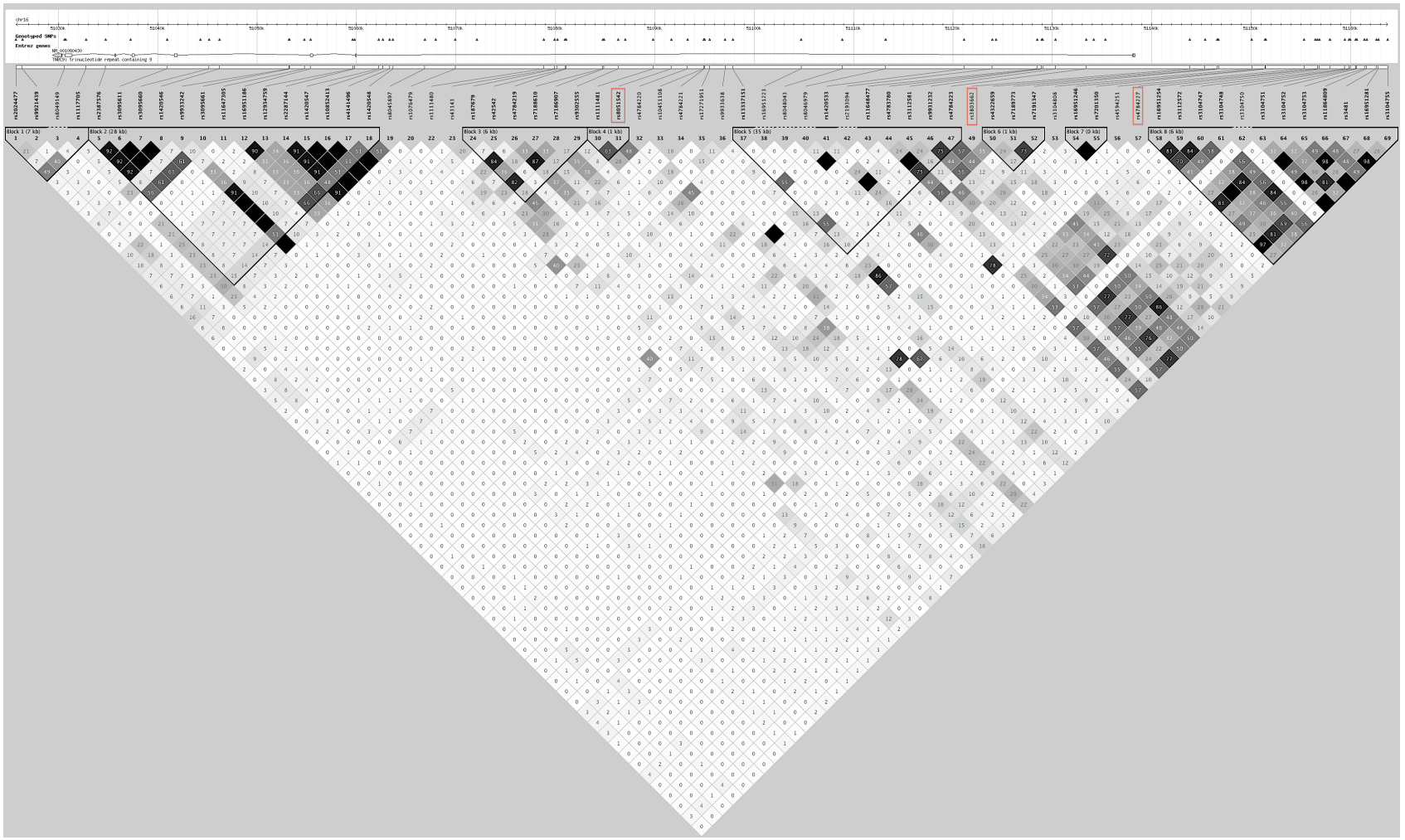


Table 1: Summary of risk factors by breast cancer subtype: Subtype vs. control

	Subtype			
Risk Factor	**Luminal A**	**Luminal B**	**HER2+/ER-**	**Basal-like or triple-negative**
Young vs. Old Age at Diagnosis	inverse	likely positive	likely positive	positive, 30-49 year olds may have highest risk
African American vs. non-Hispanic White	inverse in one study	null in one study	positive in one study	positive in one study
Hispanic vs. non-Hispanic White	not assessed	not assessed	not assessed	not assessed
Asian vs. non-Hispanic White	not assessed	not assessed	not assessed	not assessed
Positive Family History of Breast Cancer	positive	positive	positive	positive
Pre vs. Postmenopausal	positive in one study	positive in one study	null in one study	null in one study
Later Age at Menopause	positive	likely null	likely null	inconsistent
Early vs. Late Age at Menarche	positive	inconsistent	inconsistent	positive
Parity vs. Nulliparity	inverse	inverse	likely null	inconsistent
Lactation	inverse	inverse	inverse	inverse
Parity and Lactation	No interaction in one study	not assessed	not assessed	increasing parity positively associated in non-lactating women (1 study)
Later Age at First Full Term Pregnancy	any parity inversely associated with risk; later age riskier	any parity inversely associated with risk; later age riskier	inconsistent	inconsistent
Oral Contraceptive Use	possibly inverse	null	null	positively associated with long term use at younger age
Hormone Replacement Therapy Use	positive with current use	positive with current use	null	possible positive association with current use
Time Since Last Full Term Pregnancy	Inverse association with recent pregnancy	Inverse association with recent pregnancy	possibly positive	likely null

Spontaneous or Induced Abortion	inconsistent	inconsistent	inconsistent	inconsistent
High Body Mass Index: Overall	inconsistent	inverse in one study	null in one study	positive
High Body Mass Index: Premenopausal	likely inverse	possibly positive	likely null	positive
Body Mass Index: Postmenopausal	likely null	null	null	null
Waist to Hip Ratio	positive association with greater WHR in premenopausal and postmenopausal women	not assessed	not assessed	higher WHR increases risk of basal-like in pre and post menopausal women (1 study); null effect on triple-negative in postmenopausal women (1 study)
Increased Physical Activity	inverse	null in one study	inverse in one study	inverse
Increased Alcohol Consumption	positive	null	positive	inconsistent
Increased Smoking	Duration and intensity positively associated; former positively associated; current inconsistent	former smoking positively associated (1 study); current inversely associated (1 study)	former and current smoking positively associated (1 study)	null association with intensity and duration; inconsistent association with former and current smoking
History of Benign Breast Disease	positive	positive	null	inconsistent
High Breast Density	positive in one study	not assessed	not assessed	positive in one study

Table 2: Summary of risk factors by breast cancer subtype: Subtype vs. luminal A

Risk Factor	Luminal B	HER2+/ER-	Basal-like or triple-negative
Young vs. Old Age at Diagnosis	positive	positive	positive
African American vs. non-Hispanic White	null	positive	positive
Hispanic vs. non-Hispanic White	positive	positive	inconsistent
Asian vs. non-Hispanic White	positive	positive	inverse
Positive Family History of Breast Cancer	null	null	positive in pooled analysis; null in two other studies
Pre vs. Postmenopausal	inconsistent	likely inverse	null
Later Age at Menopause	not assessed	not assessed	not assessed
Early vs. Late Age at Menarche	null	inconsistent	positive
Parity vs. Nulliparity	null	null	possibly positive with a possible positive effect for increasing parity
Lactation	null	null	null
Parity and Lactation	parity positively associated in non-lactating women in one study	increasing parity positively associated in non-lactating women in one study	increasing parity positively associated in non-lactating women in one study
Later Age at First Full Term Pregnancy	parity positively associated in younger mothers, effect attenuates with age	parity positively associated in younger mothers, effect attenuates with age	parity positively associated in younger mothers, effect attenuates with age
Oral Contraceptive Use	inconsistent	null	null
Hormone Replacement Therapy Use	inconsistent	inconsistent	null
Time Since Last Full Term Pregnancy	null in one study	Shorter time associated with increased risk	positive association in one study

Spontaneous or Induced Abortion	not assessed	not assessed	not assessed
High Body Mass Index: Overall	null	likely null; inverse association in non-Africans (1 study)	positive
High Body Mass Index: Premenopausal	likely null	likely positive	positive
Body Mass Index: Postmenopausal	null	null	null
Waist to Hip Ratio	high WHR inversely associated in postmenopausal women, null in premenopausal (all from 1 study)	high WHR inversely associated in postmenopausal women, null in premenopausal (all from 1 study)	high WHR positively associated in pre and postmenopausal women (all from one study)
Other Body Size Measurements	not assessed	not assessed	not assessed
Increased Physical Activity	positive in one study	null	positive in one study
Increased Alcohol Consumption	inconsistent	likely null	likely null
Increased Smoking	null association with intensity and duration; positive association with former (1 study); null association with current (1 study)	null association with intensity and duration; positive association with current (1 study); null association with former (1 study)	inconsistent for duration and intensity; null for current smoking and former smoking (1 study)
History of Benign Breast Disease	not assessed	not assessed	not assessed
High Breast Density	not assessed	not assessed	null in one study

Table 3: Summary of candidate genes by breast cancer subtype

Chromosome	Gene	Summary of subtype findings
2	**CASP8** LD block 2	One SNP assessed: associated with ER+/PR+ and ER-/PR- disease
	CASP8 LD block 3	rs1045485 inversely associated with all subtypes; rs17468277 associated with triple-negative disease (pooled analysis)
	CTLA4	one study found an association between a SNP in the promoter region and ER+ and PR+ status
	LKB1	not assessed
5	**TERT**	Possible association with ER- and triple negative disease (pooled analysis)
8	**NBN**	not assessed
10	**PTEN**	not assessed
11	**ATM**	Several SNPs associated with PR+ status; Mostly null associations with ER status, though three SNPs associated with ER- disease
14	**XRCC3**	No association with ER or PR status
15	**CYP19A1** LD block 1	conflicting results regarding association between ER status and rs10046; possible association with HER2- status
	CYP19A1 LD block 2	No association with ER, PR or HER2 status
16	**PALB2** LD block 1	Possible association with ER+ disease (1 study)
	PALB2 1592delT	Possible association with ER and PR- disease (1 study)
17	**BRIP1**	Mostly null associations with ER and PR status
	TP53	rs1042522 possibly associated with ER+ disease; no other associations observed
22	**CHEK2** any mutation	possible positive association with ER+/Luminal A and B disease
	CHEK2 1100delC (exon 10) or del5395 (exons 9 and 10)	strong evidence of an association with ER+ disease; may also be associated with ER- disease
	CHEK2 I157T and IVS2+IG>A	possible positive association with ER+/Luminal A and B disease

Table 4: Breast cancer genome-wide association studies (GWAS)

Study	Study	Initial Sample Size	Replication Sample	Region	Reported Gene(s)	SNPs	p-Value	OR	95% CI (text)
Easton, June 2007 [194]	Genome-wide association study identifies novel breast cancer susceptibility loci.	390 cases, 364 controls	26646 cases, 24889 controls	16q12.1	TNRC9	rs3803662	1.00E-36	1.2	[1.16-1.24]
				5p12	Intergenic	rs981782	9.00E-06	1.04	[1.01-1.08]
				11p15.5	LSP1	rs3817198	3.00E-09	1.07	[1.04-1.11]
				5q11.2	MAP3K1	rs889312	7.00E-20	1.13	[1.10-1.16]
				10q26.13	FGFR2	rs2981582	2.00E-76	1.26	[1.23-1.30]
				8q24.21	Intergenic	rs13281615	5.00E-12	1.08	[1.05-1.11]
Hunter, July 2007 [201]	A genome-wide association study identifies alleles in FGFR2 associated with risk of sporadic postmenopausal breast cancer.	1145 cases, 1142 controls	1176 cases, 2072 controls	10q26.13	FGFR2	rs1219648	1.00E-10	1.2	[1.07-1.42]
Stacey, July 2007 [187]	Common variants on chromosomes 2q35 and 16q12 confer susceptibility to estrogen receptor-positive breast cancer.	1599 cases, 11546 controls	2934 cases, 5967 controls	2q35	Intergenic	rs13387042	1.00E-13	1.2	[1.14-1.26]
				16q12.1	TNRC9	rs3803662	6.00E-19	1.28	[1.21-1.35]
Murabito, July 2007 [188]	A genome-wide association study of breast and prostate cancer in the NHLBI's Framingham Heart Study.	723 cases	Not replicated	5q34	Intergenic	rs6556756	5.00E-07	NR	NR
				15q21.1	FBN1	rs1876206	6.00E-06	NR	NR
				13q32.1	ABCC4	rs1926657	2.00E-06	NR	NR
				12q21.1	Intergenic	rs1154865	7.00E-07	NR	NR
				18q21.2	Intergenic	rs1978503	1.00E-06	NR	NR
				7q11.22	Intergenic	rs10263639	3.00E-06	NR	NR
				2p16.1	Intergenic	rs10490113	5.00E-06	NR	NR
				21q21.3	GRIK1	rs458685	6.00E-06	NR	NR
				17q21.33	COL1A1	rs2075555	8.00E-08	NR	NR

Gold, Mar 2008 [196]	Genome-wide association study provides evidence for a breast cancer risk locus at 6q22.33	249 cases, 299 controls (AJ, non-BRCA1/2)	1193 cases, 1166 controls (AJ, non-BRCA1/2)	6q22.33	ECHDC1	rs2180341	3.00E-08	1.41	[1.25-1.59]
Zheng, Mar 2009 [197]	Genome-wide association study identifies a new breast cancer susceptibility locus at 6q25.1	1505 CA cases, 1522 CA controls	1554 CA cases, 1576 CA controls	6q25.1	ESR1	rs2046210	2.00E-15	1.29	[1.21-1.37]
Kibriya, April 2009 [204]	A pilot genome-wide association study of early-onset breast cancer.	30 cases, 30 controls	Not replicated	16q23.1	GLG1	rs10871290	4.00E-07	NR	NR
Thomas, May 2009 [186]	A multistage genome-wide association study in breast cancer identifies two new risk alleles at 1p11.2 and 14q24.1 (RAD51L1).	1145 cases, 1142 controls	8625 cases, 9657 controls	2q35	Intergenic	rs13387042	2.00E-08	1.25	[1.15-1.37]
				10q26.13	FGFR2	rs2981579	2.00E-10	1.17	[1.07-1.27]
				1p11.2	Intergenic	rs11249433	7.00E-10	1.16	[1.09-1.24]
				16q12.1	TOX3	rs3803662	1.00E-09	1.16	[1.07-1.27]
				5q11.2	MAP3K1	rs16886165	5.00E-07	1.23	[1.12-1.35]
				14q24.1	RAD51L1	rs999737	2.00E-07	1.06	[1.01-1.14]
Turnbull, June 2010 [189]	Genome-wide association study identifies five new breast cancer susceptibility loci.	3659 UK cases, 4897 UK controls	12576 EA cases, 12223 EA controls	8q24.21	Intergenic	rs1562430	6.00E-07	1.17	[1.10-1.25]
				2q35	Intergenic	rs13387042	2.00E-10	1.21	[1.14-1.29]
				10q21.2	ZNF365	rs10995190	5.00E-15	1.16	[1.10-1.22]
				9p21.3	CDKN2A	rs1011970	3.00E-08	1.09	[1.04-1.14]
				11q13.3	MYEOV	rs614367	3.00E-15	1.15	[1.10-1.20]
				3p24.1	SLC4A7	rs4973768	6.00E-07	1.16	[1.10-1.24]
				16q12.1	TOX3	rs3803662	3.00E-15	1.3	[1.22-1.39]
				10q26.13	FGFR2	rs2981579	4.00E-31	1.43	[1.35-1.53]
				5q11.2	MAP3K1	rs889312	5.00E-09	1.22	[1.14-1.30]
				10q22.3	ZMIZ1	rs704010	4.00E-09	1.07	[1.03-1.11]
				10p15.1	ANKRD16	rs2380205	5.00E-07	1.06	[1.02-1.10]

				6q25.1 11p15.5	ESR1 LSP1	rs3757318 rs909116	3.00E-06 7.00E-07	1.3 1.17	[1.17-1.46] [1.10-1.24]
Long, June 2010 [205]	Identification of a functional genetic variant at 16q12.1 for breast cancer risk: results from the Asia Breast Cancer Consortium.	2073 CA cases, 2084 CA controls	10598 A cases, 8254 A controls, 2797 EA cases, 2662 EA controls	16q12.1	TOX3	rs4784227	1.00E-28	1.24	[1.20-1.29]
Antoniou, Oct 2010 [206]	A locus on 19p13 modifies risk of breast cancer in BRCA1 mutation carriers and is associated with hormone receptor-negative breast cancer in the general population.	1193 EA cases, 1190 EA controls	3012 EA cases, 2974 EA controls	19p13.11	ABHD8	rs8170	2.00E-09	1.26	[1.17-1.35]
Gaudet, Oct 2010 [202]	Common genetic variants and modification of penetrance of BRCA2-associated breast cancer.	899 BRCA2+ EA cases, 804 BRCA2+ EA controls	1264 cases, 1222 controls	10q26.13	FGFR2	rs2981575	1.00E-08	1.28	[1.18-1.39]
Li, Nov 2010 [207]	A genome-wide association scan on estrogen receptor-negative breast cancer.	617 EA ER- cases, 4,583 EA controls	1001 EA ER- cases, 7604 EA controls	No SNPS met reporting requirements for statistical significance					
Fletcher, Mar 2011 [190]	Novel breast cancer susceptibility locus at 9q31.2: results of a genome-wide association study.	1694 UK cases, 2365 UK, 1145 EA cases, 1142 EA controls	7317 UK cases, 8124 UK controls	6q25.1 5p12 2q35 3p24.1 8q24.21	ESR1 Intergenic Intergenic SLC4A7 Intergenic	rs3734805 rs4415084 rs13387042 rs4973768 rs1562430	1.00E-07 8.00E-11 2.00E-10 2.00E-08 3.00E-11	1.19 1.17 1.16 1.14 1.16	[1.11-1.27] [1.11-1.22] [1.11-1.22] [1.09-1.19] [1.11-1.22]

				16q12.2	TOX3	rs3112612	4.00E-10	1.15	[1.10-1.21]
				10q26.13	FGFR2	rs1219648	1.00E-30	1.31	[1.25-1.37]
				10q26.13	FGFR2	rs10510102	2.00E-06	1.12	[1.07-1.17]
				9q31.2	RAD23B	rs865686	2.00E-10	1.12	[1.09-1.18]
Li, April 2011 [191]	A combined analysis of genome-wide association studies in breast cancer.	2702 EA women, 5726 EA controls	Up to 7386 EA cases, 7576 EA controls	10q26.13	FGFR2	rs1219648	2.00E-13	1.32	[1.22-1.42]
				5p12	MRPS30	rs7716600	7.00E-07	1.24	[1.14-1.34]
				2q35	Intergenic	rs13387042	9.00E-06	1.18	[1.10-1.27]
				16q12.1	TOX3	rs3803662	4.00E-07	1.22	[1.13-1.32]
Cai, Sept 2011 [200]	Genome-wide association study identifies breast cancer risk variant at 10q21.2: results from the Asia Breast Cancer Consortium.	2,062 EAA, 2,066 EAA controls	15091 EAA cases, 14877 EAA controls	10q21.2	ZNF365	rs10822013	6.00E-09	1.12	[1.06-1.18]
				7q32.3	NR	rs2048672	6.00E-06	1.11	[1.05-1.17]
Sehrawat, Oct 2011 [195]	Potential novel candidate polymorphisms identified in genome-wide association study for breast cancer susceptibility.	302 EA cases, 321 EA controls	1153 cases, 1,215 controls	19q13.41	ZNF577	rs10411161	7.00E-06	1.42	[1.22-1.65]
				5p15.2	ROPN1L	rs1092913	2.00E-06	1.45	[1.24-1.69]
Haiman, Oct 2011 [166]	A common variant at the TERT-CLPTM1L locus is associated with estrogen receptor-negative breast cancer.	1004 AA cases, 2745 AA controls, 1718 EA cases, 3670 EA controls	2222 EA cases, 16363 EA controls	5p15.33	TERT	rs10069690	1.00E-10	1.18	[1.13-1.25]
Long, Feb 2012 [199]	Genome-wide association study in east Asians identifies novel susceptibility loci for breast cancer	2918 CA cases, 2324 CA controls	16173 A cases, 18282 A controls	6q25.1	TAB2	rs9485372	4.00E-12	1.11	[1.09-1.15]
				11q24.3	BARX2	rs7107217	5.00E-7	1.08	[1.05-1.11]
				6q25.1	ESR1	rs9383951	2.00E-6	1.14	[1.08-1.19]
Kim,	A genome-wide	2273 KA	4049 KA	2q34	ERBB4	rs13393577	9.00E-14	1.53	[1.37-1.70]

Mar 2012 [192]	association study identifies a breast cancer risk variant in ERBB4 at 2q34: results from the Seoul Breast Cancer Study	cases, 2052 KA controls	cases, 3845 KA controls						
Chen, Jan 2013 [203]	A genome-wide association study of breast cancer in women of African ancestry	3016 AA cases, 2745 AA controls	3533 AA cases, 11046 AA controls	14q31.3	GALC	rs4322600	4.00E-6	1.18	[1.10-1.27]
Elgazzar, Dec 2012 [193]	A genome-wide association study identifies a genetic variant in the SIAH2 locus associated with hormonal receptor-positive breast cancer in Japanese	1086 JA cases, 1816 JA controls	1653 JA cases, 2797 JA controls	10q26.13	FGFR2	rs3750817	8.00E-8	1.22	NR
				3q25.1	SIAH2	rs6788895	9.00E-8	1.22	[1.13-1.31]
Siddiq, Dec 2012 [198]	A meta-analysis of genome-wide association studies of breast cancer identifies two novel susceptibility loci at 6q14 and 20q11	3666 EA cases, 28864 EA controls, 1004 AA cases, 2744 AA controls	562 EA cases, 6410 EA controls, 84 JA cases, 830 JA controls, 300 H, 1164 H controls	6q25.1	ESR1	rs9383938	2.00E-10	1.28	NR
				20q11.22	RALY	rs2284378	1.00E-8	1.16	[1.10-1.22]
				19p13.11	ANKLE1	rs8100241	4.00E-8	1.14	NR
				6q14.1	FAM46A	rs17530068	3.00E-7	1.16	[1.10-1.23]

Abbreviations: NR=Not reported, AA= African Ancestry, A=Asian, AJ= Ashkenazi Jewish, CA= Chinese ancestry, EA= European Ancestry, EAA= East Asian Ancestry, H=Hispanic, JA= Japanese Ancestry, KA= Korean ancestry, UK= United Kingdom

Table 5: Summary of effect estimates for GWAS-identified and other selected breast cancer SNPs among women of European (EA) and African American (AA) ancestry

Chro m-osome	Gene or gene region	SNP	Minor Allele	MAF: CEU	MAF: YRI	MAF: ASW	Number of studies (EA/AA)	Range of estimates for log-additive genetic model		Notes on subtype findings
								EA	AA	
1	1p12	rs11249433	G	0.43	0.07	0.09	7 / 5	1.00 - 1.18	0.93 - 1.08	strongest association in Luminal A and B cancers
2	2p	rs4666451	A	0.45	0.23	--	5 / 0	0.93 - 1.01	NE	evidence of association with ER+ and ER- disease
2	2q35	rs13387042	G	0.44	0.22	0.25	15 / 7	0.83- 0.91	0.83- 1.07	some association with all subtypes
3	SLC4A7	rs4973768	T	0.44	0.29	0.32	10 / 5	0.89- 1.16	0.92- 1.06	strongest association in Luminal A cancers; no association in triple-negative cancers
4	4p	rs12505080	C	0.29	0.10	0.17	1 / 0	0.99	NE	NE
4	TLR1	rs7696175	T	0.47	0.00	0.09	1 / 0	0.99	NE	Associated with ER+, ER- and HER2- disease in Chinese
5	MRPS30	rs4415084	T	0.38	0.66	0.68	8 / 6	1.01- 1.17	0.89- 1.13	stronger associations with ER+ or PR+ disease; mostly null associations with ER- or PR- disease
5	MRPS30	rs10941679	G	0.24	0.17	--	7 / 6	1.11- 1.19	0.94 - 1.14	strongest association in Luminal A and B cancers; null association in triple-negative cancers
5	5p12	rs981782	C	0.39	0.00	--	5 / 0	0.90- 0.97	not polymorp hic	strongest associations with ER- and HER2- disease
5	5q	rs30099	A	0.08	0.18	0.14	3 / 0	1.02- 1.05	NR	possible associations with ER+, ER- and HER2+

										disease
5	MAP3K1	rs889312	C	0.30	0.31	--	16 / 7	1.01- 1.26	0.93- 1.33	evidence of association with all subtypes
6	ECHDC1	rs2180341	G	0.26	0.36	0.25	4 / 5	0.96 - 1.41	0.98 - 1.09	predominantly null associations with ER and PR status
6	ESR1	rs2046210	A	0.29	0.68	0.60	5 / 8	0.94- 1.15	0.94 - 1.11	some association with all subtypes
6	ESR1	rs851974	G	0.47	0.14	0.24	1 / 1	0.92	1.13	NE
6	ESR1	rs2077647	C	0.43	0.48	0.59	4 / 0	0.97*	NE	NE
6	ESR1	rs2234693	C	0.41	0.53	0.54	10 / 1	0.97*	1.43#	NE
6	ESR1	rs1801132	G	0.17	0.07	0.08	5 / 0	0.95*	NE	NE
6	ESR1	rs3020314	C	0.26	0.78	0.64	2 / 0	1.12*	NE	NE
6	ESR1	rs3798577	C	0.44	0.45	0.47	4 / 0	0.98*	NE	evidence of association with ER- and PR- disease
7	RELN	rs17157903	T	0.12	0.11	0.08	1 / 0	1.11	NE	NE
8	8q24	rs13281615	G	0.45	0.42	--	15 / 6	1.07- 1.31	0.94- 1.16	strongest association in Luminal A and B; null association in triple-negative
8	8q24	rs1562430	C	0.35	0.50	0.45	4 / 0	0.85- 0.93	NE	Associated with ER+, ER-, PR+, and PR- disease in Chinese (one study)
9	CDKN2A/B	rs3731257	A	0.26	0.06	0.12	1 / 0	0.95#	NE	NE
9	CDKN2A/B	rs3731249	T	0.02	0.00	0.01	1 / 0	1.5&	NE	NE
9	CDKN2A/B	rs1011970	T	0.17	0.37	0.28	2 / 4	1.08- 1.09	0.9- 1.07	some evidence of association with ER+ and triple-negative
10	ANKRD16	rs2380205	T	0.48	0.64	0.60	2 / 4	0.94- 0.98	0.98- 1.03	possible inverse association with ER+ disease, null association

										with triple negative
10	ZNF365	rs10995190	A	0.13	0.18	0.20	2 / 2	0.85- 0.86	1.03- 1.12	inversely associated with ER+ disease
10	ZMIZ1	rs704010	T	0.43	0.02	0.07	1 / 2	1.07	0.90- 0.99	positively associated with all ER/PR subtypes, but not triple-negative disease
10	FGFR2	rs3750817	T	0.36	0.01	0.04	2 / 0	0.78- 0.86	NE	GWAS hit in study of hormone receptor positive disease
10	FGFR2	rs10736303	A	0.47	0.07	0.11	1 / 1	1.25	1.21[#]	evidence of associations with ER+ and PR+ disease
10	FGFR2	rs11200014	A	0.47	0.22	0.31	5 / 0	1.24- 1.31	NE	Associated with ER+, PR+, HER2+ and HER2- in Chinese study
10	FGFR2	rs2981579	A	0.47	0.64	0.63	7 / 3	1.20- 1.43	0.99- 1.08	associated with ER+/PR+ in AA,and most subtypes in Chinese study
10	FGFR2	rs1078806	G	0.46	0.21	0.31	1 / 1	1.26	0.95	NE
10	FGFR2	rs2981578	C	0.49	0.94	--	1 / 4	1.26	1.17- 1.24	evidence of association with ER+, ER- and PR+ disease
10	FGFR2	rs1219648	G	0.47	0.44	0.44	13 / 5	1.23- 1.33	0.84- 1.21	evidence of association with ER+, PR+ and HER2- disease
10	FGFR2	rs2912774	T	0.47	0.60	0.53	2 / 2	1.26 - 1.33	1.15- 1.19	evidence of association with ER+ and ER- disease
10	FGFR2	rs2936870	T	0.45	0.66	--	1 / 0	1.26	NR	NE
10	FGFR2	rs2420946	T	0.47	0.58	0.48	7 / 1	1.21- 1.34	1.1	NE
10	FGFR2	rs2981582	A	0.46	0.49	0.47	17 / 8	1.15- 1.59	0.80- 1.22	strongest association in Luminal A and B; null association in HER2+/ER- and triple-negative

10	FGFR2	rs3135718	G	0.43	0.37	--	1 / 0	1.07	NE	NE
10	10q	rs10510126	T	0.16	0.11	0.11	1 / 0	0.83	NE	NE
11	LSP1	rs3817198	C	0.33	0.09	0.07	16 / 6	0.96- 1.51	0.85- 1.21	Generally imprecise and inconsistent findings, possible association with unclassified disease
11	LSP1	rs909116	C	0.43	0.22	0.31	1 / 0	1.17	NE	NE
11	H19	rs2107425	T	0.30	0.58	0.51	1 / 0	0.96	NE	NE
11	MYEOV	rs614367	T	0.19	0.13	0.08	2 / 4	1.15- 1.21	0.96- 1.13	possible association with ER+ and PR+ disease
16	TNRC9	rs8051542	T	0.45	0.23	0.33	5 / 1	0.96- 1.15	0.98	evidence of association with ER+ disease
16	TNRC9	rs12443621	A	0.47	0.49	--	5 / 1	0.86 - 1.11	0.99	no evidence of association
16	TNRC9	rs3803662	A	0.25	0.55	0.54	21 / 9	0.87- 1.33	0.75- 1.04	evidence of association with all subtypes
16	TNRC9	rs4784227	T	0.23	0.05	0.08	2 / 3	1.17- 1.29	1.09- 1.23	positively associated with all ER and PR subtypes in one Korean study
16	TNRC9	rs3104746	A	0.05	0.24	--	0 / 2	NE	1.17- 1.23	evidence of association with ER+ and ER- disease
16	TNRC9	rs3112562	G	0.26	0.50	--	0 / 1	NR	1.17	NE
17	COX11	rs7222197	A	0.32	0.33	0.35	1 / 0	0.89	NE	NE
17	COX11	rs6504950	A	0.30	0.31	--	6 / 5	0.80- 0.95	0.79- 1.06	inverse association in Luminal A and B; null association in HER2+/ER- and triple-negative

NE= not evaluated; *= from Zhang et al. [20]meta-analysis; #=homozygous rare versus homozygous wildtype; &= dominant model

3. Methods

3.1 Study population

3.1.1 Case and control ascertainment

The Carolina Breast Cancer Study (CBCS) is a population-based, case-control study of invasive and *in situ* breast cancer conducted in North Carolina between 1993 and 2001. Cases from 24 central and eastern North Carolina counties (see Figure 20) were identified using the North Carolina Central Cancer Registry's Rapid Case Ascertainment program. This program offers registrars financial incentives for accelerated registration of new cancer diagnoses and thus provides investigators with more immediate access to pathology reports and contact information for potential study participants [271, 272]. Women were eligible for CBCS if they lived in one of the 24 counties, were diagnosed with invasive, primary breast cancer between 1993 and 2001, and were between the ages of 20 and 74 at the time of their diagnosis. Women diagnosed with *in situ* breast cancer between 1996 and 2001 were also eligible if they had ductal carcinoma *in situ* with microinvasion to a depth of 2 mm or lobular carcinoma *in situ*.

To ensure approximately equal representation of African Americans and non-African Americans, as well as pre and postmenopausal women, breast cancer cases were randomly sampled at disproportionate rates based on their age and race category. For Phase I of the study (1993-1996) these sampling proportions were as follows: 100% of African Americans under age 50, 75% of African Americans aged 50 or older, 67% of non-African Americans under age 50, and 20% of non-African Americans aged 50 or older.

For Phase II of the study (1996-2001), all African American invasive and *in situ* cases were recruited, as were all non-African American *in situ* cases. 50% of younger, non-African American, invasive breast cancer cases were recruited, as were 20% of older, non-African American cases. Although urban or rural status did not affect sampling probability, the 24 selected counties included geographically and socioeconomically diverse regions of the state.

Controls aged 20-64 and 65-74 were selected from North Carolina Department of Motor Vehicle and Health Care Financing Administration records, respectively. Controls were approximately frequency-matched to cases on race and 5-year age groups using another set of predetermined sampling fractions. These ranged from 0.003% of non-African Americans aged 20-24 years to 2.1% of African Americans aged 70-74 years. Women who did not live in the study area or who had previous history of breast cancer were ineligible.

3.1.2 Data collection

Once a newly diagnosed case was selected as a potential study participant, CBCS personnel contacted the woman's treating physician, who had the right to refuse any further contact with the patient. All approved cases and randomly selected controls received an introductory letter followed by a telephone call, if a phone number was available. If a number was not available, as occurred with 5% of potential cases and 40% of potential controls, a study nurse visited the potential participant's home [272]. Women who met all eligibility criteria and verbally consented to study participation were scheduled for a home interview with one of the study's registered nurses.

After obtaining written informed consent, the study nurse administered a questionnaire about demographic information and known or suspected breast cancer risk factors. The nurse also took body measurements, including height, weight and waist and hip

girths, and drew a 30 ml blood sample. For cases, the nurse asked for permission to access medical records and paraffin-embedded tumor tissue blocks. Some participants opted to have their blood drawn at a physician's office and sent into the study. Whenever possible, a University of North Carolina (UNC) pathologist used provided tumor tissue blocks to confirm the breast cancer diagnoses.

The overall response rate was 77% for cases and 57% for controls. Among cases, older African Americans had the lowest response rate, with 71% of initially contacted women successfully enrolled. Younger African American controls were the least likely to participate, with only 48% responding. The highest response rates for cases and controls were for younger non-African Americans (82%) and older non-African Americans (66%), respectively. Total enrollment included 1808 invasive cases (787 African American, 1016 non-African American), and 1564 corresponding controls (718 African American, 846 non-African American), as well as 503 *in situ* cases (107 African American, 401 non-African American) and 458 corresponding controls (70 African American, 388 non-African American).

Of those enrolled, 88% of cases (2039 of 2311) and 90% of controls (1817 of 2022) provided sufficient blood samples for genotype analysis (see Figure 21). Non- African Americans were more likely to provide blood samples than African Americans, but blood donation status did not differ by stage of disease or other breast cancer risk factors [273].

More than 98% of non-African Americans self-identified as white, with the remaining 2% identifying as multi-racial, Hispanic, Native American, Asian-American, or other race/ethnicity. Aside from the expected differences in age and race distributions, the only observed discrepancy between Phase I CBCS cases and other North Carolina Central Cancer

Registry cases was that African Americans aged 40-59 with later stage disease were underrepresented in CBCS [274].

3.1.3 SNP selection

I selected candidate single nucleotide polymorphisms (SNPs) for the proposed analysis from a list of SNPs already successfully genotyped in the CBCS population. As the purpose of my investigation is to evaluate the effect of previously established breast cancer risk variants within the CBCS population, I retained any SNPs identified as hits in genome-wide association studies (GWAS) or that had strong and consistent evidence of an association in a candidate gene meta-analysis [20]. I also included any SNPs in the same gene as a selected risk variant.

As the most recent genotype analysis of CBCS participants was completed in mid-2010, only some of the GWAS hits were included. In total, I selected 31 SNPs with genome-wide significant or nearly genome-wide significant p-values in studies published before September 2010 [186, 187, 189, 194, 196, 197, 201, 205, 208, 209], 31 SNPs from the same genes as previous GWAS hits, and 21 SNPs from 3 critical candidate genes (*ATM, CASP8*, and *TP53*). I considered 41 of these SNPs to be 'GWAS-identified', which meant that they were either GWAS hits (n=22), near GWAS hits (n=9), or on the same gene as a GWAS hit and assessed in a GWAS study (n=10). Of the remaining 21 SNPs from GWAS-identified genes, nine had never before been evaluated in whites or African Americans with breast cancer. The full list of included SNPs can be seen in Table 6. This table includes the SNPs discussed in sections 2.7.3, 5 *ATM*, *CASP8*, *TP53* or *ESR1* SNPs identified by Zhang et al. as having strong evidence of an association with breast cancer [20], and other *ATM*, *CASP8*, and *TP53* SNPs. Given that the candidate gene hits and the GWAS-identified SNPs had a

concrete reference OR (from Zhang et al. or the original GWAS, respectively), these SNPs were eligible for replication. I considered a SNP to have replicated successfully if the OR was in the same direction as the reference OR and the 95% CI or 95% posterior interval (PI) excluded the null.

3.1.4 Ancestry informative markers (AIMs)

A panel of 144 ancestry informative markers (AIMs) was also genotyped in all CBCS participants with available blood samples. To be useful as an AIM, a genetic variant (usually a SNP) must have a widely discrepant MAF in the relevant parent populations and be highly informative of ancestry [275, 276]. In CBCS, we selected AIMs that could distinguish between European and African ancestry. More specifically, we selected SNPs with MAF differences of at least 60% between the International HapMap CEU and YRI populations that had high values for Fisher's information criterion. This was completed for the following admixture proportions, all of which are likely scenarios in CBCS: 90% European/ 10% African, 90% African/ 10% European, and 50% European/ 50% African [210]. Fisher's information criterion is the inverse of the variance of the maximum likelihood estimate of the ancestral contribution, and can therefore be interpreted as a measure of precision for the ancestral proportion estimate [277]. A more detailed account of the CBCS AIMs selection process and a table of selected markers can be found in Barnholtz-Sloan et al [210]. Of note, the ASW HapMap population (African Americans living in the Southwest USA) was not available when these AIMs were selected.

3.1.5 Genotype analysis

The SNPs included in this analysis were genotyped at two different times using two different genotyping platforms. Most SNPs were evaluated with a Custom GoldenGate

Genotyping assay from Illumina (Illumina, Inc., San Diego, CA). A few SNPs that failed Illumina genotyping analysis were reassessed using a Taqman panel. The Taqman analysis also included a number of GWAS hits identified in the time since the Illumina panel was performed.

These analyses are used to determine whether an individual has two copies of the common allele, two copies of the variant allele, or one copy of each. These three possible genotypes are also described as homozygous for the major allele, homozygous for the minor allele, or heterozygous. In some cases, the allele that is more common in whites was the rare allele in African Americans.

3.1.5.1 Illumina assays

Initially, 1762 SNPs were submitted to Illumina for review and preliminary quality control assessment. Illumina validated each SNP according to its dbSNP identification number [278, 279], and provided designability scores to indicate the probability that each SNP would be genotyped successfully on their platform. 1365 SNPs passed this initial review process. To replace the failed SNPs, CBCS investigators selected the best possible substitutes and resubmitted the SNP panel to Illumina until a complete set of 1536 SNPs with passable design scores was identified.

Once the custom solid-phase bead assay was received from Illumina, the 3857 CBCS participants with available blood samples were genotyped for all 1536 SNPs in UNC's Mammalian Genotyping Core lab. At UNC, 2 μg of participant DNA were extracted, then activated and combined with assay oligonucleotides, hybridization buffer, and paramagnetic particles [280]. Each assay contained three locus-specific oligonucleotides for each SNP, two of which bonded directly to the allele sites, and a third that hybridized downstream from the

other two. The third oligonucleotide contained a unique address sequence that corresponded to a particular bead type.

After several wash steps and the addition of DNA ligase, the oligonucleotide-tagged sequences formed double-stranded DNA fragments. These fragments were amplified with dyed polymerase chain reaction (PCR) primers such that each allele of a given SNP was uniquely labeled. Lastly, the finalized, single-stranded, dye-labeled DNA products were hybridized to the bead type specified by the address sequence contained in the third oligonucleotide. In this final form, the fluorescence signal of each bead, and thus the genotype of each SNP, could be measured using Illumina's BeadArray Reader.

3.1.5.2 Illumina Quality Control

CBCS investigators used several quality control techniques to measure and improve overall data accuracy. This included the use of blind duplicate samples, lab controls, the examination of individual call rates, and careful inspection of assay intensity data and genotype clustering images.

Roughly four percent (169 of 3857) of samples underwent duplicate, blinded genotyping. Of more then 200,000 possible discrepancies, 11 genotype miscalls were identified in these samples. DNA samples from the HapMap CEU population (provided by the Coriell Institute for Medical Research) were also repeatedly genotyped as lab controls, with only 2 discrepancies in 184 replications. As these error rates were well-within reasonable limits, no SNPs were excluded based on these results. However, closer inspection of assay intensity data and genotype clustering images did lead to the exclusion of 163 of the 1536 genotyped SNPs (11%). Here, SNPs were excluded from further analyses if they showed low signal intensity or if genotype clusters were overlapping, as both are indications

of genotyping error.

Individuals were excluded from further analyses if they had call rates of less than 95% for the 1383 remaining SNPs. 103 participants met this exclusion criteria. Six other subjects were omitted, including 5 genotyped as male and 1 with conflicting duplicate samples. Case, race, age, and stage distributions were similar for excluded and included subjects [273]. After exclusions, genotyping information was available for 3748 participants (1972 cases, 1776 controls) and 1373 SNPs, including 57 of the 83 candidate SNPs selected for this analysis and 144 AIMs.

3.1.5.3 Taqman assays

The remaining 26 candidate SNPs were genotyped using Applied Biosystems' (Foster City, CA) fluorogenic 5' nuclease assay ('TaqMan®') and PRISM® 7900HT Sequence Detection System in UNC's Mammalian Genotyping Core lab. As with the Illumina assays, the TaqMan assays were customized to target specific candidate SNPs. However, instead of oligonucleotides, TaqMan assays contain site-specific primers and fluorescently labeled probes for each selected SNP. TaqMan is more appropriate for genotyping a small number of SNPs in up to several thousand samples.

The genotyping process begins when activated DNA is combined with the custom-designed SNP genotyping assay and the Taqman Universal PCR Master Mix. During the PCR process, which consists of 10 minutes at 95˚ C, followed by 40 cycles of 15 seconds at 92˚ C and 1 minute at 60˚ C, the AmpliTaq Gold® DNA polymerase contained in the master mix amplifies genomic regions where the primers and probes have annealed [281]. As the probe is digested by the polymerase, it releases a unique fluorescence signal which is read using the PRISM® 7900HT Sequence Detection System.

3.1.5.4 Taqman Quality Control

Approximately 10% of the CBCS samples were run as blind duplicates, all of which were in agreement. Although several SNPs failed the initial design phase, all of the genotyped SNPs met quality controls standards. No additional participants were removed due to low call rates.

3.1.5.5 SNP exclusions

I excluded six SNPs with MAFs less than 1% in white participants and ten SNPs with MAF less than 1% in African Americans, leaving a total of 77 evaluable SNPs in whites and 73 in African Americans. 79 SNPs with MAF greater than 1% in the non-racially stratified CBCS population were evaluated in subtype analyses.

3.1.6 IHC analysis

Tumor tissue was collected from 80% of all cases (1446 of 1808 invasive cases and 399 of 503 *in situ* cases) and sent to the UNC Immunohistochemistry Core Laboratory for storage and assaying of three or more immunohistochemical (IHC) markers. This process has also been described in previous CBCS publications [3, 4, 89].

When available, estrogen receptor (ER) and progesterone receptor (PR) status was abstracted from the patient's medical records. This information was provided for approximately 80% of all invasive cases. For the remaining 11% of invasive cases with available tissue, ER and PR IHC assays were performed at the UNC core laboratory using archived tumor tissue [282]. Tumors with more than 5% of cells showing nuclei-specific staining were considered receptor positive [283]. A random 10% sample of tumors reported as ER+ in medical records were retested in the UNC lab, as were a random 10% sample of ER- tumors. With a concordance of 81% and a kappa statistic of 0.62 between the medical

records and the UNC-run assays, CBCS investigators decided that agreement was high enough to justify the continued reliance on medical records data [3].

All tissue samples with sufficient tissue were assayed for human epidermal growth factor receptor 2 (HER2), human epidermal growth factor receptor 1 (HER1 or EGFR), and cytokeratin 5/6 (CK 5/6) in the UNC Immunohistochemistry Core Laboratory. HER2 status was detected using the CB11 monoclonal antibody [284]. A case was considered HER2+ if at least 10% of observed cells showed signs of staining. This method had high concordance (81%) with PCR-based measures of HER2 gene expression. HER1 and CK 5/6 assays were conducted according to the methods developed by Nielsen et al [42]. Here, tissue with any sign of cytoplasmic or membranous staining was considered positive for CK 5/6 or HER1, respectively.

Due to the limited amount of available tissue, IHC analyses for *in situ* tumors were handled differently. The process is fully described in Livasy et al [89]. Briefly, tumor tissue was stained using antibodies for ER, HER2, HER1 and CK 5/6, then classified as positive or negative for each marker based on pre-established criteria. Tumors with an Allred score above 2 with nuclear staining were ER+. Tumors were HER2+ if more than 10% of visible cells demonstrated membranous staining with an intensity of 3+ using DAB chromogen or 2+/3+ using SG chromogen. For CK 5/6 and HER1, any tissue with staining in the cytoplasm or membrane, respectively, was considered positive for expression. Given the high correlation between ER and PR expression [34-37] and the paucity of available tissue, PR status was not assessed in *in situ* tumors [4].

Upon completion of the IHC analyses, breast cancer cases were categorized according to their intrinsic subtype. As described in section 2.5, these subtypes are classified as follows:

luminal A (ER+ and/or PR+, HER2-), luminal B (ER+ and/or PR+, HER2+), HER2+/ER- (ER-, PR-, HER2+), and basal-like (ER-, PR-, HER2-, HER1+ and/or CK 5/6+), with those negative for all 5 markers considered "unclassified". Altogether, 62% of cases were assigned subtypes, including 1149 invasive and 275 *in situ* cases. Compared to CBCS cases without IHC marker data, subtyped cases were more likely to be African American or have a higher stage of disease [4]. There were no differences between CBCS cases with and without IHC marker data in terms of age, menopausal status, or family history.

Of the 1972 cases with genotyping data, 1220 (62%) also had complete subtype information (Figure 21). Among individuals with both genotyping and subtype data, there were 700 luminal A cases (56%), 122 luminal B cases (10%), 98 HER2+/ER- cases (8%), 207 basal-like cases (16%), and 133 unclassified cases (11%). This subtype distribution is virtually identical to the subtype distribution of all CBCS participants, as reported in Millikan et al [4]. In this regard, participants with genotyping and subtype information are well representative of all CBCS participants with subtype data.

3.2 Other covariates

3.2.1 Race and age

Prior to initial contact, potential participants' race and age were abstracted from either the pathology report (cases) or Department of Motor Vehicle or Health Care Financing Administration records (controls). Each woman's selection probability was based on this preliminary data. During the enrollment and eligibility confirmation phase of the study, each control was asked to verify her current age and each case her age at diagnosis. Additionally, nurse interviewers recorded each participant's exact birth date during the in-home interviews

and confirmed that this date was consistent with prior reports. Nurse interviewers also asked participants to describe which racial group they belonged to – White, Black/African American, American Indian/ Eskimo, Asian/ Pacific Islander, or Other – and whether they considered themselves to be Hispanic.

3.2.2 Stage at diagnosis

Breast cancer stage at diagnosis, as defined by the American Joint Committee on Cancer criteria, was abstracted from cases' medical records. Briefly, women with *in situ* disease were classified as Stage 0 and women with progressively larger primary tumors and/or more lymph node involvement had Stages I, II or III breast cancer. Women with distant metastases were diagnosed with Stage IV disease [285].

3.2.3 African and European ancestry

When examining how genetic variations affect disease status, I also had to account for differences in allele frequencies due to divergent ancestral origins. This phenomenon is known as population stratification. Self-identified race can capture some of the variability caused by population stratification, but many individuals cannot accurately classify their own ancestry. This is especially true with recently admixed populations, such as African Americans, who often have both African and European lineage.

In the CBCS, each participant's proportion of European and African ancestry was estimated using maximum likelihood estimate (MLE) methods, 144 AIMs, and allele frequencies in HapMap CEU and YRI individuals. The HapMap populations served as proxies for the European and African parent populations, respectively. By including only these two parent populations, we assumed that each individual descended from one or both of these two groups, but no others. Individual ancestry proportions were estimated using the

equations described below [286].

Given ancestral contributions m_1 and m_2, the probability of observing the kth allele at the gth locus in an admixed population A is:

$$p_{gAk} = m_1 p_{g1k} + m_2 p_{g2k}.$$

Based on our assumption that each population's ancestry is limited to only two parent groups, we know that $m_1 + m_2 = 1$. We can also define δ as the allele frequency difference in the two parent populations, such that $\delta = p_{g1k} - p_{g2k}$. With these substitutions, we now have:

$$p_{gAk} = m_1 p_{g1k} + m_2 p_{g2k} = p_{g2k} + m_1 \delta_{g1k}.$$

The likelihood of this function is equal to the product of each possible value of g and k:

$$L = \prod_g \prod_k p_{gAk} = \prod_g \prod_k (p_{g2k} + m_1 \delta_{g1k}),$$

and the log-likelihood is:

$$\ln(L) = \sum_g \sum_k \ln(p_{gAk}) = \sum_g \sum_k \ln(p_{g2k} + m_1 \delta_{g1k}).$$

By maximizing this log-likelihood, we can obtain an equation that can be used to estimate the value of m_1 that is most probable given the provided data:

$$\frac{\delta \ln(L)}{\delta m_1} = \sum_g \sum_k \frac{\delta_{g1k}}{p_{g2k} + m_1 \delta_{g1k}} = 0.$$

The MLE of m_1 was estimated using the Newton-Raphson method [286].

For our analysis, we treated each participant as a population with a sample size of one and estimated her proportion of European ancestry based on her genotype for each of the 144 AIMs. Among African American CBCS participants, the median proportion of European ancestry was 0.19 (average= 0.22), with most women in the 0 to 0.50 range [210, 287]. The large majority of self-identified whites had between 80% and 100% European ancestry, with a median proportion of 0.94 (average = 0.93). The small proportion of women who self-identified as mixed race, Hispanic, Asian/ Pacific Islander, or American Indian/ Eskimo had

median European ancestry estimates of 0.93, 0.75, 0.73 and 0.60 respectively.

3.3 Statistical methods

3.3.1 Descriptive statistics

Although previous publications have included detailed descriptions of the demographic and clinical characteristics of CBCS participants [3, 4, 282], I generated race-stratified frequency estimates for the distribution of age, subtype, and stage at diagnosis for the women included in these analyses. To account for the unequal sampling and provide population-based estimates, frequency estimates were weighted by the inverse of the probability of being selected as a CBCS participant. I also provided weighted case- and race-stratified genotype frequencies (Aim 1) and weighted race and subtype-stratified genotype frequencies (Aim 2). These descriptive statistics provided a preliminary account of the exposures and outcomes of interest and allowed comparisons between and across populations. Women who self-identified as a race other than non-Hispanic white or African American were excluded from race-stratified analyses, though they were included in some overall evaluations (Aim 2 only).

3.3.2 Hardy-Weinberg equilibrium

In conjunction with the case and race-stratified genotype frequency analysis, I assessed whether the control population for each racial group was in Hardy-Weinberg equilibrium (HWE). According to the Hardy-Weinberg Law, "the genotype and allele frequencies in a large, randomly mating population remain stable over generations and there is a fixed relationship between allele and genotype frequencies" [288]. In practice, this means that the proportion of individuals with each genotype should directly correspond to the

overall allele frequencies for the population. In other words, if alleles A and B occur with frequencies p_A and p_B, then the relative frequencies of genotypes AA, AB and BB should be roughly equal to p_A^2, $p_A p_B$ and p_B^2, respectively. Any violation of this relationship is a sign of non-random genotyping error, population stratification, natural selection, disease association, or chance. As the probability of natural selection within a single generation is quite minimal, the degree of disequilibrium due to either genotyping error or chance can be quantified by testing for HWE within racially restricted control populations.

I evaluated HWE using Pearson's chi-square test. For each SNP I compared the number of individuals with genotype *i* expected under HWE (E_i) to the number of individuals observed with genotype i (O_i).

$$\chi^2 = \sum_{i=1}^{3} \frac{(O_i - E_i)}{E_i}, \text{ with 1 degree of freedom (df)}$$

I conducted this analysis in African American and white control populations separately. Any SNP with a p-value less than 0.05 was considered in violation of HWE. Although violation of HWE may indicate genotyping error, the test has poor sensitivity and there is poor agreement on appropriate significance cut-points [289]. Therefore, rather than simply excluding those SNPs that were not in HWE, clustering images of those in Hardy-Weinberg disequilibrium were re-inspected for signs of poor genotype differentiation. SNPs were retained as long as their genotypes formed discrete clusters with minimal overlap and HWE was not violated in both population groups. As this strategy cannot rule out certain types of genotype misclassification, such as allelic drop-out, any SNPs not in HWE were noted and interpreted with caution [290].

3.3.3 Linkage disequilibrium

LD blocks were defined using Haploview (Haploview 4.2, Version 1.0, Broad Institute, Cambridge, MA, USA) [164] and haplotype block criteria established by Gabriel et al [165]. I assessed LD patterns in whites and African Americans separately. I used the ALLELE procedure in SAS v9.3 (SAS, Cary, NC) to calculate between SNP correlation measures r and D′ [288].

3.3.4 Confounding and other adjustment factors

Confounders are factors that are related to the exposure and disease of interest in such a way that the estimated effect of the exposure on the disease are difficult to separate out from the combined effect of the exposure and the confounder on the disease. To be a confounder, a factor must be (1) an extraneous risk factor for the disease, (2) associated with the exposure under study in the source population and (3) not be affected by the exposure or the disease [291]. In terms of the analyses proposed here, a true confounder must affect a woman's genotype and her breast cancer incidence (Aim 1) or a woman's genotype and her breast cancer subtype (Aim 2). As there are few factors that can affect genotype, the list of potential confounders is limited to ancestry, race, and age.

3.3.4.1 Ancestry and Race

As discussed previously, allele frequency patterns vary across populations. These divergent patterns, known as population stratification, persist as long as populations remain isolated. However, if individuals from two previously isolated populations reproduce, as occurred with African Americans and, to a lesser extent, US whites, the populations become admixed and their offspring have allele frequencies somewhere between that of the two parent populations.

Population stratification can induce confounding in genetic association studies if both the genotype frequency and underlying disease risk vary by ancestry [291, 292]. Given the racial disparities in breast cancer incidence and subtype distributions discussed in sections 2.2 and 2.5.1, respectively, and the MAF discrepancies between the CEU, YRI and ASW populations presented in Table 5, population stratification likely confounds the association between genotype and breast cancer risk and between genotype and breast cancer subtype. Stratifying by self-reported race will partially address this concern, but because admixture has produced a great deal of variability within self-identified populations, I also adjusted for proportion of African ancestry. This proportion was centered at its overall mean for the subtype-specific analyses (Aim 2), but at the race-specific means for the overall breast cancer analysis (Aim 1) and the race-stratified subtype analyses (Aim 2). The causal directed acyclic graph (DAG) included as Figure 22 illustrates the relationship between these factors.

For Aim 1, I stratified by race as both a means to control confounding and to explore effect modification between African Americans and whites. In addition to affecting breast cancer incidence, breast cancer subtype, and allele frequencies, race also affects LD patterns. As most of the candidate SNPs are probably highly correlated with causal variants, rather than causal SNPs themselves, comparing results across populations may help to pinpoint the genomic region where the true causal SNP is located.

I also conducted some race-stratified analyses within the breast cancer subtype analysis (Aim 2), but because of within-strata sample size limitations, my primary analyses were completed using the entire CBCS population. This included women who identified as something other than non-Hispanic white or African American. In these overall analyses, I adjusted for race as a dichotomous variable (African American vs. non African American).

3.3.4.2 Age at selection

Older age is strongly associated with overall breast cancer incidence, and average age at diagnosis varies by breast cancer subtype. Although one's genotype does not change with age, genotype can affect survival, and thus influences the probability that women with certain genotypes will be selected into the study. For this reason, I adjusted for age as a selection factor in both sets of analyses. In all analyses, I centered age at 50 years.

3.3.4.3 Offset term

In addition to race, proportion European ancestry, and age, all models were adjusted for an offset term to account for the uneven sampling fractions utilized during study recruitment. Inclusion of this offset term, which was equal to the case versus control ratio of the log of the selection probability, eliminated the selection bias created by the distorted sampling [293]. This method is more efficient than traditional matched designs, and allows for simple, unbiased valuation of population-based point estimates and variances [294].

3.3.5 Frequentist analysis

In addition to Bayesian statistical methods, which will be discussed in the following section, I estimated ORs and 95% CI using frequentist-based unconditional logistic regression models. Effect estimates obtained using these methods were directly comparable to previously published reports for the same SNPs and also served as reference points for the estimates obtained using Bayesian analysis methods. I did not adjust for multiple comparisons, as these were primarily candidate SNPs. Because they are calculated using maximum-likelihood estimation, I will herein refer to the frequentist estimates as the MLE ORs.

3.3.5.1 Genotype classification

For Aim 1, I estimated the association between each selected risk variant and incident breast cancer. For Aim 2, I estimated the association between each selected risk variant and breast cancer subtype. In both analyses, the risk allele for each SNP was selected *a priori* based on previously published results.

For whites, I selected risk alleles for SNPs in *ATM, CASP8*, *ESR1* and *TP53* based on the ORs reported by Zhang et al. in their comprehensive meta-analysis [20]. For the GWAS-identified SNPs, I ascertained the risk allele in the original GWAS [186, 187, 189, 194, 196, 197, 201, 205, 208, 209] and then compared these findings to subsequent replication studies [162, 163, 166, 168, 169, 188, 190, 191, 195, 199, 200, 202-207, 211-213, 216, 218-222, 224, 225, 228-230, 232-237, 241-243, 245, 246, 250-256, 295-301]. As all of the statistically significant ($p<0.05$) ORs were in the same direction as the original GWAS, all designated risk alleles are identical to those indicated in the initial assessment (see Table 7). If a candidate SNP was not assessed in a GWAS or Zhang et al., I designated the minor variant as the risk variant, using the HapMap CEU population as a reference. The only exception to this rule was for the major allele of rs3750817 (*FGFR2*), which was associated with an increased risk of breast cancer in two large case-control studies and an increased risk of ER or PR+ disease in a later GWAS [169, 193, 252].

I assumed the risk allele was the same for African Americans as it was for whites unless the majority of reported statistically significant associations were contradictory to those seen in whites. Two SNPs met these criteria (rs3803662 on *TNRC9/TOX3* and rs1045485 on *CASP8*) [11, 12, 161, 169, 214-216, 220, 302-311].

I estimated three MLE ORs for each SNP-outcome combination. Two of these ORs

were estimated assuming a general genetic model and the last was estimated assuming an additive genetic model. Either genetic model can be used to generate estimates for the relative effect of having one or two copies versus no copies of the risk allele, but the additive model makes the assumption that the relationship between the log ORs is linear. In other words, when assuming a general model, I treated genotype as a three-level categorical variable and estimated two ORs, one for the effect of being heterozygous versus homozygous for the risk allele, and one for the effect of being homozygous for the risk allele versus homozygous for the referent allele. With the additive model, I treated genotype as an ordinal variable with three levels corresponding to the number of copies of the risk allele (0, 1, or 2). The OR estimated using this model corresponded to the effect of being heterozygous versus homozygous for the referent allele, while $e^{2*\ln(OR)}$ was the estimated effect of being homozygous for the risk allele versus homozygous for the referent allele. The additive model includes one less parameter than the general model, but it is also more restrictive and may fail to capture the relationship between the SNP and the outcome.

Alternatively, SNPs may follow dominant or recessive genetic patterns. Under either model assumption, two of the genotypes are grouped together, such that individuals with one or two copies of the risk allele have equal risk of the outcome (dominant model) or individuals with one or no copies of the risk allele have equal risk of the outcome (recessive model). These genetic models appear much less frequently in the literature, but are useful replacements for the general or additive model when the proportion of individuals homozygous for the risk allele is especially low or especially high. Accordingly, I estimated the OR using the dominant model for any SNPs with risk allele frequency (RAF) of less than 5% in either cases or controls and used the recessive model when the RAF was greater than

95%.

3.3.5.2 Outcome classification

For my first aim, the outcome of interest was breast cancer, which included both invasive and *in situ* disease. This outcome was coded as a dichotomous variable, with each SNP-specific OR estimating the odds of being a case rather than a control among women of one genotype, relative to the odds of being a case rather than a control among women of the referent genotype.

To address my second aim, I treated breast cancer subtype as a categorical variable with six levels (control, luminal A, luminal B, HER2+/ER-, basal-like or unclassified) and modeled ORs using polytomous logistic regression models. With each polytomous model, I simultaneously estimated the effect of a given risk variant on the odds of having each subtype relative to being a control.

3.3.5.3 Logistic regression

As this was a retrospective case-control study with unequal sampling of cases and controls, logistic regression was an appropriate method to model the effect of genotype on breast cancer risk in the presence of known and measured confounders and selection factors. Assuming a general genetic model and stratification by race, the estimated odds of having the outcome of interest (Y=1) given genotype g took the following form:

$$\text{Odds}\,(\text{Y} = 1 \,|\text{G} = \text{g}, \text{race} = \text{r}) = \exp\,(\alpha + \beta_1 x_1 + \beta_2 x_2 + \beta_3 age + \beta_4 ancestry + offset),$$

where x_1=1 for heterozygotes, and 0 otherwise and x_2 = 1 for risk allele homozygotes and 0 otherwise. β_1 and β_2 estimated the log ORs for the effect of a heterozygous versus homozygous referent genotype and the homozygous risk genotype versus homozygous

referent genotype on breast cancer risk, respectively. The additive model took a similar form:

$$\text{Odds}\,(\text{Y} = 1 \,|\text{G} = \text{g}, \text{race} = \text{r}) = \exp(\alpha + \beta_1 x_1 + \beta_2 age + \beta_3 ancestry + offset),$$

where x_1 was equal to the number of copies of the risk allele and β_1 was the log OR for the effect of each additional copy of the risk allele on the odds of having breast cancer. ORs and 95% CIs were estimated using the LOGISTIC procedure in SAS v9.3 (SAS, Cary, NC).

3.3.5.4 Polytomous logistic regression

To create the polytomous logistic regression model, I expanded the above binary form to include unique coefficients for each level of the outcome. This included a unique intercept and SNP, age, and ancestry parameter for each of the five subtypes. I selected this approach based on evidence that effect estimates generated using polytomous logistic regression models have lower standard errors than effect estimates fit using five separate binary logistic regression models [312]. I assessed the polytomous MLE ORs using the 'glogit' link function in PROC LOGISTIC (SAS, Cary, NC).

3.3.6 Bayesian analysis

At the cost of a few additional assumptions about the data and associations of interest, Bayesian statistical methods offer a powerful and more easily interpreted alternative to frequentist methods. Although many investigators avoid these methods because of added complexity and subjectivity, the added complexity is only moderate and frequentist analyses are also subjective. Moreover, my research questions could be addressed more effectively using Bayesian techniques, as such methods allowed me to use previous research to produce more accurate and precise effect estimates while implicitly controlling for false positive associations.

3.3.6.1 Bayes law

The difference between Bayesian and frequentist methods can best be described using Bayes' Law, a principle of conditional probability:

$$p(\theta|D) = \frac{p(\theta)}{p(D)}\, p(D|\theta)$$

Here θ is the parameter of interest, D are the observed data, and p(θ|D), the probability of observing an effect θ given the observed data, is the desired estimate. P(θ|D) is often referred to as the posterior probability and written in the form π(θ|D). Although the above form of Bayes' Law only allows for a single parameter, it can easily be expanded to accommodate more variables.

As θ is an unknown random variable, the probability of observing data D given parameter θ, p(D|θ), is best estimated using a likelihood function, L(θ|D). Although the form of the function depends on the investigator's research question and modeling assumptions, he or she would utilize the same likelihood function in a Bayesian or frequentist analysis.

Additional subjectivity is introduced into Bayesian analysis with the specification of p(θ), which is the probability of observing parameter θ independently of, or prior to, observing data D. As this quantity cannot be directly measured, it must be assigned by the investigator based on his or her prior knowledge about the nature of the association of interest. If the investigator is unwilling to assign an informed probability distribution to p(θ), the prior is given an infinite, uniform distribution. In this way, a frequentist analysis is equivalent to a Bayesian analysis where all values of θ, from -∞ to +∞, are equally likely.

The last term of Bayes' Law is p(D), the marginal distribution of the data. This quantity is directly dependent on the values of p(θ) and p(D|θ) and takes the following form:

$\int_\theta p(\theta)L(\theta|D)d\theta$. As this quantity is not dependent on θ, we can ignore its exact specifications and state Bayes' Law as a simple proportional relationship between the posterior probability and the product of the prior probability and the likelihood function: $\pi(\theta|D) \propto p(\theta)L(\theta|D)$. This proportion conveys the simple nature of the mathematical relationship between frequentist methods, which rely solely on the likelihood function, and Bayesian methods, which are additionally dependent on the specification of a prior probability estimate.

3.3.6.2 Priors

A simple way to incorporate the prior is to add a second stage to the likelihood model. If the likelihood takes the form of a logistic regression, with the log OR of being a case given exposure X_j and covariates W modeled as: $\ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_j X_j + W\gamma$, a second level can be added to the coefficient of the exposure β_j: $\beta_j = z_j\pi + \delta_j$. Here, z_j is a $j\times1$ matrix of 1's, π is the prior log OR, **T** is a $j\times j$ identity matrix and the δ_j are assumed to be normally distributed with mean 0 and variance $\tau^2\mathbf{T}$. The prior mean (π), variance ($\tau^2\mathbf{T}$), or both can either be estimated directly from the data or assigned by the investigator. I use the term 'full-Bayes' to indicate that priors were assigned independently of the data at hand, 'empirical Bayes' when both were estimated from the data, and 'semi-Bayes' when one prior was assigned and the other was estimated [14].

3.3.6.3 General benefits of Bayesian analysis

Essentially, posterior probabilities are the result of shrinking the MLE towards the specified prior estimate, with the degree of shrinkage determined by the variance of the MLE relative to the variance of the prior [14, 313, 314]. In this way, including any prior with a finite variance will reduce the variance of the posterior probability relative to the MLE,

though the discrepancy will be minimal if the sample size is large and the prior is diffuse. If the MLE is between the true parameter value and the mean of the prior distribution, the shrinkage process will generate a biased effect measure. However, if the prior is well-selected, any increase in bias due to shrinkage will be offset by an increase in precision, and, on average, allow more accurate effect estimation.

Bayesian effect estimates are also easier to interpret than frequentist estimates, as they can be phrased in terms of probability rather than the idea of repeated, unbiased sampling of the target population. For example, a frequentist 95% confidence interval refers to the region that will include the true parameter in 95% of infinite replications of unbiased data collection and measurement [291]. Alternatively, a 95% posterior interval (95% PI) is the region where the probability of covering the true estimate is 95% [315]. In terms of the actual effect estimates, a frequentist-determined estimate is the parameter that maximizes the likelihood of the observed data, while a Bayesian effect estimate reflects the odds that an investigator would place on the estimated parameter versus an alternative parameter, given his or her prior knowledge and the observed data.

Admittedly, the relative benefits of Bayesian analysis are dependent on the investigator's choice of prior. While this requires a certain degree of subjectivity, prior selection is not arbitrary. The best priors are selected based on previously published findings or because of certain inherent properties. For example, null-centered priors will shrink all estimates towards 0, thereby attenuating the effect sizes of extreme observations and reducing overall type I error [14, 15]. If an investigator is concerned about the bias or generalizability of others' results, he or she can accommodate this uncertainty by selecting a prior with a large variance. Additionally, while it is not possible to verify how a prior affects

the validity of an estimate, one can easily compare the relative influences of a variety of priors using sensitivity analyses.

3.3.6.4 Application to replication of breast cancer susceptibility loci

For many of the reasons described above, Bayesian methods were well-suited to address the described research questions [316, 317]. First of all, most of the selected SNPs have been studied previously, many of them in 5 or more separate investigations. Nearly all of these replication studies were conducted using a case-control study designs and well-defined, race-specific populations. While these studies may suffer from selection biases or other sources of error that affect generalizability or comparability, the assessment of both genotype and breast cancer status was extremely consistent across studies, with most previous studies of the association between known susceptibility variants and breast cancer producing ORs in the range of 1.1-1.3 [18, 20, 318].

Additionally, Bayesian methods provide a means to advance research on the less studied topics of genetic risk factors for African Americans and breast cancer subtypes. Previous investigations have evaluated these SNPs in African Americans or within specific subtypes, but such analyses often have small sample sizes and thus produce only imprecise effect measures. By constructing priors based on previous knowledge about SNPs' effects on overall breast cancer in previous studies of white or Asian populations, we can conduct better-informed statistical analyses and obtain valid and precise race-specific and subtype-specific ORs.

When examining numerous SNPs in the same study, most investigators apply Bonferroni corrections to control for false positive associations and ensure that the overall α-level remains at an acceptable level (usually 0.05). Because SNPs are often highly correlated

with one another, multiple comparisons correction via Bonferroni methods are overly-conservative and may fail to detect true associations [15, 291]. Alternatively, by applying Bayesian methods to genetic association studies and shrinking all effect estimates toward a well-justified, preconceived prior, we can limit the number of false positive associations [13-16, 319-321]. As discussed previously, a null-centered prior should achieve this result.

3.3.6.5 Bayes regression analysis

For both aims, I conducted full-Bayes analyses, where I assigned priors for the mean (π) and the variance ($\tau^2\mathbf{T}$). For Aim 1 I also conducted semi-Bayes analyses, where I assigned a fixed τ^2 but used LD-block (i.e. haplotype) level OR estimates to inform π. I did not use empirical Bayes methods, as the near-zero τ^2 generated from this rich data set would cause over-shrinkage of the SNP-specific effects [322]. For all Bayesian analyses I assessed all SNP-outcome associations using additive genetic models.

All 83 candidate SNPs (Table 6) were individually assessed using Bayesian regression analysis methods with priors specified for the intercept, SNP, age, and ancestry parameters (Aim 1) or the intercept, SNP, age, ancestry and race parameters (Aim 2).

3.3.6.5.1 Full Bayes analysis for Aim 1: Race-stratified estimates for SNP effects on overall breast cancer

Given the probable SNP-overall breast cancer OR range of 1.1-1.3, I assigned each SNP a null-centered, lognormal prior with a mean of 0 and variance $\tau^2 \sim \Gamma^{-1}(3, 0.2)$, such that 95% of the prior mass for each SNP-breast cancer OR lay between 0.64 and 1.55 when τ^2 was equal to 0.05, the mode of the specified distribution. I also assigned null-centered, lognormal priors for age, ancestry and the intercept term. I assigned moderately strong priors to age and ancestry (τ^2=0.68), both of which were mean-centered variables. Because the

intercept is difficult to define or interpret in a case-control study with weighted sampling, I assigned only a vague prior, with τ^2=1000. I assumed that all priors were independent.

Bayesian ORs and 95% PIs were computed using the MCMC procedure in SAS v9.3 (SAS, Cary, NC). PROC MCMC uses Markov-chain Monte Carlo (MCMC) methods with a random walk Metropolis algorithm to obtain posterior probability estimates [323]. This iterative, conditionally dependent stochastic process searches the space defined by the joint distribution of the specified parameters to find the region with the highest density [315]. I ran 30,000 samples for each SNP model, discarding the first 1000 draws as a burn-in and retaining every fifth draw. I inspected autocorrelation, trace, and density plots for signs of poor mixing or non-convergence.

3.3.6.5.2 Full Bayes analysis for Aim 2: SNP effects on breast cancer subtype

Given that subtype-specific SNP associations are poorly understood, I selected more diffuse hyperpriors when estimating subtype-specific effects. Each SNP was again assigned a null-centered, lognormal prior with a mean of 0, but this time the variance was $\tau^2 \sim \Gamma^{-1}(4, 0.5)$. As such, 95% of the prior mass for each SNP-subtype OR lay between 0.54 and 1.86 when τ^2 was equal to its mode of 0.1. I used the same age, ancestry and intercept priors as I had in Aim 1, but also included a lognormal, null-centered prior for race (τ^2=1.0). I opted for a more diffuse prior for race because it likely has a larger effect than the other mean-centered, continuous covariates. As with the frequentist analysis, I used polytomous regression to model all five subtype versus control comparisons simultaneously.

For these complex models, I ran 50,000 samples, discarding the first 1000 draws as a burn-in and thinning by retaining every tenth sample, such that the results are based on 4990 draws. I again inspected autocorrelation, trace, and density plots for signs of poor mixing or

non-convergence.

3.3.6.6 Hierarchical models

Sixty-six of the 83 candidate SNPs were located in the same gene or gene region as at least one other candidate SNP (see Table 6). As many of the SNPs within these regions were highly correlated with one another, I implemented a different Bayesian approach, known as hierarchical modeling, in hopes of generating more informative and precise effect estimates than either the frequentist or full Bayes methods [324-332]. I used these semi-Bayesian methods to integrate linkage disequilibrium (LD) information into SNP-breast cancer association analyses. I did not evaluate SNP-subtype associations with these methods.

Like other Bayesian regression models, hierarchical models shrink MLEs towards a prior estimate, with the degree of shrinkage determined by the relative variances of the MLE and the prior. However, by selecting a semi-Bayes rather than a full Bayes approach, I was able to use the observed data to help inform prior selection. More specifically, I used the data to determine the mean of the prior for each SNP log OR (π) but selected priors for SNP variance parameters ($\tau^2\mathbf{T}$) based on existing knowledge. This application of semi-Bayes methods did not require prior specification for all included parameters.

The prior mean was modeled using the joint effect of all of the SNPs in an LD block. In terms of the framework specified in Section 3.3.6.2, the first level of the hierarchy was used to model the joint association between the outcome and all of the SNPs included in each LD block, and the second level was used to model the association between the outcome and the haplotype [16, 328, 333, 334]. For example, if there were j=3 SNPs (x_1, x_2 and x_3) in the LD block, the first level of the hierarchical model took the following form:

$$\ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_j X_j + W\gamma = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + W\gamma.$$

The second level was used to model β_j: $\beta_j = [\beta_1 \quad \beta_2 \quad \beta_3]' = z_j\pi + \delta_j$, where z_j is a 3×1 matrix of 1's and π is both the haplotype effect and the prior log OR for each of the individual SNPs. In its simplest form, **T** is a 3x3 identity matrix and the δ_j are assumed to be normally distributed with mean 0 and variance $\tau^2\mathbf{T}$.

I assigned fixed prior variances of 0.05 for each SNP log OR. As noted earlier, this variance corresponds to a prior OR with 95% mass between 0.64 and 1.55 when the prior mean is equal to 1. Although the estimated prior means were not set to zero in this scenario, this approximate range seemed reasonable given the usual size of individual SNP effects [18, 20, 318]. Of note, if there was only one genotyped SNP in an LD block, π and β_j were identical.

The above grouping approach is valid as long as an exchangeability assumption is met. This assumption states that before evaluating the relationship between the exposures (SNPs) and the outcome (breast cancer) in this study, there was no reason to suspect that any one SNP in an LD block had a true log OR different from the others in that LD block. As none of the included SNPs are known causal variants and all effects are evaluated in terms of risk alleles, I believe this assumption is acceptable in this setting.

Further specification of the correlation structure within the LD block may offer additional increases in precision. Here, I explored two such methods, both of which required alternative **T** matrices. For the first variation, I modeled correlation as an exponential decay function of the base pair distance between any two SNPs (d_{ij}). This takes the form of $t_{ij} = \exp\left[-\left(\frac{d_{ij}}{1000}\right)\right]$ [328]. I used their suggested values of 1000 and 1 for θ_1 and θ_2, respectively. For the second variation, I modeled the correlation structure directly, with t_{ij} = D′, a measure of LD. Only the 66 candidate SNPs located in the same gene or gene region as

at least one other candidate SNP were evaluated using these alternative T matrices. All hierarchical models were evaluated using the GLIMMIX procedure in SAS (Cary, NC).

3.3.6.7 Sensitivity Analyses

Given that few studies have examined the effects of these previously established risk variants in breast cancer subtypes, I selected more diffuse priors for the SNP-subtype association analysis than I did for the SNP-overall breast cancer association analysis. However, I also conducted sensitivity analyses to explore the robustness of the SNP effect estimates under a variety of model assumptions. Therefore, in addition to the MLE and previously described Bayesian analysis [SNP~N(0, τ^2), $\tau^2 \sim \Gamma^{-1}(4, 0.5)$, with mode at 0.1], I also estimated another set of Bayesian ORs and 95% PIs with more informative prior distributions [SNP~N(0, τ^2), $\tau^2 \sim \Gamma^{-1}(3, 0.2)$, with mode at 0.05].

Figure 20: Carolina Breast Cancer Study (CBCS) Study area

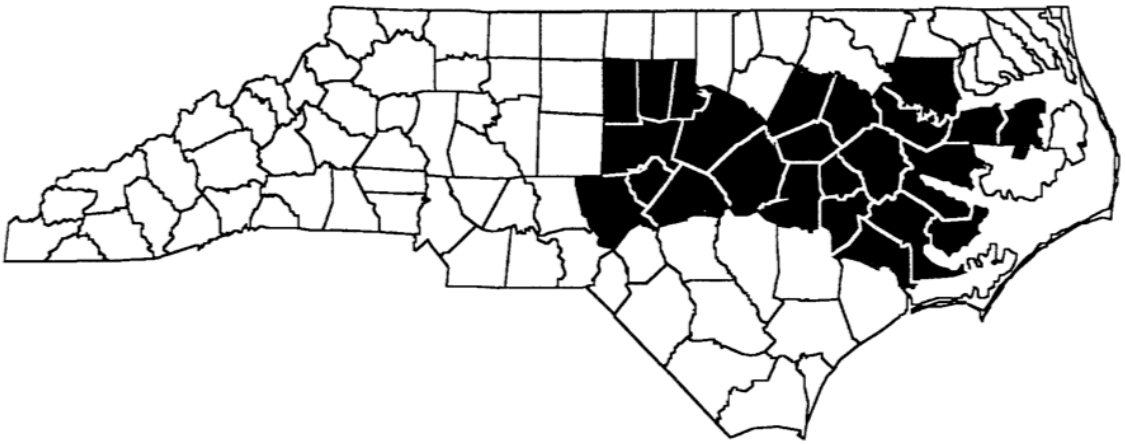


Figure 21: Flow chart for Carolina Breast Cancer Study participants

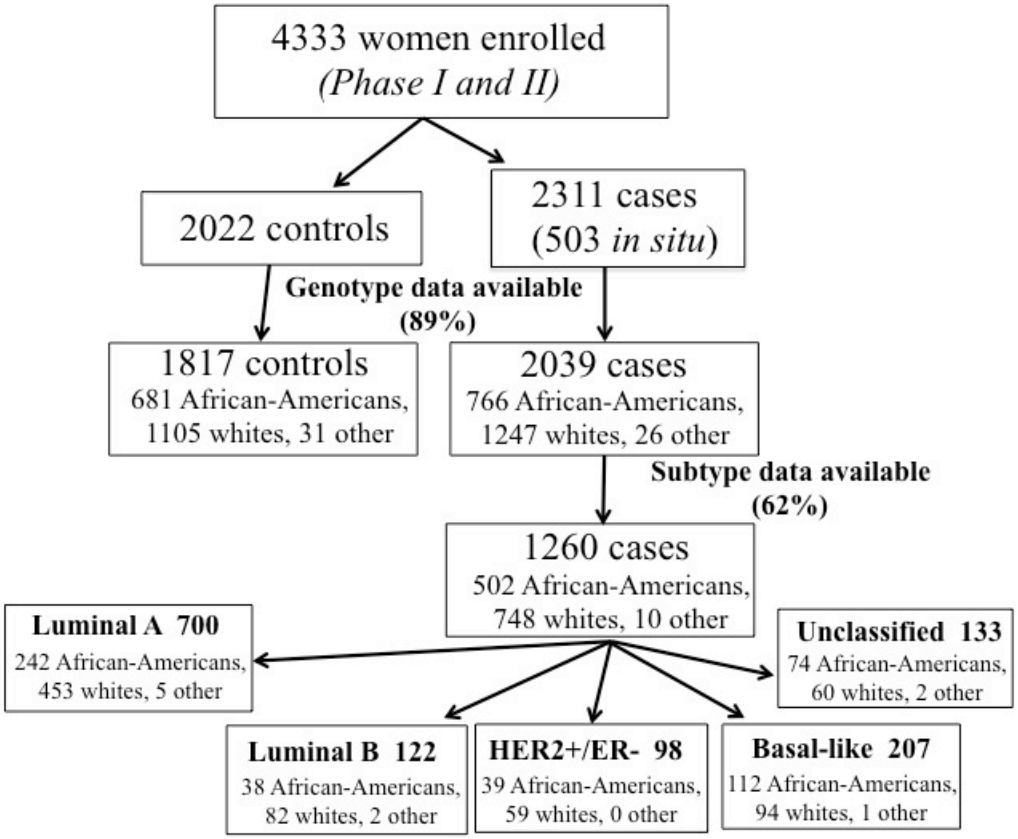


Figure 22: Directed acyclic graph (DAG) for the relationship between genotype and breast cancer or breast cancer subtype

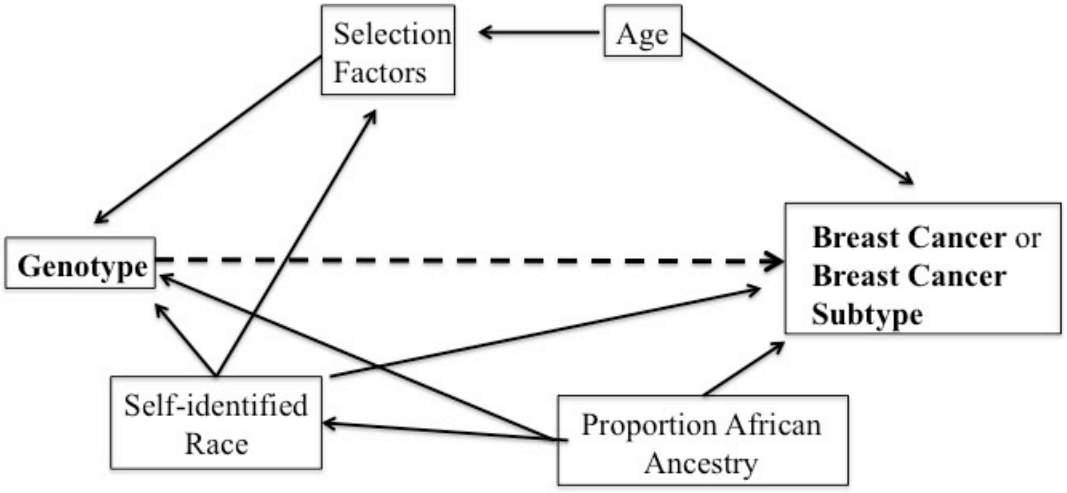


Table 6: Included SNPs

Chromosome	Gene	SNP
1	1p12	rs11249433*
2	CASP8	rs1045485§ and rs17468277
2	2q35	rs13387042*
2	2p	rs4666451†
3	SLC4A7	rs4973768*
4	4p	rs12505080†
4	TLR1	rs7696175†
5	MRPS30	rs4415084* and rs10941679†
5	5p12	rs981782*
5	5q	rs30099†
5	MAP3K1	rs889312*
6	ESR1	rs2046210*, rs851974, rs2077647, rs2234693, rs1801132§, rs3020314§, and rs3798577
6	ECHDC1	rs2180341*
7	RELN	rs17157903†
8	8q24	rs13281615* and rs1562430*
9	CDKN2A/ CDKN2B	rs10757278, rs1011970*, rs3731249, rs3731257, rs10811661, rs518394 and rs564398
10	ANKRD16	rs2380205*
10	ZNF365	rs10995190*
10	ZMIZ1	rs704010*
10	FGFR2	rs2981579*, rs1219648*, rs2981582*, rs1896395, rs3750817‡, rs10736303‡, rs11200014, rs1078806‡, rs2981578‡, rs2912774‡, rs2936870‡, rs2420946‡, rs2162540, and rs3135718
10	10q	rs10510126†
11	ATM	rs3092992, rs664143, rs170548, rs3092993, rs1800054, rs4986761, rs1800056, rs1800057§, rs1800058, and rs1801516
11	LSP1	rs3817198* and rs909116*
11	MYEOV/CCND1	rs614367*
11	H19	rs2107425†
16	TNRC9/TOX3	rs3803662*, rs4784227*, rs8049149, rs12443621‡, rs16951186, rs8051542‡, rs3104746, rs3112562, and rs9940048
17	TP53	rs1042522, rs9894946, rs1614984, rs4968187, rs12951053§, rs17880604, rs1800372, rs2909430, and rs8079544
17	COX11	rs7222197‡ and rs6504950†

*GWAS hit (p-value $<1.0 \times 10^{-5}$ in a GWAS)

†near-GWAS hit (p-value $<1.0 \times 10^{-5}$ in preliminary GWAS or follow-up study)

‡GWAS-identified SNP (OR estimated in one or more GWAS)

§Strong evidence of an association with breast cancer in Zhang et al. meta-analysis

Table 7: Selection of risk variants among women of European (EA) and African American (AA) ancestry*

Chrom-osome	Gene or gene region	SNP	Risk Allele	Number of studies (EA/AA)	Statistically significant studies (EA/AA)	Proportion of consistent findings (EA/AA)
1	1p12	rs11249433	G	7 / 5	4 / 0	1.00 / NA
2	2p	rs4666451	G	5 / 0	2 / 0	1.00 / NA
2	2q35	rs13387042	A	15 / 7	12 / 2	1.00 / 1.00
3	SLC4A7	rs4973768	T	10 / 5	7 / 0	1.00 / NA
4	4p	rs12505080	C	1 / 0	0 / 0	NA / NA
4	TLR1	rs7696175	T	1 / 0	0 / 0	NA / NA
5	MRPS30	rs4415084	T	8 / 6	4 / 0	1.00 / NA
5	MRPS30	rs10941679	G	7 / 6	5 / 1	1.00 / 1.00
5	5p12	rs981782	T	5 / 0	2 / 0	1.00 / NA
5	5q	rs30099	A	3 / 0	0 / 0	NA / NA
5	MAP3K1	rs889312	C	16 / 7	9 / 0	1.00 / NA
6	ECHDC1	rs2180341	G	4 / 5	1 / 0	1.00/ NA
6	ESR1	rs2046210	A	5 / 8	2 / 0	1.00 / NA
7	RELN	rs17157903	T	1 / 0	0 / 0	NA / NA
8	8q24	rs13281615	G	15 / 6	9 / 0	1.00 / NA
8	8q24	rs1562430	T	4 / 0	2 / 0	1.00 / NA
9	CDKN2A/B	rs3731257	T	1 / 0	0 / 0	NA / NA
9	CDKN2A/B	rs3731249	A	1 / 0	1/ 0	1.00 / NA
9	CDKN2A/B	rs1011970	T	2 / 4	1 / 0	1.00 / NA
10	ANKRD16	rs2380205	C	2 / 4	1 / 0	1.00 / NA
10	ZNF365	rs10995190	G	2 / 2	2 / 0	1.00 / NA
10	ZMIZ1	rs704010	T	1 / 2	1 / 1	1.00 / 1.00
10	FGFR2	rs3750817	C	2 / 0	2 / 0	1.00 / NA
10	FGFR2	rs10736303	G	1 / 1	0 / 0	NA / NA
10	FGFR2	rs11200014	A	5 / 0	4 / 0	1.00 / NA
10	FGFR2	rs2981579	T	7 / 3	5 / 0	1.00 / NA
10	FGFR2	rs1078806	G	1 / 1	0 / 0	NA / NA
10	FGFR2	rs2981578	C	1 / 4	0 / 3	NA / 1.00
10	FGFR2	rs1219648	G	13 / 5	9 / 2	1.00 / 1.00
10	FGFR2	rs2912774	T	2 / 2	1 / 0	1.00 / NA
10	FGFR2	rs2936870	T	1 / 0	0 / 0	NA / NA
10	FGFR2	rs2420946	T	7 / 1	5 / 0	NA / NA
10	FGFR2	rs2981582	A	17 / 8	15 / 1	1.00 / 1.00
10	FGFR2	rs3135718	G	1 / 0	0 / 0	NA / NA
10	10q	rs10510126	C	1 / 0	0 / 0	NA / NA
11	LSP1	rs3817198	C	16 / 6	7 / 2	1.00 / 0.50
11	LSP1	rs909116	T	1 / 0	0 / 0	NA / NA
11	H19	rs2107425	C	1 / 0	0 / 0	NA / NA

11	MYEOV	rs614367	T	2 / 4	1 / 0	1.00 / NA
16	TNRC9	rs16951186	T	0 / 1	0 / 0	NA/ NA
16	TNRC9	rs8051542	T	5 / 1	1 / 0	1.00 / NA
16	TNRC9	rs12443621	G	5 / 1	2 / 0	1.00 / NA
16	TNRC9	rs3803662	A	21 / 9	16 / 4	1.00 / 0.00
16	TNRC9	rs4784227	T	2 / 3	1 / 0	1.00 / NA
16	TNRC9	rs3104746	A	0 / 2	0 / 2	NA/ 1.00
16	TNRC9	rs3112562	G	0 / 1	0 / 1	NA / 1.00
17	COX11	rs7222197	G	1 / 1	0 / 0	NA / NA
17	COX11	rs6504950	G	6 / 5	3 / 1	1.00 / 1.00

*Excludes previously unstudied variants and *CASP8*, *ESR1*, *ATM*, and *TP53* variants, which were assessed using the Zhang et al. meta-analysis. NA= not applicable

4. Replication of Breast Cancer Susceptibility Loci in Whites and African Americans Using a Bayesian Approach

4.1 Overview

Genome-wide association studies (GWAS) and candidate gene analyses have led to the discovery of several dozen genetic polymorphisms associated with breast cancer susceptibility, many of which are now considered well-established risk factors for the disease in women of European descent. Despite attempts to replicate these same variant-disease associations in African Americans, the evaluable populations are often too small to produce precise or consistent results. I estimated the association between 83 previously identified single nucleotide polymorphisms (SNPs) and breast cancer in whites and African Americans from the Carolina Breast Cancer Study (1993-2001) using maximum likelihood, Bayesian, and hierarchical modeling methods. The selected SNPs were previous GWAS hits (n=22) or near-hits (n=19), otherwise well-established risk loci (n=5), or in the same gene as another selected variant (n=37). I successfully replicated eighteen GWAS-identified SNPs in whites and ten GWAS-identified SNPs in African Americans. SNPs in *FGFR2* and *TNRC9/TOX3* were strongly associated with breast cancer in both races. Additionally, SNPs in *MRPS30*, *MAP3K1, CDKN2A/B, ZM1Z1, LSP1, H19,* and *TP53* were associated with breast cancer in whites and SNPs in *TLR1*, *ESR1*, and *H19* were associated with breast cancer in African Americans. I provided precise and well-informed race-stratified ORs for several key breast cancer-related SNPs. My results demonstrate the utility of Bayesian methods in genetic

epidemiology and provide support for their application in relatively small, etiologically driven investigations.

4.2 Introduction

As of January 2013, fifty-eight single nucleotide polymorphisms (SNPs) have met the criterion for genome-wide statistical significance of $p<10^{-5}$ in at least one of twenty-three genome-wide association studies (GWAS) of breast cancer [18]. Most of these SNPs were consistently associated with the disease in subsequent investigations among women of European [162, 163, 166, 168, 169, 186-191, 193, 195, 198-207, 209, 211-213, 215, 216, 218-225, 228-230, 232-237, 241-243, 245-247, 250-256, 258, 267, 295-301, 335] or Asian descent [168, 169, 187, 192, 194, 197, 199, 205, 209, 212, 217, 221, 225-227, 230, 232, 238, 240-244, 246, 247, 262, 265, 336-339], but attempts to replicate these findings in African American women have been largely unsuccessful [11, 12, 169, 214-216, 231, 236, 243, 259, 263]. In general, the search for African American specific risk variants has made slow progress, with few insights to explain the tendency for African Americans to have more aggressive, less-treatable disease subtypes [3, 4, 8, 50] and markedly higher breast-cancer specific mortality than whites (32.7 versus 23.7 deaths per 100,000 US women with breast cancer per year, 2000-2009) [2].

Allele frequencies and linkage disequilibrium (LD) structure vary by race, with African Americans exhibiting generally weaker between-SNP correlations and smaller LD blocks [269, 270]. Because each SNP represents all correlated variants, SNPs associated with breast cancer in African Americans correspond to a narrower genomic region than SNPs associated with the disease in whites. Therefore,

studying African Americans can improve our understanding of disease etiology. Unfortunately, most of the existing studies of genetic risk factors for breast cancer in African Americans are small and therefore underpowered to reliably differentiate between null effects and ORs of 1.1-1.3, which is the typical range for GWAS-identified risk variants in other populations [18, 318].

Given this existing knowledge about association size, as well as information about the genome's physical structure, Bayesian statistical methods may be useful tools for enhancing the analysis of race-specific genetic risk factors for breast cancer. A variety of Bayesian-based methods have been proposed for use in genetic epidemiology, but I focused on two of the most basic applications, namely hierarchical modeling and Bayesian regression. My goal was to obtain valid and precise effect estimates by capitalizing on existing information.

To further our understanding of genetic risk factors for the disease, I examined the association between breast cancer and several candidate SNPs using traditional maximum likelihood methods and both Bayesian approaches. All selected SNPs were located on genes with strong prior evidence of association with breast cancer, including both GWAS and candidate gene investigations. I assessed the relationship between these SNPs and breast cancer risk using the Carolina Breast Cancer Study, a population-based, case-control study with large samples of both white and African American women.

4.3 Methods

4.3.1 Study population

Cases were identified using the North Carolina Central Cancer Registry's rapid case ascertainment program [271]. Women were eligible for the study if they were diagnosed with

invasive breast cancer between 1993 and 2001, were between 20 and 74 years of age at the time of their diagnosis, and were living in one of 24 selected North Carolina counties. Women diagnosed with *in situ* breast cancer between 1996 and 2001 were also eligible if they had ductal carcinoma *in situ* with microinvasion to a depth of 2 mm or lobular carcinoma *in situ*. To ensure approximately equal representation of African Americans and non African Americans, as well as pre- and postmenopausal women, breast cancer cases were randomly sampled at disproportionate rates based on race and age.

Controls aged 20-64 and 65-74 were selected from North Carolina Department of Motor Vehicles and Health Care Financing Administration records, respectively. Controls were probability-matched to cases on race and 5-year age group [293]. Women with a history of breast cancer were ineligible. All participants provided informed consent and all study procedures were approved by the Institutional Review Board at the University of North Carolina.

A study nurse administered a questionnaire to each participant during an in-home visit. The survey included questions on basic demographics, including age and self-reported race, known or suspected breast cancer risk factors, and medical and family history. Additionally, the nurse drew 30 ml of blood. The overall response rate was 77% for cases and 57% for controls. Of those enrolled, 88% of cases and 90% of controls provided sufficient blood samples for inclusion in genotype analyses, leaving a total sample size of 2013 cases (1247 white, 766 African American) and 1786 controls (1105 white, 681 African American). I excluded 166 individuals who did not self-identify as African American or non-Hispanic white.

4.3.2 SNP selection

Initially, I selected candidate SNPs with genome-wide significant p-values in any of eight early breast cancer GWAS or two GWAS follow-up studies [186, 187, 189, 194, 196, 197, 201, 205, 208, 209]. I also evaluated several SNPs from these initial studies that had GWAS-significant p-values only in discovery phase analyses. Lastly, I retained any SNPs already genotyped in the study population that Zhang et al. [20] determined had "cumulative evidence of an association" with breast cancer, or that were in the same gene as a previously selected variant. In total, I selected 41 GWAS-identified SNPs, including 22 GWAS hits and 19 other strongly associated SNPs, as wells as 5 SNPs from the Zhang et al. meta-analysis, and 37 SNPs from included genes. I later excluded six SNPs with minor allele frequencies (MAFs) less than 1% in white participants and ten SNPs with MAF<1% in African Americans, leaving a total of 77 SNPs in whites and 73 in African Americans.

All study participants were genotyped for 144 ancestry informative markers. I then estimated each participant's proportion of African ancestry using maximum likelihood methods and allele frequency data from HapMap YRI and CEU populations. As ancestry can affect both allele frequency and breast cancer incidence, inclusion of this variable in regression models should reduce confounding due to population stratification [275, 292].

4.3.3 Genotype analysis

The SNPs included in this analysis were genotyped using either a Custom GoldenGate Genotyping assay (Illumina, Inc., San Diego, CA) or a Taqman panel (Applied Biosystems, Inc., Foster City, CA). The Taqman panel included a few SNPs that failed the Illumina assay, as well as several more recently discovered GWAS hits. Both genotyping processes have been described previously [287, 340]. All of the SNPs included in this

analysis passed quality control evaluations, including the examination of individual call rates and inspection of assay intensity data and genotype clustering images. 109 women were assigned missing values for all of the Illumina SNPs due to poor genotyping quality, but were successfully genotyped on the Taqman SNPs.

4.3.3 Statistical methods

I calculated risk allele frequencies (RAFs) and age and African ancestry distributions for whites and African Americans separately. To account for the sampling mechanism, I weighted these estimates by the inverse of the probability of being selected as a study participant. I tested for departures from Hardy-Weinberg equilibrium (HWE) in white and African American controls using Pearson's chi-squared test. As low HWE p-values can indicate genotyping error, I re-inspected the genotype clustering images of SNPs with $p<0.05$ in either population for signs of poor genotype differentiation. SNPs were retained if their genotypes formed discrete clusters with minimal overlap.

The relationship between each risk variant and incident breast cancer was assessed using logistic regression. I estimated ORs and 95% CIs assuming additive genetic models. I also assessed the ORs and 95% CIs for all SNPs under general genetic modeling assumptions. The risk allele for each SNP was selected *a priori* based on previously published results. For whites, I selected risk alleles for SNPs in *ATM, CASP8*, *ESR1* and *TP53* based on the ORs reported by Zhang et al [20]. For the remaining SNPs, I ascertained the risk alleles from the original GWAS [186, 187, 189, 194, 196, 197, 201, 205, 208, 209] and subsequent replication studies [11, 12, 162, 163, 166, 168, 169, 186-191, 193, 195, 198-200, 202-209, 211-216, 218-225, 228-230, 232-237, 241-243, 245-247, 250-256, 258, 262, 267, 295-301, 335]. As all of the statistically significant ($p<0.05$) ORs were in the same

direction, all designated risk alleles are identical to those indicated in the initial assessment. For novel SNPs and SNPs with no prior statistically significant findings, I designated the HapMap CEU minor variant as the risk variant.

I assumed the risk allele was the same for African Americans as it was for whites unless the majority of reported statistically significant associations were contradictory to those seen in whites. Two SNPs met these criteria (rs3803662 on *TNRC9/TOX3* and rs1045485 on *CASP8*) [11, 12, 169, 214-216].

All models were stratified by race and adjusted for proportion of African ancestry and age at diagnosis or selection. I centered age at 50 years and ancestry at the race-specific means. An offset term was included to account for unequal sampling by race and age group. As these ORs were generated using maximum-likelihood estimation, I will herein refer to these frequentist estimates as the MLE ORs. GWAS-identified SNPs were considered successfully replicated if their entire 95% CI fell above the null, as were SNPs identified using the candidate gene meta-analysis. More formal homogeneity tests comparing my findings to the original GWAS studies or meta-analysis estimates were not appropriate, as these studies did not consistently report ORs from additive genetic models.

4.3.4 Bayesian analysis

Bayes's theorem states that the posterior probability distribution for the parameter of interest given the observed data, ƒ(β|D), is proportional to the likelihood of the observed data, L(β|D), multiplied by the prior probability distribution ƒ(β) [314, 315]. Here, the likelihood function is the same as the MLE likelihood, with the log OR of being a case given exposure X_j and covariates W modeled as: $\ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_j X_j + W\gamma$, where p is the probability of being a case. In this application, β_j is the estimated log OR of being a breast cancer case for

each copy of the risk allele at SNP j and γ is a vector of estimated log ORs for age and ancestry.

To incorporate the priors, I added a second stage to the model: $\beta_j = z_j\pi + \delta_j$, where z_j is a $j\times1$ matrix of 1's, π is the prior log OR, $\mathbf{T}$ is a $j\times j$ identity matrix and the δ_j are assumed to be normally distributed with mean 0 and variance $\tau^2\mathbf{T}$. The prior mean (π), variance ($\tau^2\mathbf{T}$), or both can either be estimated directly from the data or assigned by the investigator. I use the term 'full-Bayes' to indicate that priors were assigned independently of the data at hand, 'empirical Bayes' when both were estimated from the data, and 'semi-Bayes' when one prior was assigned and the other was estimated [14].

I conducted both a full-Bayes analysis, where I assigned priors for the mean and τ^2, and a semi-Bayes analysis, where I assigned a fixed τ^2 but used LD-block (i.e. haplotype) level OR estimates to inform π. I did not use empirical Bayes methods, as the near-zero τ^2 generated from this rich data set caused over-shrinkage of the SNP-specific effects [322].

To obtain Bayesian (i.e. full-Bayes) log OR estimates, I assigned a null-centered, lognormal prior with a mean of 0 and variance $\tau^2 \sim \Gamma^{-1}(3, 0.2)$ to each SNP, such that 95% of the prior mass for each SNP-breast cancer OR lay between 0.64 and 1.55 when τ^2 was equal to 0.05, the mode of the specified distribution. As discussed previously, this range likely includes the true value for any single SNP-breast cancer association. Each Bayesian model included exactly one SNP (j=1).

I also assigned null-centered, lognormal priors for both age and ancestry, giving both parameters prior variances of 0.68. These priors reflect my belief, with moderate uncertainty, that these mean-centered covariates are weakly associated with breast cancer. Because standard implementation of MCMC requires prior distributions for every parameter, I placed

a $\beta_0 \sim N(0,1000)$ prior on the intercept. In the absence of other information, this vague prior should generate posterior intercept estimates nearly identical to the MLE intercept parameter estimate. I assumed that all priors were independent.

Additionally, I used semi-Bayes analysis (i.e. hierarchical modeling) to integrate haplotype information [13, 326, 328, 333]. Specifically, I used the estimated joint effect of all of the SNPs in an LD block to inform the prior mean (π). If there was only one genotyped SNP in an LD block, π and β_j were identical. I assigned fixed prior variances of 0.05 for each SNP, but did not assign priors for the intercept, age, or ancestry.

The above grouping approach is valid as long as an exchangeability assumption is met. This assumption states that before evaluating the relationship between the exposures (SNPs) and the outcome (breast cancer), there was no reason to suspect that any one SNP in an LD block had a true log OR different from the others in that LD block. As none of the included SNPs are known causal variants and all effects are evaluated in terms of risk alleles, I believe this assumption is acceptable in this setting.

Further specification of the correlation structure within the LD block may offer additional increases in precision. Here, I explored two such methods, both of which required alternative **T** matrices. For the first variation, I modeled correlation as an exponential decay function of the base pair distance between any two SNPs (d_{ij}). This takes the form of $t_{ij} = \exp\left[-\left(\frac{d_{ij}}{1000}\right)\right]$ [328]. For the second variation, I modeled the correlation structure directly, with $t_{ij} = D'$, a measure of LD.

For Bayes methods I presented posterior geometric mean ORs (i.e., antilogs of posterior mean log ORs) and 95% posterior intervals (PIs). For the Bayesian analyses I ran 30,000 samples for each SNP model, discarding the first 1000 draws as a burn-in and

retaining every fifth draw. I inspected autocorrelation, trace and density plots to verify that all posterior estimates converged appropriately.

LD blocks were determined using methods proposed by Gabriel et al. [165] and conducted in Haploview (Haploview 4.2, Version 1.0, Broad Institute, Cambridge, MA, USA) [164]. Bayesian models were analyzed using PROC MCMC or PROC GLIMMIX (SAS v9.3, Cary, NC). Example code is provided in the appendix.

4.4 Results

As expected, age distributions were similar for cases and controls, regardless of race (Table 8). White cases and controls were 52 and 53 years old at time of selection, on average, and African American cases and controls were both 52 years old. Whites had approximately 7% African ancestry and African Americans had 77%. More detailed descriptions of the study population, without the exclusions for missing SNP data, can be found in previous Carolina Breast Cancer Study publications [282].

Table 9 shows the RAFs for white and African American cases and controls and HWE p-values for white and African American controls. Seven SNPs were not in HWE by the established criterion ($p<0.05$). I retained 6 of the 7 SNPs, as their clustering images indicated good differentiation with no overlap between genotypes and none failed HWE tests in both races. I excluded the seventh SNP, rs614367 (*MYEOV*), after observing disparate clusters for the homozygous rare genotype and finding evidence of allelic dropout.

MLE ORs and 95% CIs for general and additive models for whites and African Americans can be seen in Tables 10 and 11, respectively. Tables 12 and 13 show the MLE ORs and confidence limit ratios (CLRs) for the additive models, as well as the ORs and

posterior limit ratios (PLRs) for the Bayesian models. The CLRs and PLRs are displayed to facilitate comparisons of model precision.

Among study whites, 18 of the GWAS-identified SNPs successfully replicated according to both MLE CIs and Bayesian PIs. Notably, all the *FGFR2* SNPs had relatively strong, positive associations with breast cancer (ORs>1.15), as did both of the *MRPS30* SNPs, two of the *TNRC9/TOX3* SNPs (rs3803662 and rs4784227), rs889312 in *MAP3K1,* rs704010 in *ZMIZ1*, and rs2107425 in *H19*. rs909116 in *LSP1* replicated in MLE, but not Bayesian assessments. Three other *FGFR2* SNPs (rs3750817, rs11200014, and rs2162540) were strongly associated with breast cancer (OR>1.2) in study whites.

None of the SNPs selected from the candidate gene meta-analysis replicated, though several SNPs in *ATM* and *TP53* were strongly associated with disease (|ln OR| > 0.15). The original GWAS and meta-analysis ORs are provided in Table 2 for further reference.

The most extreme example of the difference between MLE and Bayesian estimates in study whites was for rs3104746 in *TNRC9/TOX3,* the SNP with the highest MLE OR (1.66, 95% CI: 1.10, 2.51). Here, the Bayesian estimate was closer to the null (OR=1.42, 95% PI: 0.97, 1.94) and more precise (MLE CLR=2.29 vs. Bayesian PLR=2.01).

Ten of the GWAS-identified SNPs successfully replicated in African Americans (Table 3). This included nine *FGFR2* SNPs (ORs >1.15) and rs2046210 in *ESR1*. Two other *TNRC9/TOX3* SNPs (rs3104746 and rs3112562) had ORs >1.25 via either analysis method. rs7696175 in *TLR1* replicated when MLE methods were used, but not when Bayesian methods were used.

Two other SNPs, rs2107425 (*H19*) and rs12443621 (*TNRC9/TOX3*), were statistically significant, but the SNPs were inversely associated with breast cancer and thus inconsistent

with the original reports. Some of the *ATM* and *ESR1* MLE ORs were relatively strong, but none of the SNPs from the candidate gene meta-analysis successfully replicated.

The 77 SNPs evaluable in whites separated into 55 unique LD blocks. The thirteen SNPs in *FGFR2* formed the largest block (Figure 23), followed by *ATM* (5 SNPs) (Figure 24) and *TP53* (3 SNPs) (Figure 25). LD blocks consisting of two highly correlated SNPs were also genotyped in *CASP8, CDKN2A/B* (Figure 26)*, TNRC9/TOX3* (Figure 27**)** and *COX11*. In African Americans the *FGFR2* SNPs formed three separate blocks of 5, 4, and 2 SNPs respectively (Figure 23), while the other 3 SNPs were not strongly linked. One of the unlinked SNPs, rs1896395, was not evaluable in whites (RAF=0%). *TNRC9/TOX3* contained two 2-SNP LD-blocks (Figure 27) and *TP53* (Figure 25) contained a single 3-SNP block. The two SNPs within *CASP8, CDKN2A* (Figure 26)*,* and *COX11* were again in high LD. None of the *ATM* SNPs were strongly correlated in African Americans (Figure 24). None of the *MRPS30*, *ESR1*, 8q24, or *LSP1* SNPs were in LD in either race. In total, the 73 SNPs evaluable in African Americans formed 58 LD blocks.

Semi-Bayes ORs and 95% PIs for the identity matrix-based hierarchical models are also presented in Tables 12 and 13. For both African Americans and whites, the hierarchical-based estimates had comparable or slightly lower precision than the MLE ORs, and consistently lower precision than the Bayesian estimates. According to hierarchically derived estimates, many of the SNPs in the larger LD blocks were not associated with breast cancer. For example, MLE and Bayesian ORs indicated that all thirteen of the highly correlated *FGFR2* SNPs were strongly associated with breast cancer among whites, while the hierarchical model generated mostly near-null estimates for these SNPs. Of the thirteen, rs2981579 had the strongest effect (OR=1.20, 95% PI: 0.85, 1.72). Similarly, MLE and

Bayesian models indicated that 10 of the 14 *FGFR2* SNPs were associated with breast cancer in African Americans, while hierarchical modeling produced elevated associations for one SNP in each of the three LD blocks (rs3750817: OR=1.38, 95% CI: 1.05, 1.83, rs2981578: OR=1.23, 95% CI: 0.99, 1.53, and rs2420946: OR=1.27, 95% CI: 0.96, 1.66) and for 2 of the 3 independent SNPs.

Direct comparisons of all 3 hierarchical models (identity, exponential decay by spatial distance, and correlation) for both races are shown in Figures 28 and 29. Further specification of the **T** covariance matrix did not substantially affect estimates or model precision, though PLRs were consistently smaller when the correlation structure was incorporated in some form **(**Tables 14 and 15). Haplotype ORs are also captured in these figures.

4.5 Discussion

As several of the SNPs analyzed here were previously reported for this study population [210], I will limit my discussion to novel findings. Among whites, statistically significant associations for rs10757278 in *CDKN2A/B* and rs3104746 in *TNRC9/TOX3* have never before been reported. I also corroborated previously observed associations between breast cancer and several well-validated GWAS-identified SNPs, including two *MRPS30* SNPs (rs4415084 and rs10941679) [169, 186, 190, 191, 208, 211, 220, 223, 228, 230], rs1562430 in 8q24 [186, 189, 190], and rs4784227 in *TNRC9/TOX3* [205, 262]. Additionally, I replicated several less-established GWAS-identified SNPs, including rs704010 in *ZMIZ1* [189], and rs3750817, rs10736303, rs1078806, and rs2981578 in *FGFR2* [194, 196]. The only *CASP8*, *ATM*, or *TP53* SNP to demonstrate a statistically significant association (rs9894986 in *TP53*) was not associated with disease in Zhang et al [20].

I am the first to report a statistically significant association for rs3750817 in *FGFR2* in African Americans. Previous investigations of rs2046210 (*ESR1*) in African Americans produced mostly near-null ORs [11, 12, 169, 214, 216, 242, 243], but several of the *FGFR2* and *TNRC9/TOX3* SNPs were associated with breast cancer in one or more prior investigations. This includes rs10736303 and rs2981578 (*FGFR2*) [12, 259] and rs3104746 and rs3112562 (*TNRC9/TOX3)* [231]. rs2981578 and rs3104746 were both positively associated with disease in a pooled analysis by Chen et al. [214], but approximately 20% of these participants were drawn from the CBCS population.

Because the hierarchical models included more parameters than the MLE or Bayesian models, they did not improve precision. However, my results do lend support to previous claims that these methods can help differentiate individual effects of highly correlated SNPs [326, 328, 330].

The closely linked *FGFR2* SNPs provide the best example of these potential benefits. Even though all thirteen evaluable SNPs were strongly associated with breast cancer in whites, it is implausible that each association is independent. A more likely explanation is that one or two causal variants within the LD block drive all of the observed associations. In this scenario, models that evaluate all the SNPs simultaneously in a single-level model will often produce unstable estimates. Rather than limiting analyses to haplotype effects, hierarchical models can effectively accommodate correlated exposures and provide stable SNP and haplotype-level ORs.

Unfortunately, even though the estimated OR for one *FGFR2* SNP, rs2981579, was notably higher than the rest, I did not have sufficient precision to reliably differentiate between its relatively weak OR and the null when so many SNPs were assessed

simultaneously. Analyses of SNPs in the other multi-SNP LD blocks were relatively more precise, but also largely inconclusive.

The hierarchical models performed better in African Americans, with rs3750817, rs2981578, rs2420946, and rs3104746 demonstrating notably stronger associations than the other SNP(s) in *FGFR2* block 1, *FGFR2* block 2, *FGFR2* block 3, and *TNRC9/TOX3* block 2, respectively. This performance improvement is likely attributable to the anticipated racial differences in LD block size. More explicit specifications of the incorporated covariance matrices had little impact on point estimates or precision in either racial group.

I believe my specifications of prior means and variances are reasonable. First, aside from mutations in *BRCA1/2,* it is unlikely that a single SNP has a large effect on breast cancer risk [318]. Second, as long as the covariate priors are appropriately specified, Bayes analysis with null-centered priors should bias effect estimates towards the null [17]. In this way, Bayesian analysis also reduces the probability of observing a false positive association. Lastly, I believe that correlated SNPs within an LD-block meet the criteria for exchangeability.

After accounting for the sampling mechanisms, the only observed discrepancy between study cases and other North Carolina cases was that African Americans with later stage disease were underrepresented in CBCS [274]. Therefore, ORs could be biased if the evaluated SNP is related to disease aggressiveness or medical care utilization. With regard to genotyping, whites were more likely to provide blood samples than African Americans, but blood donation status did not differ by stage of disease or other breast cancer risk factors. Case, race, age, and stage distributions were similar for genotyped and non-genotyped subjects.

The inclusion of *in situ* cases could bias estimates of SNPs associated with disease aggressiveness or progression. However, given the strong evidence that *in situ* tumors have similar risk profiles to invasive cases [4, 341], I chose to retain these cases and corresponding controls.

A major strength of this study is the population itself, which is a large, racially diverse, population-based sample with well-validated data. The inclusion of a large sample of African American women allowed me to investigate racial differences in genetic risk factors and, accordingly, provide information that may help pinpoint causal variants. Both age and case status should be accurately captured and I had information on both self-reported race and proportion of African ancestry. While factors such as allelic dropout cannot be ruled out for SNPs demonstrating Hardy-Weinberg disequilibrium, the quality control measures employed during the genotyping process should have reduced the number and impact of genotype misclassification. Nonetheless, results for SNPs that violated HWE should be interpreted with caution.

In this analysis I replicated several previously identified breast cancer susceptibility loci in whites and African Americans using both MLE and Bayesian methods. My findings offer additional evidence that the regions containing these replicated SNPs play an important role in breast cancer etiology. The SNPs that replicated in African Americans are especially instructive, as they narrow or refine the genomic region containing the causal variant. My use of Bayesian methods to incorporate external information about the likely effect size and correlation structure further augments the utility of these results. I believe that fine-mapping studies and smaller etiologically-driven investigations may derive even greater benefit from these better-informed, more stable approaches.

Table 8: Descriptive statistics for Whites and African Americans in the Carolina Breast Cancer Study (1993-2001)

	Cases		Controls	
	Whites (%)* N=1247	**African Americans (%)* N=766**	**Whites (%)* N=1105**	**African Americans (%)* N=681**
Age (years); mean	52.2 (11.7)	51.6 (11.7)	53.0 (11.2)	51.9 (11.3)
Proportion African	0.06 (0.07)	0.78 (0.13)	0.07 (0.07)	0.77 (0.14)
Postmenopausal; N	686 (69)	430 (58)	640 (37)	377 (41)
Stage of Disease; N				
In situ	356 (12)	89 (12)		
Stage I	405 (46)	224 (31)		
Stage II	346 (34)	309 (43)		
Stage III	72 (6)	79 (11)		
Stage IV	16 (1)	27 (4)		
missing	52	38		
Subtype; N (%)				
Luminal A	453 (64)	242 (49)		
Luminal B	82 (11)	38 (8)		
HER2+/ER-	59 (6)	39 (8)		
Basal-like	94 (11)	112 (22)		
Unclassified	60 (8)	71 (14)		
Missing	499	264		
ER status; N (%)				
Positive	697 (69)	334 (50)		
Negative	359 (31)	352 (50)		
Missing	191	80		
PR status; N (%)				
Positive	515 (64)	272 (44)		
Negative	312 (36)	356 (56)		
Missing	420	138		
HER2 status; N				
Positive	180 (18)	94 (16)		
Negative	746 (82)	514 (84)		
Missing	321	158		

*weighted by inverse sampling probability

Table 9: Risk allele frequencies (RAF) by race and case status, whites and African Americans in the Carolina Breast Cancer Study

		Whites				African Americans			
Gene	**Locus**	**Risk allele**	**RAF cases[a]**	**RAF controls[a]**	**HWE p-value**	**Risk allele**	**RAF cases[a]**	**RAF controls[a]**	**HWE p-value**
1p12	rs11249433	G	0.44	0.41	0.54	G	0.14	0.10	0.01
CASP8	rs1045485	G	0.88	0.87	0.63	C	0.06	0.05	0.74
CASP8	rs17468277	C	0.87	0.87	0.63	C	0.95	0.95	0.95
2q35	rs13387042	A	0.54	0.47	0.83	A	0.74	0.73	1.00
2p	rs4666451	G	0.60	0.63	0.30	G	0.78	0.77	0.12
SLC4A	rs4973768	T	0.48	0.42	0.22	T	0.36	0.40	0.05
4p	rs12505080	C	0.29	0.24	0.80	C	0.17	0.17	0.64
TLR1	rs7696175	T	0.45	0.45	0.91	T	0.08	0.06	0.54
MRPS30	rs4415084	T	0.43	0.42	0.18	T	0.64	0.58	0.70
MRPS30	rs10941679	G	0.29	0.30	0.76	G	0.19	0.19	0.17
5p12	rs981782	T	0.53	0.59	0.26	T	0.92	0.91	0.60
5q	rs30099	T	0.10	0.10	0.40	T	0.16	0.12	0.75
MAP3K	rs889312	C	0.32	0.34	0.85	C	0.33	0.36	0.08
ESR1	rs2046210	A	0.36	0.35	0.48	A	0.64	0.61	0.15
ESR1	rs851974	G	0.42	0.43	0.28	G	0.17	0.17	0.46
ESR1	rs2077647	A	0.51	0.49	0.64	A	0.52	0.51	0.16
ESR1	rs2234693	T	0.53	0.57	0.45	T	0.47	0.48	0.63
ESR1	rs1801132	C	0.76	0.76	0.43	C	0.90	0.88	0.36
ESR1	rs3020314	C	0.36	0.34	0.15	C	0.69	0.71	0.75
ESR1	rs3798577	T	0.52	0.53	0.43	T	0.57	0.54	0.27
ECHDC	rs2180341	G	0.25	0.27	0.55	G	0.31	0.33	0.83
RELN	rs17157903	T	0.13	0.12	0.06	T	0.11	0.10	0.08
8q24	rs13281615	G	0.43	0.42	0.17	G	0.44	0.43	0.58
8q24	rs1562430	T	0.59	0.57	0.78	T	0.54	0.53	0.61
CDKN2A/B	rs3731257	T	0.23	0.23	0.24	T	0.09	0.11	0.89
CDKN2A/B	rs3731249	A	0.03	0.03	0.90	A	0.01	0.00	0.95
CDKN2A/B	rs518394	G	0.44	0.48	0.17	G	0.08	0.08	0.06

CDKN2A/B	rs564398	G	0.42	0.47	0.29	G	0.08	0.08	0.02
CDKN2A/B	rs1011970	T	0.19	0.15	0.62	T	0.33	0.34	0.14
CDKN2A/B	rs10757278	A	0.54	0.55	0.18	A	0.81	0.82	0.77
CDKN2A/B	rs10811661	C	0.17	0.20	0.02	C	0.07	0.07	0.24
ANKRD	rs2380205	C	0.56	0.60	0.88	C	0.42	0.46	0.72
ZNF365	rs10995190	G	0.86	0.82	0.76	G	0.83	0.83	0.90
ZMIZ1	rs704010	T	0.43	0.42	0.93	T	0.11	0.08	0.82
FGFR2	rs1896395	A	0.00	0.00	0.96	A	0.20	0.20	0.04
FGFR2	rs3750817	C	0.65	0.60	0.16	C	0.91	0.88	0.83
FGFR2	rs10736303	G	0.54	0.49	0.19	G	0.87	0.84	0.75
FGFR2	rs11200014	A	0.46	0.41	0.65	A	0.20	0.21	0.75
FGFR2	rs2981579	T	0.47	0.41	0.51	T	0.62	0.61	0.10
FGFR2	rs1078806	G	0.45	0.41	0.53	G	0.21	0.21	0.99
FGFR2	rs2981578	C	0.54	0.49	0.09	C	0.87	0.84	0.45
FGFR2	rs1219648	G	0.45	0.39	0.35	G	0.44	0.41	0.57
FGFR2	rs2912774	A	0.45	0.40	0.26	A	0.59	0.55	0.07
FGFR2	rs2936870	T	0.45	0.40	0.25	T	0.60	0.56	0.14
FGFR2	rs2420946	T	0.44	0.39	0.21	T	0.54	0.52	0.03
FGFR2	rs2162540	G	0.44	0.39	0.28	G	0.54	0.52	0.41
FGFR2	rs2981582	T	0.44	0.39	0.30	T	0.49	0.49	0.96
FGFR2	rs3135718	G	0.44	0.39	0.23	G	0.58	0.54	0.65
10q	rs10510126	C	0.89	0.89	0.38	C	0.89	0.90	0.21
ATM	rs1800054	G	0.02	0.02	0.34	G	0.00	0.00	0.94
ATM	rs4986761	C	0.02	0.01	0.68	C	0.00	0.00	0.98
ATM	rs1800056	C	0.02	0.01	0.67	C	0.00	0.00	0.95
ATM	rs1800057	G	0.03	0.02	0.90	G	0.01	0.01	0.91
ATM	rs1800058	T	0.02	0.02	0.06	T	0.01	0.01	0.91
ATM	rs1801516	A	0.15	0.14	0.17	A	0.03	0.02	0.48
ATM	rs3092992	C	0.06	0.04	0.13	C	0.01	0.01	0.77
ATM	rs664143	C	0.58	0.57	0.70	C	0.66	0.66	0.45
ATM	rs170548	G	0.31	0.37	0.88	G	0.09	0.12	0.07
ATM	rs3092993	A	0.15	0.14	0.19	A	0.03	0.02	0.48
LSP1	rs3817198	C	0.33	0.34	0.18	C	0.17	0.17	0.16
LSP1	rs909116	T	0.54	0.52	0.20	T	0.71	0.72	0.96

MYEOV	rs614367	T	0.18	0.11	0.05	T	0.13	0.15	0.33
H19	rs2107425	C	0.71	0.68	0.74	C	0.48	0.53	0.42
TNRC9/TOX3	rs8049149	T	0.00	0.00	0.98	T	0.02	0.02	0.32
TNRC9/TOX3	rs16951186	T	0.01	0.01	0.75	T	0.17	0.19	0.95
TNRC9/TOX3	rs8051542	T	0.46	0.44	0.43	T	0.35	0.30	0.12
TNRC9/TOX3	rs12443621	G	0.51	0.41	0.39	G	0.47	0.51	1.00
TNRC9/TOX3	rs3803662	T	0.32	0.24	0.73	C	0.48	0.46	0.65
TNRC9/TOX3	rs4784227	T	0.29	0.22	0.62	T	0.08	0.07	0.59
TNRC9/TOX3	rs3104746	A	0.03	0.02	0.48	A	0.26	0.18	0.87
TNRC9/TOX3	rs3112562	G	0.22	0.20	0.45	G	0.52	0.46	0.88
TNRC9/TOX3	rs9940048	A	0.26	0.24	0.50	A	0.31	0.30	0.64
TP53	rs9894946	G	0.82	0.84	0.48	G	0.95	0.95	0.25
TP53	rs1614984	T	0.41	0.39	0.22	T	0.40	0.40	0.03
TP53	rs4968187	T	0.00	0.00	0.93	T	0.01	0.00	0.92
TP53	rs12951053	C	0.07	0.06	0.47	C	0.11	0.11	0.09
TP53	rs17880604	C	0.02	0.01	0.21	C	0.00	0.00	0.95
TP53	rs1800372	G	0.02	0.02	0.54	G	0.00	0.00	0.98
TP53	rs2909430	G	0.15	0.13	0.66	G	0.27	0.24	0.64
TP53	rs1042522	C	0.75	0.77	0.64	C	0.39	0.43	0.77
TP53	rs8079544	C	0.95	0.95	1.00	C	0.89	0.89	0.83
COX11	rs7222197	G	0.71	0.75	0.60	G	0.66	0.65	0.70
COX11	rs6504950	G	0.71	0.75	0.59	G	0.67	0.65	0.66

Table 10: SNP genotype distributions and associations with incident breast cancer for White women in the Carolina Breast Cancer Study (1993-2000)

Gene	SNP	genotype	Cases (%)* n=1247	Controls (%)* n=1105	General model: OR (95% CI)†	Additive model: OR (95% CI)†
1p12	rs11249433	A/A	382 (31.7)	361 (36.8)	1.00	1.00
		A/G	601 (47.7)	547 (43.8)	0.99 (0.81-1.21)	1.09 (0.96-1.24)
		G/G	251 (20.6)	192 (19.4)	1.22 (0.95-1.58)	P-trend= 0.2
		missing	13	5		
CASP8	rs1045485	C/C	20 (2.0)	23 (1.6)	1.00	1.00
		C/G	264 (21.0)	257 (23.2)	1.21 (0.63-2.32)	1.13 (0.94-1.35)
		G/G	919 (77.0)	809 (75.2)	1.34 (0.71-2.54)	P-trend= 0.2
		missing	44	16		
CASP8	rs17468277	T/T	20 (2.0)	23 (1.6)	1.00	1.00
		C/T	266 (21.2)	257 (23.2)	1.22 (0.63-2.34)	1.12 (0.93-1.34)
		C/C	918 (76.8)	809 (75.2)	1.34 (0.71-2.53)	P-trend= 0.2
		missing	43	16		
2q35	rs13387042	G/G	245 (21.2)	242 (29.7)	1.00	1.00
		A/G	590 (49.2)	546 (46.6)	1.04 (0.83-1.30)	1.08 (0.96-1.22)
		A/A	369 (29.6)	300 (23.7)	1.16 (0.91-1.49)	P-trend= 0.2
		missing	43	17		
2p	rs4666451	A/A	188 (15.7)	186 (14.3)	1.00	1.00
		A/G	582 (49.5)	508 (45.7)	1.10 (0.85-1.41)	1.02 (0.90-1.16)
		G/G	432 (34.8)	395 (39.9)	1.07 (0.82-1.38)	P-trend= 0.8
		missing	45	16		
SLC4A7	rs4973768	C/C	326 (27.4)	309 (33.6)	1.00	1.00
		C/T	636 (49.9)	526 (47.9)	1.17 (0.95-1.44)	1.04 (0.92-1.17)
		T/T	271 (22.8)	260 (18.5)	1.07 (0.83-1.37)	P-trend= 0.5
		missing	14	10		
4p	rs12505080	T/T	600 (49.4)	569 (58.5)	1.00	1.00
		C/T	520 (42.6)	421 (34.4)	1.15 (0.96-1.39)	1.06 (0.92-1.23)
		C/C	92 (7.9)	81 (7.1)	0.99 (0.70-1.40)	P-trend= 0.4
		missing	35	34		
TLR1	rs7696175	C/C	376 (32.0)	347 (28.1)	1.00	1.00

		C/T	565 (46.7)	537 (54.8)	0.97 (0.80-1.19)	1.09 (0.96-1.23)
		T/T	262 (21.3)	205 (17.1)	1.22 (0.95-1.56)	P-trend= 0.2
		missing	44	16		
		C/C	391 (32.5)	410 (34.9)	1.00	1.00
MRPS30	rs4415084	C/T	631 (49.9)	535 (46.4)	1.22 (1.00-1.48)	1.23 (1.08-1.40)
		T/T	207 (17.7)	147 (18.7)	1.52 (1.16-2.00)	P-trend= 0.002
		missing	18	13		
		A/A	634 (50.7)	608 (51.9)	1.00	1.00
MRPS30	rs10941679	A/G	507 (41.0)	417 (35.5)	1.20 (1.00-1.44)	1.18 (1.03-1.36)
		G/G	98 (8.3)	68 (12.5)	1.37 (0.96-1.95)	P-trend= 0.02
		missing	8	12		
		G/G	254 (22.5)	224 (15.7)	1.00	1.00
5p12	rs981782	G/T	611 (49.4)	560 (50.7)	0.92 (0.73-1.15)	0.98 (0.86-1.11)
		T/T	339 (28.2)	305 (33.5)	0.95 (0.74-1.22)	P-trend= 0.7
		missing	43	16		
		C/C	975 (81.3)	896 (81.4)	1.00	1.00
5q	rs30099	C/T	219 (18.0)	181 (17.8)	1.08 (0.86-1.36)	1.04 (0.85-1.28)
		T/T	10 (0.8)	12 (0.8)	0.83 (0.34-2.05)	P-trend= 0.7
		missing	43	16		
		A/A	559 (47.4)	551 (45.9)	1.00	1.00
MAP3K1	rs889312	A/C	512 (41.4)	449 (39.4)	1.09 (0.91-1.31)	1.19 (1.04-1.35)
		C/C	132 (11.1)	89 (14.8)	1.56 (1.15-2.13)	P-trend= 0.01
		missing	44	16		
		G/G	507 (39.4)	480 (40.9)	1.00	1.00
ESR1	rs2046210	A/G	581 (48.3)	482 (47.7)	1.19 (0.99-1.44)	1.09 (0.96-1.24)
		A/A	144 (12.3)	133 (11.4)	1.10 (0.82-1.46)	P-trend= 0.2
		missing	15	10		
		A/A	404 (32.5)	338 (29.6)	1.00	1.00
ESR1	rs851974	A/G	611 (50.6)	557 (54.1)	0.93 (0.76-1.13)	0.91 (0.80-1.03)
		G/G	217 (16.9)	201 (16.3)	0.82 (0.63-1.06)	P-trend= 0.1
		missing	15	9		
		A/A	277 (23.0)	239 (28.4)	1.00	1.00
ESR1	rs2077647	A/G	602 (51.1)	550 (45.5)	0.98 (0.78-1.22)	0.97 (0.86-1.10)
		A/A	325 (25.9)	299 (26.2)	0.94 (0.74-1.21)	P-trend= 0.6
		missing	43	17		

		C/C	264 (22.6)	214 (18.3)	1.00	1.00
ESR1	rs2234693	C/T	579 (48.3)	551 (50.2)	0.85 (0.67-1.06)	0.95 (0.84-1.07)
		T/T	361 (29.1)	323 (31.5)	0.88 (0.69-1.13)	P-trend= 0.4
		missing	43	17		
		G/G	78 (6.9)	51 (3.8)	1.00	1.00
ESR1	rs1801132	C/G	423 (35.0)	390 (39.6)	0.66 (0.45-0.99)	0.92 (0.80-1.06)
		C/C	703 (58.1)	648 (56.6)	0.68 (0.46-1.00)	P-trend= 0.3
		missing	43	16		
		T/T	519 (42.6)	460 (40.3)	1.00	1.00
ESR1	rs3020314	C/T	512 (41.8)	512 (51.3)	0.81 (0.67-0.97)	1.05 (0.92-1.19)
		C/C	173 (15.6)	117 (8.5)	1.37 (1.03-1.81)	P-trend= 0.5
		missing	43	16		
		C/C	271 (22.3)	264 (21.4)	1.00	1.00
ESR1	rs3798577	C/T	598 (50.6)	531 (51.1)	1.12 (0.90-1.39)	1.03 (0.91-1.17)
		T/T	334 (27.0)	294 (27.5)	1.07 (0.84-1.37)	P-trend= 0.6
		missing	44	16		
		A/A	698 (56.1)	642 (49.8)	1.00	1.00
ECHDC1	rs2180341	A/G	462 (37.8)	399 (45.4)	1.03 (0.86-1.24)	1.04 (0.90-1.20)
		G/G	67 (6.0)	56 (4.8)	1.10 (0.74-1.63)	P-trend= 0.6
		missing	20	8		
		C/C	924 (76.1)	806 (77.8)	1.00	1.00
RELN	rs17157903	C/T	252 (21.9)	252 (20.0)	0.88 (0.71-1.08)	0.87 (0.73-1.04)
		T/T	27 (2.0)	30 (2.2)	0.76 (0.43-1.33)	P-trend= 0.1
		missing	44	17		
		A/A	383 (32.0)	398 (32.0)	1.00	1.00
8q24	rs13281615	A/G	594 (49.4)	502 (51.4)	1.19 (0.98-1.45)	1.11 (0.98-1.26)
		G/G	221 (18.5)	188 (16.6)	1.21 (0.94-1.56)	P-trend= 0.1
		missing	49	17		
		C/C	192 (15.8)	204 (15.7)	1.00	1.00
8q24	rs1562430	C/T	602 (49.4)	543 (54.8)	1.14 (0.89-1.46)	1.13 (0.99-1.28)
		T/T	435 (34.8)	349 (29.5)	1.28 (0.98-1.65)	P-trend= 0.1
		missing	18	9		
		C/C	710 (60.3)	621 (61.2)	1.00	1.00
CDKN2A/B	rs3731257	C/T	412 (33.3)	393 (32.4)	0.90 (0.75-1.09)	0.93 (0.81-1.07)
		T/T	82 (6.4)	75 (6.4)	0.92 (0.65-1.31)	P-trend= 0.3

		missing	43	16		
		G/G	1134 (94.5)	1020 (94.1)	1.00	1.00
CDKN2A/B	rs3731249[c]	A/G or A/A	70 (5.5)	69 (5.9)	0.00 (0.00-0.00)	0.90 (0.63-1.28)
		missing	43	16		
		C/C	402 (32.3)	371 (30.4)	1.00	1.00
CDKN2A/B	rs518394	C/G	572 (46.6)	510 (42.3)	1.04 (0.85-1.26)	1.03 (0.91-1.16)
		G/G	229 (21.1)	208 (27.3)	1.05 (0.82-1.35)	P-trend= 0.7
		missing	44	16		
		A/A	429 (34.8)	395 (32.1)	1.00	1.00
CDKN2A/B	rs564398	G/A	566 (45.7)	507 (42.4)	1.02 (0.84-1.24)	1.04 (0.92-1.17)
		G/G	207 (19.5)	186 (25.5)	1.08 (0.84-1.40)	P-trend= 0.6
		missing	45	17		
		G/G	832 (66.6)	752 (72.3)	1.00	1.00
CDKN2A/B	rs1011970	G/T	344 (29.4)	299 (25.4)	1.07 (0.88-1.30)	1.13 (0.96-1.33)
		T/T	47 (4.1)	33 (2.2)	1.51 (0.92-2.45)	P-trend= 0.1
		missing	24	21		
		G/G	268 (20.9)	260 (20.2)	1.00	1.00
CDKN2A/B	rs10757278	A/G	604 (50.1)	567 (49.6)	0.98 (0.79-1.22)	1.17 (1.04-1.33)
		A/A	351 (29.0)	263 (30.1)	1.36 (1.06-1.74)	P-trend= 0.01
		missing	24	15		
		T/T	844 (69.7)	773 (64.9)	1.00	1.00
CDKN2A/B	rs10811661	C/T	325 (27.5)	276 (30.7)	1.07 (0.88-1.31)	1.00 (0.85-1.18)
		C/C	35 (2.8)	40 (4.4)	0.79 (0.49-1.30)	P-trend= 0.99
		missing	43	16		
		T/T	247 (20.3)	203 (16.6)	1.00	1.00
ANKRD16	rs2380205	C/T	579 (46.5)	532 (46.0)	0.92 (0.73-1.16)	1.01 (0.89-1.14)
		C/C	394 (33.2)	355 (37.3)	1.00 (0.78-1.29)	P-trend= 0.9
		missing	27	15		
		A/A	24 (2.0)	21 (1.7)	1.00	1.00
ZNF365	rs10995190	A/G	303 (24.6)	270 (31.6)	0.82 (0.43-1.58)	1.00 (0.84-1.20)
		G/G	909 (73.4)	804 (66.7)	0.85 (0.45-1.61)	P-trend= 1
		missing	11	10		
		C/C	406 (31.9)	426 (36.6)	1.00	1.00
ZMIZ1	rs704010	C/T	601 (50.7)	510 (42.1)	1.30 (1.07-1.57)	1.24 (1.09-1.41)
		T/T	211 (17.3)	151 (21.4)	1.50 (1.15-1.96)	P-trend= 0.001

		missing	29	18		
		T/T	142 (11.1)	180 (14.5)	1.00	1.00
FGFR2	rs3750817	C/T	576 (47.5)	500 (51.3)	1.56 (1.19-2.03)	1.24 (1.09-1.40)
		C/C	513 (41.4)	414 (34.1)	1.68 (1.28-2.20)	P-trend= 0.001
		missing	16	11		
		A/A	258 (20.5)	313 (23.5)	1.00	1.00
FGFR2	rs10736303	A/G	627 (51.2)	523 (54.9)	1.46 (1.18-1.82)	1.33 (1.17-1.50)
		G/G	351 (28.3)	256 (21.6)	1.76 (1.37-2.26)	P-trend= 1×10^{-5}
		missing	11	13		
		G/G	358 (29.2)	406 (32.8)	1.00	1.00
FGFR2	rs11200014	A/G	593 (50.0)	512 (52.2)	1.31 (1.08-1.60)	1.30 (1.15-1.48)
		A/A	253 (20.9)	171 (14.9)	1.70 (1.31-2.19)	P-trend= 3×10^{-5}
		missing	43	16		
		C/C	346 (28.2)	401 (32.4)	1.00	1.00
FGFR2	rs2981579	C/T	594 (50.1)	511 (52.3)	1.36 (1.11-1.65)	1.33 (1.18-1.51)
		T/T	264 (21.7)	177 (15.3)	1.77 (1.38-2.28)	P-trend= 1×10^{-5}
		missing	43	16		
		A/A	376 (30.0)	411 (33.0)	1.00	1.00
FGFR2	rs1078806	A/G	596 (49.1)	514 (51.9)	1.29 (1.06-1.57)	1.29 (1.14-1.46)
		G/G	258 (20.9)	174 (15.1)	1.67 (1.29-2.16)	P-trend= 1×10^{-4}
		missing	17	6		
		T/T	261 (20.6)	321 (24.4)	1.00	1.00
FGFR2	rs2981578	C/T	625 (51.0)	520 (53.4)	1.49 (1.20-1.85)	1.32 (1.17-1.50)
		C/C	354 (28.3)	259 (22.1)	1.76 (1.37-2.25)	P-trend= 1×10^{-5}
		missing	7	5		
		A/A	374 (30.9)	425 (36.2)	1.00	1.00
FGFR2	rs1219648	A/G	588 (49.0)	499 (49.1)	1.34 (1.11-1.63)	1.31 (1.16-1.48)
		G/G	241 (20.0)	165 (14.7)	1.69 (1.31-2.19)	P-trend= 2×10^{-5}
		missing	44	16		
		C/C	366 (30.4)	420 (35.9)	1.00	1.00
FGFR2	rs2912774	A/C	594 (49.7)	497 (48.9)	1.37 (1.13-1.67)	1.30 (1.15-1.47)
		A/A	242 (19.9)	170 (15.2)	1.65 (1.28-2.13)	P-trend= 4×10^{-5}
		missing	45	18		
FGFR2	rs2936870	C/C	366 (30.3)	420 (35.8)	1.00	1.00
		C/T	594 (49.5)	498 (49.0)	1.37 (1.13-1.66)	1.30 (1.15-1.47)

		T/T	244 (20.1)	171 (15.2)	1.66 (1.29-2.14)	P-trend= 3 x 10^{-5}
		missing	43	16		
		C/C	380 (31.6)	433 (36.7)	1.00	1.00
FGFR2	rs2420946	C/T	587 (49.0)	490 (48.8)	1.37 (1.13-1.66)	1.30 (1.15-1.48)
		T/T	235 (19.4)	163 (14.6)	1.66 (1.28-2.15)	P-trend= 3 x 10^{-5}
		missing	45	19		
		A/A	385 (32.0)	436 (36.8)	1.00	1.00
FGFR2	rs2162540	A/G	583 (48.5)	493 (49.1)	1.34 (1.11-1.63)	1.31 (1.15-1.48)
		G/G	234 (19.4)	160 (14.2)	1.69 (1.30-2.18)	P-trend= 3 x 10^{-5}
		missing	45	16		
		C/C	385 (31.9)	437 (36.8)	1.00	1.00
FGFR2	rs2981582	C/T	587 (48.9)	493 (49.1)	1.35 (1.12-1.64)	1.30 (1.15-1.48)
		T/T	232 (19.2)	159 (14.1)	1.67 (1.29-2.17)	P-trend= 3 x 10^{-5}
		missing	43	16		
		A/A	376 (31.2)	432 (36.5)	1.00	1.00
FGFR2	rs3135718	A/G	592 (49.2)	493 (48.8)	1.38 (1.14-1.67)	1.31 (1.16-1.48)
		G/G	236 (19.6)	164 (14.6)	1.68 (1.30-2.17)	P-trend= 2 x 10^{-5}
		missing	43	16		
		T/T	16 (1.5)	13 (0.9)	1.00	1.00
10q	rs10510126	C/T	231 (19.4)	239 (19.4)	0.70 (0.31-1.55)	1.11 (0.91-1.35)
		C/C	957 (79.1)	837 (79.7)	0.82 (0.38-1.79)	P-trend= 0.3
		missing	43	16		
		C/C	1160 (96.3)	1048 (96.9)	1.00	1.00
ATM	rs1800054[c]	C/G or G/G	43 (3.7)	41 (3.1)	1.04 (0.66-1.64)	1.01 (0.65-1.58)
		missing	44	16		
		C/C	1130 (93.9)	1028 (95.4)	1.00	1.00
ATM	rs1800057[c]	C/G or G/G	74 (6.1)	61 (4.6)	1.10 (0.76-1.60)	1.09 (0.76-1.56)
		missing	43	16		
		C/C	1160 (96.2)	1040 (96.2)	1.00	1.00
ATM	rs1800058[c]	C/T or T/T	44 (3.8)	49 (3.8)	0.82 (0.53-1.28)	0.82 (0.54-1.25)
		missing	43	16		
		G/G	877 (71.4)	792 (72.8)	1.00	1.00
ATM	rs1801516	A/G	310 (27.2)	279 (26.1)	0.98 (0.80-1.19)	0.98 (0.82-1.17)
		A/A	17 (1.4)	17 (1.2)	0.98 (0.48-1.99)	P-trend= 0.8
		missing	43	17		

ATM	rs3092992[c]	A/A	1081 (89.1)	993 (91.7)	1.00	1.00
		A/C or C/C	123 (10.9)	96 (8.3)	1.17 (0.87-1.58)	1.19 (0.89-1.60)
		missing	43	16		
ATM	rs664143	T/T	213 (17.2)	203 (18.4)	1.00	1.00
		C/T	578 (50.0)	527 (43.3)	1.07 (0.84-1.36)	1.02 (0.90-1.15)
		C/C	412 (32.9)	359 (38.3)	1.05 (0.82-1.36)	P-trend= 0.8
		missing	44	16		
ATM	rs170548	T/T	538 (45.5)	499 (44.6)	1.00	1.00
		G/T	553 (46.6)	478 (36.9)	1.10 (0.91-1.32)	0.98 (0.86-1.12)
		G/G	113 (7.9)	112 (18.5)	0.83 (0.61-1.13)	P-trend= 0.7
		missing	43	16		
ATM	rs3092993	C/C	877 (71.4)	793 (72.9)	1.00	1.00
		A/C	310 (27.2)	278 (26.0)	0.98 (0.80-1.20)	0.98 (0.82-1.18)
		A/A	17 (1.4)	17 (1.2)	0.98 (0.48-1.99)	P-trend= 0.8
		missing	43	17		
LSP1	rs3817198	T/T	537 (45.3)	502 (40.7)	1.00	1.00
		C/T	529 (43.9)	488 (50.3)	1.01 (0.84-1.22)	1.08 (0.95-1.24)
		C/C	136 (10.8)	98 (9.0)	1.26 (0.93-1.71)	P-trend= 0.2
		missing	45	17		
	rs909116	C/C	257 (21.1)	251 (19.7)	1.00	1.00
		C/T	608 (50.5)	566 (56.0)	1.03 (0.83-1.29)	1.14 (1.01-1.30)
		T/T	358 (28.3)	273 (24.4)	1.30 (1.01-1.66)	P-trend= 0.04
		missing	24	15		
H19	rs2107425	T/T	101 (7.6)	111 (8.5)	1.00	1.00
		C/T	505 (42.7)	466 (46.7)	1.24 (0.90-1.70)	1.15 (1.00-1.31)
		C/C	593 (49.7)	512 (44.7)	1.38 (1.01-1.89)	P-trend= 0.04
		missing	48	16		
TNRC9/ TOX3	rs8051542	C/C	372 (31.8)	350 (30.4)	1.00	1.00
		C/T	580 (44.9)	546 (51.9)	0.95 (0.78-1.15)	1.12 (0.99-1.26)
		T/T	252 (23.3)	193 (17.7)	1.30 (1.01-1.67)	P-trend= 0.1
		missing	43	16		
TNRC9/ TOX3	rs12443621	A/A	306 (25.0)	296 (33.9)	1.00	1.00
		A/G	563 (48.2)	557 (49.2)	1.01 (0.82-1.25)	1.17 (1.04-1.33)
		G/G	335 (26.8)	236 (16.9)	1.38 (1.08-1.77)	P-trend= 0.01
		missing	43	16		

TNRC9/ TOX3	rs3803662	C/C	572 (47.5)	579 (58.4)	1.00	1.00
		C/T	502 (41.2)	427 (35.1)	1.22 (1.01-1.46)	1.27 (1.11-1.46)
		T/T	130 (11.4)	83 (6.5)	1.73 (1.26-2.37)	P-trend= 4 x 10^{-3}
		missing	43	16		
TNRC9/ TOX3	rs4784227	C/C	637 (52.0)	631 (61.6)	1.00	1.00
		C/T	486 (38.4)	398 (32.9)	1.20 (1.00-1.44)	1.26 (1.09-1.44)
		T/T	103 (9.6)	68 (5.5)	1.70 (1.21-2.41)	P-trend= 0.001
		missing	21	8		
TNRC9/ TOX3	rs3104746[c]	T/T	1160 (94.6)	1039 (96.9)	1.00	1.00
		A/T or A/A	69 (5.4)	46 (3.1)	1.66 (1.10-2.51)	1.66 (1.10-2.51)
		missing	18	16		
TNRC9/ TOX3	rs3112562	C/C	741 (61.0)	644 (64.0)	1.00	1.00
		C/G	416 (33.9)	398 (32.5)	0.90 (0.75-1.08)	0.99 (0.86-1.15)
		G/G	70 (5.1)	54 (3.5)	1.22 (0.82-1.81)	P-trend= 0.9
		missing	20	9		
TNRC9/ TOX3	rs9940048	G/G	656 (54.0)	602 (57.1)	1.00	1.00
		A/G	483 (40.5)	421 (37.6)	1.04 (0.87-1.25)	1.03 (0.89-1.19)
		A/A	65 (5.5)	66 (5.3)	1.02 (0.70-1.49)	P-trend= 0.7
		missing	43	16		
TP53	rs9894946	A/A	44 (3.5)	31 (2.3)	1.00	1.00
		A/G	346 (32.0)	286 (24.3)	0.76 (0.46-1.27)	0.84 (0.72-0.99)
		G/G	814 (64.5)	770 (73.4)	0.66 (0.40-1.07)	P-trend= 0.04
		missing	43	18		
TP53	rs1614984	C/C	418 (35.1)	354 (36.4)	1.00	1.00
		C/T	564 (47.0)	551 (49.0)	0.84 (0.69-1.02)	1.03 (0.91-1.17)
		T/T	221 (17.9)	184 (14.6)	1.15 (0.89-1.48)	P-trend= 0.6
		missing	44	16		
TP53	rs12951053	A/A	1047 (86.9)	953 (87.7)	1.00	1.00
		A/C	146 (12.5)	133 (12.2)	1.01 (0.77-1.32)	1.09 (0.85-1.39)
		C/C	10 (0.5)	3 (0.2)	2.71 (0.71-10.34)	P-trend= 0.5
		missing	44	16		
TP53	rs17880604[c]	G/G	1168 (97.0)	1053 (97.1)	1.00	1.00
		C/G or C/C	36 (3.0)	36 (2.9)	0.82 (0.49-1.35)	0.82 (0.51-1.33)
		missing	43	16		
TP53	rs1800372[c]	A/A	1165 (97.1)	1044 (97.0)	1.00	1.00

		A/G or G/G	38 (2.9)	40 (3.0)	0.84 (0.52-1.35)	0.88 (0.55-1.40)
		missing	44	21		
		A/A	887 (72.4)	814 (76.1)	1.00	1.00
TP53	rs2909430	A/G	291 (25.5)	257 (22.6)	1.04 (0.85-1.28)	1.11 (0.93-1.33)
		G/G	26 (2.2)	18 (1.3)	1.67 (0.89-3.14)	P-trend= 0.2
		missing	43	16		
		G/G	90 (7.8)	73 (6.5)	1.00	1.00
TP53	rs1042522	C/G	445 (35.1)	406 (33.7)	0.98 (0.68-1.41)	0.98 (0.85-1.13)
		C/C	680 (57.2)	608 (59.8)	0.96 (0.68-1.37)	P-trend= 0.8
		missing	32	18		
		T/T or C/T	123 (9.0)	128 (9.7)	1.00	1.00
TP53	rs8079544[d]	CC	1081 (91.0)	961 (90.3)	1.21 (0.92-1.60)	1.24 (0.95-1.63)
		missing	43	16		
		A/A	95 (7.4)	85 (6.1)	1.00	1.00
		A/G	525 (43.2)	454 (37.2)	1.07 (0.76-1.51)	0.98 (0.85-1.12)
COX11	rs7222197	G/G	613 (49.4)	560 (56.6)	1.01 (0.72-1.42)	P-trend= 0.7
		missing	14	6		
		A/A	94 (7.4)	86 (6.2)	1.00	1.00
		A/G	522 (43.1)	457 (37.3)	1.08 (0.77-1.52)	0.98 (0.85-1.12)
COX11	rs6504950	G/G	614 (49.5)	560 (56.5)	1.02 (0.73-1.43)	P-trend= 0.8
		missing	17	2		

[a]weighted by inverse sampling probability

[b]adjusted for age at diagnosis (cases) or selection (controls) and proportion of African ancestry

[c]Assessed using dominant and additive model (MAF<5%)

[d]Assessed using recessive and additive model (MAF>95%)

Table 11: SNP genotype distributions and associations with incident breast cancer for African American women in the Carolina Breast Cancer Study (1993-2000)

Gene	SNP	genotype	Cases (%)* n=742	Controls (%)* n=681	General model: OR (95% CI)†	Additive model: OR (95% CI)†
1p12	rs11249433	A/A	565 (74.0)	536 (81.4)	1.00	1.00
		A/G	190 (25.0)	121 (16.9)	1.51 (1.15-1.99)	1.26 (0.99-1.60)
		G/G	8 (1.0)	16 (1.7)	0.58 (0.24-1.44)	P-trend= 0.1
		missing	3	8		
CASP8	rs1045485	G/G	660 (89.0)	579 (89.2)	1.00	1.00
		C/G	80 (10.7)	77 (10.6)	0.96 (0.68-1.35)	0.93 (0.67-1.29)
		C/C	2 (0.3)	2 (0.2)	0.45 (0.05-3.74)	P-trend= 0.7
		missing	24	23		
CASP8	rs17468277	T/T	1 (0.1)	2 (0.2)	1.00	1.00
		C/T	74 (9.8)	70 (10.0)	3.97 (0.29-54.81)	1.09 (0.78-1.54)
		C/C	667 (90.0)	586 (89.9)	4.14 (0.31-56.07)	P-trend= 0.6
		missing	24	23		
2q35	rs13387042	G/G	47 (6.2)	45 (8.4)	1.00	1.00
		A/G	292 (39.3)	254 (37.0)	1.10 (0.70-1.75)	1.02 (0.86-1.22)
		A/A	403 (54.4)	358 (54.6)	1.10 (0.70-1.73)	P-trend= 0.8
		missing	24	24		
2p	rs4666451	A/A	26 (3.5)	43 (8.6)	1.00	1.00
		A/G	267 (36.1)	221 (29.6)	2.16 (1.25-3.72)	1.15 (0.96-1.39)
		G/G	449 (60.5)	394 (61.9)	2.10 (1.23-3.57)	P-trend= 0.1
		missing	24	23		
SLC4A7	rs4973768	C/C	322 (42.6)	250 (36.2)	1.00	1.00
		C/T	334 (43.8)	340 (46.6)	0.77 (0.61-0.97)	0.90 (0.77-1.06)
		T/T	103 (13.7)	83 (17.2)	0.91 (0.65-1.29)	P-trend= 0.2
		missing	7	8		
4p	rs12505080	T/T	514 (68.1)	461 (68.2)	1.00	1.00
		C/T	218 (28.9)	184 (30.0)	1.09 (0.86-1.39)	1.09 (0.88-1.34)
		C/C	22 (3.0)	16 (1.7)	1.17 (0.58-2.37)	P-trend= 0.4
		missing	12	20		
TLR1	rs7696175	C/C	630 (85.0)	573 (88.0)	1.00	1.00
		C/T	99 (13.3)	81 (11.6)	1.22 (0.87-1.71)	1.39 (1.04-1.86)

		T/T	13 (1.7)	4 (0.4)	4.11 (1.27-13.24)	P-trend= 0.03
		missing	24	23		
		C/C	97 (12.8)	100 (19.7)	1.00	1.00
MRPS30	rs4415084	C/T	343 (45.9)	312 (44.0)	1.33 (0.95-1.86)	1.13 (0.97-1.33)
		T/T	310 (41.3)	259 (36.3)	1.37 (0.97-1.94)	P-trend= 0.1
		missing	16	10		
		A/A	497 (65.5)	436 (65.5)	1.00	1.00
MRPS30	rs10941679	A/G	234 (30.8)	219 (31.7)	0.95 (0.75-1.20)	1.00 (0.82-1.22)
		G/G	28 (3.7)	19 (2.8)	1.24 (0.67-2.30)	P-trend= 1
		missing	7	7		
		G/G	11 (1.4)	7 (1.0)	1.00	1.00
5p12	rs981782	G/T	105 (14.1)	110 (15.0)	0.78 (0.27-2.26)	1.11 (0.84-1.46)
		T/T	625 (84.5)	540 (84.0)	0.91 (0.32-2.59)	P-trend= 0.5
		missing	25	24		
		C/C	527 (70.9)	493 (76.2)	1.00	1.00
5q	rs30099	C/T	199 (26.9)	152 (22.7)	1.26 (0.98-1.62)	1.22 (0.98-1.52)
		T/T	16 (2.2)	13 (1.1)	1.28 (0.59-2.74)	P-trend= 0.1
		missing	24	23		
		A/A	316 (42.7)	281 (37.9)	1.00	1.00
MAP3K1	rs889312	A/C	364 (49.2)	313 (53.1)	1.00 (0.79-1.25)	0.95 (0.80-1.13)
		C/C	62 (8.2)	64 (9.0)	0.84 (0.57-1.25)	P-trend= 0.5
		missing	24	23		
		G/G	96 (12.7)	99 (13.4)	1.00	1.00
ESR1	rs2046210	A/G	350 (46.1)	340 (50.9)	1.06 (0.76-1.48)	1.22 (1.04-1.43)
		A/A	312 (41.2)	231 (35.6)	1.40 (1.00-1.98)	P-trend= 0.02
		missing	8	11		
		A/G	523 (69.3)	444 (68.0)	1.00	1.00
ESR1	rs851974	A/G	208 (27.5)	208 (29.4)	0.87 (0.68-1.11)	0.93 (0.76-1.14)
		G/G	25 (3.2)	20 (2.6)	1.06 (0.57-2.00)	P-trend= 0.5
		missing	10	9		
		A/A	170 (22.9)	155 (24.6)	1.00	1.00
ESR1	rs2077647	A/G	372 (50.2)	346 (48.2)	0.97 (0.74-1.27)	1.07 (0.92-1.25)
		A/A	200 (26.9)	155 (27.2)	1.14 (0.84-1.56)	P-trend= 0.4
		missing	24	25		
ESR1	rs2234693	C/C	198 (26.8)	182 (26.1)	1.00	1.00

		C/T	396 (53.3)	334 (52.2)	1.09 (0.84-1.41)	0.96 (0.82-1.13)
		T/T	148 (19.9)	142 (21.7)	0.91 (0.66-1.25)	P-trend= 0.6
		missing	24	23		
		G/G	8 (1.0)	7 (0.9)	1.00	1.00
ESR1	rs1801132	C/G	134 (18.4)	144 (23.0)	0.76 (0.26-2.19)	1.19 (0.93-1.52)
		C/C	599 (80.6)	507 (76.1)	0.95 (0.33-2.68)	P-trend= 0.2
		missing	25	23		
		T/T	63 (8.5)	65 (7.8)	1.00	1.00
ESR1	rs3020314	C/T	339 (45.5)	278 (43.2)	1.20 (0.81-1.77)	1.00 (0.84-1.19)
		C/C	339 (46.0)	315 (49.0)	1.11 (0.75-1.65)	P-trend= 1
		missing	25	23		
		C/C	140 (19.0)	130 (22.7)	1.00	1.00
ESR1	rs3798577	C/T	367 (49.0)	309 (46.6)	1.18 (0.88-1.58)	1.02 (0.88-1.19)
		T/T	235 (32.0)	219 (30.7)	1.08 (0.79-1.47)	P-trend= 0.8
		missing	24	23		
		A/A	354 (47.7)	317 (42.5)	1.00	1.00
ECHDC1	rs2180341	A/G	321 (42.6)	288 (48.8)	0.94 (0.75-1.19)	0.98 (0.83-1.15)
		G/G	75 (9.7)	68 (8.7)	1.00 (0.68-1.46)	P-trend= 0.8
		missing	16	8		
		C/C	599 (80.7)	523 (79.4)	1.00	1.00
RELN	rs17157903	C/T	129 (17.4)	132 (20.3)	0.87 (0.66-1.16)	1.07 (0.83-1.37)
		T/T	14 (1.9)	3 (0.3)	5.21 (1.48-18.29)	P-trend= 0.6
		missing	24	23		
		A/A	219 (29.2)	204 (33.1)	1.00	1.00
8q24	rs13281615	A/G	387 (53.1)	331 (48.7)	1.14 (0.89-1.46)	1.00 (0.86-1.18)
		G/G	130 (17.7)	123 (18.2)	0.97 (0.70-1.34)	P-trend= 1
		missing	30	23		
		C/C	158 (20.7)	147 (23.5)	1.00	1.00
8q24	rs1562430	C/T	375 (50.3)	327 (47.4)	1.11 (0.84-1.47)	1.00 (0.86-1.17)
		T/T	219 (29.1)	197 (29.1)	1.02 (0.75-1.39)	P-trend= 1
		missing	14	10		
		C/C	617 (83.3)	535 (79.2)	1.00	1.00
CDKN2A/B	rs3731257	C/T	119 (15.9)	117 (19.8)	0.87 (0.65-1.16)	0.88 (0.67-1.15)
		T/T	6 (0.8)	6 (1.1)	0.86 (0.26-2.82)	P-trend= 0.3
		missing	24	23		

CDKN2A/B	rs518394	C/C	623 (84.1)	559 (84.4)	1.00	1.00
		C/G	112 (14.9)	91 (14.4)	1.05 (0.77-1.45)	1.01 (0.76-1.35)
		G/G	7 (1.0)	8 (1.2)	0.81 (0.27-2.42)	P-trend= 0.9
		missing	24	23		
CDKN2A/B	rs564398	A/A	628 (84.7)	566 (85.5)	1.00	1.00
		G/A	109 (14.5)	84 (13.3)	1.11 (0.80-1.54)	1.01 (0.75-1.35)
		G/G	5 (0.7)	8 (1.2)	0.52 (0.16-1.74)	P-trend= 0.9
		missing	24	23		
CDKN2A/B	rs1011970	G/G	349 (46.6)	314 (47.6)	1.00	1.00
		G/T	307 (40.9)	277 (37.1)	0.92 (0.73-1.16)	0.95 (0.81-1.11)
		T/T	93 (12.5)	79 (15.3)	0.93 (0.65-1.32)	P-trend= 0.5
		missing	17	11		
CDKN2A/B	rs10757278	G/G	27 (3.7)	22 (3.3)	1.00	1.00
		A/G	241 (31.2)	192 (29.0)	0.96 (0.52-1.79)	0.91 (0.75-1.12)
		A/A	490 (65.1)	452 (67.7)	0.87 (0.47-1.59)	P-trend= 0.4
		missing	8	15		
ANKRD16	rs2380205	T/T	259 (34.1)	226 (29.9)	1.00	1.00
		C/T	360 (48.1)	329 (48.9)	0.95 (0.74-1.21)	0.97 (0.83-1.13)
		C/C	132 (17.8)	113 (21.2)	0.95 (0.69-1.31)	P-trend= 0.7
		missing	15	13		
ZNF365	rs10995190	A/A	28 (3.6)	22 (3.9)	1.00	1.00
		A/G	206 (27.0)	202 (26.6)	0.75 (0.41-1.38)	1.06 (0.87-1.29)
		G/G	526 (69.4)	449 (69.5)	0.86 (0.48-1.55)	P-trend= 0.6
		missing	6	8		
ZMIZ1	rs704010	C/C	601 (79.5)	532 (85.4)	1.00	1.00
		C/T	148 (19.4)	128 (13.7)	1.07 (0.80-1.42)	1.05 (0.80-1.36)
		T/T	8 (1.0)	7 (0.9)	0.94 (0.32-2.75)	P-trend= 0.7
		missing	9	14		
FGFR2	rs1896395	C/C	474 (63.9)	408 (61.4)	1.00	1.00
		A/C	235 (31.6)	231 (36.4)	0.87 (0.69-1.10)	1.01 (0.83-1.23)
		A/A	33 (4.5)	19 (2.2)	1.69 (0.94-3.07)	P-trend= 0.9
		missing	24	23		
FGFR2	rs3750817	T/T	6 (0.9)	11 (1.9)	1.00	1.00
		C/T	120 (15.9)	155 (20.7)	1.83 (0.59-5.73)	1.74 (1.34-2.26)
		C/C	635 (83.2)	506 (77.4)	3.16 (1.03-9.71)	P-trend= $3x10^{-5}$

		missing	5	9		
		A/A	15 (2.1)	20 (3.8)	1.00	1.00
FGFR2	rs10736303	A/G	168 (22.1)	186 (23.7)	1.29 (0.62-2.72)	1.39 (1.12-1.74)
		G/G	580 (75.8)	471 (72.5)	1.83 (0.89-3.77)	P-trend= 0.003
		missing	3	4		
		G/G	476 (64.3)	425 (61.4)	1.00	1.00
FGFR2	rs11200014	A/G	232 (31.2)	206 (35.1)	1.00 (0.79-1.27)	1.04 (0.86-1.26)
		A/A	34 (4.5)	27 (3.5)	1.23 (0.71-2.11)	P-trend= 0.7
		missing	24	23		
		C/C	110 (14.9)	129 (18.1)	1.00	1.00
FGFR2	rs2981579	C/T	341 (46.1)	301 (42.4)	1.39 (1.02-1.90)	1.22 (1.04-1.42)
		T/T	291 (39.0)	228 (39.5)	1.54 (1.12-2.13)	P-trend= 0.01
		missing	24	23		
		A/A	480 (63.5)	433 (61.3)	1.00	1.00
FGFR2	rs1078806	A/G	240 (31.8)	216 (35.6)	1.01 (0.80-1.28)	1.06 (0.88-1.29)
		G/G	36 (4.7)	27 (3.1)	1.30 (0.75-2.25)	P-trend= 0.5
		missing	10	5		
		T/T	15 (2.1)	22 (3.9)	1.00	1.00
FGFR2	rs2981578	C/T	165 (21.8)	184 (23.7)	1.42 (0.68-2.93)	1.42 (1.14-1.77)
		C/C	580 (76.1)	470 (72.4)	2.02 (0.99-4.09)	P-trend= 0.002
		missing	6	5		
		A/A	230 (31.2)	227 (32.9)	1.00	1.00
FGFR2	rs1219648	A/G	363 (48.8)	325 (52.0)	1.13 (0.89-1.45)	1.19 (1.02-1.39)
		G/G	149 (20.0)	106 (15.1)	1.45 (1.05-1.99)	P-trend= 0.03
		missing	24	23		
		C/C	134 (18.0)	131 (19.6)	1.00	1.00
FGFR2	rs2912774	A/C	347 (46.7)	350 (50.4)	1.05 (0.78-1.41)	1.27 (1.09-1.49)
		A/A	261 (35.3)	176 (30.0)	1.55 (1.13-2.13)	P-trend= 0.003
		missing	24	24		
		C/C	129 (17.4)	126 (18.8)	1.00	1.00
FGFR2	rs2936870	C/T	340 (45.9)	345 (50.2)	1.05 (0.78-1.41)	1.27 (1.09-1.48)
		T/T	272 (36.7)	187 (31.0)	1.53 (1.11-2.11)	P-trend= 0.003
		missing	25	23		
		C/C	158 (21.3)	146 (21.6)	1.00	1.00
FGFR2	rs2420946	C/T	366 (49.4)	356 (52.3)	1.02 (0.77-1.35)	1.17 (1.00-1.37)

		T/T	218 (29.4)	156 (26.1)	1.36 (0.99-1.86)	P-trend= 0.05
		missing	24	23		
		A/A	161 (21.7)	166 (24.0)	1.00	1.00
FGFR2	rs2162540	A/G	367 (49.5)	339 (48.8)	1.19 (0.91-1.56)	1.23 (1.05-1.44)
		G/G	214 (28.8)	152 (27.2)	1.51 (1.11-2.06)	P-trend= 0.01
		missing	24	24		
		C/C	186 (25.1)	200 (27.7)	1.00	1.00
FGFR2	rs2981582	C/T	382 (51.3)	325 (47.6)	1.30 (1.00-1.68)	1.19 (1.02-1.39)
		T/T	174 (23.6)	133 (24.7)	1.40 (1.02-1.91)	P-trend= 0.03
		missing	24	23		
		A/A	147 (19.7)	148 (24.3)	1.00	1.00
FGFR2	rs3135718	A/G	326 (44.0)	321 (44.3)	1.11 (0.83-1.47)	1.26 (1.08-1.46)
		G/G	269 (36.3)	187 (31.4)	1.54 (1.14-2.10)	P-trend= 0.003
		missing	24	25		
		T/T	16 (2.2)	6 (0.6)	1.00	1.00
10q	rs10510126	C/T	133 (18.2)	145 (18.5)	0.33 (0.12-0.88)	1.04 (0.82-1.32)
		C/C	593 (79.6)	507 (80.9)	0.40 (0.15-1.05)	P-trend= 0.8
		missing	24	23		
		G/G	700 (94.3)	623 (95.2)	1.00	1.00
ATM	rs1801516‡	A/G or A/A	42 (5.7)	35 (4.8)	1.16 (0.72-1.88)	1.22 (0.77-1.95)
		missing	24	23		
		T/T	96 (13.0)	70 (9.6)	1.00	1.00
ATM	rs664143	C/T	316 (42.6)	302 (44.9)	0.77 (0.54-1.10)	0.95 (0.81-1.11)
		C/C	330 (44.4)	285 (45.5)	0.81 (0.57-1.16)	P-trend= 0.5
		missing	24	24		
		T/T	622 (84.0)	535 (78.5)	1.00	1.00
ATM	rs170548	G/T	105 (14.1)	112 (19.9)	0.78 (0.57-1.05)	0.86 (0.67-1.11)
		G/G	14 (1.9)	11 (1.6)	1.13 (0.49-2.58)	P-trend= 0.3
		missing	25	23		
		C/C	700 (94.3)	623 (95.2)	1.00	1.00
ATM	rs3092993‡	A/C or A/A	42 (5.7)	35 (4.8)	1.16 (0.72-1.88)	1.22 (0.77-1.95)
		missing	24	23		
		T/T	506 (68.2)	459 (69.7)	1.00	1.00
LSP1	rs3817198	C/T	224 (30.1)	175 (26.3)	1.19 (0.93-1.52)	1.01 (0.82-1.24)
		C/C	12 (1.7)	24 (4.0)	0.47 (0.23-0.99)	P-trend= 0.9

		missing	24	23		
		C/C	59 (7.8)	54 (9.4)	1.00	1.00
LSP1	rs909116	C/T	309 (41.5)	271 (36.4)	1.13 (0.74-1.73)	1.00 (0.84-1.19)
		T/T	381 (50.7)	343 (54.2)	1.08 (0.71-1.65)	P-trend= 1.00
		missing	17	13		
		T/T	186 (25.1)	149 (21.1)	1.00	1.00
H19	rs2107425	C/T	390 (53.1)	339 (52.3)	0.92 (0.70-1.20)	0.84 (0.71-0.98)
		C/C	161 (21.8)	170 (26.6)	0.70 (0.51-0.96)	P-trend= 0.03
		missing	29	23		
		C/C	712 (96.0)	627 (95.5)	1.00	1.00
TNRC9/TOX3	rs8049149‡	C/T or T/T	30 (4.0)	31 (4.5)	0.92 (0.54-1.56)	0.92 (0.54-1.56)
		missing	24	23		
		C/C	511 (69.3)	438 (64.8)	1.00	1.00
TNRC9/TOX3	rs16951186	C/T	202 (26.9)	198 (32.5)	0.81 (0.63-1.03)	0.90 (0.74-1.09)
		T/T	29 (3.8)	22 (2.7)	1.10 (0.61-1.97)	P-trend= 0.3
		missing	24	23		
		C/C	313 (42.5)	295 (46.0)	1.00	1.00
TNRC9/TOX3	rs8051542	C/T	342 (45.8)	304 (47.6)	1.04 (0.83-1.31)	1.14 (0.97-1.35)
		T/T	87 (11.7)	59 (6.5)	1.45 (0.99-2.11)	P-trend= 0.1
		missing	24	23		
		A/A	208 (28.3)	164 (25.7)	1.00	1.00
TNRC9/TOX3	rs12443621	A/G	370 (49.8)	329 (47.1)	0.90 (0.69-1.16)	0.86 (0.74-1.01)
		G/G	164 (21.9)	165 (27.2)	0.74 (0.55-1.01)	P-trend= 0.1
		missing	24	23		
		T/T	196 (26.2)	182 (29.5)	1.00	1.00
TNRC9/TOX3	rs3803662	C/T	378 (51.3)	333 (49.6)	1.08 (0.84-1.40)	1.06 (0.90-1.23)
		C/C	166 (22.4)	142 (21.0)	1.11 (0.81-1.52)	P-trend= 0.5
		missing	26	24		
		C/C	636 (84.2)	580 (86.6)	1.00	1.00
TNRC9/TOX3	rs4784227	C/T	112 (14.6)	83 (13.0)	1.17 (0.85-1.62)	1.25 (0.93-1.67)
		T/T	9 (1.2)	4 (0.4)	2.70 (0.69-10.65)	P-trend= 0.1
		missing	9	14		
		T/T	411 (54.3)	432 (66.4)	1.00	1.00
TNRC9/TOX3	rs3104746	A/T	297 (39.5)	208 (30.4)	1.58 (1.25-2.00)	1.54 (1.27-1.86)
		A/A	46 (6.2)	24 (3.2)	2.22 (1.31-3.76)	P-trend= 1×10^{-5}

		missing	12	17		
		C/C	166 (21.6)	187 (29.6)	1.00	1.00
TNRC9/TOX3	rs3112562	C/G	391 (52.1)	332 (48.8)	1.39 (1.06-1.81)	1.28 (1.09-1.50)
		G/G	199 (26.3)	144 (21.6)	1.64 (1.19-2.25)	P-trend= 0.002
		missing	10	18		
		G/G	352 (47.4)	341 (47.5)	1.00	1.00
TNRC9/TOX3	rs9940048	A/G	327 (44.1)	262 (45.2)	1.20 (0.96-1.51)	1.10 (0.93-1.31)
		A/A	63 (8.6)	55 (7.3)	1.07 (0.71-1.61)	P-trend= 0.3
		missing	24	23		
		A/A or A/G	71 (9.6)	62 (7.9)	1.00	1.00
TP53	rs9894946[d]	G/G	670 (90.4)	596 (92.1)	0.94 (0.65-1.36)	0.96 (0.68-1.36)
		missing	25	23		
		C/C	271 (36.4)	264 (38.1)	1.00	1.00
TP53	rs1614984	C/T	347 (46.7)	285 (43.0)	1.21 (0.95-1.53)	1.07 (0.92-1.25)
		T/T	124 (16.9)	109 (18.9)	1.07 (0.78-1.48)	P-trend= 0.4
		missing	24	23		
		A/A	581 (78.2)	526 (80.2)	1.00	1.00
TP53	rs12951053	A/C	158 (21.5)	119 (18.3)	1.24 (0.94-1.63)	1.03 (0.80-1.32)
		C/C	2 (0.3)	12 (1.5)	0.16 (0.04-0.72)	P-trend= 0.8
		missing	25	24		
		A/A	404 (54.1)	354 (59.8)	1.00	1.00
TP53	rs2909430	A/G	273 (37.2)	254 (32.7)	1.00 (0.79-1.26)	1.06 (0.89-1.25)
		G/G	65 (8.7)	50 (7.5)	1.22 (0.81-1.83)	P-trend= 0.5
		missing	24	23		
		G/G	282 (37.8)	247 (34.7)	1.00	1.00
TP53	rs1042522	C/G	353 (46.9)	310 (44.8)	0.95 (0.74-1.20)	0.98 (0.83-1.15)
		C/C	117 (15.3)	102 (20.5)	0.97 (0.69-1.36)	P-trend= 0.8
		missing	14	22		
		T/T	8 (1.0)	7 (1.0)	1.00	1.00
TP53	rs8079544	C/T	142 (19.1)	117 (20.0)	1.09 (0.36-3.27)	0.89 (0.69-1.15)
		C/C	592 (79.9)	534 (79.0)	0.95 (0.32-2.79)	P-trend= 0.4
		missing	24	23		
		A/A	93 (12.3)	87 (12.8)	1.00	1.00
COX11	rs7222197	A/G	324 (42.5)	305 (45.1)	1.02 (0.72-1.45)	1.14 (0.97-1.33)
		G/G	343 (45.3)	285 (42.1)	1.23 (0.87-1.75)	P-trend= 0.1

		missing	6	4		
COX11	rs6504950	A/A	94 (12.3)	87 (12.8)	1.00	1.00
		A/G	325 (42.3)	305 (44.8)	1.01 (0.71-1.43)	1.13 (0.96-1.32)
		G/G	346 (45.4)	288 (42.4)	1.21 (0.86-1.71)	P-trend= 0.1
		missing	1	1		

[a]weighted by inverse sampling probability
[b]adjusted for age at diagnosis (cases) or selection (controls) and proportion of African ancestry
[c]Assessed using dominant and additive model (MAF<5%)
[d]Assessed using recessive and additive model (MAF>95%)

Table 12: Comparison of odds ratios (ORs) and confidence limit ratios (CLRs) or posterior limit ratios (PLRs) for MLE, Bayesian and hierarchical regression models among white

Gene and LD block	SNP	Reference OR[a]	MLE OR[b] (95% CI)	MLE CLR	Bayesian OR[b] (95% PI)	Bayes PLR	Hierarchical OR[e] (95% PI)	Hierarchical PLR
1p12	rs11249433[d]	1.14 (1.10, 1.19) [186]	1.09 (0.96, 1.24)	1.28	1.09 (0.96, 1.22)	1.28	1.09 (0.96, 1.24)	1.28
CASP8 block1	rs1045485[f]	1.12 (1.08, 1.18) [20]	1.13 (0.94, 1.35)	1.43	1.11 (0.93, 1.29)	1.39	1.08 (0.78, 1.49)	1.90
CASP8 block1	rs17468277		1.12 (0.93, 1.34)	1.43	1.11 (0.93, 1.29)	1.38	1.04 (0.76, 1.44)	1.90
2q35	rs13387042[d]	1.20 (1.14, 1.26) [187]	1.08 (0.96, 1.22)	1.28	1.08 (0.96, 1.21)	1.27	1.08 (0.96, 1.22)	1.28
2p	rs4666451[e]	1.03 (1.00, 1.06) [194]	1.02 (0.90, 1.16)	1.28	1.02 (0.90, 1.14)	1.27	1.02 (0.90, 1.16)	1.28
SLC4A7	rs4973768[d]	1.16 (1.10, 1.24) [189]	1.04 (0.92, 1.17)	1.28	1.04 (0.92, 1.17)	1.28	1.04 (0.92, 1.17)	1.28
4p	rs12505080[e]	1.15 (1.03, 1.28)[g] [201]	1.06 (0.92, 1.23)	1.33	1.06 (0.92, 1.21)	1.31	1.06 (0.92, 1.23)	1.33
TLR1	rs7696175[e]	1.12 (1.00, 1.26)[g] [201]	1.09 (0.96, 1.23)	1.28	1.09 (0.96, 1.22)	1.27	1.09 (0.96, 1.23)	1.28
MRPS30	rs4415084[d]	1.16 (1.10, 1.21) [208]	1.23 (1.08, 1.40)	1.30	1.22 (1.07, 1.37)	1.28	1.23 (1.08, 1.40)	1.30
MRPS30	rs10941679[e]	1.19 (1.13, 1.26) [208]	1.18 (1.03, 1.36)	1.32	1.17 (1.01, 1.33)	1.31	1.18 (1.03, 1.36)	1.32
5p12	rs981782[d]	1.04 (1.01, 1.08) [194]	0.98 (0.86, 1.11)	1.28	0.98 (0.87, 1.10)	1.27	0.98 (0.86, 1.11)	1.28
5q	rs30099[e]	1.05 (1.01, 1.10) [194]	1.04 (0.85, 1.28)	1.52	1.04 (0.84, 1.24)	1.48	1.04 (0.85, 1.28)	1.52
MAP3K1	rs889312[d]	1.13 (1.10, 1.16) [194]	1.19 (1.04, 1.35)	1.30	1.18 (1.03, 1.32)	1.29	1.19 (1.04, 1.35)	1.30
ESR1	rs2046210[d]	1.29 (1.21, 1.37) [197]	1.09 (0.96, 1.24)	1.30	1.08 (0.95, 1.22)	1.28	1.09 (0.96, 1.24)	1.30
ESR1	rs851974		0.91 (0.80, 1.03)	1.29	0.91 (0.81, 1.03)	1.27	0.91 (0.80, 1.03)	1.29
ESR1	rs2077647		0.97 (0.86, 1.10)	1.28	0.97 (0.86, 1.10)	1.27	0.97 (0.86, 1.10)	1.28
ESR1	rs2234693		0.95 (0.84, 1.07)	1.28	0.95 (0.84, 1.06)	1.27	0.95 (0.84, 1.07)	1.28
ESR1	rs1801132[f]	1.05 (1.00, 1.11) [20]	0.92 (0.80, 1.06)	1.34	0.93 (0.80, 1.05)	1.31	0.92 (0.80, 1.06)	1.34
ESR1	rs3020314[f]	1.12 (1.06,1.18) [20]	1.05 (0.92, 1.19)	1.29	1.05 (0.93, 1.18)	1.27	1.05 (0.92, 1.19)	1.29
ESR1	rs3798577		1.03 (0.91, 1.17)	1.28	1.03 (0.91, 1.16)	1.27	1.03 (0.91, 1.17)	1.28
ECHDC1	rs2180341[d]	1.41 (1.25, 1.59) [196]	1.04 (0.90, 1.20)	1.34	1.04 (0.89, 1.19)	1.33	1.04 (0.90, 1.20)	1.34
RELN	rs17157903[e]	1.11 (1.00, 1.23) [201]	0.87 (0.73, 1.04)	1.42	0.89 (0.75, 1.05)	1.40	0.87 (0.73, 1.04)	1.42
8q24	rs13281615[d]	1.08 (1.05, 1.11) [194]	1.11 (0.98, 1.26)	1.28	1.11 (0.98, 1.24)	1.26	1.11 (0.98, 1.26)	1.28
8q24	rs1562430[d]	1.17 (1.10, 1.25) [189]	1.13 (0.99, 1.28)	1.29	1.12 (0.99, 1.26)	1.27	1.13 (0.99, 1.28)	1.29
CDKN2A/B	rs3731257		0.93 (0.81, 1.07)	1.32	0.94 (0.81, 1.07)	1.31	0.93 (0.81, 1.07)	1.32
CDKN2A/B	rs3731249		0.90 (0.63, 1.28)	2.04	0.94 (0.68, 1.20)	1.78	0.90 (0.63, 1.29)	2.04
CDKN2A/B block 1	rs518394		1.03 (0.91, 1.16)	1.28	1.03 (0.91, 1.15)	1.26	1.03 (0.91, 1.16)	1.28

CDKN2A/B block 1	rs564398		1.04 (0.92, 1.17)	1.28	1.04 (0.91, 1.17)	1.28	1.05 (0.82, 1.36)	1.67
CDKN2A/B	rs1011970d	1.20 (1.11, 1.30) [189]	1.13 (0.96, 1.33)	1.38	1.12 (0.95, 1.30)	1.36	1.13 (0.96, 1.33)	1.38
CDKN2A/B	rs10757278		1.17 (1.04, 1.33)	1.28	1.16 (1.01, 1.30)	1.28	1.17 (1.04, 1.33)	1.28
CDKN2A/B	rs10811661		1.00 (0.85, 1.18)	1.38	1.01 (0.85, 1.16)	1.36	1.00 (0.85, 1.18)	1.38
ANKRD16	rs2380205[d]	1.06 (1.02, 1.10) [189]	1.01 (0.89, 1.14)	1.28	1.01 (0.89, 1.14)	1.27	1.01 (0.89, 1.14)	1.28
ZNF365	rs10995190[d]	1.16 (1.10, 1.22) [189]	1.00 (0.84, 1.20)	1.43	1.00 (0.85, 1.17)	1.38	1.00 (0.84, 1.20)	1.43
ZMIZ1	rs704010[d]	1.07 (1.03, 1.11) [189]	1.24 (1.09, 1.41)	1.29	1.23 (1.08, 1.39)	1.28	1.24 (1.09, 1.41)	1.29
FGFR2 block 1	rs3750817		1.24 (1.09, 1.40)	1.29	1.22 (1.08, 1.37)	1.28	0.96 (0.81, 1.15)	1.43
FGFR2 block 1	rs10736303[e]	1.25 (1.18, 1.32) [194]	1.33 (1.17, 1.50)	1.28	1.31 (1.15, 1.47)	1.28	1.08 (0.79, 1.48)	1.88
FGFR2 block 1	rs11200014		1.30 (1.15, 1.48)	1.28	1.29 (1.13, 1.44)	1.27	0.94 (0.66, 1.35)	2.05
FGFR2 block 1	rs2981579[d]	1.17 (1.07, 1.27)[g] [186]	1.33 (1.18, 1.51)	1.28	1.31 (1.16, 1.48)	1.27	1.20 (0.85, 1.72)	2.03
FGFR2 block 1	rs1078806[e]	1.26 (1.13, 1.40) [196]	1.29 (1.14, 1.46)	1.28	1.28 (1.14, 1.44)	1.26	0.95 (0.67, 1.34)	1.99
FGFR2 block 1	rs2981578[e]	1.26 (1.19, 1.34) [194]	1.32 (1.17, 1.50)	1.28	1.30 (1.15, 1.45)	1.26	1.11 (0.81, 1.51)	1.86
FGFR2 block 1	rs1219648[d]	1.27 (1.18, 1.36) [201]	1.31 (1.16, 1.48)	1.28	1.29 (1.14, 1.45)	1.27	1.04 (0.72, 1.51)	2.10
FGFR2 block 1	rs2912774[e]	1.26 (1.19, 1.34) [194]	1.30 (1.15, 1.47)	1.28	1.28 (1.13, 1.44)	1.27	0.96 (0.66, 1.40)	2.13
FGFR2 block 1	rs2936870[e]	1.26 (1.19, 1.34) [194]	1.30 (1.15, 1.47)	1.28	1.29 (1.13, 1.44)	1.28	0.98 (0.67, 1.43)	2.14
FGFR2 block 1	rs2420946[e]	1.25 (1.18, 1.36) [201]	1.30 (1.15, 1.48)	1.28	1.28 (1.14, 1.45)	1.27	0.99 (0.67, 1.46)	2.16
FGFR2 block 1	rs2162540		1.31 (1.15, 1.48)	1.28	1.29 (1.13, 1.44)	1.27	1.04 (0.72, 1.50)	2.10
FGFR2 block 1	rs2981582[d]	1.26 (1.23, 1.30) [194]	1.30 (1.15, 1.48)	1.28	1.29 (1.13, 1.44)	1.28	1.01 (0.70, 1.46)	2.09
FGFR2 block 1	rs3135718[e]	1.15 (1.07, 1.23) [194]	1.31 (1.16, 1.48)	1.28	1.29 (1.14, 1.45)	1.27	1.04 (0.73, 1.49)	2.04
10q	rs10510126[e]	1.20 (1.08, 1.35) [201]	1.11 (0.91, 1.35)	1.47	1.10 (0.91, 1.31)	1.43	1.11 (0.91, 1.35)	1.47
ATM	rs1800054		1.01 (0.65, 1.58)	2.45	1.03 (0.70, 1.42)	2.03	1.01 (0.65, 1.58)	2.45
ATM	rs1800057[f]	1.20 (1.01, 1.44)[h] [20]	1.09 (0.76, 1.56)	2.06	1.08 (0.78, 1.39)	1.79	1.09 (0.76, 1.56)	2.06
ATM	rs1800058		0.82 (0.54, 1.25)	2.33	0.90 (0.62, 1.18)	1.90	0.82 (0.54, 1.25)	2.34
ATM block 1	rs1801516		0.98 (0.82, 1.17)	1.43	0.99 (0.84, 1.16)	1.39	0.94 (0.68, 1.31)	1.94
ATM block 1	rs3092992		1.19 (0.89, 1.60)	1.80	1.16 (0.87, 1.46)	1.67	1.13 (0.87, 1.46)	1.68
ATM block 1	rs664143		1.02 (0.90, 1.15)	1.28	1.02 (0.90, 1.14)	1.27	1.10 (0.92, 1.31)	1.42
ATM block 1	rs170548		0.98 (0.86, 1.12)	1.31	0.98 (0.86, 1.11)	1.29	0.91 (0.74, 1.11)	1.50
ATM block 1	rs3092993		0.98 (0.82, 1.18)	1.43	0.99 (0.83, 1.15)	1.39	0.96 (0.69, 1.34)	1.94
LSP1	rs3817198[d]	1.07 (1.04, 1.11) [194]	1.08 (0.95, 1.24)	1.30	1.08 (0.95, 1.22)	1.28	1.08 (0.95, 1.24)	1.30
LSP1	rs909116[d]	1.17 (1.10, 1.24) [189]	1.14 (1.01, 1.30)	1.28	1.13 (0.99, 1.27)	1.28	1.14 (1.01, 1.30)	1.28
H19	rs2107425[e]	1.04 (1.01, 1.08) [194]	1.15 (1.00, 1.31)	1.31	1.14 (1.01, 1.30)	1.29	1.15 (1.00, 1.31)	1.31
TNRC9/TOX3	rs16951186		1.17 (0.62, 2.18)	3.49	1.10 (0.69, 1.56)	2.26	1.17 (0.62, 2.18)	3.49

TNRC9/TOX3	rs8051542[e]	1.09 (1.06, 1.13) [194]	1.12 (0.99, 1.26)	1.28	1.11 (0.97, 1.23)	1.27	1.12 (0.99, 1.26)	1.28
TNRC9/TOX3	rs1244362[e]	1.11 (1.08, 1.14) [194]	1.17 (1.04, 1.33)	1.28	1.16 (1.03, 1.31)	1.27	1.17 (1.04, 1.33)	1.28
TNRC9/TOX3 block 1	rs3803662[d]	1.20 (1.16, 1.24) [194]	1.27 (1.11, 1.46)	1.31	1.25 (1.10, 1.42)	1.29	1.14 (0.91, 1.43)	1.58
TNRC9/TOX3 block 1	rs4784227[d]	1.25 (1.20, 1.31) [199]	1.26 (1.09, 1.44)	1.32	1.23 (1.07, 1.41)	1.31	1.11 (0.88, 1.41)	1.60
TNRC9/TOX3	rs3104746		1.66 (1.10, 2.51)	2.29	1.42 (0.97, 1.94)	2.01	1.66 (1.10, 2.51)	2.29
TNRC9/TOX3	rs3112562		0.99 (0.86, 1.15)	1.34	0.99 (0.86, 1.13)	1.32	0.99 (0.86, 1.15)	1.34
TNRC9/TOX3	rs9940048		1.03 (0.89, 1.19)	1.33	1.03 (0.89, 1.17)	1.32	1.03 (0.89, 1.19)	1.33
TP53	rs9894946		0.84 (0.72, 0.99)	1.38	0.86 (0.73, 1.00)	1.36	0.84 (0.72, 0.99)	1.38
TP53	rs1614984		1.03 (0.91, 1.17)	1.28	1.03 (0.92, 1.15)	1.26	1.03 (0.91, 1.17)	1.28
TP53	rs12951053[f]	1.15 (1.04, 1.26) [20]	1.09 (0.85, 1.39)	1.63	1.08 (0.83, 1.32)	1.58	1.09 (0.85, 1.39)	1.63
TP53	rs17880604		0.82 (0.51, 1.33)	2.62	0.92 (0.61, 1.26)	2.06	0.82 (0.51, 1.33)	2.62
TP53 block 1	rs1800372		0.88 (0.55, 1.40)	2.57	0.95 (0.64, 1.30)	2.03	1.01 (0.72, 1.40)	1.94
TP53 block 1	rs2909430		1.11 (0.93, 1.33)	1.43	1.10 (0.92, 1.29)	1.41	1.10 (0.89, 1.35)	1.51
TP53 block 1	rs1042522		0.98 (0.85, 1.13)	1.33	0.99 (0.85, 1.12)	1.31	1.09 (0.92, 1.29)	1.40
TP53	rs8079544		1.24 (0.95, 1.63)	1.72	1.19 (0.92, 1.50)	1.63	1.24 (0.95, 1.63)	1.72
COX11 block 1	rs7222197[e]	1.12 (1.04, 1.20) [189]	0.98 (0.85, 1.12)	1.32	0.98 (0.86, 1.12)	1.29	0.99 (0.72, 1.36)	1.89
COX11 block 1	rs6504950[e]	1.05 (1.03, 1.09) [209]	0.98 (0.85, 1.12)	1.32	0.98 (0.84, 1.11)	1.31	0.99 (0.72, 1.36)	1.89

[a]OR from initial GWAS or candidate gene meta-analysis (if met criteria for cumulative evidence of association); all ORs for log-additive genetic models, unless otherwise specified

[b]adjusted for age at diagnosis (case) or selection (controls) and proportion of African ancestry

[c]adjusted for age at diagnosis (case) or selection (controls), proportion of African ancestry and other SNPs in LD block

[d]previous GWAS hit

[e]Other GWAS-identified gene

[f]cumulative evidence of an association in Zhang et al. meta-analysis

[g]OR estimated using general genetic model

[h]OR estimated using dominant genetic model

Table 13: Comparison of odds ratios (ORs) and confidence limit ratios (CLRs) or posterior limit ratios (PLRs) for frequentist, basic hierarchical and Bayesian regression models among African American women in the Carolina Breast Cancer Study

Gene and LD block	SNP	Reference OR[a]	MLE OR[b] (95% CI)	MLE CLR	Bayesian OR[b] (95% PI)	Bayes PLR	Hierarchical OR[c] (95% PI)	Hierarchical PLR
1p12	rs11249433[d]	1.14 (1.10, 1.19) [186]	1.26 (0.99, 1.60)	1.61	1.22 (0.96, 1.48)	1.53	1.26 (0.99, 1.60)	1.61
CASP8 block 1	rs1045485[f]	1.12 (1.08, 1.18) [20]	0.93 (0.67, 1.29)	1.93	0.96 (0.69, 1.22)	1.77	1.12 (0.39, 3.17)	8.05
CASP8 block 1	rs17468277		1.09 (0.78, 1.54)	1.99	1.07 (0.78, 1.38)	1.78	1.20 (0.41, 3.53)	8.68
2q35	rs13387042[d]	1.20 (1.14, 1.26) [187]	1.02 (0.86, 1.22)	1.43	1.02 (0.86, 1.20)	1.39	1.02 (0.86, 1.22)	1.43
2p	rs4666451[e]	1.03 (1.00, 1.06) [194]	1.15 (0.96, 1.39)	1.45	1.13 (0.95, 1.34)	1.41	1.15 (0.96, 1.39)	1.45
SLC4A7	rs4973768[d]	1.16 (1.10, 1.24) [189]	0.90 (0.77, 1.06)	1.38	0.91 (0.77, 1.04)	1.35	0.90 (0.77, 1.06)	1.38
4p	rs12505080[e]	1.15 (1.03, 1.28)[g] [201]	1.09 (0.88, 1.34)	1.52	1.08 (0.88, 1.30)	1.48	1.09 (0.88, 1.34)	1.52
TLR1	rs7696175[e]	1.12 (1.00, 1.26)[g] [201]	1.39 (1.04, 1.86)	1.79	1.29 (0.99, 1.66)	1.68	1.39 (1.04, 1.86)	1.80
MRPS30	rs4415084[d]	1.16 (1.10, 1.21) [208]	1.13 (0.97, 1.33)	1.38	1.13 (0.96, 1.30)	1.35	1.13 (0.97, 1.33)	1.38
MRPS30	rs10941679[e]	1.19 (1.13, 1.26) [208]	1.00 (0.82, 1.22)	1.49	1.01 (0.83, 1.19)	1.43	1.00 (0.82, 1.22)	1.49
5p12	rs981782[d]	1.04 (1.01, 1.08) [194]	1.11 (0.84, 1.46)	1.74	1.09 (0.84, 1.36)	1.61	1.11 (0.84, 1.46)	1.74
5q	rs30099[e]	1.05 (1.01, 1.10) [194]	1.22 (0.98, 1.52)	1.55	1.19 (0.96, 1.44)	1.50	1.22 (0.98, 1.52)	1.55
MAP3K1	rs889312[d]	1.13 (1.10, 1.16) [194]	0.95 (0.80, 1.13)	1.41	0.96 (0.81, 1.11)	1.37	0.95 (0.80, 1.13)	1.41
ESR1	rs2046210[d]	1.29 (1.21, 1.37) [197]	1.22 (1.04, 1.43)	1.38	1.20 (1.03, 1.39)	1.35	1.22 (1.04, 1.43)	1.38
ESR1	rs851974		0.93 (0.76, 1.14)	1.50	0.94 (0.78, 1.13)	1.45	0.93 (0.75, 1.14)	1.50
ESR1	rs2077647		1.07 (0.92, 1.25)	1.37	1.07 (0.92, 1.23)	1.34	1.07 (0.92, 1.25)	1.37
ESR1	rs2234693		0.96 (0.82, 1.13)	1.37	0.97 (0.83, 1.12)	1.35	0.96 (0.82, 1.13)	1.37
ESR1	rs1801132[f]	1.05 (1.00, 1.11) [20]	1.19 (0.93, 1.52)	1.64	1.16 (0.91, 1.42)	1.55	1.19 (0.93, 1.52)	1.64
ESR1	rs3020314[f]	1.12 (1.06,1.18) [20]	1.00 (0.84, 1.19)	1.41	1.01 (0.85, 1.17)	1.37	1.00 (0.84, 1.19)	1.41
ESR1	rs3798577		1.02 (0.88, 1.19)	1.36	1.02 (0.89, 1.18)	1.33	1.02 (0.87, 1.19)	1.36
ECHDC1	rs2180341[d]	1.41 (1.25, 1.59) [196]	0.98 (0.83, 1.15)	1.39	0.98 (0.84, 1.14)	1.36	0.98 (0.83, 1.15)	1.39
RELN	rs17157903[e]	1.11 (1.00, 1.23) [201]	1.07 (0.83, 1.37)	1.64	1.07 (0.85, 1.31)	1.54	1.07 (0.83, 1.37)	1.64
8q24	rs13281615[d]	1.08 (1.05, 1.11) [194]	1.00 (0.86, 1.18)	1.38	1.01 (0.85, 1.17)	1.37	1.00 (0.86, 1.18)	1.38
8q24	rs1562430[d]	1.17 (1.10, 1.25) [189]	1.00 (0.86, 1.17)	1.36	1.00 (0.87, 1.16)	1.34	1.00 (0.86, 1.17)	1.36
CDKN2A/B	rs3731257		0.88 (0.67, 1.15)	1.71	0.91 (0.71, 1.14)	1.60	0.88 (0.67, 1.15)	1.71
CDKN2A/B block 1	rs518394		1.01 (0.76, 1.35)	1.76	1.02 (0.78, 1.28)	1.64	1.01 (0.76, 1.35)	1.76
CDKN2A/B block 1	rs564398		1.01 (0.75, 1.35)	1.79	1.01 (0.77, 1.27)	1.66	1.00 (0.72, 1.40)	1.96

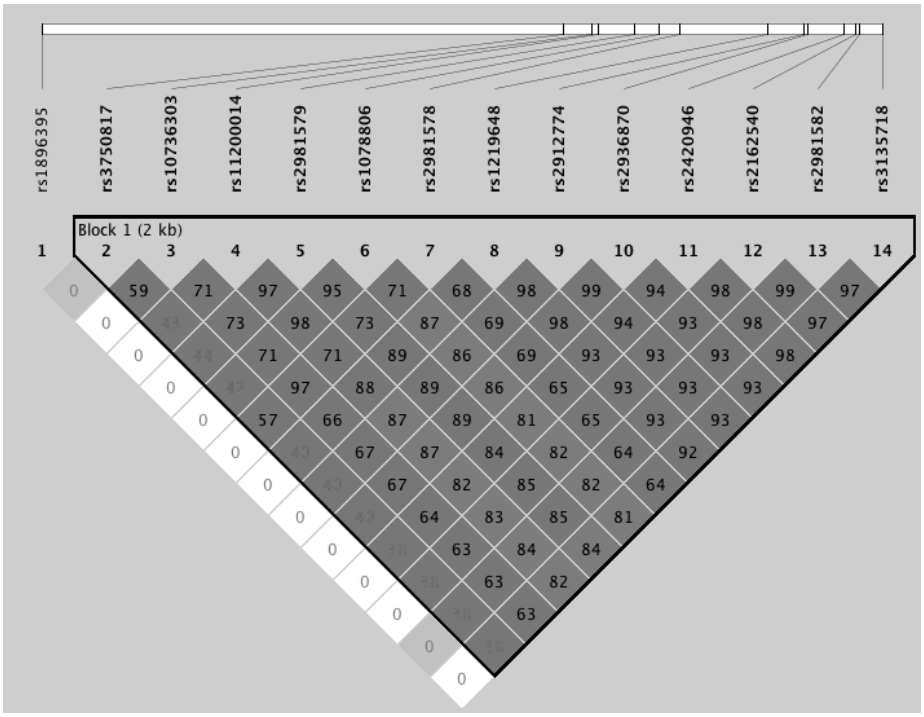
CDKN2A/B	rs1011970[d]	1.20 (1.11, 1.30) [189]	0.95 (0.81, 1.11)	1.38	0.96 (0.82, 1.10)	1.34	0.95 (0.81, 1.11)	1.38
CDKN2A/B	rs10757278		0.91 (0.75, 1.12)	1.50	0.93 (0.76, 1.09)	1.44	0.91 (0.75, 1.12)	1.50
CDKN2A/B	rs10811661		1.00 (0.74, 1.35)	1.83	1.01 (0.76, 1.28)	1.67	1.00 (0.74, 1.35)	1.83
ANKRD16	rs2380205[d]	1.06 (1.02, 1.10) [189]	0.97 (0.83, 1.13)	1.37	0.98 (0.84, 1.13)	1.34	0.97 (0.83, 1.13)	1.37
ZNF365	rs10995190[d]	1.16 (1.10, 1.22) [189]	1.06 (0.87, 1.29)	1.49	1.05 (0.86, 1.23)	1.44	1.06 (0.87, 1.29)	1.49
ZMIZ1	rs704010[d]	1.07 (1.03, 1.11) [189]	1.05 (0.80, 1.36)	1.69	1.04 (0.80, 1.28)	1.61	1.05 (0.80, 1.36)	1.69
FGFR2	rs1896395		1.01 (0.83, 1.23)	1.48	1.02 (0.84, 1.20)	1.44	1.01 (0.83, 1.23)	1.48
FGFR2 block 1	rs3750817		1.74 (1.34, 2.26)	1.69	1.61 (1.22, 2.02)	1.66	1.38 (1.05, 1.83)	1.74
FGFR2 block 1	rs10736303[e]	1.25 (1.18, 1.32) [194]	1.39 (1.12, 1.74)	1.56	1.33 (1.07, 1.61)	1.51	1.08 (0.83, 1.39)	1.67
FGFR2 block 1	rs11200014		1.04 (0.86, 1.26)	1.47	1.04 (0.85, 1.23)	1.45	0.97 (0.70, 1.34)	1.92
FGFR2 block 1	rs2981579[d]	1.17 (1.07, 1.27)[g] [186]	1.22 (1.04, 1.42)	1.36	1.19 (1.03, 1.37)	1.33	1.10 (0.93, 1.31)	1.41
FGFR2 block 1	rs1078806[e]	1.26 (1.13, 1.40) [196]	1.06 (0.88, 1.29)	1.47	1.06 (0.87, 1.25)	1.43	1.00 (0.72, 1.38)	1.92
FGFR2 block 2	rs2981578[e]	1.26 (1.19, 1.34) [194]	1.42 (1.14, 1.77)	1.56	1.36 (1.10, 1.65)	1.51	1.23 (0.99, 1.53)	1.54
FGFR2 block 2	rs1219648[d]	1.27 (1.18, 1.36) [201]	1.19 (1.02, 1.39)	1.37	1.18 (1.01, 1.35)	1.33	1.01 (0.82, 1.24)	1.51
FGFR2 block 2	rs2912774[e]	1.26 (1.19, 1.34) [194]	1.27 (1.09, 1.49)	1.37	1.25 (1.06, 1.43)	1.35	1.09 (0.80, 1.49)	1.85
FGFR2 block 2	rs2936870[e]	1.26 (1.19, 1.34) [194]	1.27 (1.09, 1.48)	1.37	1.25 (1.07, 1.44)	1.34	1.09 (0.80, 1.48)	1.84
FGFR2 block 3	rs2420946[e]	1.25 (1.18, 1.36) [201]	1.17 (1.00, 1.37)	1.37	1.16 (1.00, 1.35)	1.35	0.95 (0.72, 1.25)	1.73
FGFR2 block 3	rs2162540		1.23 (1.05, 1.44)	1.36	1.21 (1.04, 1.40)	1.34	1.23 (1.05, 1.44)	1.37
FGFR2	rs2981582[d]	1.26 (1.23, 1.30) [194]	1.19 (1.02, 1.39)	1.37	1.18 (1.00, 1.35)	1.35	1.19 (1.02, 1.39)	1.37
FGFR2	rs3135718[e]	1.15 (1.07, 1.23) [194]	1.26 (1.08, 1.46)	1.35	1.24 (1.07, 1.43)	1.33	1.26 (1.08, 1.46)	1.35
10q	rs10510126[e]	1.20 (1.08, 1.35) [201]	1.04 (0.82, 1.32)	1.61	1.04 (0.82, 1.25)	1.53	1.04 (0.82, 1.32)	1.61
ATM	rs1801516		1.22 (0.77, 1.95)	2.54	1.14 (0.76, 1.58)	2.08	1.22 (0.77, 1.95)	2.54
ATM	rs664143		0.95 (0.81, 1.11)	1.38	0.96 (0.82, 1.10)	1.35	0.95 (0.81, 1.11)	1.38
ATM	rs170548		0.86 (0.67, 1.11)	1.66	0.90 (0.70, 1.10)	1.57	0.86 (0.67, 1.11)	1.66
ATM	rs3092993		1.22 (0.77, 1.95)	2.54	1.14 (0.76, 1.58)	2.08	1.22 (0.77, 1.95)	2.54
LSP1	rs3817198[d]	1.07 (1.04, 1.11) [194]	1.01 (0.82, 1.24)	1.52	1.01 (0.82, 1.24)	1.47	1.01 (0.82, 1.24)	1.52
LSP1	rs909116[d]	1.17 (1.10, 1.24) [189]	1.00 (0.84, 1.19)	1.42	1.01 (0.84, 1.17)	1.39	1.00 (0.84, 1.19)	1.42
H19	rs2107425[e]	1.04 (1.01, 1.08) [194]	0.84 (0.71, 0.98)	1.38	0.86 (0.73, 0.98)	1.35	0.84 (0.71, 0.98)	1.38
TNRC9/TOX3	rs8049149		0.92 (0.54, 1.56)	2.87	0.98 (0.64, 1.35)	2.10	0.92 (0.54, 1.56)	2.87
TNRC9/TOX3	rs16951186		0.90 (0.74, 1.09)	1.49	0.92 (0.76, 1.10)	1.45	0.90 (0.74, 1.09)	1.49
TNRC9/TOX3	rs8051542[e]	1.09 (1.06, 1.13) [194]	1.14 (0.97, 1.35)	1.39	1.13 (0.97, 1.30)	1.35	1.14 (0.97, 1.35)	1.39
TNRC9/TOX3	rs12443621[e]	1.11 (1.08, 1.14) [194]	0.86 (0.74, 1.01)	1.36	0.88 (0.75, 1.00)	1.34	0.86 (0.74, 1.01)	1.36
TNRC9/TOX3 block 1	rs3803662[d]	1.20 (1.16, 1.24) [194]	1.06 (0.90, 1.23)	1.37	1.05 (0.91, 1.21)	1.34	1.11 (0.94, 1.31)	1.39
TNRC9/TOX3 block 1	rs4784227[d]	1.25 (1.20, 1.31) [199]	1.25 (0.93, 1.67)	1.80	1.19 (0.90, 1.51)	1.67	1.28 (0.96, 1.71)	1.77
TNRC9/ TOX3	rs3104746		1.54 (1.27, 1.86)	1.46	1.49 (1.22, 1.75)	1.43	1.39 (1.14, 1.71)	1.50

block 2								
TNRC9/ TOX3 block 2	rs3112562		1.28 (1.09, 1.50)	1.37	1.26 (1.09, 1.46)	1.35	1.12 (0.94, 1.33)	1.41
TNRC9/TOX3	rs9940048		1.10 (0.93, 1.31)	1.40	1.10 (0.92, 1.29)	1.39	1.10 (0.93, 1.31)	1.41
TP53	rs9894946		0.96 (0.68, 1.36)	2.01	0.93 (0.70, 1.28)	1.82	0.96 (0.68, 1.37)	2.01
TP53	rs1614984		1.07 (0.92, 1.25)	1.36	1.07 (0.91, 1.22)	1.34	1.07 (0.92, 1.25)	1.36
TP53 block 1	rs12951053[f]	1.15 (1.04, 1.26) [20]	1.03 (0.80, 1.32)	1.64	1.03 (0.81, 1.25)	1.55	1.01 (0.77, 1.32)	1.71
TP53 block 1	rs2909430		1.06 (0.89, 1.25)	1.40	1.06 (0.89, 1.22)	1.37	1.04 (0.84, 1.28)	1.52
TP53 block 1	rs1042522		0.98 (0.83, 1.15)	1.38	0.98 (0.83, 1.13)	1.36	1.00 (0.82, 1.22)	1.50
TP53	rs8079544		0.89 (0.69, 1.15)	1.66	0.92 (0.72, 1.12)	1.57	0.89 (0.69, 1.15)	1.66
COX 11 block 1	rs7222197[e]	1.12 (1.04, 1.20) [189]	1.14 (0.97, 1.33)	1.38	1.12 (0.95, 1.29)	1.36	1.07 (0.77, 1.47)	1.90
COX 11 block 1	rs6504950[e]	1.05 (1.03, 1.09) [209]	1.13 (0.96, 1.32)	1.37	1.12 (0.95, 1.28)	1.35	1.07 (0.77, 1.47)	1.90

[a]OR from initial GWAS or candidate gene meta-analysis (if met criteria for cumulative evidence of association); all ORs for log-additive genetic
[b]adjusted for age at diagnosis (case) or selection (controls) and proportion of African ancestry
[c]adjusted for age at diagnosis (case) or selection (controls), proportion of African ancestry and other SNPs in the LD block
[d]previous GWAS hit
[e]Other GWAS-identified gene
[f]cumulative evidence of an association in Zhang et al. meta-analysis
[g]OR estimated using general genetic model

Figure 23: *FGFR2* linkage disequilibrium (LD) patterns, CBCS whites (A) and African Americans (B)

A



B

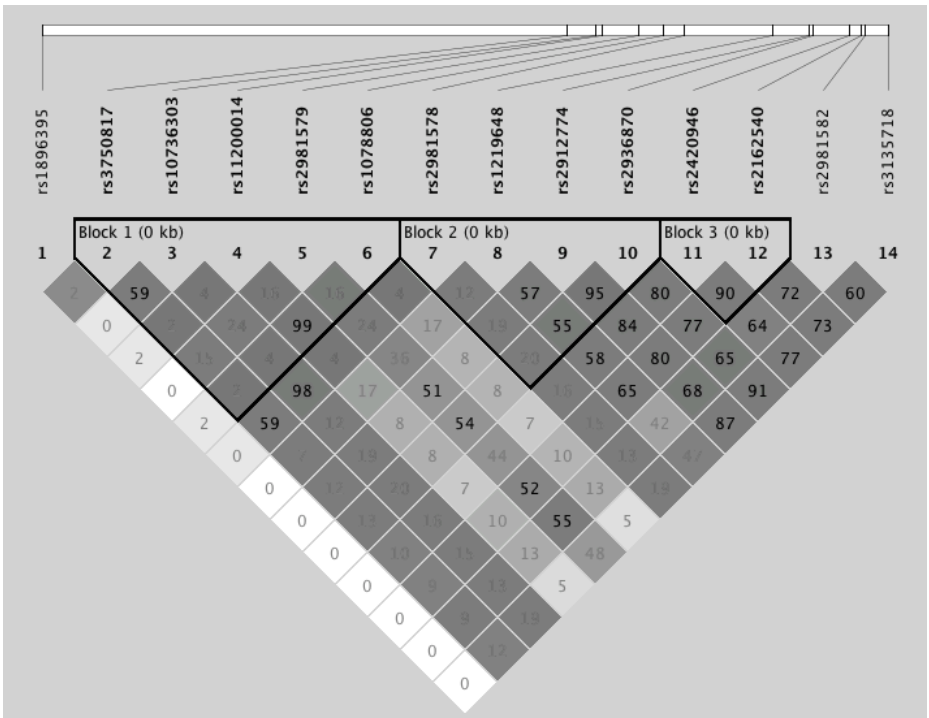
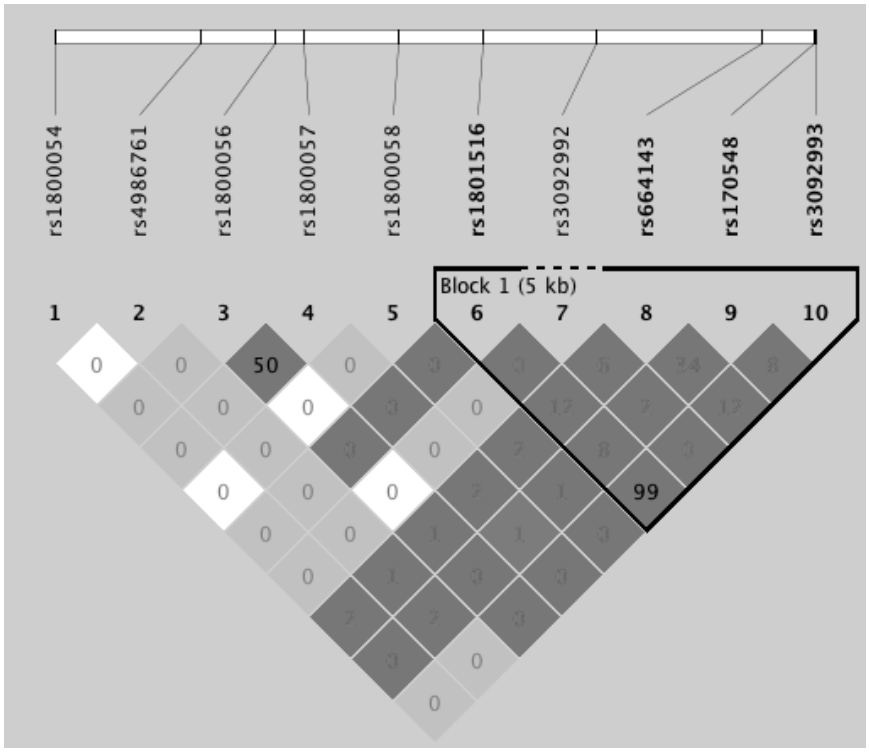


Figure 24: *ATM* linkage disequilibrium (LD) patterns, CBCS whites (A) and African Americans (B)

A



B

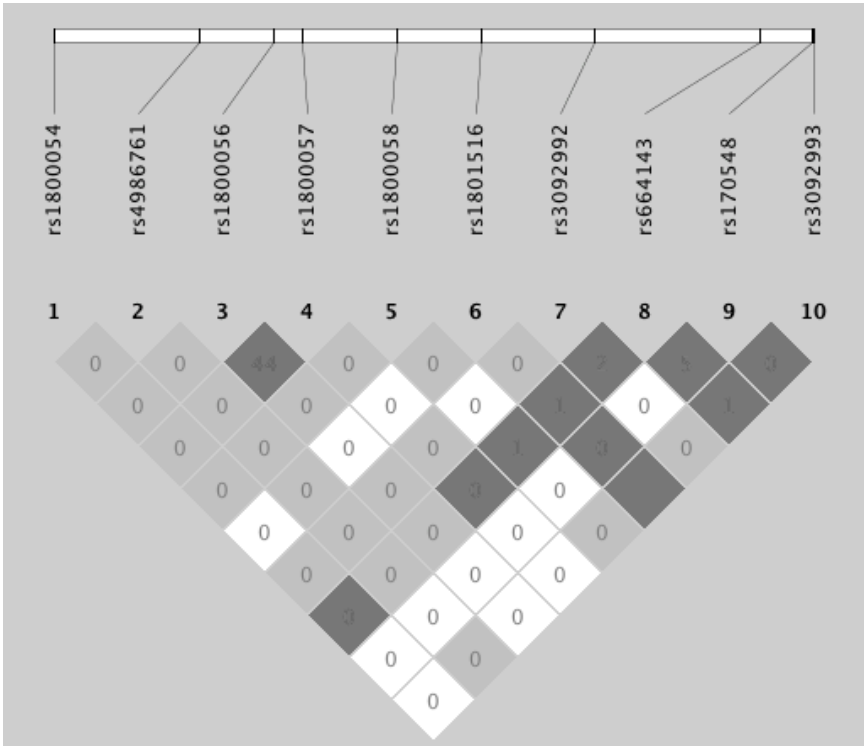
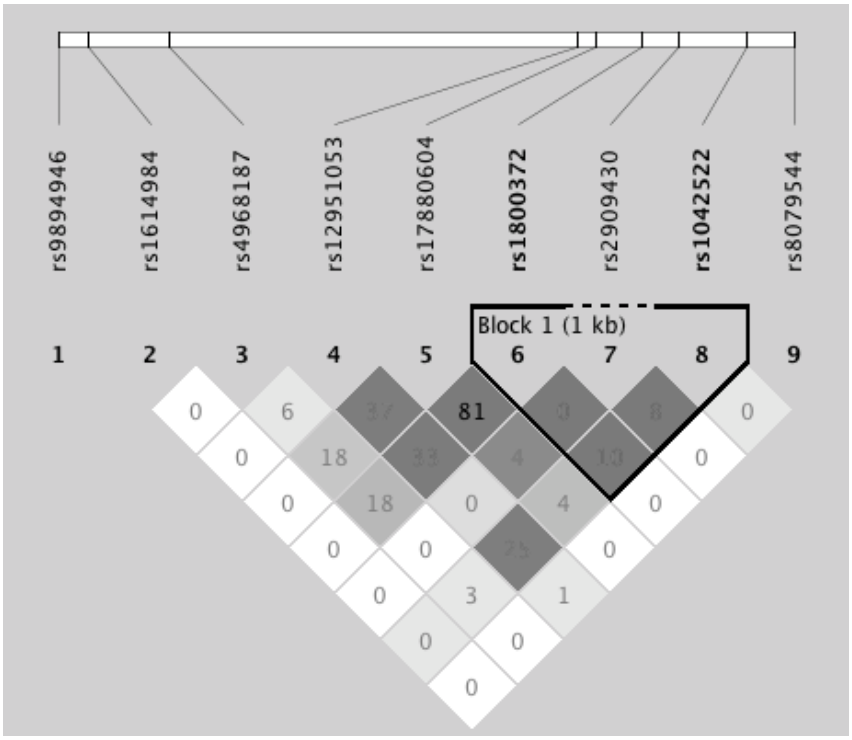


Figure 25: *TP53* linkage disequilibrium (LD) patterns, CBCS whites (A) and African Americans (B)

A



B

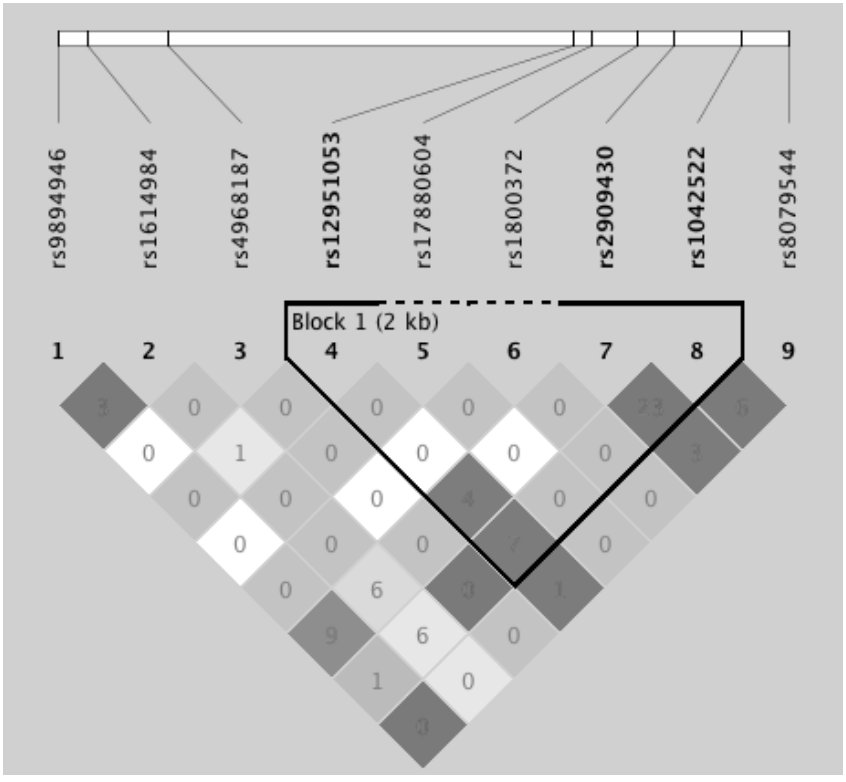
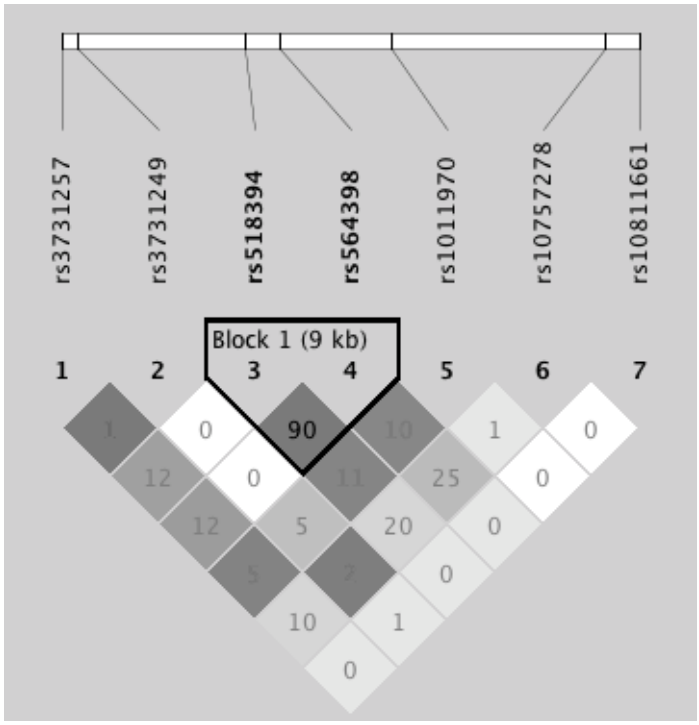


Figure 26: *CDNK2A/B* linkage disequilibrium (LD) patterns, CBCS whites (A) and African Americans (B)

A



B

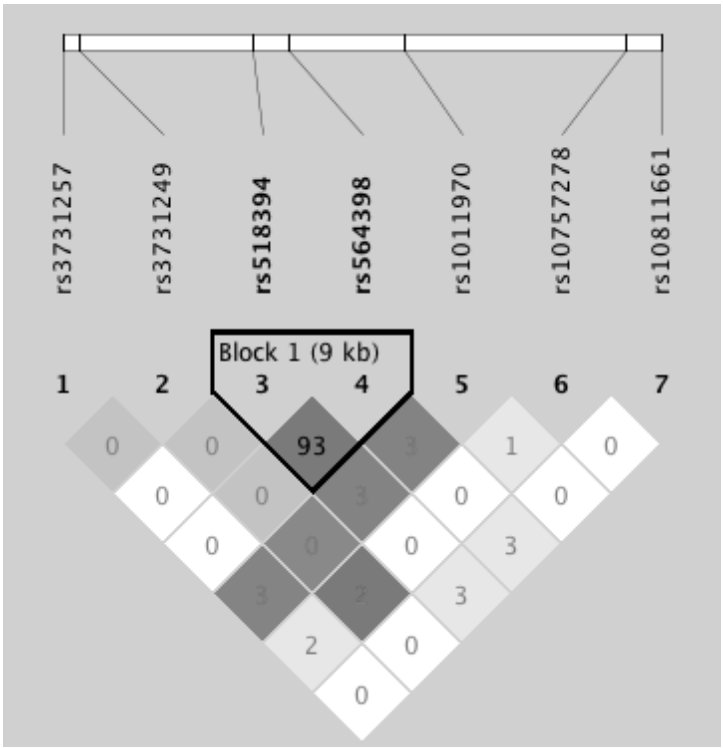
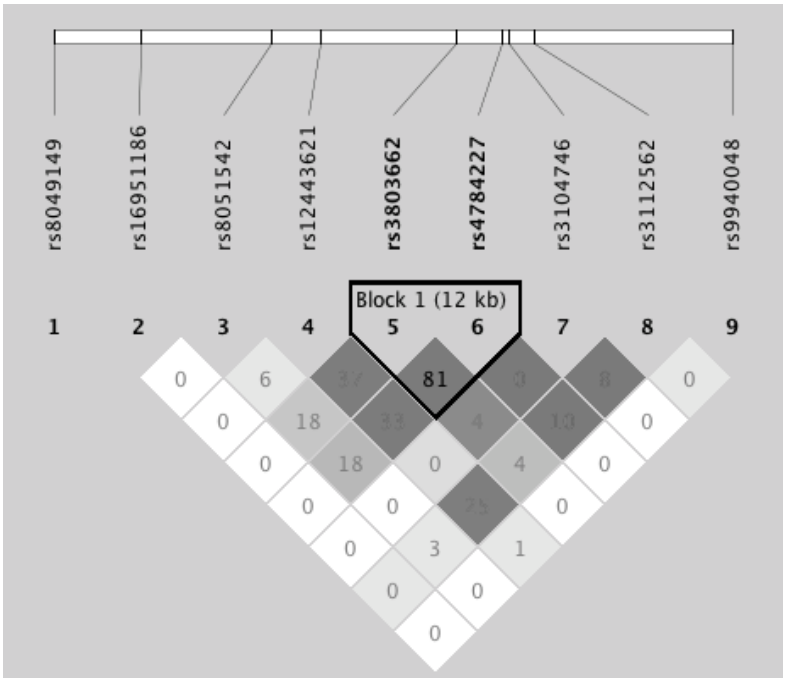


Figure 27: *TNRC9/TOX3* linkage disequilibrium (LD) patterns, CBCS whites (A) and African Americans (B)

A



B

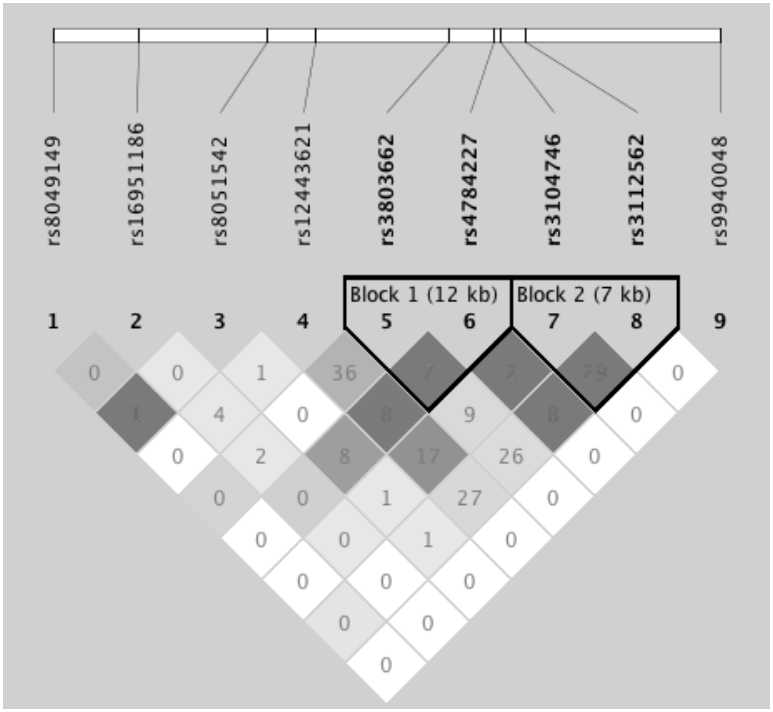


Figure 28: Comparison of point estimates and 95% PIs for hierarchical models with investigator specified covariance matrices, CBCS whites

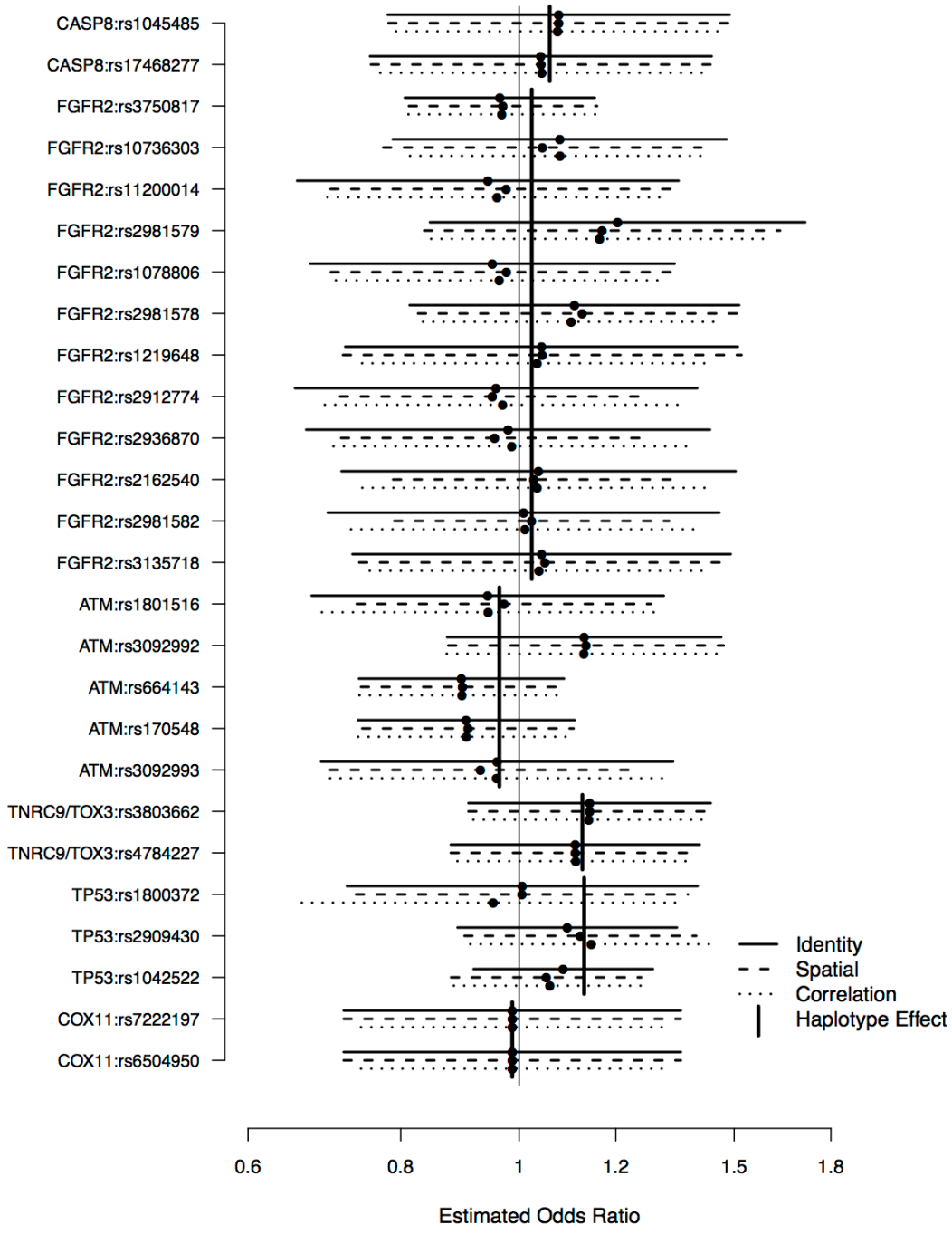


Figure 29: Comparison of point estimates and 95% PIs for hierarchical models with investigator specified covariance matrices, CBCS African Americans

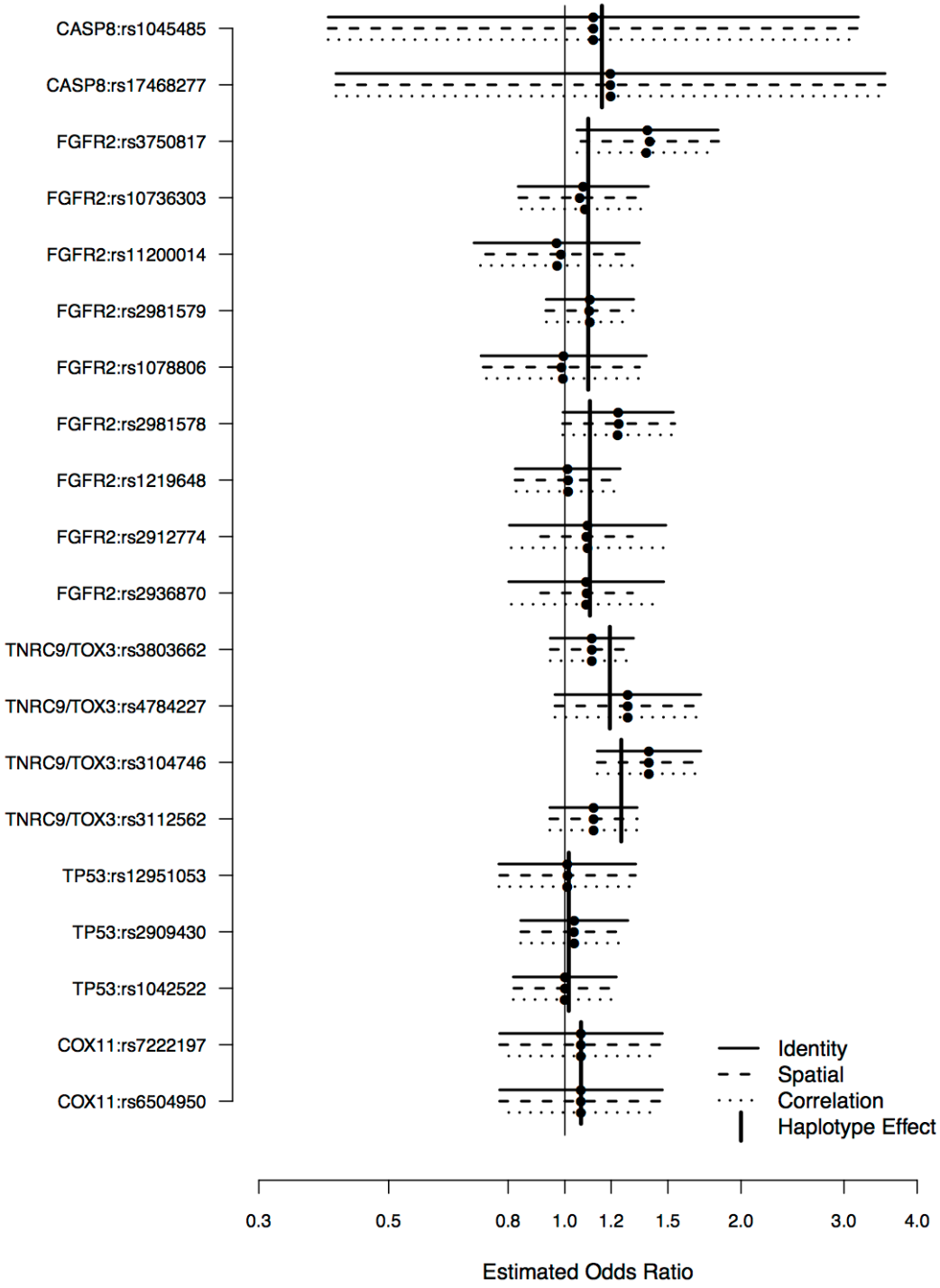


Table 14: Comparison of posterior limit ratios (PLRs) for hierarchical regression models among white women in the Carolina Breast Cancer Study

Gene and LD block	SNP	Identity Covariance Matrix	Exponential Decay by Spatial Distance Covariance Matrix	Correlation-based Covariance Matrix
CASP8 block1	rs1045485	1.90	1.90	1.84
CASP8 block1	rs17468277	1.90	1.90	1.84
CDKN2A/B block 1	rs518394	1.66	1.66	1.59
CDKN2A/B block 1	rs564398	1.67	1.67	1.60
FGFR2 block 1	rs3750817	1.43	1.43	1.42
FGFR2 block 1	rs10736303	1.88	1.82	1.76
FGFR2 block 1	rs11200014	2.05	1.94	1.89
FGFR2 block 1	rs2981579	2.03	1.96	1.89
FGFR2 block 1	rs1078806	1.99	1.94	1.85
FGFR2 block 1	rs2981578	1.86	1.86	1.75
FGFR2 block 1	rs1219648	2.10	2.12	1.94
FGFR2 block 1	rs2912774	2.14	1.79	1.96
FGFR2 block 1	rs2936870	2.15	1.80	1.97
FGFR2 block 1	rs2420946	2.16	1.97	1.97
FGFR2 block 1	rs2162540	2.15	1.67	1.96
FGFR2 block 1	rs2981582	2.13	1.70	1.96
FGFR2 block 1	rs3135718	2.09	2.08	1.93
ATM block 1	rs1801516	1.93	1.73	1.86
ATM block 1	rs3092992	1.68	1.69	1.69
ATM block 1	rs664143	1.42	1.43	1.41
ATM block 1	rs170548	1.45	1.45	1.44
ATM block 1	rs3092993	1.93	1.76	1.86
TNRC9/ TOX3 block 1	rs3803662	1.58	1.58	1.55
TNRC9/ TOX3 block 1	rs4784227	1.60	1.60	1.56
TP53 block 1	rs1800372	1.93	1.87	2.06
TP53 block 1	rs2909430	1.63	1.55	1.58
TP53 block 1	rs1042522	1.41	1.43	1.44
COX11 block 1	rs7222197	1.89	1.89	1.77
COX11 block 1	rs6504950	1.89	1.89	1.77

Table 15: Comparison of posterior limit ratios (PLRs) for hierarchical regression models among African American women in the Carolina Breast Cancer Study

Gene and LD block	SNP	Identity Covariance Matrix	Exponential Decay by Spatial Distance Covariance Matrix	Correlation-based Covariance Matrix
CASP8 block 1	rs1045485	8.05	8.05	8.05
CASP8 block 1	rs1746827	8.68	8.68	8.69
CDKN2A/B block 1	rs518394	1.93	1.93	1.93
CDKN2A/B block 2	rs564398	1.96	1.96	1.96
FGFR2 block 1	rs3750817	1.74	1.72	1.74
FGFR2 block 1	rs1073630	1.67	1.62	1.66
FGFR2 block 1	rs1120001	1.92	1.81	1.87
FGFR2 block 1	rs2981579	1.41	1.41	1.41
FGFR2 block 1	rs1078806	1.92	1.85	1.87
FGFR2 block 2	rs2981578	1.54	1.55	1.54
FGFR2 block 2	rs1219648	1.51	1.52	1.51
FGFR2 block 2	rs2912774	1.85	1.44	1.83
FGFR2 block 2	rs2936870	1.84	1.44	1.83
FGFR2 block 3	rs2420946	1.73	1.57	1.73
FGFR2 block 3	rs2162540	1.72	1.56	1.72
TNRC9/ TOX3 block 1	rs3803662	1.39	1.39	1.39
TNRC9/ TOX3 block 1	rs4784227	1.77	1.77	1.77
TNRC9/ TOX3 block 2	rs3104746	1.50	1.50	1.50
TNRC9/ TOX3 block 2	rs3112562	1.41	1.41	1.41
TP53 block 1	rs1295105	1.71	1.71	1.71
TP53 block 1	rs2909430	1.52	1.52	1.52
TP53 block 1	rs1042522	1.50	1.50	1.50
COX 11 block 1	rs7222197	1.90	1.90	1.77
COX 11 block 1	rs6504950	1.90	1.90	1.77

5. Breast Cancer Subtypes and Previously Established Genetic Risk Factors: A Bayesian Approach

5.1 Overview

Gene expression analyses indicate that breast cancer is a heterogeneous disease with at least five immunohistologic subtypes. Despite growing evidence that these subtypes are etiologically and prognostically distinct, few studies have investigated whether they have divergent genetic risk factors. To help fill in this gap in our understanding, I examined associations between breast cancer subtypes and previously established susceptibility loci among white and African American women in the Carolina Breast Cancer Study. I used Bayesian polytomous logistic regression to estimate odds ratios (ORs) and 95% posterior intervals (PIs) for the association between each of 78 single nucleotide polymorphisms (SNPs) and 5 breast cancer subtypes. Subtypes were defined using 5 immunohistochemical markers: estrogen receptors (ER), progesterone receptors (PR), human epidermal growth factor receptors 1 and 2 (HER1/2) and cytokeratin (CK) 5/6. Several SNPs in *TNRC9/TOX3* were associated with luminal A (ER/PR+, HER2-) or basal-like breast cancer (ER-, PR-, HER2-, HER1 or CK 5/6+), and one SNP (rs3104746) was associated with both. SNPs in *FGFR2* were associated with luminal A, luminal B (ER/PR+, HER2+), or HER2+/ER- disease, but none were associated with basal-like disease. I also observed subtype differences in the effects of SNPs in 2q35, 4p, *TLR1*, *MAP3K1*, *ESR1*, *CDKN2A/B*, *ANKRD16*, and *ZM1Z1*. I found evidence that genetic risk factors for breast cancer vary by subtype and further clarified the role of several key susceptibility genes.

5.2 Introduction

Researchers have long recognized that breast cancer is a heterogeneous disease with variable prognoses and clinical characteristics. Further, epidemiologic investigations have discovered evidence of divergent etiologic processes [29, 89], with some key differences in risk factors across disease subgroups [4, 101-103]. While these findings have led to advancements in our understanding of the disease, inconsistent subtype definitions and imprecise estimates have hampered progress. Attempts to identify subtype-specific genetic risk factors have been especially discouraging, with little consistency across study populations [167, 168, 226].

Most investigators rely on immunohistochemical (IHC) analysis of estrogen receptors (ER), progesterone receptors (PR) and human epidermal growth factor receptors-2 (HER2) to define breast cancer subtypes. These markers are included in most routine clinical evaluations of breast tumors, as they are predictive of response to targeted therapies such as tamoxifen and trastuzumab. Based on concerns that these three markers did not adequately capture disease heterogeneity, researchers turned to gene expression analysis for more in-depth assessments. In one of the first large-scale gene expression analyses of breast tissue, Perou et al. [39] observed that tumors with similar expression patterns also had similar IHC subtypes. The only major exception was triple-negative tumors (i.e. ER-, PR- and HER2-), which clustered into two separate groups with different cytokeratin 5/6 (CK 5/6) and human epidermal growth factor receptor-1 (HER1) expression patterns.

This research led to a new classification system with 5 IHC markers serving as adequate, inexpensive surrogates for more complex gene expression profiles [3, 41, 42].

Because the CK 5/6 protein is usually present in basal epithelial cells but not in more differentiated luminal epithelial cells, the subtypes were designated as follows: Luminal A (ER or PR+, HER2-)**,** Luminal B (ER/PR+, HER2+), HER2+/ER-, and Basal-like (ER-, PR-, HER2-, HER1+ or CK 5/6+).

This subtype classification system has led to insights into racial disparities and furthered understanding of etiologic and prognostic differences between disease subgroups. Luminal A is the most common subtype, but subtype prevalence varies according to the age and race of the population [3, 5, 50, 53]. Notably, basal-like and other triple-negative tumors are more common in women of African descent [3, 4, 8, 50, 53, 63, 64]. For women diagnosed before 2000, those with HER2+/ER- and basal-like breast cancers had the poorest prognoses [3, 5, 53, 58]. The development and FDA approval of trastuzumab has since improved survival rates for women with HER2+ disease, but women with basal-like or other types of triple-negative disease still experience high short-term mortality [60, 86, 95]. This phenomenon likely explains some of the racial disparity in mortality between US African Americans and whites (30.5 versus 21.6 deaths per 100,000 women with breast cancer per year, 2009) [2].

In previous studies of subtype-specific determinants, luminal A breast cancer was associated with most established breast cancer risk factors, including family history of breast cancer, reproductive factors, decreased physical activity, increased alcohol consumption and high breast density [4, 62, 65, 66, 79, 82, 83, 103, 108, 109, 110, 111, 114, 115, 117, 118, 122, 342]. In case-only risk ratio analyses, women with a family history of the disease, a younger age at diagnosis, or an earlier age at menarche were more likely to have triple-negative than luminal A tumors. Triple-negative tumors were also relatively more common

in African Americans and in women who had more children but did not breastfeed. Risk factors for the rarer luminal B and HER2+/ER- subtypes are less well-established, but evidence suggests that African American race, family history of breast cancer, lack of breastfeeding and high alcohol consumption are risk factors for HER2+/ER- disease. Luminal B breast cancers are more common in younger women, but otherwise have similar risk profiles to luminal A tumors.

The ground-breaking discovery of the rare but highly penetrant *BRCA1* and *BRCA2* genes [130, 132] opened a floodgate of linkage analyses, candidate gene studies, and later, genome-wide association studies (GWAS). Since then, 52 single nucleotide polymorphisms (SNPs) have met the criteria for genome-wide "discovery" [18] and variants on six candidate genes (*ATM*, *CASP8*, *CHEK2*, *CTLA4*, *NBN*, and *TP53*) have "cumulative evidence of an association" [20]. Of the aforementioned variants, only *BRCA1* has been consistently linked to a particular subtype, with numerous studies observing associations between *BRCA1* mutations and triple-negative disease [41, 140, 141, 143] or increased basal marker expression [41, 148].

In an attempt to elucidate subtype-specific genetic risk factors for breast cancer and further our understanding of disease etiology, I estimated associations between breast cancer subtypes and several previously identified candidate gene and GWAS hits using women from the Carolina Breast Cancer Study (CBCS). This population is well-suited to answer this research question, as it is one of the few studies to have both a large proportion of African American participants and information on basal IHC markers. This evaluation is further enhanced by the use of Bayesian statistical methods, which improve effect estimate accuracy through the incorporation of prior information.

5.3 Methods

5.3.1 Study population

The CBCS is a population-based, case-control study of invasive and *in situ* breast cancer. The study was conducted in 24 North Carolina counties between 1993 and 2001. To be eligible, cases had to be between 20 and 74 years of age at the time of their diagnosis, with no prior history of breast cancer. Women with *in situ* breast cancer were eligible if they were diagnosed with ductal carcinoma *in situ* with microinvasion to a depth of 2 mm or lobular carcinoma *in situ* between 1996 and 2001.

Both invasive and *in situ* cases were identified using the North Carolina Central Cancer Registry's rapid case ascertainment program [272]. A main objective of the CBCS was to collect information on traditionally under-researched populations. Therefore, cases were randomly sampled at disproportionate rates based on race and age. This sampling strategy ensured approximately equal representation of African American and non-African American women, as well as younger (age<50) and older women (age 50+).

Throughout the study period, controls aged 20-64 years were selected from North Carolina Department of Motor Vehicles records and were probability matched to cases based on race and age group [293]. Controls aged 65-74 were selected form Health Care Financing Administration records in a similar fashion. Women with a history of breast cancer were excluded.

A study nurse conducted detailed in-home interviews of all cases and controls. During the interview, each participant answered questions about her reproductive, medical, and family history, and her exposure to several known or suspected breast cancer risk factors.

Each participant was also asked to confirm her age and race and provide a 30 ml blood sample. All participants provided written informed consent and cases were asked to release their medical records and tumor tissue. The Institutional Review Board at the University of North Carolina (UNC) approved this study.

The overall response rate was 77% for cases and 57% for controls. 90% of controls, or 1816 women, provided sufficient blood samples for inclusion in genotype analyses (1105 whites, 681 African Americans, 30 other race). 88% of cases provided blood samples (2039 women), but only 55% of cases provided both blood and tumor samples (748 whites, 502 African Americans, 10 other race). This included 247 *in situ* cases. Individuals who self-identified as a race other than white or African American were included in overall analyses but excluded from race-specific assessments because of small numbers.

5.3.2 IHC analysis

Tumor tissue and medical records were collected from area hospitals and sent to UNC. ER and PR status was abstracted from the patient's medical records, when available. If not available, ER and PR IHC assays were performed at the UNC Immunohistochemistry Core Laboratory. Tumors with more than 5% of cells showing nuclei-specific staining were considered receptor positive [283]. Agreement between medical records reports and UNC-run assays in 10% random samples of ER+ and ER- tumors was high (concordance = 81%, kappa = 0.62) [3].

All tumor samples with sufficient tissue were assayed for HER2, HER1 and CK 5/6. A case was considered HER2+ if at least 10% of observed cells showed signs of CB11 monoclonal antibody staining [284]. Tissue with any sign of cytoplasmic or membranous staining was considered positive for CK 5/6 or HER1, respectively [4, 42]. Due to the limited

amount of available tissue, *in situ* tumors were not evaluated for PR status and staining techniques for ER, HER2, HER1 and CK 5/6 status were slightly modified (see Livasy et al. [89]).

As described above, these subtypes were classified as follows: luminal A (ER+ and/or PR+, HER2-), luminal B (ER+ and/or PR+, HER2+), HER2+/ER- (ER-, PR-, HER2+), and basal-like (ER-, PR-, HER2-, HER1+ and/or CK 5/6+). Additionally, tumors negative for all five markers were grouped together as the 'unclassified' subtype.

5.3.3 SNP selection

Single nucleotide polymorphisms (SNPs) from ten early breast cancer GWAS [186, 187, 189, 194, 196, 197, 201, 205] or GWAS follow-up studies [208, 209] were selected for inclusion in this subtype evaluation study. I included SNPs from these studies that had genome-wide p-values below 10^{-5} in preliminary or pooled analyses. I also retained SNPs in *CASP8*, *ATM*, and *TP53*, some of the key genes identified in a recent comprehensive meta-analysis [20]. Lastly, I included a number of SNPs in the same gene as GWAS selected variants, most of which were originally selected to enhance coverage of these regions. In total, this analysis included 22 GWAS hits, 19 other GWAS-identified variants that fell short of genome-wide significance criteria, 21 SNPs from *CASP8*, *ATM*, or *TP53*, and 21 tag SNPs from select GWAS genes.

Each CBCS participant was genotyped at 144 ancestry informative markers. This genotype information was used to estimate each participant's proportion of African ancestry. When included in regression models, this ancestry proportion estimate should control confounding due to population stratification [275, 292].

5.3.4 Genotype analysis

The included SNPs were genotyped using either a Taqman panel (Applied Biosystems, Inc., Foster City, CA) or a Custom GoldenGate Genotyping assay (Illumina, Inc., San Diego, CA). The majority of SNPs were genotyped on the Illumina panel, as described previously [287]. The Taqman panel [340] included SNPs that had low Illumina design scores, failed the Illumina assay, or were identified as GWAS hits after the Illumina assays were performed. Eighty-one women with poor genotyping quality on the Illumina panel were assigned missing values for those SNPs. All of the SNPs selected for inclusion in this subtype analysis passed quality control tests, including those for call rate, assay intensity, and genotype clustering.

For each SNP I examined published studies to determine which allele was associated with an increased risk of breast cancer in previous analyses. This allele was designated as the risk allele. For whites, I selected risk alleles for all *ATM*, *CASP8*, and *TP53* SNPs based on the Zhang et al. meta-analysis [20]. For the remaining SNPs, I ascertained the risk allele in the initial GWAS [186, 187, 189, 194, 196, 197, 201, 205, 208, 209] and subsequent replication studies [11, 12, 162, 163, 166, 168, 169, 186-191, 193, 195, 198-200, 202-209, 211-216, 218-225, 228-230, 232-237, 241-243, 245-247, 250-256, 258, 262, 267, 295-301, 335]. In each case, if the 95% CI limits excluded the null, the OR for the specified allele was in the same direction as the initial study. Despite some minor discrepancies in the direction of the ORs in African American only studies, I assigned the same risk allele for both racial groups to allow pooling and facilitate comparisons. For novel SNPs and SNPs with no prior statistically significant findings, I designated the minor variant as the risk variant, using the HapMap CEU population as a reference.

5.3.5 Statistical methods

I calculated case-stratified descriptive statistics for age, proportion of African ancestry, and menopausal status, and then repeated these analyses for white and African American participants separately. I also examined overall and race-stratified distributions of stage at diagnosis, breast cancer subtype, and ER, PR, and HER2 status. Participants were weighted according to their inverse sampling probability. Similarly, all regression models included an offset term to account for the weighted sampling procedures.

For all SNPs, I calculated overall and race-stratified risk allele frequencies (RAFs). I tested for departures from Hardy-Weinberg equilibrium (HWE) separately in white and African American controls using Pearson's chi-squared test. If a SNP had a HWE p-value less than 0.05 in either population, I re-inspected the SNP's genotype clustering images for indications of poor genotype differentiation or other lab error.

I calculated ORs and 95% posterior intervals (PIs) for the association between each subtype and SNP using Bayesian polytomous logistic regression models. I assumed additive genetic models and adjusted for self-reported race (African American or non-African American), proportion of African ancestry, and age at diagnosis or selection. I centered age at 50 years and ancestry at its mean value. I also calculated race-specific ORs and 95% PIs, adjusting for age and ancestry.

Previous studies of the association between known susceptibility variants and breast cancer have produced ORs in the range of 1.1-1.3 [18, 20, 318], but subtype-specific associations are less well characterized. Bearing this in mind, I assigned each SNP log OR a null-centered prior with a mean of 0, but selected a variance of $\tau^2 \sim 1/\Gamma(4, 0.5)$ to reflect the likely effect size. These parameters correspond to prior SNP-subtype ORs with 95% mass

between 0.54 and 1.86 when τ^2 is equal to the mode of the distribution (0.1). As a full Bayes approach requires priors for all parameters, I also assigned null-centered, lognormal priors for age, ancestry, race and the intercept term. I assigned relatively informative priors to age and ancestry (τ^2=0.68), which were both mean-centered variables, but a larger variance to race (τ^2=1.0). Because the intercept is difficult to define or interpret in a case-control study with weighted sampling, I assigned a vague prior, with τ^2=1000. I assumed that all priors were independent.

Priors were incorporated into regression models using Bayes' theorem. Briefly, Bayes' theorem states the posterior probability distribution for the parameter of interest given the observed data, $f(\beta|D)$, is proportional to the likelihood of the observed data, $L(\beta;D)$, multiplied by the prior probability distribution $f(\beta)$ [14, 315, 343]. The aforementioned likelihood is identical to the likelihood used to obtain the maximum likelihood estimate (MLE) in a standard frequentist logistic regression model. Put another way, the posterior OR is essentially an inverse-variance weighted combination of the likelihood and prior distribution. Further, the variance of the resulting, normal posterior distribution is the inverse of the sum of the weights.

I also conducted sensitivity analyses, estimating MLE of ORs and 95% CIs and another set of Bayesian ORs and 95% PIs given a more informative, but still null-centered prior [$SNP \sim N(0, \tau^2)$, $\tau^2 \sim 1/\Gamma(3, 0.2)$, with mode at 0.05]. For each Bayesian model, I took 50,000 samples, discarding the first 1000 draws as a burn in, and thinning by retaining every tenth draw, such that the results are based on 4990 samples. Autocorrelation, trace, and density plots indicated adequate mixing and model convergence. All analyses were conducted using the SAS procedure MCMC (v9.3, Cary, NC). Example code is provided in

the appendix.

5.4 Results

As seen in Table 16, white and African American participants in the CBCS population differed in a few key ways. African Americans were more likely to have later stage disease, with 63% presenting at stage II or higher, relative to 48% of whites. African Americans were also less likely to be postmenopausal at the time of their diagnosis and were more likely to have basal-like (22% vs. 11%), unclassified (14% vs. 8%) or HER2+/ER- disease (8% vs. 6%). Luminal A breast cancer was the most common breast cancer subtype overall (60%). Seven SNPs had HWE p-values<0.05 (Table 17), though no SNPs failed HWE tests in both whites and African Americans. Upon re-inspection of genotype clustering images, I found that six of the seven SNPs showed good differentiation with no overlap between genotypes. I excluded the seventh SNP, rs614367 (*MYEOV*), after discovering evidence of allelic dropout and observing disparate clustering within the homozygous rare genotype. I also excluded SNPs with minor allele frequencies less than 1% in this sample. This left me with 78 SNPs in the overall analysis, 76 in the white only analysis and 73 in the African American only analysis.

Subtype-specific ORs and 95% PIs for all participants are presented in Table 18. Several SNPs were associated with luminal A breast cancer, including 13 of 14 evaluated *FGFR2* SNPs (ORs≈1.25) and several SNPs in *TNRC9/TOX3* (see Figure 30). The strongest association was seen for rs3104746 on *TNRC9/TOX3* (OR=1.58, 95% PI: 1.24, 1.94). Other noteworthy associations included rs13387042 (2q35), rs12505080 (4p), rs7696175 (*TLR1*), rs889312 (*MAP3K1*), rs851974 (*ESR1*), rs1011970 (*CDKN2A/B*), and rs9894946 (*TP53*).

HER2+/ER- disease and unclassified disease were also strongly associated with several *FGFR2* and *TNRC9/TOX3* SNPs. For both subtypes the OR estimates for the FGFR2 SNPs were high, with many at or near 1.4. Beyond these key genes, HER2+/ER- disease was positively correlated with the designated risk variant at rs2046210 on *ESR1*, and rs704010 on *ZMIZ1*, but negatively correlated with the risk variant at rs7696175 on *TLR1*, rs3798577 on *ESR1*, and rs518394 on *CDKN2A/B*. The C allele at rs2380205 (*ANKRD16*) was inversely associated with the risk of unclassified breast cancer.

I identified relatively few susceptibility variants for luminal B breast cancer. All but one *FGFR2* SNP was associated with increased disease risk, but the observed effects were weaker than the other non basal-like subtype ORs and only one had a posterior interval that excluded the null (rs2981578). The risk allele at rs704010 on *ZMIZ1* was also associated with luminal B disease (OR=1.34, 95% PI: 0.96, 1.70).

None of the *FGFR2* SNPs were associated with an increased risk of basal-like breast cancer. In fact, most of the *FGFR2* ORs for basal-like disease were less than one. Risk variants at two *TNRC9*/*TOX3* SNPs (rs3014746 and rs3112562) were positively associated with basal-like disease, as were risk variants at rs704010 on *ZMIZ1* and rs2046210 on *ESR1*. Additionally, rs7696175 on *TLR1* and rs10941679 on *MRPS30* each had ORs greater than 1.2 for basal-like breast cancer, relative to controls.

Race-stratified subtype analyses revealed a few additional insights (Tables 19 and 20). The most striking was for rs10757278 on *CDKN2A/B*, where the A allele was positively associated with basal-like disease in whites (OR=1.19, 95% PI: 1.02, 1.39) but negatively associated with disease in African Americans (OR=0.75, 95% PI: 0.58, 0.94). Race-specific *FGFR2* and *TNRC9/TOX3* results can be seen in Figure 31.

The two 8q24 SNPs were strongly associated with luminal A breast cancer only among whites (OR=1.16, 95% PI: 0.98, 1.35 and OR=1.17, 95% PI: 1.00, 1.37 for rs13281615 and rs1562430, respectively). The same was true for a *TNRC9/TOX3* SNP (rs8051542 OR=1.16, 95% PI: 0.99, 1.35) and a *LSP1 SNP* (rs909116 OR=1.17, 95% PI: 0.99, 1.37). As for the other subtypes, rs3112562 and rs12443621 (*TNRC9/TOX3*) were strongly associated with luminal B (OR=0.64, 95% PI: 0.39, 0.89) and HER2+/ER- breast cancer (OR=1.53, 95% PI: 1.05, 2.09), respectively, only among whites. I observed no noteworthy findings in the African American only analyses.

Results from the MLE analysis and alternate Bayes analysis are presented in Tables 21 and 22. Compared with the MLE results, the ORs and PIs presented here are attenuated towards the null and are more precise. The ORs from the Bayesian analysis with more informative priors were further attenuated. The SNP-subtype association patterns were consistent across all methods.

5.5 Discussion

In this study of breast cancer subtypes and previously established susceptibility variants, I observed critical differences in subtype-specific genetic risk factors. The most conspicuous differences involved the *FGFR2* gene, where most of the 14 evaluated SNPs were associated with luminal A, HER2+/ER- and unclassified disease, but not basal-like disease. I also found evidence that SNPs on or near *TNRC9/TOX3* are differentially related to breast cancer subtype and that rs10757278 (*CDKN2A*/*B*) is differentially related to basal-like disease by race. SNPs in 2q35, 4p, *TLR1*, *MRPS30, MAP3K1*, *ESR1*, *ANKRD16*, *ZM1Z1*, and *TP53* may also be related to subtype-specific etiology.

As few other studies have employed these enhanced subtype definitions, it is difficult to compare my results with previous reports. Most prior investigations of this topic were limited to comparisons of a single hormone receptor, usually ER+ versus ER- disease [162, 169, 187, 192, 208, 216, 217, 222, 232, 236, 247, 259]. A few have looked at risk factors for combined ER, PR, HER2 status [166, 167, 226, 344], but to my knowledge, only one other study by Broeks et al. [168] has examined genetic risk factors according to all five IHC markers. Broeks et al. [168], Stevens et al. [167], and Han et al. [226] examined some of the SNPs included in this analysis, with some consistencies across populations.

The only *FGFR2* SNP examined by Broeks et al. [168] was rs2981582. They also observed positive associations between the T allele and luminal A disease and no association between the SNP and basal-like breast cancer. Their luminal B OR was in the same direction I observed, but of much greater magnitude. Stevens et al. and Han et al. also found near-null associations between rs2981582 and triple negative disease. Contrary to my findings, however, rs2981582 was not associated with HER2+/ER- disease in either study, nor was it associated with unclassified disease in Stevens et al. The effect estimates for rs2981582 and luminal A and B disease reported by Han et al. are similar to those seen in my study. I am the first to report subtype-specific estimates for any other *FGFR2* SNPs.

Broeks et al., Stevens et al., and Han et al. also evaluated one *TNRC9/TOX3* SNP, rs3803662. Both Broeks et al. and Han et al. observed a positive association with the T allele and luminal A breast cancer, as I did. However, these authors also observed associations between the T allele and luminal B and HER2+/ER- disease, where I found only a weak association with Luminal B and a near-null association with HER2+/ER- disease. Lastly,

Broeks et al. observed an association between rs3803662 and basal-like disease, which I did not observe.

These three study groups also assessed other SNPs included in this panel. While it is difficult to draw clear inferences from individual SNP-subtype analyses, these studies, together with mine, suggest that some important differences by subtype do exist. In addition to *FGFR2* and *TNRC9/TOX3*, the effects of rs2046210 (*ESR1*), rs13387042 (2q35), and rs889312 (*MAP3K1*) seem to vary according to subtype. Additional studies are needed to further clarify the role of these SNPs and the other potentially important genes identified in my investigation.

While this is one of the first studies to look at genetic risk factors for specific subtypes, breast cancer susceptibility loci are a commonly studied topic. Bayesian methods allowed me to use this plethora of prior information to generate more precise estimates. Assuming I selected reasonable priors, the results presented here will also be more accurate, on average, than those produced using frequentist methods that do not incorporate the wealth of information from prior studies. Further, by selecting null-centered, highly informative priors, bias resulting from these methods is likely to be towards the null [17]. In this way, this application of Bayesian methods also reduces the probability of observing false positive associations. I believe the priors specified here are reasonable given existing knowledge of breast cancer susceptibility variants, but I also provide alternate analyses that demonstrate the influence of my assumptions.

Within the CBCS population, African Americans were less likely than non-African Americans to provide blood for genotyping, but were more likely to have tumor tissue available for IHC analysis. Women with advanced disease were also more likely to provide

tumor tissue. These trends may result in biased effect estimates for SNPs related to race or disease aggressiveness. Controlling for self-reported race and ancestry should alleviate some of this bias. Though not included in this analysis, I could have used inverse-probability of selection weighting or Bayesian imputation methods to further address this issue.

There is some disagreement in the field as to how best to classify breast cancer subtypes. As discussed, the IHC markers used here are only proxies for more complex gene expression profiles, and thus may not sufficiently capture tumor heterogeneity [44-47]. While my approach is likely more informative than one using three or fewer markers, poor subtype specification may attenuate effects and underestimate subtype differences. Misclassification due to inaccurate medical records or IHC evaluations could also bias effects. Other potential sources of misclassification include allelic drop-out and other genotyping errors, though thorough quality control checks likely limited the impact of such errors.

I included *in situ* cases to increase sample size and improve precision. While this could bias effect estimates of SNPs associated with disease aggressiveness or progression, shared risk profiles [4, 341, 345] and subtype distributions [4, 89, 346] suggest this bias would be small.

The diverse composition of the CBCS population is a major strength of this study. By recruiting a large proportion of African Americans, study investigators generated a population uniquely suited to answer questions about race and subtype differences in risk factors. To date, this is the largest study to evaluate breast cancer subtypes using a five-marker panel and one of the largest population-based studies of breast cancer in African Americans.

This analysis of previously established breast cancer susceptibility loci provides strong evidence of etiologic heterogeneity across breast cancer subtypes. Though likely only a small part of the carcinogenic process, the risk variants identified here offer valuable clues about the nature of these diverse pathways. In turn, this vital information may help to advance disease prevention and control efforts.

Table 16: Descriptive statistics for Carolina Breast Cancer Study participants included in subtype analysis

	Cases			Controls		
	Overall (%)* N=1260**	**White (%)* N=748**	**African Americans (%)* N=502**	**Overall (%)* N=1816****	**White (%)* N=1105**	**African Americans (%)* N=681**
Age (years); mean (std)	51.5 (11.6)	52.1 (11.8)	50.8 (11.4)	52.5 (11.3)	53.0 (11.2)	51.9 (11.3)
Proportion African Ancestry; mean (std)	0.35 (0.36)	0.06 (0.06)	0.77 (0.13)	0.33 (0.36)	0.07 (0.09)	0.77 (0.14)
Postmenopausal; N (%)	691 (67)	417 (70)	269 (56)	1032 (38)	640 (37)	377 (41)
Stage of Disease; N (%)						
In situ	247 (10)	192 (10)	52 (10)			
Stage I	369 (39)	232 (42)	136 (28)			
Stage II	492 (42)	250 (40)	237 (50)			
Stage III	100 (7)	46 (6)	54 (11)			
Stage IV	24 (2)	12 (2)	11 (2)			
missing	28	16	12			
Subtype; N (%)						
Luminal A	700 (60)	453 (64)	242 (49)			
Luminal B	122 (11)	82 (11)	38 (8)			
HER2+/ER-	98 (6)	59 (6)	39 (8)			
Basal-like	207 (13)	94 (11)	112 (22)			
Unclassified	133 (9)	60 (8)	71 (14)			
ER+; N (%)	745 (66)	497 (70)	243 (50)			
PR+; N (%)	543 (60)	341 (64)	197 (45)			
missing	247	192	52			
HER2+; N (%)	220 (17)	141 (17)	77 (16)			

*percentages weighted by inverse sampling probability

**includes those who self-identified as a race other than white or African American

Table 17: Risk allele frequencies (RAF) by race and case status, African Americans and non African Americans in the Carolina Breast Cancer Study

			All (1260 cases, 1817 controls)**		Whites (748 cases, 1105 controls)			African Americans (502 cases, 681 controls)		
Gene	Locus	Risk allele	RAF cases*	RAF controls*	RAF cases*	RAF controls*	HWE p-value	RAF cases*	RAF controls*	HWE p-value
1p12	rs11249433	G	0.37	0.36	0.44	0.41	0.54	0.14	0.10	0.01
CASP8	rs1045485	G	0.89	0.88	0.88	0.87	0.63	0.94	0.95	0.74
CASP8	rs17468277	C	0.89	0.88	0.88	0.87	0.63	0.95	0.95	0.95
2q35	rs13387042	A	0.59	0.52	0.55	0.47	0.83	0.73	0.73	1.00
2p	rs4666451	G	0.64	0.65	0.60	0.63	0.30	0.78	0.77	0.12
SLC4A7	rs4973768	T	0.44	0.42	0.47	0.42	0.22	0.35	0.40	0.05
4p	rs12505080	C	0.27	0.23	0.30	0.24	0.80	0.18	0.17	0.64
TLR1	rs7696175	T	0.38	0.38	0.46	0.45	0.91	0.09	0.06	0.54
MRPS30	rs4415084	T	0.46	0.45	0.42	0.42	0.18	0.64	0.58	0.70
MRPS30	rs10941679	G	0.27	0.28	0.29	0.30	0.76	0.19	0.19	0.17
5p12	rs981782	T	0.60	0.65	0.51	0.59	0.26	0.91	0.91	0.60
5q	rs30099	T	0.11	0.10	0.10	0.10	0.40	0.15	0.12	0.75
MAP3K1	rs889312	C	0.33	0.35	0.33	0.34	0.85	0.33	0.36	0.08
ESR1	rs2046210	A	0.43	0.40	0.38	0.35	0.48	0.62	0.61	0.15
ESR1	rs851974	G	0.36	0.39	0.41	0.43	0.28	0.18	0.17	0.46
ESR1	rs2077647	A	0.50	0.49	0.50	0.49	0.64	0.51	0.51	0.16
ESR1	rs2234693	T	0.50	0.55	0.51	0.57	0.45	0.46	0.48	0.63
ESR1	rs1801132	C	0.79	0.78	0.76	0.76	0.43	0.90	0.88	0.36
ESR1	rs3020314	C	0.44	0.41	0.38	0.34	0.15	0.68	0.71	0.75
ESR1	rs3798577	T	0.52	0.53	0.50	0.53	0.43	0.58	0.54	0.27
ECHDC1	rs2180341	G	0.26	0.28	0.24	0.27	0.55	0.32	0.33	0.83
RELN	rs17157903	T	0.14	0.12	0.14	0.12	0.06	0.11	0.10	0.08
8q24	rs13281615	G	0.43	0.42	0.43	0.42	0.17	0.43	0.43	0.58
8q24	rs1562430	T	0.59	0.56	0.60	0.57	0.78	0.53	0.53	0.61
CDKN2A/B	rs3731257	T	0.20	0.21	0.23	0.23	0.24	0.08	0.11	0.89

CDKN2A/B	rs3731249	A	0.02	0.02	0.03	0.03	0.90	0.01	0.00	0.95
CDKN2A/B	rs518394	G	0.36	0.41	0.44	0.48	0.17	0.09	0.08	0.06
CDKN2A/B	rs564398	G	0.35	0.40	0.42	0.47	0.29	0.08	0.08	0.02
CDKN2A/B	rs1011970	T	0.22	0.18	0.19	0.15	0.62	0.33	0.34	0.14
CDKN2A/B	rs10757278	A	0.60	0.60	0.55	0.55	0.18	0.80	0.82	0.77
CDKN2A/B	rs10811661	C	0.14	0.18	0.16	0.20	0.02	0.07	0.07	0.24
ANKRD16	rs2380205	C	0.54	0.58	0.57	0.60	0.88	0.41	0.46	0.72
ZNF365	rs10995190	G	0.85	0.83	0.86	0.82	0.76	0.82	0.83	0.90
ZMIZ1	rs704010	T	0.35	0.36	0.42	0.42	0.93	0.11	0.08	0.82
FGFR2	rs1896395	A	0.05	0.04	0.00	0.00	0.96	0.21	0.20	0.04
FGFR2	rs3750817	C	0.71	0.65	0.65	0.60	0.16	0.91	0.88	0.83
FGFR2	rs10736303	G	0.61	0.55	0.54	0.49	0.19	0.86	0.84	0.75
FGFR2	rs11200014	A	0.40	0.38	0.45	0.41	0.65	0.22	0.21	0.75
FGFR2	rs2981579	T	0.49	0.45	0.46	0.41	0.51	0.61	0.61	0.10
FGFR2	rs1078806	G	0.40	0.38	0.45	0.41	0.53	0.22	0.21	0.99
FGFR2	rs2981578	C	0.61	0.55	0.54	0.49	0.09	0.86	0.84	0.45
FGFR2	rs1219648	G	0.44	0.40	0.44	0.39	0.35	0.47	0.41	0.57
FGFR2	rs2912774	A	0.47	0.42	0.44	0.40	0.26	0.58	0.55	0.07
FGFR2	rs2936870	T	0.47	0.43	0.44	0.40	0.25	0.59	0.56	0.14
FGFR2	rs2420946	T	0.45	0.41	0.43	0.39	0.21	0.54	0.52	0.03
FGFR2	rs2162540	G	0.45	0.41	0.42	0.39	0.28	0.54	0.52	0.41
FGFR2	rs2981582	T	0.44	0.40	0.42	0.39	0.30	0.50	0.49	0.96
FGFR2	rs3135718	G	0.46	0.42	0.43	0.39	0.23	0.58	0.54	0.65
10q	rs10510126	C	0.89	0.90	0.89	0.89	0.38	0.89	0.90	0.21
ATM	rs1800054	G	0.02	0.01	0.02	0.02	0.34	0.00	0.00	0.94
ATM	rs4986761	C	0.01	0.01	0.01	0.01	0.68	0.00	0.00	0.98
ATM	rs1800056	C	0.02	0.01	0.02	0.01	0.67	0.00	0.00	0.95
ATM	rs1800057	G	0.03	0.02	0.04	0.02	0.90	0.01	0.01	0.91
ATM	rs1800058	T	0.02	0.02	0.02	0.02	0.06	0.01	0.01	0.91
ATM	rs1801516	A	0.12	0.12	0.15	0.14	0.17	0.03	0.02	0.48
ATM	rs3092992	C	0.04	0.04	0.05	0.04	0.13	0.01	0.01	0.77
ATM	rs664143	C	0.60	0.61	0.58	0.60	0.70	0.67	0.68	0.45

ATM	rs170548	G	0.27	0.33	0.32	0.37	0.88	0.09	0.12	0.07
ATM	rs3092993	A	0.12	0.12	0.15	0.14	0.19	0.03	0.02	0.48
LSP1	rs3817198	C	0.29	0.31	0.32	0.34	0.18	0.17	0.17	0.16
LSP1	rs909116	T	0.57	0.56	0.53	0.52	0.20	0.72	0.72	0.96
MYEOV	rs614367	T	0.16	0.12	0.17	0.11	0.05	0.14	0.15	0.33
H19	rs2107425	C	0.65	0.65	0.70	0.68	0.74	0.50	0.53	0.42
TNRC9/TOX3	rs8049149	T	0.00	0.00	0.00	0.00	0.98	0.01	0.02	0.32
TNRC9/TOX3	rs16951186	T	0.05	0.04	0.01	0.01	0.75	0.17	0.19	0.95
TNRC9/TOX3	rs8051542	T	0.43	0.41	0.46	0.44	0.43	0.33	0.30	0.12
TNRC9/TOX3	rs12443621	G	0.50	0.43	0.51	0.41	0.39	0.48	0.51	1.00
TNRC9/TOX3	rs3803662	T	0.36	0.29	0.32	0.24	0.73	0.52	0.54	0.65
TNRC9/TOX3	rs4784227	T	0.24	0.19	0.29	0.22	0.62	0.09	0.07	0.59
TNRC9/TOX3	rs3104746	A	0.08	0.05	0.03	0.02	0.48	0.26	0.18	0.87
TNRC9/TOX3	rs3112562	G	0.28	0.25	0.22	0.20	0.45	0.51	0.46	0.88
TNRC9/TOX3	rs9940048	A	0.26	0.25	0.25	0.24	0.50	0.31	0.30	0.64
TP53	rs9894946	G	0.84	0.87	0.81	0.86	0.48	0.96	0.96	0.25
TP53	rs1614984	T	0.41	0.39	0.41	0.39	0.22	0.39	0.40	0.03
TP53	rs4968187	T	0.00	0.00	0.00	0.00	0.93	0.01	0.00	0.92
TP53	rs12951053	C	0.08	0.07	0.08	0.06	0.47	0.11	0.11	0.09
TP53	rs17880604	C	0.01	0.01	0.01	0.01	0.21	0.00	0.00	0.95
TP53	rs1800372	G	0.01	0.01	0.02	0.02	0.54	0.01	0.00	0.98
TP53	rs2909430	G	0.17	0.14	0.14	0.13	0.66	0.28	0.24	0.64
TP53	rs1042522	C	0.67	0.71	0.75	0.77	0.64	0.40	0.43	0.77
TP53	rs8079544	C	0.94	0.94	0.96	0.95	1.00	0.89	0.89	0.83
COX11	rs7222197	G	0.70	0.73	0.72	0.75	0.60	0.66	0.65	0.70
COX11	rs6504950	G	0.70	0.73	0.72	0.75	0.59	0.66	0.65	0.66

*Weighted by inverse sampling probability

**Includes individuals who identified as a race other than white or African American

Table 18: Odds ratios and 95% posterior intervals for the association between the selected single nucleotide polymorphisms (SNPs) and each breast cancer subtype, relative to controls [SNP log OR~N(0,τ^2), $\tau^2 \sim \Gamma^{-1}(4, 0.5)$ with mode=0.10]*

Gene	SNP	Luminal A N=700	Luminal B N=122	HER2+/ER- N=98	Unclassified N=133	Basal-like N=207
1p12	rs11249433	1.09 (0.94, 1.25)	1.04 (0.78, 1.33)	1.11 (0.79, 1.45)	1.24 (0.93, 1.58)	1.09 (0.85, 1.35)
CASP8	rs1045485	1.10 (0.88, 1.32)	1.00 (0.67, 1.35)	0.97 (0.62, 1.33)	1.24 (0.80, 1.73)	1.15 (0.80, 1.56)
CASP8	rs17468277	1.11 (0.89, 1.35)	0.96 (0.64, 1.34)	0.94 (0.60, 1.32)	1.29 (0.84, 1.82)	1.10 (0.79, 1.47)
2q35	rs13387042	1.18 (1.04, 1.35)	0.99 (0.76, 1.25)	0.98 (0.69, 1.25)	1.10 (0.85, 1.40)	0.92 (0.73, 1.12)
2p	rs4666451	0.97 (0.85, 1.11)	1.03 (0.76, 1.29)	1.14 (0.84, 1.48)	1.24 (0.93, 1.56)	1.16 (0.91, 1.41)
SLC4A7	rs4973768	0.98 (0.86, 1.09)	1.01 (0.78, 1.27)	0.91 (0.68, 1.17)	0.93 (0.70, 1.15)	0.99 (0.78, 1.19)
4p	rs12505080	1.18 (1.02, 1.36)	1.14 (0.84, 1.45)	1.15 (0.82, 1.54)	0.88 (0.61, 1.13)	0.98 (0.75, 1.21)
TLR1	rs7696175	1.20 (1.02, 1.38)	0.99 (0.74, 1.29)	0.79 (0.54, 1.04)	1.08 (0.77, 1.38)	1.27 (0.96, 1.59)
MRPS30	rs4415084	1.11 (0.96, 1.25)	1.06 (0.80, 1.32)	1.15 (0.85, 1.49)	1.08 (0.84, 1.35)	1.17 (0.93, 1.41)
MRPS30	rs10941679	1.10 (0.93, 1.27)	0.94 (0.69, 1.22)	1.04 (0.73, 1.36)	0.95 (0.69, 1.23)	1.21 (0.95, 1.45)
5p12	rs981782	0.90 (0.78, 1.03)	0.88 (0.64, 1.12)	1.08 (0.78, 1.41)	0.99 (0.72, 1.28)	1.07 (0.82, 1.33)
5q	rs30099	1.09 (0.89, 1.31)	1.03 (0.66, 1.37)	1.13 (0.75, 1.56)	0.91 (0.60, 1.20)	0.99 (0.73, 1.28)
MAP3K1	rs889312	1.17 (1.02, 1.33)	1.04 (0.80, 1.32)	1.00 (0.74, 1.30)	1.08 (0.82, 1.35)	0.89 (0.70, 1.09)
ESR1	rs2046210	1.03 (0.90, 1.16)	1.07 (0.82, 1.36)	1.29 (0.97, 1.67)	1.15 (0.88, 1.44)	1.29 (1.01, 1.55)
ESR1	rs851974	0.89 (0.77, 1.02)	1.13 (0.84, 1.43)	0.86 (0.62, 1.12)	0.88 (0.64, 1.10)	0.96 (0.76, 1.18)
ESR1	rs2077647	0.95 (0.83, 1.08)	1.00 (0.76, 1.25)	0.98 (0.72, 1.26)	0.94 (0.71, 1.16)	0.96 (0.76, 1.17)
ESR1	rs2234693	0.91 (0.79, 1.04)	0.96 (0.75, 1.20)	1.02 (0.75, 1.27)	0.86 (0.67, 1.08)	0.99 (0.79, 1.18)
ESR1	rs1801132	1.09 (0.92, 1.28)	0.91 (0.66, 1.20)	0.98 (0.67, 1.31)	1.02 (0.74, 1.32)	0.85 (0.64, 1.05)
ESR1	rs3020314	1.02 (0.89, 1.17)	1.23 (0.93, 1.54)	1.08 (0.80, 1.37)	0.99 (0.77, 1.23)	1.12 (0.90, 1.34)
ESR1	rs3798577	0.97 (0.86, 1.10)	1.05 (0.79, 1.32)	0.82 (0.62, 1.02)	1.10 (0.85, 1.35)	1.00 (0.82, 1.21)
ECHDC1	rs2180341	0.99 (0.84, 1.13)	1.01 (0.75, 1.31)	1.04 (0.74, 1.36)	0.95 (0.70, 1.20)	1.01 (0.80, 1.23)
RELN	rs17157903	1.00 (0.83, 1.18)	1.07 (0.74, 1.41)	0.77 (0.48, 1.09)	0.99 (0.70, 1.32)	1.09 (0.78, 1.39)
8q24	rs13281615	1.04 (0.90, 1.17)	0.99 (0.75, 1.25)	1.13 (0.85, 1.43)	1.08 (0.85, 1.35)	0.98 (0.79, 1.17)
8q24	rs1562430	1.06 (0.93, 1.19)	0.96 (0.72, 1.22)	1.20 (0.90, 1.55)	1.06 (0.83, 1.32)	1.07 (0.85, 1.28)
CDKN2A/B	rs3731257	0.92 (0.78, 1.06)	0.89 (0.64, 1.17)	0.96 (0.66, 1.26)	0.91 (0.66, 1.19)	0.91 (0.70, 1.16)

CDKN2A/B	rs3731249	0.99 (0.65, 1.33)	1.08 (0.55, 1.68)	1.11 (0.56, 1.76)	0.95 (0.49, 1.46)	0.96 (0.51, 1.48)
CDKN2A/B	rs518394	0.99 (0.86, 1.13)	1.00 (0.73, 1.26)	0.77 (0.52, 1.00)	1.04 (0.76, 1.34)	1.14 (0.88, 1.38)
CDKN2A/B	rs564398	1.01 (0.85, 1.15)	1.01 (0.76, 1.28)	0.81 (0.56, 1.08)	1.08 (0.80, 1.40)	1.09 (0.83, 1.36)
CDKN2A/B	rs1011970	1.12 (0.96, 1.29)	0.91 (0.65, 1.15)	0.90 (0.63, 1.15)	0.87 (0.63, 1.10)	1.12 (0.86, 1.38)
CDKN2A/B	rs10757278	1.05 (0.90, 1.18)	1.06 (0.80, 1.35)	1.02 (0.75, 1.31)	1.22 (0.93, 1.52)	1.10 (0.86, 1.34)
CDKN2A/B	rs10811661	0.95 (0.78, 1.13)	1.01 (0.71, 1.35)	0.82 (0.54, 1.15)	0.84 (0.56, 1.12)	1.10 (0.81, 1.39)
ANKRD16	rs2380205	1.03 (0.90, 1.15)	1.05 (0.79, 1.30)	1.10 (0.82, 1.40)	0.80 (0.61, 0.99)	0.94 (0.77, 1.13)
ZNF365	rs10995190	1.03 (0.84, 1.22)	0.83 (0.59, 1.09)	0.83 (0.59, 1.11)	1.24 (0.88, 1.64)	0.95 (0.74, 1.21)
ZMIZ1	rs704010	1.09 (0.94, 1.25)	1.34 (0.96, 1.70)	1.36 (0.97, 1.77)	0.96 (0.70, 1.25)	1.34 (1.03, 1.66)
FGFR2	rs1896395	1.05 (0.80, 1.30)	0.92 (0.54, 1.33)	1.31 (0.73, 1.89)	1.06 (0.70, 1.44)	1.06 (0.71, 1.40)
FGFR2	rs3750817	1.33 (1.13, 1.53)	1.27 (0.92, 1.64)	1.22 (0.86, 1.60)	1.26 (0.92, 1.64)	1.01 (0.79, 1.25)
FGFR2	rs10736303	1.32 (1.14, 1.52)	1.31 (0.98, 1.65)	1.26 (0.90, 1.64)	1.37 (0.99, 1.77)	0.99 (0.78, 1.22)
FGFR2	rs11200014	1.26 (1.10, 1.43)	1.05 (0.80, 1.31)	1.19 (0.86, 1.52)	1.41 (1.08, 1.77)	0.88 (0.70, 1.07)
FGFR2	rs2981579	1.26 (1.10, 1.42)	1.11 (0.84, 1.41)	1.34 (1.00, 1.68)	1.44 (1.11, 1.80)	0.92 (0.74, 1.10)
FGFR2	rs1078806	1.24 (1.08, 1.42)	1.07 (0.79, 1.36)	1.19 (0.86, 1.53)	1.40 (1.06, 1.76)	0.87 (0.68, 1.07)
FGFR2	rs2981578	1.33 (1.15, 1.51)	1.31 (1.01, 1.67)	1.34 (0.98, 1.76)	1.38 (1.03, 1.84)	1.01 (0.79, 1.25)
FGFR2	rs1219648	1.29 (1.13, 1.47)	1.10 (0.83, 1.37)	1.24 (0.89, 1.59)	1.62 (1.24, 2.04)	0.95 (0.76, 1.15)
FGFR2	rs2912774	1.26 (1.10, 1.41)	1.11 (0.84, 1.39)	1.42 (1.06, 1.80)	1.47 (1.14, 1.85)	0.92 (0.73, 1.09)
FGFR2	rs2936870	1.26 (1.09, 1.41)	1.13 (0.86, 1.43)	1.38 (1.02, 1.76)	1.50 (1.16, 1.89)	0.91 (0.72, 1.09)
FGFR2	rs2420946	1.22 (1.06, 1.38)	1.06 (0.80, 1.32)	1.40 (1.04, 1.79)	1.46 (1.12, 1.83)	0.89 (0.71, 1.07)
FGFR2	rs2162540	1.28 (1.11, 1.45)	1.08 (0.82, 1.36)	1.42 (1.05, 1.83)	1.52 (1.17, 1.90)	0.91 (0.72, 1.10)
FGFR2	rs2981582	1.27 (1.09, 1.43)	1.10 (0.87, 1.40)	1.39 (1.02, 1.76)	1.28 (1.00, 1.57)	0.92 (0.73, 1.10)
FGFR2	rs3135718	1.26 (1.10, 1.42)	1.13 (0.85, 1.40)	1.35 (1.01, 1.71)	1.51 (1.17, 1.90)	0.91 (0.73, 1.09)
10q	rs10510126	1.09 (0.88, 1.30)	1.07 (0.73, 1.46)	1.08 (0.72, 1.49)	0.94 (0.63, 1.26)	1.14 (0.81, 1.48)
ATM	rs1800054	1.10 (0.66, 1.59)	1.10 (0.54, 1.76)	1.12 (0.51, 1.78)	1.00 (0.49, 1.62)	1.20 (0.58, 1.91)
ATM	rs1800057	1.19 (0.81, 1.64)	0.95 (0.49, 1.48)	1.33 (0.68, 2.13)	1.01 (0.50, 1.60)	1.10 (0.56, 1.67)
ATM	rs1800058	1.05 (0.67, 1.44)	0.94 (0.45, 1.46)	1.03 (0.51, 1.67)	0.98 (0.47, 1.56)	0.98 (0.49, 1.56)
ATM	rs1801516	1.03 (0.82, 1.24)	1.05 (0.70, 1.43)	0.97 (0.64, 1.37)	0.95 (0.63, 1.32)	0.99 (0.70, 1.32)
ATM	rs3092992	0.97 (0.68, 1.30)	0.99 (0.57, 1.49)	1.19 (0.62, 1.85)	1.05 (0.61, 1.57)	1.38 (0.79, 1.97)
ATM	rs664143	1.09 (0.95, 1.23)	1.00 (0.77, 1.25)	1.06 (0.79, 1.35)	0.94 (0.74, 1.17)	0.98 (0.79, 1.19)

ATM	rs170548	0.98 (0.84, 1.13)	0.90 (0.66, 1.16)	1.00 (0.72, 1.31)	0.94 (0.69, 1.22)	1.00 (0.77, 1.24)
ATM	rs3092993	1.03 (0.83, 1.25)	1.06 (0.69, 1.48)	0.94 (0.60, 1.35)	0.96 (0.61, 1.29)	1.03 (0.73, 1.37)
LSP1	rs3817198	1.02 (0.88, 1.18)	0.87 (0.62, 1.10)	1.19 (0.86, 1.52)	1.21 (0.92, 1.55)	1.02 (0.79, 1.26)
LSP1	rs909116	1.09 (0.94, 1.23)	0.92 (0.69, 1.15)	1.23 (0.90, 1.58)	1.03 (0.79, 1.30)	1.09 (0.86, 1.32)
H19	rs2107425	1.03 (0.90, 1.17)	0.93 (0.71, 1.17)	0.99 (0.74, 1.24)	0.94 (0.72, 1.18)	1.00 (0.81, 1.19)
TNRC9/TOX3	rs16951186	1.02 (0.78, 1.25)	1.30 (0.79, 1.88)	0.84 (0.48, 1.21)	0.83 (0.52, 1.12)	0.97 (0.67, 1.31)
TNRC9/TOX3	rs8051542	1.10 (0.97, 1.24)	0.96 (0.74, 1.20)	1.12 (0.84, 1.43)	1.31 (1.02, 1.64)	0.84 (0.66, 1.03)
TNRC9/TOX3	rs12443621	1.06 (0.94, 1.20)	0.93 (0.71, 1.16)	1.21 (0.92, 1.55)	0.95 (0.74, 1.18)	1.00 (0.82, 1.21)
TNRC9/TOX3	rs3803662	1.16 (1.01, 1.33)	1.09 (0.83, 1.35)	1.01 (0.77, 1.29)	1.13 (0.88, 1.41)	0.96 (0.78, 1.16)
TNRC9/TOX3	rs4784227	1.32 (1.13, 1.54)	1.09 (0.76, 1.43)	1.10 (0.76, 1.46)	1.29 (0.94, 1.67)	0.90 (0.66, 1.17)
TNRC9/TOX3	rs3104746	1.58 (1.24, 1.94)	1.05 (0.60, 1.50)	1.31 (0.80, 1.85)	1.12 (0.76, 1.58)	1.49 (1.06, 1.98)
TNRC9/TOX3	rs3112562	1.07 (0.93, 1.22)	0.88 (0.62, 1.14)	1.46 (1.06, 1.87)	0.80 (0.60, 1.03)	1.33 (1.06, 1.62)
TNRC9/TOX3	rs9940048	1.13 (0.97, 1.28)	0.84 (0.61, 1.08)	1.17 (0.86, 1.50)	0.98 (0.72, 1.25)	0.91 (0.72, 1.11)
TP53	rs9894946	0.86 (0.72, 1.02)	0.83 (0.59, 1.12)	1.01 (0.65, 1.36)	1.05 (0.71, 1.43)	1.12 (0.81, 1.49)
TP53	rs1614984	1.01 (0.88, 1.13)	0.94 (0.71, 1.16)	1.01 (0.78, 1.28)	1.04 (0.79, 1.30)	1.13 (0.92, 1.36)
TP53	rs12951053	0.95 (0.75, 1.18)	1.32 (0.86, 1.84)	1.16 (0.71, 1.62)	1.25 (0.82, 1.73)	0.99 (0.70, 1.30)
TP53	rs17880604	0.87 (0.49, 1.24)	0.84 (0.35, 1.38)	1.21 (0.54, 1.94)	0.96 (0.46, 1.57)	1.12 (0.57, 1.76)
TP53	rs1800372	0.97 (0.55, 1.38)	1.24 (0.57, 2.00)	0.93 (0.39, 1.53)	1.34 (0.59, 2.22)	1.04 (0.46, 1.71)
TP53	rs2909430	1.12 (0.96, 1.31)	1.09 (0.78, 1.45)	1.06 (0.73, 1.38)	0.88 (0.63, 1.13)	1.10 (0.85, 1.37)
TP53	rs1042522	1.03 (0.89, 1.18)	0.99 (0.73, 1.26)	0.82 (0.61, 1.06)	0.94 (0.70, 1.17)	0.97 (0.77, 1.18)
TP53	rs8079544	0.98 (0.78, 1.23)	1.38 (0.83, 2.07)	0.76 (0.45, 1.07)	0.87 (0.59, 1.20)	1.05 (0.71, 1.44)
COX11	rs7222197	1.06 (0.92, 1.22)	1.08 (0.80, 1.35)	1.01 (0.74, 1.29)	1.09 (0.82, 1.37)	1.05 (0.85, 1.27)
COX11	rs6504950	1.05 (0.91, 1.20)	1.07 (0.80, 1.37)	1.00 (0.73, 1.28)	1.08 (0.81, 1.35)	1.05 (0.84, 1.28)

*Estimates generated using polytomous logistic regression adjusting for age at diagnosis/selection, proportion African ancestry and self-reported race.

Figure 30: Odds ratios and 95% posterior intervals for FGFR2 and TNRC9/TOX3 SNPs, All CBCS participants

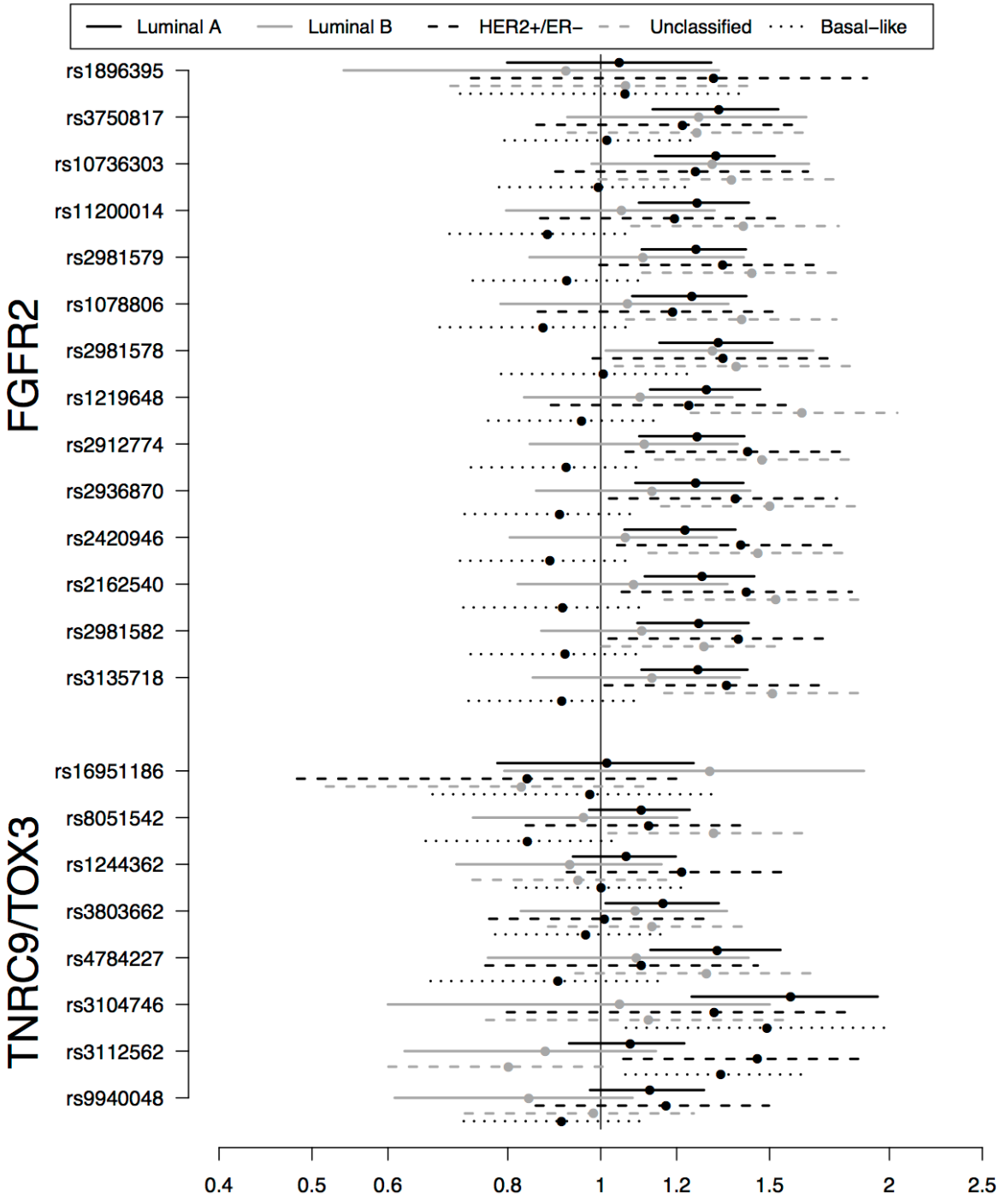


Table 19: Odds ratios and 95% posterior intervals for SNP-subtype associations in Carolina Breast Cancer Study whites, [SNP log OR~N(0,τ^2), $\tau^2 \sim \Gamma^{-1}(4, 0.5)$ with mode=0.10]*

Gene	SNP	Luminal A N=453	Luminal B N=82	HER2+/ER- N=59	Unclassified N=60	Basal-like N=94
1p12	rs11249433	1.04 (0.87, 1.20)	1.10 (0.81, 1.44)	0.98 (0.69, 1.30)	1.21 (0.83, 1.63)	1.05 (0.76, 1.35)
CASP8	rs1045485	1.13 (0.90, 1.40)	0.95 (0.62, 1.32)	0.93 (0.57, 1.30)	1.13 (0.70, 1.62)	1.39 (0.89, 1.99)
CASP8	rs17468277	1.12 (0.87, 1.39)	0.94 (0.60, 1.29)	0.94 (0.58, 1.33)	1.16 (0.70, 1.71)	1.36 (0.87, 1.94)
2q35	rs13387042	1.18 (1.00, 1.37)	1.07 (0.76, 1.37)	1.02 (0.72, 1.36)	1.02 (0.71, 1.36)	1.05 (0.78, 1.37)
2p	rs4666451	0.98 (0.83, 1.14)	1.12 (0.79, 1.47)	1.05 (0.72, 1.39)	1.11 (0.76, 1.46)	1.13 (0.81, 1.47)
SLC4A7	rs4973768	1.04 (0.87, 1.20)	1.10 (0.80, 1.43)	0.91 (0.62, 1.20)	0.97 (0.68, 1.27)	0.93 (0.67, 1.19)
4p	rs12505080	1.17 (0.97, 1.37)	1.04 (0.72, 1.37)	1.20 (0.80, 1.63)	0.91 (0.60, 1.25)	1.06 (0.75, 1.37)
TLR1	rs7696175	1.17 (0.99, 1.35)	0.98 (0.71, 1.27)	0.81 (0.53, 1.07)	1.11 (0.77, 1.45)	1.21 (0.89, 1.53)
MRPS30	rs4415084	1.19 (0.99, 1.39)	0.88 (0.62, 1.15)	1.28 (0.88, 1.71)	1.00 (0.68, 1.35)	1.16 (0.83, 1.50)
MRPS30	rs10941679	1.18 (0.99, 1.40)	0.93 (0.65, 1.26)	1.17 (0.77, 1.62)	0.97 (0.63, 1.32)	1.19 (0.88, 1.54)
5p12	rs981782	0.87 (0.73, 1.01)	0.80 (0.58, 1.05)	1.15 (0.81, 1.54)	0.94 (0.65, 1.26)	1.08 (0.77, 1.37)
5q	rs30099	1.05 (0.80, 1.31)	0.99 (0.61, 1.41)	0.95 (0.55, 1.42)	0.91 (0.52, 1.32)	0.94 (0.59, 1.33)
MAP3K1	rs889312	1.25 (1.04, 1.47)	1.19 (0.86, 1.60)	1.07 (0.73, 1.47)	1.21 (0.84, 1.64)	0.97 (0.70, 1.26)
ESR1	rs2046210	1.03 (0.87, 1.20)	1.12 (0.83, 1.47)	1.07 (0.75, 1.42)	1.18 (0.79, 1.56)	1.25 (0.91, 1.58)
ESR1	rs851974	0.88 (0.75, 1.03)	1.07 (0.74, 1.36)	0.86 (0.58, 1.16)	0.85 (0.58, 1.15)	0.94 (0.69, 1.20)
ESR1	rs2077647	0.97 (0.82, 1.13)	0.96 (0.69, 1.25)	0.97 (0.70, 1.29)	0.87 (0.61, 1.16)	0.93 (0.69, 1.19)
ESR1	rs2234693	0.88 (0.74, 1.02)	0.95 (0.69, 1.23)	1.05 (0.70, 1.40)	0.80 (0.54, 1.06)	1.01 (0.74, 1.28)
ESR1	rs1801132	1.05 (0.86, 1.25)	0.90 (0.62, 1.20)	0.88 (0.58, 1.21)	0.85 (0.58, 1.15)	0.82 (0.59, 1.08)
ESR1	rs3020314	1.02 (0.86, 1.19)	1.10 (0.77, 1.43)	1.16 (0.79, 1.57)	1.06 (0.74, 1.41)	1.34 (0.97, 1.73)
ESR1	rs3798577	0.96 (0.81, 1.09)	0.99 (0.73, 1.30)	0.77 (0.52, 1.04)	1.11 (0.78, 1.44)	0.97 (0.72, 1.26)
ECHDC1	rs2180341	0.94 (0.77, 1.13)	1.00 (0.67, 1.35)	0.95 (0.61, 1.34)	1.24 (0.80, 1.67)	0.90 (0.62, 1.21)
RELN	rs17157903	0.92 (0.74, 1.14)	0.97 (0.64, 1.37)	0.78 (0.43, 1.12)	1.12 (0.74, 1.59)	1.15 (0.78, 1.54)
8q24	rs13281615	1.16 (0.98, 1.35)	0.97 (0.70, 1.26)	1.11 (0.76, 1.47)	1.04 (0.73, 1.38)	0.99 (0.75, 1.28)
8q24	rs1562430	1.17 (1.00, 1.37)	0.95 (0.69, 1.23)	1.17 (0.79, 1.59)	1.34 (0.89, 1.79)	1.02 (0.76, 1.32)
CDKN2A/B	rs3731257	0.99 (0.82, 1.16)	0.94 (0.64, 1.24)	1.07 (0.74, 1.44)	0.95 (0.63, 1.29)	0.93 (0.63, 1.22)

CDKN2A/B	rs3731249	0.86 (0.53, 1.19)	1.11 (0.55, 1.72)	1.14 (0.59, 1.77)	0.97 (0.47, 1.53)	0.88 (0.44, 1.36)
CDKN2A/B	rs518394	1.00 (0.84, 1.16)	0.99 (0.71, 1.29)	0.76 (0.53, 1.01)	1.11 (0.78, 1.49)	1.17 (0.87, 1.50)
CDKN2A/B	rs564398	1.02 (0.86, 1.18)	1.04 (0.76, 1.33)	0.82 (0.53, 1.09)	1.16 (0.79, 1.54)	1.10 (0.82, 1.40)
CDKN2A/B	rs1011970	1.10 (0.89, 1.32)	0.99 (0.62, 1.37)	1.08 (0.66, 1.54)	0.98 (0.62, 1.40)	1.19 (0.82, 1.63)
CDKN2A/B	rs10757278	1.19 (1.02, 1.39)	1.07 (0.79, 1.41)	1.03 (0.71, 1.35)	1.08 (0.78, 1.45)	1.13 (0.85, 1.45)
CDKN2A/B	rs10811661	0.91 (0.73, 1.11)	1.01 (0.68, 1.38)	0.91 (0.58, 1.28)	0.89 (0.56, 1.28)	1.12 (0.78, 1.51)
ANKRD16	rs2380205	1.12 (0.94, 1.30)	1.08 (0.76, 1.39)	1.00 (0.70, 1.33)	0.81 (0.57, 1.08)	1.01 (0.73, 1.29)
ZNF365	rs10995190	1.03 (0.81, 1.28)	0.91 (0.58, 1.27)	0.80 (0.51, 1.14)	1.17 (0.73, 1.66)	1.10 (0.70, 1.53)
ZMIZ1	rs704010	1.12 (0.95, 1.32)	1.37 (0.96, 1.79)	1.33 (0.89, 1.77)	0.95 (0.64, 1.27)	1.26 (0.92, 1.61)
FGFR2	rs3750817	1.29 (1.08, 1.49)	1.26 (0.89, 1.65)	1.10 (0.78, 1.49)	1.11 (0.77, 1.50)	0.92 (0.69, 1.19)
FGFR2	rs10736303	1.36 (1.14, 1.57)	1.36 (0.91, 1.75)	1.08 (0.72, 1.45)	1.31 (0.92, 1.78)	0.90 (0.65, 1.16)
FGFR2	rs11200014	1.30 (1.10, 1.52)	1.24 (0.88, 1.62)	1.14 (0.76, 1.50)	1.55 (1.06, 2.06)	0.84 (0.61, 1.08)
FGFR2	rs2981579	1.36 (1.14, 1.57)	1.21 (0.85, 1.57)	1.15 (0.79, 1.53)	1.53 (1.06, 2.08)	0.84 (0.60, 1.06)
FGFR2	rs1078806	1.29 (1.09, 1.51)	1.23 (0.89, 1.58)	1.14 (0.75, 1.54)	1.51 (1.04, 2.02)	0.83 (0.61, 1.08)
FGFR2	rs2981578	1.38 (1.17, 1.62)	1.36 (0.96, 1.76)	1.20 (0.83, 1.58)	1.32 (0.88, 1.75)	0.90 (0.66, 1.17)
FGFR2	rs1219648	1.32 (1.13, 1.54)	1.21 (0.85, 1.58)	1.13 (0.74, 1.49)	1.42 (0.97, 1.92)	0.82 (0.61, 1.07)
FGFR2	rs2912774	1.31 (1.11, 1.51)	1.18 (0.83, 1.54)	1.14 (0.76, 1.51)	1.39 (0.96, 1.85)	0.81 (0.59, 1.05)
FGFR2	rs2936870	1.31 (1.11, 1.52)	1.18 (0.86, 1.57)	1.15 (0.81, 1.58)	1.36 (0.94, 1.78)	0.83 (0.61, 1.08)
FGFR2	rs2420946	1.31 (1.10, 1.51)	1.14 (0.82, 1.49)	1.14 (0.79, 1.53)	1.32 (0.90, 1.74)	0.83 (0.62, 1.08)
FGFR2	rs2162540	1.32 (1.10, 1.52)	1.13 (0.81, 1.46)	1.14 (0.78, 1.53)	1.34 (0.94, 1.78)	0.80 (0.58, 1.04)
FGFR2	rs2981582	1.32 (1.11, 1.53)	1.16 (0.83, 1.49)	1.16 (0.78, 1.56)	1.32 (0.89, 1.74)	0.82 (0.60, 1.07)
FGFR2	rs3135718	1.31 (1.10, 1.51)	1.13 (0.83, 1.46)	1.12 (0.76, 1.47)	1.31 (0.90, 1.71)	0.83 (0.61, 1.06)
10q	rs10510126	1.17 (0.89, 1.45)	1.09 (0.67, 1.51)	0.98 (0.58, 1.40)	0.97 (0.61, 1.37)	1.11 (0.71, 1.57)
ATM	rs1800054	1.01 (0.62, 1.45)	1.15 (0.55, 1.90)	1.19 (0.50, 1.99)	0.94 (0.42, 1.47)	1.31 (0.66, 2.07)
ATM	rs1800057	1.12 (0.75, 1.55)	0.89 (0.43, 1.42)	1.40 (0.68, 2.26)	0.99 (0.44, 1.60)	1.22 (0.67, 1.95)
ATM	rs1800058	1.08 (0.67, 1.53)	0.90 (0.43, 1.42)	1.01 (0.45, 1.62)	1.09 (0.52, 1.76)	1.00 (0.52, 1.59)
ATM	rs1801516	0.98 (0.77, 1.19)	1.07 (0.69, 1.49)	0.89 (0.54, 1.27)	0.91 (0.54, 1.29)	1.05 (0.69, 1.43)
ATM	rs3092992	0.93 (0.64, 1.25)	1.07 (0.57, 1.61)	1.17 (0.60, 1.85)	1.05 (0.54, 1.65)	1.37 (0.77, 2.04)
ATM	rs664143	1.07 (0.91, 1.24)	1.05 (0.77, 1.34)	1.25 (0.88, 1.66)	0.89 (0.63, 1.17)	1.00 (0.75, 1.26)
ATM	rs170548	1.03 (0.87, 1.22)	1.01 (0.72, 1.32)	1.17 (0.77, 1.57)	0.90 (0.61, 1.22)	0.98 (0.72, 1.29)

ATM	rs3092993	0.98 (0.76, 1.20)	1.06 (0.68, 1.46)	0.89 (0.54, 1.30)	0.90 (0.54, 1.28)	1.04 (0.70, 1.43)
LSP1	rs3817198	1.07 (0.90, 1.25)	0.84 (0.60, 1.10)	1.20 (0.81, 1.63)	1.07 (0.71, 1.43)	1.07 (0.79, 1.39)
LSP1	rs909116	1.17 (0.99, 1.37)	0.90 (0.65, 1.15)	1.06 (0.75, 1.43)	1.08 (0.72, 1.45)	1.07 (0.79, 1.38)
H19	rs2107425	1.11 (0.93, 1.31)	0.96 (0.68, 1.25)	1.10 (0.76, 1.48)	1.02 (0.67, 1.38)	1.19 (0.82, 1.54)
TNRC9/TOX3	rs8051542	1.16 (0.99, 1.35)	1.05 (0.75, 1.34)	1.07 (0.75, 1.42)	1.10 (0.76, 1.49)	0.91 (0.67, 1.16)
TNRC9/TOX3	rs12443621	1.22 (1.02, 1.40)	0.86 (0.62, 1.09)	1.53 (1.05, 2.09)	0.89 (0.61, 1.18)	1.18 (0.86, 1.51)
TNRC9/TOX3	rs3803662	1.34 (1.14, 1.59)	1.12 (0.79, 1.47)	1.16 (0.79, 1.58)	0.99 (0.67, 1.34)	0.99 (0.72, 1.28)
TNRC9/TOX3	rs4784227	1.30 (1.08, 1.53)	1.15 (0.80, 1.54)	1.13 (0.75, 1.56)	1.12 (0.76, 1.53)	1.00 (0.70, 1.32)
TNRC9/TOX3	rs3104746	1.41 (0.86, 2.01)	0.97 (0.40, 1.55)	1.15 (0.55, 1.90)	1.02 (0.44, 1.68)	1.08 (0.50, 1.75)
TNRC9/TOX3	rs3112562	0.96 (0.78, 1.15)	0.64 (0.39, 0.89)	1.58 (1.03, 2.16)	0.75 (0.45, 1.07)	1.32 (0.92, 1.75)
TNRC9/TOX3	rs9940048	1.07 (0.89, 1.26)	0.85 (0.59, 1.14)	1.04 (0.66, 1.41)	0.86 (0.57, 1.19)	0.97 (0.69, 1.27)
TP53	rs9894946	0.85 (0.68, 1.02)	0.91 (0.61, 1.27)	1.05 (0.66, 1.51)	1.05 (0.63, 1.46)	1.01 (0.69, 1.36)
TP53	rs1614984	1.03 (0.87, 1.19)	0.93 (0.67, 1.23)	1.09 (0.77, 1.49)	1.04 (0.71, 1.38)	1.10 (0.83, 1.42)
TP53	rs12951053	0.94 (0.70, 1.20)	1.19 (0.66, 1.73)	1.51 (0.83, 2.32)	1.04 (0.55, 1.54)	1.28 (0.74, 1.89)
TP53	rs17880604	0.83 (0.49, 1.23)	0.86 (0.32, 1.42)	1.25 (0.56, 2.13)	0.92 (0.31, 1.53)	1.10 (0.51, 1.84)
TP53	rs1800372	0.88 (0.51, 1.31)	1.25 (0.57, 2.07)	0.96 (0.41, 1.65)	1.12 (0.47, 1.87)	1.09 (0.48, 1.75)
TP53	rs2909430	1.15 (0.91, 1.41)	0.85 (0.49, 1.17)	0.96 (0.56, 1.38)	0.94 (0.60, 1.37)	1.09 (0.69, 1.51)
TP53	rs1042522	0.97 (0.81, 1.15)	1.25 (0.85, 1.69)	0.86 (0.59, 1.17)	1.10 (0.74, 1.52)	0.93 (0.67, 1.22)
TP53	rs8079544	1.01 (0.73, 1.33)	1.46 (0.76, 2.38)	1.03 (0.51, 1.61)	1.18 (0.60, 1.80)	1.21 (0.69, 1.88)
COX11	rs7222197	1.03 (0.85, 1.21)	1.01 (0.72, 1.33)	1.01 (0.67, 1.37)	1.22 (0.83, 1.66)	0.92 (0.66, 1.18)
COX11	rs6504950	1.02 (0.85, 1.20)	0.99 (0.69, 1.31)	1.02 (0.69, 1.37)	1.24 (0.84, 1.72)	0.92 (0.64, 1.19)

*Estimates generated using polytomous logistic regression adjusting for age at diagnosis/selection and proportion African ancestry.

Table 20: Odds ratios and 95% posterior intervals for SNP-subtype associations in Carolina Breast Cancer Study African Americans [SNP log OR~N(0,τ^2), $\tau^2 \sim \Gamma^{-1}(4, 0.5)$ with mode=0.10]*

Gene	SNP	Luminal A N=242	Luminal B N=38	HER2+/ER- N=39	Unclassified N=71	Basal-like N=112
1p12	rs11249433	1.23 (0.89, 1.61)	0.79 (0.41, 1.25)	1.36 (0.71, 2.07)	1.38 (0.86, 1.97)	1.16 (0.74, 1.58)
CASP8	rs1045485	1.04 (0.68, 1.43)	1.18 (0.58, 1.89)	1.08 (0.57, 1.71)	1.27 (0.69, 1.96)	0.83 (0.50, 1.21)
CASP8	rs17468277	1.12 (0.72, 1.56)	1.15 (0.60, 1.88)	1.06 (0.51, 1.67)	1.37 (0.71, 2.13)	0.82 (0.49, 1.22)
2q35	rs13387042	1.16 (0.89, 1.43)	0.92 (0.55, 1.32)	0.91 (0.57, 1.28)	1.15 (0.78, 1.56)	0.80 (0.59, 1.04)
2p	rs4666451	0.98 (0.76, 1.20)	0.84 (0.51, 1.22)	1.35 (0.82, 1.98)	1.32 (0.83, 1.81)	1.18 (0.85, 1.57)
SLC4A7	rs4973768	0.89 (0.71, 1.08)	0.87 (0.55, 1.26)	0.95 (0.61, 1.35)	0.92 (0.64, 1.21)	1.06 (0.78, 1.35)
4p	rs12505080	1.18 (0.91, 1.51)	1.25 (0.78, 1.81)	1.09 (0.64, 1.57)	0.95 (0.61, 1.32)	0.93 (0.65, 1.24)
TLR1	rs7696175	1.43 (0.95, 1.94)	1.09 (0.57, 1.69)	0.88 (0.39, 1.38)	1.07 (0.59, 1.58)	1.39 (0.86, 2.05)
MRPS30	rs4415084	1.01 (0.81, 1.22)	1.36 (0.88, 1.95)	1.01 (0.66, 1.42)	1.12 (0.79, 1.48)	1.18 (0.89, 1.52)
MRPS30	rs10941679	0.97 (0.74, 1.23)	0.95 (0.55, 1.39)	0.87 (0.50, 1.27)	0.96 (0.63, 1.32)	1.23 (0.87, 1.59)
5p12	rs981782	1.16 (0.79, 1.53)	1.23 (0.66, 1.88)	0.95 (0.53, 1.43)	1.00 (0.62, 1.49)	1.05 (0.67, 1.47)
5q	rs30099	1.08 (0.80, 1.36)	1.10 (0.64, 1.59)	1.30 (0.78, 1.90)	0.96 (0.63, 1.36)	1.07 (0.71, 1.46)
MAP3K1	rs889312	1.08 (0.85, 1.32)	0.79 (0.46, 1.13)	0.98 (0.60, 1.39)	0.94 (0.62, 1.29)	0.87 (0.63, 1.12)
ESR1	rs2046210	0.99 (0.78, 1.20)	1.05 (0.68, 1.48)	1.52 (0.95, 2.16)	1.10 (0.75, 1.46)	1.26 (0.93, 1.64)
ESR1	rs851974	0.96 (0.73, 1.20)	1.13 (0.66, 1.60)	0.96 (0.57, 1.40)	1.05 (0.66, 1.42)	1.03 (0.74, 1.38)
ESR1	rs2077647	0.96 (0.77, 1.14)	1.08 (0.69, 1.48)	0.99 (0.65, 1.36)	1.08 (0.73, 1.40)	1.00 (0.75, 1.28)
ESR1	rs2234693	0.96 (0.79, 1.16)	0.99 (0.64, 1.35)	1.00 (0.64, 1.37)	0.98 (0.71, 1.31)	0.96 (0.71, 1.21)
ESR1	rs1801132	1.21 (0.86, 1.60)	1.00 (0.56, 1.56)	1.26 (0.67, 1.99)	1.36 (0.81, 2.04)	0.95 (0.61, 1.34)
ESR1	rs3020314	0.98 (0.77, 1.20)	1.50 (0.85, 2.15)	0.96 (0.60, 1.34)	0.87 (0.61, 1.16)	0.88 (0.65, 1.14)
ESR1	rs3798577	1.02 (0.81, 1.22)	1.22 (0.77, 1.68)	0.97 (0.61, 1.32)	1.16 (0.84, 1.53)	1.05 (0.77, 1.33)
ECHDC1	rs2180341	1.03 (0.82, 1.26)	1.06 (0.69, 1.50)	1.13 (0.73, 1.61)	0.79 (0.55, 1.06)	1.13 (0.85, 1.48)
RELN	rs17157903	1.16 (0.83, 1.50)	1.23 (0.63, 1.84)	0.92 (0.45, 1.40)	0.91 (0.54, 1.35)	1.06 (0.67, 1.47)
8q24	rs13281615	0.85 (0.68, 1.03)	1.02 (0.63, 1.43)	1.09 (0.71, 1.51)	1.07 (0.77, 1.40)	0.99 (0.74, 1.26)
8q24	rs1562430	0.90 (0.72, 1.08)	0.97 (0.62, 1.37)	1.20 (0.78, 1.62)	0.85 (0.62, 1.12)	1.10 (0.81, 1.38)
CDKN2A/B	rs3731257	0.76 (0.49, 1.02)	0.91 (0.45, 1.38)	0.88 (0.45, 1.33)	1.02 (0.60, 1.47)	0.95 (0.62, 1.35)
CDKN2A/B	rs518394	0.98 (0.65, 1.28)	0.97 (0.51, 1.45)	0.94 (0.49, 1.42)	0.96 (0.54, 1.37)	1.10 (0.70, 1.55)

CDKN2A/B	rs564398	0.97 (0.67, 1.30)	1.02 (0.58, 1.61)	0.90 (0.43, 1.38)	0.95 (0.53, 1.36)	1.11 (0.68, 1.56)
CDKN2A/B	rs1011970	1.15 (0.92, 1.38)	0.92 (0.59, 1.28)	0.79 (0.48, 1.11)	0.80 (0.57, 1.07)	1.07 (0.80, 1.38)
CDKN2A/B	rs10757278	0.75 (0.58, 0.94)	1.11 (0.65, 1.68)	1.05 (0.62, 1.51)	1.35 (0.82, 1.91)	0.99 (0.69, 1.32)
CDKN2A/B	rs10811661	1.09 (0.74, 1.44)	1.00 (0.49, 1.58)	0.79 (0.37, 1.25)	0.91 (0.48, 1.33)	1.08 (0.67, 1.52)
ANKRD16	rs2380205	0.91 (0.72, 1.10)	1.08 (0.64, 1.49)	1.23 (0.80, 1.70)	0.85 (0.60, 1.09)	0.90 (0.67, 1.14)
ZNF365	rs10995190	1.04 (0.78, 1.30)	0.85 (0.49, 1.21)	0.92 (0.59, 1.36)	1.25 (0.79, 1.75)	0.87 (0.62, 1.15)
ZMIZ1	rs704010	0.90 (0.64, 1.20)	0.96 (0.50, 1.49)	1.19 (0.63, 1.84)	1.00 (0.57, 1.44)	1.40 (0.96, 1.95)
FGFR2	rs1896395	1.01 (0.77, 1.27)	0.93 (0.54, 1.35)	1.32 (0.79, 1.90)	1.03 (0.67, 1.38)	1.05 (0.73, 1.37)
FGFR2	rs3750817	1.38 (0.93, 1.87)	1.36 (0.67, 2.16)	1.53 (0.70, 2.54)	1.44 (0.81, 2.23)	1.29 (0.82, 1.84)
FGFR2	rs10736303	1.13 (0.83, 1.47)	1.14 (0.63, 1.75)	1.60 (0.82, 2.53)	1.27 (0.77, 1.81)	1.14 (0.75, 1.56)
FGFR2	rs11200014	1.16 (0.89, 1.44)	0.83 (0.49, 1.21)	1.25 (0.79, 1.76)	1.19 (0.77, 1.63)	0.96 (0.66, 1.28)
FGFR2	rs2981579	1.10 (0.90, 1.33)	1.02 (0.68, 1.44)	1.59 (0.96, 2.27)	1.29 (0.91, 1.72)	1.00 (0.74, 1.24)
FGFR2	rs1078806	1.18 (0.89, 1.44)	0.83 (0.50, 1.21)	1.26 (0.79, 1.80)	1.21 (0.82, 1.66)	0.98 (0.68, 1.28)
FGFR2	rs2981578	1.15 (0.86, 1.48)	1.18 (0.67, 1.82)	1.62 (0.84, 2.60)	1.34 (0.82, 1.91)	1.13 (0.76, 1.53)
FGFR2	rs1219648	1.19 (0.95, 1.45)	0.95 (0.60, 1.35)	1.38 (0.87, 1.96)	1.71 (1.17, 2.32)	1.09 (0.81, 1.39)
FGFR2	rs2912774	1.14 (0.91, 1.41)	1.04 (0.66, 1.47)	1.80 (1.15, 2.71)	1.44 (1.00, 1.93)	1.01 (0.74, 1.30)
FGFR2	rs2936870	1.15 (0.93, 1.38)	1.09 (0.69, 1.55)	1.73 (1.01, 2.56)	1.49 (1.04, 2.02)	0.97 (0.73, 1.23)
FGFR2	rs2420946	1.06 (0.85, 1.29)	0.99 (0.62, 1.39)	1.70 (1.04, 2.44)	1.50 (1.00, 1.98)	0.95 (0.71, 1.21)
FGFR2	rs2162540	1.14 (0.90, 1.37)	1.03 (0.65, 1.48)	1.71 (1.04, 2.48)	1.58 (1.08, 2.11)	1.02 (0.75, 1.28)
FGFR2	rs2981582	1.16 (0.94, 1.42)	1.05 (0.67, 1.46)	1.65 (1.08, 2.33)	1.19 (0.82, 1.60)	1.00 (0.75, 1.28)
FGFR2	rs3135718	1.14 (0.92, 1.38)	1.18 (0.73, 1.66)	1.64 (1.00, 2.36)	1.58 (1.11, 2.13)	1.00 (0.75, 1.26)
10q	rs10510126	0.97 (0.71, 1.25)	1.15 (0.58, 1.74)	1.24 (0.70, 1.92)	0.94 (0.60, 1.32)	1.17 (0.78, 1.69)
ATM	rs1801516	1.19 (0.69, 1.79)	0.88 (0.38, 1.43)	1.16 (0.52, 1.89)	1.16 (0.51, 1.82)	0.94 (0.49, 1.47)
ATM	rs664143	1.15 (0.91, 1.39)	0.99 (0.61, 1.34)	0.85 (0.54, 1.18)	1.03 (0.73, 1.37)	0.95 (0.71, 1.22)
ATM	rs170548	0.92 (0.62, 1.20)	0.78 (0.39, 1.28)	0.73 (0.36, 1.16)	1.07 (0.63, 1.56)	1.01 (0.65, 1.44)
ATM	rs3092993	1.19 (0.69, 1.79)	0.88 (0.38, 1.43)	1.16 (0.52, 1.89)	1.16 (0.51, 1.82)	0.94 (0.49, 1.47)
LSP1	rs3817198	0.87 (0.65, 1.11)	0.97 (0.56, 1.43)	1.13 (0.69, 1.61)	1.30 (0.90, 1.79)	1.00 (0.66, 1.35)
LSP1	rs909116	0.91 (0.72, 1.12)	0.98 (0.58, 1.40)	1.47 (0.88, 2.16)	0.95 (0.65, 1.26)	1.10 (0.77, 1.43)
H19	rs2107425	0.94 (0.75, 1.15)	0.92 (0.60, 1.30)	0.88 (0.57, 1.22)	0.93 (0.64, 1.22)	0.87 (0.65, 1.11)
TNRC9/TOX3	rs8049149	0.98 (0.49, 1.52)	1.10 (0.44, 1.89)	1.11 (0.47, 1.90)	0.73 (0.23, 1.20)	0.81 (0.34, 1.33)
TNRC9/TOX3	rs16951186	1.00 (0.74, 1.27)	1.32 (0.77, 1.87)	0.85 (0.48, 1.27)	0.75 (0.44, 1.05)	0.93 (0.64, 1.24)

TNRC9/TOX3	rs8051542	1.05 (0.81, 1.29)	0.90 (0.56, 1.28)	1.21 (0.74, 1.69)	1.44 (1.03, 1.96)	0.81 (0.57, 1.06)
TNRC9/TOX3	rs12443621	0.88 (0.70, 1.05)	1.17 (0.72, 1.65)	0.88 (0.57, 1.23)	1.01 (0.73, 1.34)	0.88 (0.65, 1.12)
TNRC9/TOX3	rs3803662	0.94 (0.76, 1.15)	1.10 (0.68, 1.53)	0.87 (0.56, 1.23)	1.19 (0.84, 1.58)	0.96 (0.73, 1.24)
TNRC9/TOX3	rs4784227	1.36 (0.92, 1.81)	0.89 (0.43, 1.43)	1.02 (0.49, 1.61)	1.65 (0.92, 2.47)	0.77 (0.43, 1.12)
TNRC9/TOX3	rs3104746	1.52 (1.14, 1.90)	1.12 (0.65, 1.64)	1.26 (0.73, 1.83)	1.16 (0.77, 1.60)	1.51 (1.02, 1.96)
TNRC9/TOX3	rs3112562	1.21 (0.93, 1.47)	1.27 (0.80, 1.75)	1.17 (0.77, 1.63)	0.87 (0.60, 1.14)	1.31 (0.96, 1.68)
TNRC9/TOX3	rs9940048	1.21 (0.94, 1.46)	0.93 (0.60, 1.33)	1.38 (0.88, 1.97)	1.14 (0.79, 1.54)	0.89 (0.64, 1.15)
TP53	rs9894946	1.03 (0.64, 1.44)	0.88 (0.42, 1.40)	0.99 (0.47, 1.67)	1.18 (0.57, 1.82)	1.32 (0.67, 2.11)
TP53	rs1614984	0.97 (0.77, 1.16)	1.03 (0.66, 1.41)	0.93 (0.60, 1.28)	1.00 (0.70, 1.30)	1.15 (0.84, 1.45)
TP53	rs12951053	0.97 (0.69, 1.29)	1.34 (0.73, 2.06)	0.81 (0.40, 1.26)	1.31 (0.82, 1.86)	0.85 (0.51, 1.21)
TP53	rs2909430	1.05 (0.82, 1.28)	1.29 (0.84, 1.79)	1.10 (0.69, 1.54)	0.89 (0.59, 1.18)	1.07 (0.79, 1.36)
TP53	rs1042522	1.16 (0.92, 1.41)	0.77 (0.48, 1.08)	0.87 (0.55, 1.23)	0.84 (0.57, 1.11)	1.03 (0.76, 1.31)
TP53	rs8079544	0.95 (0.69, 1.26)	1.12 (0.60, 1.75)	0.70 (0.37, 1.04)	0.75 (0.44, 1.05)	0.99 (0.61, 1.37)
COX11	rs7222197	1.08 (0.87, 1.32)	1.19 (0.76, 1.68)	1.05 (0.69, 1.45)	1.02 (0.71, 1.35)	1.17 (0.86, 1.49)
COX11	rs6504950	1.08 (0.87, 1.29)	1.21 (0.77, 1.71)	1.06 (0.70, 1.46)	0.99 (0.69, 1.30)	1.16 (0.87, 1.51)

*Estimates generated using polytomous logistic regression adjusting for age at diagnosis/selection and proportion African ancestry.

Figure 31: Odds ratios and 95% posterior intervals for FGFR2 and TNRC9/TOX3 SNPs for CBCS whites (left) and African Americans (right)

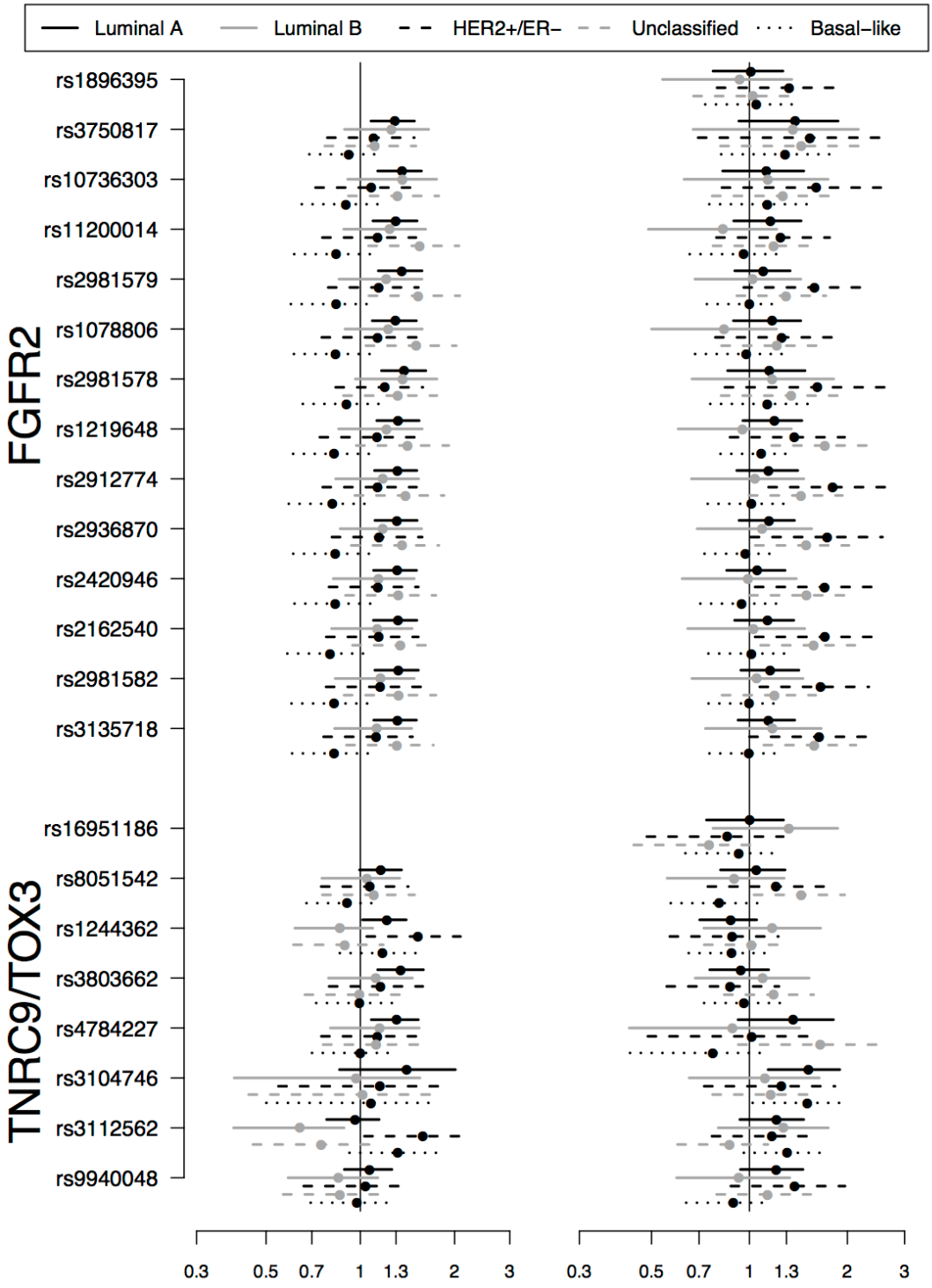


Table 21: Maximum likelihood odds ratios and 95% confidence intervals for the association between the selected single nucleotide polymorphisms (SNPs) and each breast cancer subtype, relative to controls*

Gene	SNP	Luminal A N=700	Luminal B N=122	HER2+/ER- N=98	Unclassified N=133	Basal-like N=207
1p12	rs11249433	1.10 (0.95, 1.28)	1.05 (0.78, 1.42)	1.15 (0.82, 1.61)	1.35 (1.00, 1.82)	1.12 (0.87, 1.44)
CASP8	rs1045485	1.12 (0.91, 1.39)	0.98 (0.64, 1.49)	0.92 (0.58, 1.45)	1.38 (0.85, 2.24)	1.16 (0.80, 1.69)
CASP8	rs17468277	1.14 (0.92, 1.41)	0.93 (0.61, 1.40)	0.89 (0.56, 1.42)	1.41 (0.86, 2.33)	1.13 (0.78, 1.65)
2q35	rs13387042	1.19 (1.04, 1.37)	0.98 (0.74, 1.29)	0.96 (0.70, 1.30)	1.10 (0.84, 1.45)	0.91 (0.73, 1.14)
2p	rs4666451	0.98 (0.85, 1.12)	1.03 (0.77, 1.37)	1.18 (0.85, 1.64)	1.28 (0.96, 1.72)	1.19 (0.94, 1.51)
SLC4A7	rs4973768	0.97 (0.85, 1.11)	1.01 (0.77, 1.32)	0.88 (0.65, 1.19)	0.91 (0.70, 1.18)	0.97 (0.78, 1.20)
4p	rs12505080	1.20 (1.03, 1.39)	1.18 (0.87, 1.60)	1.22 (0.87, 1.72)	0.86 (0.62, 1.20)	0.99 (0.76, 1.28)
TLR1	rs7696175	1.21 (1.04, 1.41)	0.99 (0.73, 1.36)	0.72 (0.49, 1.05)	1.10 (0.79, 1.53)	1.33 (1.02, 1.73)
MRPS30	rs4415084	1.12 (0.98, 1.29)	1.06 (0.80, 1.40)	1.20 (0.88, 1.64)	1.10 (0.84, 1.43)	1.20 (0.96, 1.51)
MRPS30	rs10941679	1.11 (0.95, 1.30)	0.91 (0.66, 1.27)	1.04 (0.73, 1.49)	0.95 (0.69, 1.31)	1.24 (0.97, 1.59)
5p12	rs981782	0.89 (0.77, 1.04)	0.83 (0.61, 1.13)	1.09 (0.76, 1.56)	0.93 (0.67, 1.29)	1.08 (0.82, 1.41)
5q	rs30099	1.08 (0.88, 1.33)	1.05 (0.69, 1.61)	1.17 (0.75, 1.83)	0.85 (0.56, 1.30)	0.99 (0.71, 1.38)
MAP3K1	rs889312	1.18 (1.03, 1.36)	1.04 (0.78, 1.40)	1.00 (0.72, 1.39)	1.09 (0.82, 1.44)	0.88 (0.69, 1.12)
ESR1	rs2046210	1.04 (0.90, 1.19)	1.09 (0.82, 1.44)	1.36 (1.00, 1.86)	1.18 (0.90, 1.54)	1.33 (1.07, 1.67)
ESR1	rs851974	0.89 (0.77, 1.02)	1.16 (0.87, 1.55)	0.81 (0.58, 1.14)	0.86 (0.64, 1.16)	0.93 (0.73, 1.19)
ESR1	rs2077647	0.95 (0.83, 1.08)	0.99 (0.75, 1.30)	0.95 (0.71, 1.29)	0.92 (0.71, 1.20)	0.95 (0.77, 1.18)
ESR1	rs2234693	0.89 (0.78, 1.02)	0.94 (0.72, 1.24)	1.01 (0.75, 1.36)	0.81 (0.62, 1.05)	0.96 (0.78, 1.20)
ESR1	rs1801132	1.09 (0.92, 1.29)	0.86 (0.62, 1.20)	0.93 (0.64, 1.37)	1.00 (0.70, 1.41)	0.81 (0.62, 1.06)
ESR1	rs3020314	1.02 (0.89, 1.17)	1.26 (0.95, 1.68)	1.09 (0.79, 1.50)	0.98 (0.74, 1.29)	1.13 (0.90, 1.42)
ESR1	rs3798577	0.97 (0.85, 1.10)	1.04 (0.80, 1.36)	0.77 (0.57, 1.03)	1.10 (0.86, 1.42)	0.99 (0.81, 1.23)
ECHDC1	rs2180341	0.98 (0.85, 1.14)	0.99 (0.73, 1.36)	1.03 (0.73, 1.44)	0.94 (0.70, 1.26)	1.01 (0.79, 1.29)
RELN	rs17157903	0.99 (0.82, 1.21)	1.07 (0.73, 1.58)	0.65 (0.39, 1.10)	0.98 (0.66, 1.45)	1.13 (0.83, 1.54)
8q24	rs13281615	1.04 (0.91, 1.18)	0.98 (0.75, 1.29)	1.14 (0.84, 1.54)	1.08 (0.84, 1.40)	0.99 (0.79, 1.22)
8q24	rs1562430	1.06 (0.93, 1.22)	0.93 (0.71, 1.22)	1.24 (0.92, 1.68)	1.09 (0.84, 1.41)	1.08 (0.87, 1.34)

CDKN2A/B	rs3731257	0.92 (0.77, 1.08)	0.85 (0.60, 1.21)	0.96 (0.66, 1.41)	0.89 (0.62, 1.26)	0.88 (0.66, 1.18)
CDKN2A/B	rs3731249	0.97 (0.61, 1.53)	1.14 (0.49, 2.65)	1.30 (0.52, 3.27)	0.69 (0.22, 2.20)	0.89 (0.38, 2.07)
CDKN2A/B	rs518394	0.98 (0.84, 1.14)	1.00 (0.74, 1.36)	0.69 (0.48, 1.00)	1.05 (0.76, 1.45)	1.16 (0.89, 1.51)
CDKN2A/B	rs564398	1.00 (0.86, 1.16)	1.03 (0.75, 1.40)	0.73 (0.50, 1.06)	1.09 (0.79, 1.52)	1.11 (0.85, 1.45)
CDKN2A/B	rs1011970	1.12 (0.96, 1.30)	0.88 (0.63, 1.24)	0.87 (0.60, 1.25)	0.82 (0.60, 1.12)	1.13 (0.89, 1.44)
CDKN2A/B	rs10757278	1.05 (0.91, 1.21)	1.07 (0.80, 1.44)	1.02 (0.74, 1.41)	1.26 (0.93, 1.70)	1.11 (0.87, 1.42)
CDKN2A/B	rs10811661	0.93 (0.77, 1.13)	1.00 (0.68, 1.48)	0.71 (0.43, 1.17)	0.77 (0.50, 1.18)	1.10 (0.80, 1.51)
ANKRD16	rs2380205	1.02 (0.90, 1.17)	1.07 (0.81, 1.40)	1.10 (0.82, 1.48)	0.76 (0.59, 0.98)	0.93 (0.75, 1.16)
ZNF365	rs10995190	1.02 (0.85, 1.22)	0.76 (0.54, 1.08)	0.75 (0.52, 1.09)	1.30 (0.89, 1.90)	0.92 (0.70, 1.22)
ZMIZ1	rs704010	1.11 (0.95, 1.29)	1.46 (1.08, 1.97)	1.49 (1.06, 2.08)	0.98 (0.71, 1.35)	1.41 (1.09, 1.82)
FGFR2	rs1896395	1.05 (0.80, 1.39)	0.82 (0.44, 1.54)	1.53 (0.89, 2.62)	1.05 (0.68, 1.63)	1.08 (0.74, 1.56)
FGFR2	rs3750817	1.37 (1.17, 1.59)	1.34 (0.98, 1.83)	1.29 (0.90, 1.85)	1.31 (0.94, 1.81)	1.04 (0.80, 1.34)
FGFR2	rs10736303	1.35 (1.17, 1.57)	1.38 (1.02, 1.86)	1.33 (0.95, 1.86)	1.45 (1.07, 1.98)	1.00 (0.78, 1.27)
FGFR2	rs11200014	1.28 (1.11, 1.47)	1.09 (0.82, 1.45)	1.25 (0.91, 1.72)	1.51 (1.15, 1.98)	0.87 (0.68, 1.11)
FGFR2	rs2981579	1.28 (1.12, 1.46)	1.13 (0.86, 1.48)	1.42 (1.05, 1.91)	1.54 (1.19, 1.99)	0.92 (0.75, 1.14)
FGFR2	rs1078806	1.26 (1.10, 1.45)	1.09 (0.82, 1.46)	1.26 (0.92, 1.73)	1.50 (1.14, 1.97)	0.87 (0.68, 1.11)
FGFR2	rs2981578	1.37 (1.18, 1.58)	1.39 (1.03, 1.87)	1.44 (1.03, 2.03)	1.48 (1.09, 2.01)	1.01 (0.79, 1.29)
FGFR2	rs1219648	1.31 (1.15, 1.50)	1.12 (0.85, 1.46)	1.31 (0.98, 1.77)	1.76 (1.36, 2.28)	0.96 (0.78, 1.20)
FGFR2	rs2912774	1.28 (1.12, 1.46)	1.13 (0.86, 1.47)	1.51 (1.12, 2.04)	1.55 (1.19, 2.00)	0.91 (0.73, 1.13)
FGFR2	rs2936870	1.28 (1.12, 1.46)	1.15 (0.88, 1.50)	1.49 (1.10, 2.00)	1.58 (1.22, 2.05)	0.90 (0.73, 1.12)
FGFR2	rs2420946	1.24 (1.09, 1.41)	1.07 (0.82, 1.40)	1.48 (1.10, 1.99)	1.54 (1.19, 2.00)	0.88 (0.71, 1.09)
FGFR2	rs2162540	1.29 (1.13, 1.47)	1.09 (0.83, 1.43)	1.50 (1.11, 2.01)	1.60 (1.24, 2.06)	0.90 (0.73, 1.12)
FGFR2	rs2981582	1.28 (1.12, 1.46)	1.13 (0.86, 1.47)	1.49 (1.11, 2.00)	1.33 (1.03, 1.72)	0.91 (0.73, 1.12)
FGFR2	rs3135718	1.28 (1.13, 1.46)	1.15 (0.88, 1.51)	1.44 (1.08, 1.94)	1.58 (1.23, 2.04)	0.91 (0.74, 1.13)
10q	rs10510126	1.11 (0.90, 1.36)	1.09 (0.71, 1.68)	1.11 (0.69, 1.79)	0.91 (0.62, 1.34)	1.18 (0.83, 1.68)
ATM	rs1800054	1.19 (0.69, 2.06)	1.37 (0.49, 3.88)	1.41 (0.43, 4.60)	0.81 (0.20, 3.40)	1.57 (0.66, 3.75)
ATM	rs1800057	1.31 (0.86, 2.01)	0.86 (0.31, 2.37)	2.02 (0.91, 4.46)	0.99 (0.36, 2.75)	1.24 (0.59, 2.64)
ATM	rs1800058	1.02 (0.62, 1.68)	0.76 (0.24, 2.43)	1.00 (0.31, 3.23)	0.85 (0.26, 2.75)	0.88 (0.35, 2.26)

ATM	rs1801516	1.01 (0.81, 1.26)	1.04 (0.66, 1.64)	0.89 (0.51, 1.53)	0.92 (0.55, 1.52)	1.00 (0.67, 1.49)
ATM	rs3092992	1.01 (0.70, 1.46)	1.04 (0.49, 2.21)	1.45 (0.68, 3.09)	1.23 (0.58, 2.62)	1.74 (1.00, 3.01)
ATM	rs664143	1.09 (0.95, 1.24)	1.00 (0.76, 1.31)	1.04 (0.77, 1.41)	0.93 (0.72, 1.20)	0.97 (0.78, 1.20)
ATM	rs170548	0.97 (0.83, 1.14)	0.87 (0.63, 1.21)	1.01 (0.71, 1.45)	0.93 (0.67, 1.30)	0.98 (0.75, 1.28)
ATM	rs3092993	1.01 (0.81, 1.27)	1.05 (0.67, 1.64)	0.89 (0.51, 1.53)	0.92 (0.56, 1.52)	1.00 (0.67, 1.49)
LSP1	rs3817198	1.02 (0.88, 1.19)	0.84 (0.61, 1.16)	1.25 (0.90, 1.75)	1.28 (0.96, 1.72)	1.04 (0.81, 1.35)
LSP1	rs909116	1.09 (0.95, 1.25)	0.88 (0.67, 1.17)	1.27 (0.92, 1.74)	1.02 (0.78, 1.34)	1.10 (0.88, 1.39)
H19	rs2107425	1.03 (0.89, 1.18)	0.90 (0.68, 1.19)	0.97 (0.71, 1.32)	0.92 (0.71, 1.21)	0.99 (0.79, 1.24)
TNRC9/TOX3	rs16951186	1.01 (0.77, 1.31)	1.47 (0.88, 2.45)	0.66 (0.34, 1.27)	0.70 (0.43, 1.13)	0.94 (0.64, 1.37)
TNRC9/TOX3	rs8051542	1.11 (0.97, 1.27)	0.95 (0.72, 1.26)	1.16 (0.85, 1.57)	1.36 (1.05, 1.77)	0.82 (0.65, 1.03)
TNRC9/TOX3	rs12443621	1.07 (0.94, 1.22)	0.91 (0.70, 1.20)	1.26 (0.94, 1.70)	0.93 (0.72, 1.20)	1.00 (0.81, 1.23)
TNRC9/TOX3	rs3803662	1.17 (1.02, 1.35)	1.10 (0.83, 1.46)	1.00 (0.73, 1.36)	1.12 (0.86, 1.47)	0.97 (0.77, 1.21)
TNRC9/TOX3	rs4784227	1.35 (1.15, 1.59)	1.12 (0.80, 1.58)	1.15 (0.79, 1.69)	1.42 (1.01, 1.98)	0.88 (0.65, 1.21)
TNRC9/TOX3	rs3104746	1.71 (1.35, 2.16)	1.08 (0.63, 1.84)	1.49 (0.90, 2.47)	1.20 (0.79, 1.82)	1.66 (1.20, 2.29)
TNRC9/TOX3	rs3112562	1.08 (0.94, 1.25)	0.82 (0.60, 1.13)	1.56 (1.13, 2.14)	0.75 (0.56, 1.01)	1.37 (1.09, 1.72)
TNRC9/TOX3	rs9940048	1.13 (0.98, 1.31)	0.81 (0.58, 1.12)	1.24 (0.89, 1.72)	0.97 (0.72, 1.30)	0.90 (0.70, 1.15)
TP53	rs9894946	0.85 (0.70, 1.03)	0.76 (0.52, 1.11)	0.97 (0.61, 1.55)	1.06 (0.68, 1.65)	1.11 (0.77, 1.61)
TP53	rs1614984	1.01 (0.88, 1.15)	0.92 (0.70, 1.21)	1.01 (0.75, 1.37)	1.03 (0.80, 1.34)	1.14 (0.93, 1.41)
TP53	rs12951053	0.96 (0.75, 1.22)	1.48 (0.96, 2.29)	1.28 (0.78, 2.10)	1.39 (0.92, 2.08)	1.01 (0.69, 1.49)
TP53	rs17880604	0.75 (0.40, 1.40)	0.31 (0.04, 2.29)	1.71 (0.60, 4.85)	0.77 (0.18, 3.22)	1.22 (0.47, 3.15)
TP53	rs1800372	0.96 (0.52, 1.77)	1.81 (0.69, 4.79)	0.48 (0.06, 3.57)	2.27 (0.86, 6.02)	1.14 (0.39, 3.30)
TP53	rs2909430	1.13 (0.96, 1.34)	1.09 (0.77, 1.55)	1.06 (0.73, 1.54)	0.85 (0.61, 1.18)	1.10 (0.85, 1.42)
TP53	rs1042522	1.02 (0.88, 1.18)	0.99 (0.73, 1.33)	0.78 (0.57, 1.07)	0.91 (0.69, 1.20)	0.95 (0.75, 1.19)
TP53	rs8079544	0.96 (0.75, 1.23)	1.71 (0.89, 3.30)	0.63 (0.39, 1.03)	0.80 (0.52, 1.23)	1.03 (0.69, 1.53)
COX11	rs7222197	1.06 (0.92, 1.22)	1.09 (0.81, 1.47)	1.01 (0.73, 1.39)	1.11 (0.84, 1.47)	1.05 (0.84, 1.32)
COX11	rs6504950	1.05 (0.91, 1.22)	1.08 (0.80, 1.45)	1.01 (0.73, 1.39)	1.08 (0.82, 1.42)	1.04 (0.83, 1.31)

*Estimates generated using polytomous logistic regression adjusting for age at diagnosis/selection, proportion African ancestry and self-reported race.

Table 22: Odds ratios and 95% posterior intervals for the association between the selected single nucleotide polymorphisms (SNPs) and each breast cancer subtype, relative to controls [SNP log OR~N(0,τ^2), $\tau^2 \sim \Gamma^{-1}(3, 0.2)$ with mode=0.05]*

Gene	SNP	Luminal A N=700	Luminal B N=122	HER2+/ER- N=98	Unclassified N=133	Basal-like N=207
1p12	rs11249433	1.08 (0.93, 1.22)	1.03 (0.77, 1.28)	1.09 (0.82, 1.40)	1.21 (0.90, 1.50)	1.09 (0.84, 1.32)
CASP8	rs1045485	1.09 (0.89, 1.30)	0.98 (0.71, 1.30)	0.97 (0.64, 1.27)	1.17 (0.82, 1.55)	1.11 (0.79, 1.40)
CASP8	rs17468277	1.11 (0.91, 1.33)	0.97 (0.68, 1.26)	0.97 (0.66, 1.28)	1.21 (0.83, 1.65)	1.07 (0.78, 1.39)
2q35	rs13387042	1.17 (1.02, 1.31)	1.00 (0.76, 1.24)	0.98 (0.73, 1.20)	1.09 (0.86, 1.34)	0.93 (0.75, 1.11)
2p	rs4666451	0.96 (0.85, 1.10)	1.02 (0.80, 1.28)	1.11 (0.83, 1.42)	1.19 (0.92, 1.50)	1.14 (0.92, 1.39)
SLC4A7	rs4973768	0.98 (0.86, 1.10)	1.01 (0.80, 1.24)	0.94 (0.70, 1.18)	0.94 (0.72, 1.15)	1.00 (0.82, 1.18)
4p	rs12505080	1.17 (1.00, 1.34)	1.11 (0.83, 1.42)	1.13 (0.81, 1.48)	0.89 (0.68, 1.14)	0.98 (0.77, 1.20)
TLR1	rs7696175	1.19 (1.03, 1.37)	1.01 (0.76, 1.25)	0.82 (0.59, 1.06)	1.04 (0.77, 1.33)	1.23 (0.96, 1.54)
MRPS30	rs4415084	1.10 (0.96, 1.24)	1.04 (0.80, 1.29)	1.14 (0.85, 1.46)	1.07 (0.84, 1.32)	1.15 (0.93, 1.39)
MRPS30	rs10941679	1.10 (0.95, 1.27)	0.94 (0.71, 1.17)	1.03 (0.77, 1.33)	0.96 (0.70, 1.21)	1.17 (0.94, 1.45)
5p12	rs981782	0.91 (0.78, 1.04)	0.91 (0.68, 1.12)	1.08 (0.81, 1.40)	0.98 (0.74, 1.23)	1.07 (0.85, 1.33)
5q	rs30099	1.09 (0.90, 1.29)	1.04 (0.76, 1.38)	1.08 (0.75, 1.44)	0.93 (0.64, 1.24)	1.01 (0.74, 1.26)
MAP3K1	rs889312	1.16 (1.00, 1.31)	1.03 (0.80, 1.31)	1.00 (0.75, 1.27)	1.06 (0.81, 1.33)	0.90 (0.73, 1.10)
ESR1	rs2046210	1.02 (0.89, 1.14)	1.06 (0.82, 1.33)	1.26 (0.94, 1.59)	1.13 (0.87, 1.39)	1.25 (0.99, 1.51)
ESR1	rs851974	0.91 (0.79, 1.03)	1.11 (0.85, 1.39)	0.88 (0.63, 1.11)	0.91 (0.71, 1.15)	0.95 (0.76, 1.15)
ESR1	rs2077647	0.95 (0.83, 1.06)	1.00 (0.78, 1.22)	0.98 (0.75, 1.20)	0.96 (0.76, 1.16)	0.97 (0.80, 1.16)
ESR1	rs2234693	0.91 (0.79, 1.03)	0.97 (0.77, 1.21)	1.01 (0.76, 1.25)	0.86 (0.65, 1.04)	0.97 (0.80, 1.16)
ESR1	rs1801132	1.09 (0.93, 1.27)	0.94 (0.71, 1.20)	0.97 (0.69, 1.25)	1.02 (0.73, 1.30)	0.87 (0.66, 1.07)
ESR1	rs3020314	1.01 (0.88, 1.13)	1.18 (0.92, 1.48)	1.06 (0.80, 1.37)	0.98 (0.76, 1.20)	1.10 (0.88, 1.31)
ESR1	rs3798577	0.97 (0.85, 1.10)	1.05 (0.84, 1.29)	0.84 (0.63, 1.05)	1.09 (0.87, 1.33)	1.00 (0.82, 1.18)
ECHDC1	rs2180341	0.98 (0.86, 1.12)	1.01 (0.78, 1.25)	1.03 (0.75, 1.31)	0.97 (0.76, 1.19)	1.01 (0.81, 1.22)
RELN	rs17157903	1.01 (0.83, 1.19)	1.06 (0.76, 1.37)	0.83 (0.57, 1.10)	0.99 (0.73, 1.30)	1.10 (0.86, 1.36)
8q24	rs13281615	1.03 (0.91, 1.16)	0.99 (0.77, 1.23)	1.09 (0.82, 1.36)	1.06 (0.84, 1.30)	0.98 (0.81, 1.18)
8q24	rs1562430	1.05 (0.93, 1.18)	0.96 (0.76, 1.19)	1.15 (0.87, 1.44)	1.05 (0.83, 1.29)	1.06 (0.88, 1.27)

CDKN2A/B	rs3731257	0.94 (0.80, 1.09)	0.92 (0.67, 1.15)	0.99 (0.70, 1.28)	0.94 (0.68, 1.19)	0.93 (0.71, 1.16)
CDKN2A/B	rs3731249	0.99 (0.67, 1.31)	1.06 (0.56, 1.55)	1.13 (0.67, 1.70)	0.93 (0.54, 1.38)	0.97 (0.60, 1.42)
CDKN2A/B	rs518394	0.99 (0.86, 1.12)	1.01 (0.76, 1.27)	0.81 (0.60, 1.06)	1.03 (0.77, 1.30)	1.13 (0.90, 1.38)
CDKN2A/B	rs564398	1.00 (0.87, 1.14)	1.02 (0.78, 1.27)	0.83 (0.59, 1.07)	1.06 (0.81, 1.35)	1.08 (0.84, 1.34)
CDKN2A/B	rs1011970	1.11 (0.95, 1.28)	0.93 (0.70, 1.17)	0.93 (0.66, 1.20)	0.88 (0.67, 1.11)	1.11 (0.88, 1.35)
CDKN2A/B	rs10757278	1.04 (0.90, 1.18)	1.06 (0.82, 1.34)	1.02 (0.77, 1.29)	1.17 (0.90, 1.46)	1.07 (0.86, 1.30)
CDKN2A/B	rs10811661	0.96 (0.80, 1.13)	1.01 (0.72, 1.31)	0.88 (0.61, 1.17)	0.87 (0.63, 1.14)	1.08 (0.83, 1.36)
ANKRD16	rs2380205	1.03 (0.90, 1.16)	1.05 (0.82, 1.29)	1.08 (0.83, 1.35)	0.82 (0.63, 1.00)	0.95 (0.78, 1.13)
ZNF365	rs10995190	1.03 (0.86, 1.21)	0.86 (0.61, 1.12)	0.85 (0.58, 1.13)	1.21 (0.90, 1.59)	0.94 (0.74, 1.20)
ZMIZ1	rs704010	1.08 (0.93, 1.23)	1.29 (0.97, 1.62)	1.26 (0.91, 1.64)	0.97 (0.72, 1.24)	1.30 (1.02, 1.60)
FGFR2	rs1896395	1.02 (0.79, 1.26)	0.95 (0.61, 1.30)	1.22 (0.79, 1.67)	1.05 (0.73, 1.37)	1.05 (0.74, 1.36)
FGFR2	rs3750817	1.30 (1.12, 1.51)	1.23 (0.90, 1.58)	1.18 (0.87, 1.52)	1.22 (0.91, 1.57)	1.02 (0.78, 1.26)
FGFR2	rs10736303	1.29 (1.13, 1.48)	1.26 (0.95, 1.59)	1.20 (0.90, 1.54)	1.31 (0.98, 1.68)	0.99 (0.79, 1.21)
FGFR2	rs11200014	1.24 (1.08, 1.41)	1.06 (0.83, 1.34)	1.17 (0.86, 1.48)	1.35 (1.04, 1.69)	0.88 (0.70, 1.06)
FGFR2	rs2981579	1.25 (1.09, 1.41)	1.11 (0.86, 1.39)	1.31 (0.97, 1.67)	1.41 (1.09, 1.74)	0.92 (0.75, 1.10)
FGFR2	rs1078806	1.23 (1.08, 1.39)	1.07 (0.83, 1.36)	1.17 (0.87, 1.48)	1.36 (1.04, 1.71)	0.89 (0.70, 1.08)
FGFR2	rs2981578	1.32 (1.13, 1.51)	1.27 (0.94, 1.62)	1.27 (0.95, 1.66)	1.35 (1.01, 1.75)	1.00 (0.81, 1.21)
FGFR2	rs1219648	1.28 (1.12, 1.45)	1.08 (0.83, 1.35)	1.24 (0.92, 1.59)	1.58 (1.22, 1.97)	0.96 (0.77, 1.16)
FGFR2	rs2912774	1.25 (1.08, 1.40)	1.08 (0.85, 1.35)	1.38 (1.03, 1.74)	1.41 (1.10, 1.74)	0.91 (0.74, 1.09)
FGFR2	rs2936870	1.25 (1.08, 1.41)	1.12 (0.85, 1.42)	1.35 (1.01, 1.72)	1.44 (1.12, 1.80)	0.91 (0.73, 1.09)
FGFR2	rs2420946	1.22 (1.07, 1.37)	1.05 (0.80, 1.31)	1.36 (1.02, 1.71)	1.42 (1.11, 1.77)	0.89 (0.71, 1.06)
FGFR2	rs2162540	1.26 (1.10, 1.41)	1.07 (0.82, 1.32)	1.36 (1.01, 1.75)	1.47 (1.14, 1.83)	0.90 (0.75, 1.08)
FGFR2	rs2981582	1.24 (1.08, 1.40)	1.08 (0.83, 1.35)	1.33 (1.01, 1.68)	1.25 (0.99, 1.55)	0.91 (0.75, 1.09)
FGFR2	rs3135718	1.25 (1.10, 1.41)	1.12 (0.87, 1.40)	1.31 (0.97, 1.66)	1.44 (1.12, 1.82)	0.91 (0.74, 1.08)
10q	rs10510126	1.08 (0.90, 1.29)	1.07 (0.73, 1.39)	1.06 (0.75, 1.43)	0.95 (0.68, 1.24)	1.12 (0.81, 1.43)
ATM	rs1800054	1.10 (0.74, 1.49)	1.09 (0.61, 1.59)	1.09 (0.52, 1.72)	0.97 (0.54, 1.45)	1.12 (0.64, 1.68)
ATM	rs1800057	1.15 (0.83, 1.54)	0.97 (0.56, 1.39)	1.20 (0.67, 1.80)	0.99 (0.58, 1.46)	1.08 (0.63, 1.54)
ATM	rs1800058	1.04 (0.70, 1.40)	0.96 (0.57, 1.42)	1.03 (0.58, 1.58)	1.02 (0.58, 1.50)	1.00 (0.59, 1.46)

ATM	rs1801516	1.02 (0.83, 1.22)	1.02 (0.71, 1.34)	0.96 (0.64, 1.30)	0.95 (0.64, 1.28)	1.01 (0.73, 1.29)
ATM	rs3092992	0.96 (0.70, 1.26)	1.00 (0.55, 1.42)	1.10 (0.65, 1.58)	1.08 (0.66, 1.53)	1.26 (0.80, 1.75)
ATM	rs664143	1.08 (0.94, 1.22)	1.00 (0.80, 1.26)	1.03 (0.78, 1.30)	0.96 (0.75, 1.16)	0.99 (0.80, 1.17)
ATM	rs170548	0.98 (0.85, 1.12)	0.92 (0.70, 1.18)	1.02 (0.79, 1.33)	0.95 (0.70, 1.19)	1.00 (0.78, 1.22)
ATM	rs3092993	1.01 (0.81, 1.20)	1.04 (0.72, 1.38)	0.95 (0.66, 1.27)	0.96 (0.66, 1.26)	1.01 (0.71, 1.32)
LSP1	rs3817198	1.01 (0.88, 1.15)	0.89 (0.66, 1.12)	1.15 (0.85, 1.47)	1.18 (0.92, 1.46)	1.02 (0.82, 1.26)
LSP1	rs909116	1.08 (0.93, 1.22)	0.93 (0.72, 1.14)	1.17 (0.87, 1.48)	1.03 (0.79, 1.27)	1.08 (0.87, 1.30)
H19	rs2107425	1.04 (0.91, 1.17)	0.94 (0.73, 1.18)	0.98 (0.75, 1.25)	0.96 (0.76, 1.18)	1.00 (0.81, 1.19)
TNRC9/TOX3	rs16951186	1.01 (0.78, 1.24)	1.23 (0.77, 1.74)	0.87 (0.55, 1.22)	0.84 (0.55, 1.11)	0.96 (0.69, 1.26)
TNRC9/TOX3	rs8051542	1.11 (0.98, 1.25)	0.97 (0.74, 1.23)	1.11 (0.83, 1.42)	1.26 (0.99, 1.56)	0.84 (0.68, 1.00)
TNRC9/TOX3	rs12443621	1.06 (0.94, 1.18)	0.94 (0.74, 1.14)	1.19 (0.90, 1.47)	0.95 (0.76, 1.17)	1.00 (0.81, 1.19)
TNRC9/TOX3	rs3803662	1.14 (0.99, 1.30)	1.09 (0.85, 1.36)	0.99 (0.74, 1.23)	1.10 (0.85, 1.36)	0.97 (0.79, 1.16)
TNRC9/TOX3	rs4784227	1.30 (1.11, 1.51)	1.07 (0.76, 1.37)	1.08 (0.78, 1.41)	1.24 (0.88, 1.61)	0.90 (0.66, 1.12)
TNRC9/TOX3	rs3104746	1.54 (1.20, 1.89)	1.02 (0.68, 1.38)	1.23 (0.80, 1.74)	1.11 (0.75, 1.50)	1.43 (1.02, 1.85)
TNRC9/TOX3	rs3112562	1.07 (0.92, 1.22)	0.88 (0.67, 1.13)	1.38 (0.99, 1.80)	0.81 (0.61, 1.01)	1.31 (1.05, 1.60)
TNRC9/TOX3	rs9940048	1.13 (0.98, 1.27)	0.87 (0.66, 1.10)	1.15 (0.86, 1.47)	0.98 (0.78, 1.22)	0.93 (0.74, 1.12)
TP53	rs9894946	0.87 (0.73, 1.03)	0.87 (0.62, 1.15)	1.01 (0.69, 1.34)	1.06 (0.75, 1.39)	1.08 (0.80, 1.39)
TP53	rs1614984	1.01 (0.89, 1.14)	0.96 (0.75, 1.18)	1.01 (0.79, 1.27)	1.02 (0.81, 1.24)	1.12 (0.91, 1.33)
TP53	rs12951053	0.94 (0.74, 1.16)	1.21 (0.84, 1.62)	1.12 (0.74, 1.51)	1.20 (0.85, 1.58)	1.00 (0.70, 1.30)
TP53	rs17880604	0.91 (0.56, 1.28)	0.90 (0.48, 1.38)	1.16 (0.59, 1.79)	0.99 (0.53, 1.46)	1.08 (0.59, 1.57)
TP53	rs1800372	0.97 (0.61, 1.35)	1.13 (0.62, 1.68)	0.95 (0.54, 1.45)	1.22 (0.66, 1.92)	1.02 (0.60, 1.51)
TP53	rs2909430	1.11 (0.95, 1.29)	1.07 (0.78, 1.36)	1.04 (0.76, 1.33)	0.91 (0.68, 1.14)	1.09 (0.87, 1.33)
TP53	rs1042522	1.04 (0.90, 1.19)	1.00 (0.76, 1.26)	0.85 (0.63, 1.07)	0.95 (0.74, 1.17)	0.97 (0.77, 1.18)
TP53	rs8079544	0.97 (0.75, 1.20)	1.26 (0.81, 1.79)	0.80 (0.53, 1.10)	0.90 (0.61, 1.24)	1.04 (0.74, 1.39)
COX11	rs7222197	1.05 (0.92, 1.19)	1.07 (0.81, 1.33)	1.00 (0.76, 1.26)	1.07 (0.84, 1.30)	1.04 (0.85, 1.24)
COX11	rs6504950	1.05 (0.91, 1.19)	1.06 (0.83, 1.32)	1.02 (0.77, 1.29)	1.05 (0.83, 1.32)	1.04 (0.84, 1.26)

*Estimates generated using polytomous logistic regression adjusting for age at diagnosis/selection, proportion African ancestry and self-reported race.

6. Discussion

6.1 Summary of findings

In this study of race and subtype differences in the effects of previously established susceptibility variants on breast cancer risk, I replicated several SNPs in white or African American CBCS participants and observed notable differences in subtype-specific genetic risk factors. I used Bayesian methods to incorporate prior information and obtain more precise effect estimates than would be possible with standard approaches, such as maximum likelihood.

6.1.1 Racial differences in breast cancer susceptibility loci

According to the ORs and 95% PIs estimated using Bayesian methods, 18 of 41 GWAS-identified SNPs replicated in whites and ten of 41 replicated in African Americans. None of the candidate gene hits replicated in either race. These findings are summarized in Table 23. This table also includes information on whether the ORs reported for CBCS were in the same direction as previous studies, regardless of their confidence limits. Because I evaluated all SNPs using the *a priori* risk variant as the index variant, any SNPs with ORs greater than one were consistent with previous findings.

In whites, I found evidence of positive associations between breast cancer and several well-validated GWAS-identified risk variants, including two *MRPS30* SNPs (rs4415084 and rs10941679) [169, 186, 190, 191, 208, 211, 220, 223, 228, 230], rs889312 on *MAP3K1* [162, 168, 169, 189, 194, 213, 218, 220, 232, 235, 239], several *FGFR2* SNPs [162, 168, 169, 186,

189-191, 194, 196, 201, 208, 213, 218-220, 222, 223, 229, 232-234, 236, 237, 251-258, 339, 347, 348], and several *TNRC9/TOX3* SNPs [11, 162, 163, 168, 169, 186, 187, 189, 191, 194, 202, 205, 217-220, 222, 229, 232, 234, 235, 237, 262]. Additionally, I replicated two less-established GWAS-identified SNPs, rs704010 in *ZMIZ1* [189, 247], and rs2107425 on *H19* [194]. As for the remaining SNPs, the large majority of estimated ORs were positive, indicating general agreement with the existing literature. This included rs909116 (*LSP1*) and rs1562430 (8q24), two GWAS-identified SNPs that fell just short of replication criteria in CBCS whites (Table 23).

A few of the non-GWAS, non-candidate gene SNPs were positively associated with breast cancer in CBCS whites (Table 24). This included rs10757278 in *CDKN2A/B*, rs3750817, rs11200014, and rs2162540 in *FGFR2* and rs3104746 in *TNRC9/TOX3*. Of these, rs2162540, rs10757278 and rs3104746 were novel findings, though rs2162540 was reported in a previous CBCS paper [210]. Results for rs3750817 agreed with those from two previous case-controls studies [169, 252]. The only *CASP8*, *ATM*, or *TP53* SNP associated with breast cancer was rs9894986 in *TP53*, which was negatively associated with the disease. The same variant allele had a near-null, positive association with disease in Zhang et al [20].

Nine of the ten SNPs that replicated in CBCS African Americans were in *FGFR2* (Table 23). This included rs1219648 [11, 12, 201, 215, 216, 236, 257] and rs2981578 [12, 214, 215, 258, 259], both of which were associated with the disease in several previous studies of African Americans. I also replicated rs2981579, rs2981582 and rs2420946. These three variants are known to increase breast cancer susceptibility in whites, but have rarely replicated in African Americans [11, 12, 169, 186, 194, 201, 214-216, 236, 257, 258, 259]. Barnholtz-Sloan et al. [210] previously reported the associations between breast cancer and

rs1219648, rs2981579, rs2981582 and rs2420946 in CBCS African Americans. Barnholtz-Sloan et al. were also the first to report ORs for rs2912774, rs2936870, and rs3135718, three *FGFR2* SNPs identified in a GWAS by Easton et al [194]. The ninth *FGFR2* SNP to replicate in CBCS African Americans was rs10736303. This SNP was also identified in Easton et al., but was not included in the previous CBCS analysis. It was positively associated with disease in one previously studied African American population [259].

The only other SNP to replicate in African Americans was rs2046210 in *ESR1*. Most previous evaluations of this SNP in African Americans generated near-null ORs for log-additive models [11, 12, 169, 214-216, 242, 243]. Two other SNPs, rs2107425 (*H19)* and rs12443621 (*TNRC9/TOX3*), had 95% PIs that excluded the null, but were inversely associated with breast cancer and thus inconsistent with the original reports. The ORs for the remaining GWAS-identified or candidate gene SNPs were again predominantly positive (Table 23), indicating concordance with previous findings. This included both of the SNPs with *a priori* racial differences in risk alleles (rs1045485 on *CASP8* and rs3803662 on *TNRC9/TOX3*) and a few SNPs that fell just short of the selected cut-point for replication (rs7696175 on *TLR1* and rs30099 on chromosome 5q).

CBCS is the first study to report statistically significant associations for rs3750817 and rs2162540 (*FGFR2*) in African Americans. rs2162540 was previously reported by Barnholtz-Sloan et al. [210], but I am the first to do so for rs3750817. I also corroborated previous evidence that rs3104746 and rs3112562 (*TNRC9/TOX3)* are associated with breast cancer in African Americans [214, 231] (Table 24).

Despite their relative imprecision in this application, the results from the hierarchical models support previous claims that these methods can help differentiate individual effects of

highly correlated SNPs [326, 328, 330]. Single-level models with the same number of parameters will often produce unstable estimates, but hierarchical models can simultaneously and stably estimate both haplotype-level and single-SNP effect estimates.

For example, even though none of the hierarchically-estimated *FGFR2* met the criteria for replication, the OR for rs2981579 in whites was notably higher than the rest of the SNPs in the 13-SNP *FGFR2* LD block. As it is unlikely that each of the *FGFR2* SNPs is independently associated with disease, the hierarchical results better reflect the presence of one or two causal variants in high LD with the other evaluated SNPs.

The *FGFR2* hierarchical models performed slightly better in African Americans, with rs3750817, rs2981578, and rs2420946 demonstrating notably stronger associations than the other SNP(s) in block 1, 2 and 3, respectively. This performance improvement may be attributable to the anticipated racial differences in LD block size. ORs for SNPs in the other, smaller LD blocks were relatively more precise, but contributed little new information. More explicit specifications of the incorporated covariance matrices had little impact on point estimates or precision in either racial group.

6.1.2 Subtype differences in breast cancer susceptibility loci

I observed some crucial differences in subtype-specific genetic risk factors. The most conspicuous differences involved the *FGFR2* gene, where most of the 14 evaluated SNPs were strongly associated with luminal A, HER2+/ER- and unclassified disease, but none were associated with basal-like breast cancer. In the only other study to examine the relationship between any *FGFR2* SNP and all five IHC subtypes, Broeks et al. [168] also observed a null association between rs2981582-T and basal-like disease and a positive association between rs2981582-T and luminal A disease. Their luminal B OR was in the

same direction as CBCS, but of much greater magnitude. Stevens et al. [167] and Han et al. [226] also found near-null associations between rs2981582 and triple negative disease, and Han et al. estimated ORs for luminal A and B disease similar to those seen in CBCS. Contrary to my findings, however, rs2981582 was not associated with HER2+/ER- disease in either study, nor was it associated with unclassified disease in Stevens et al.

None of the other *FGFR2* SNPs were evaluated using a 5-marker or even a 3-marker IHC panel, but many were assessed within categories of a single marker (e.g. ER+ cases relative to controls). In general, these studies reported positive associations between the *FGFR2* SNPs and ER+, ER-, PR+ and HER2- disease [162, 163, 168, 169, 192, 193, 198, 205, 208, 214, 216, 219, 222, 227, 232, 236, 244, 253, 255, 259, 349]
(Table 23 and 24). This included rs3750817, which met the criteria for genome-wide significance among women with hormone receptor positive disease [193].

Broeks et al. [168], Stevens et al. [167], and Han et al. [226] also evaluated rs3803662 in *TNRC9/TOX3*. This SNP had a positive association with luminal A breast cancer in CBCS, which was consistent with the ORs reported by Broeks et al. and Han et al. However, these authors also observed associations between rs3803662-T and luminal B and HER2+/ER- disease, where I found only a weak association with Luminal B and a near-null association with HER2+/ER- disease. Broeks et al. observed an association between rs3803662 and basal-like disease, which I did not replicate.

I also observed a positive association between rs4784227 (*TNRC9/TOX3*) and luminal A disease, which was consistent with Kim et al.'s [192] assessment of disease risk by ER and PR status. Similarly, I found that rs3104746 (*TNRC9/TOX3*) was associated with luminal A and basal-like disease, while Chen et al. [214] observed positive associations between

rs3104746-A and both ER+ and ER- breast cancer, relative to controls. The only other *TNRC9/TOX3* SNP with prior evidence of an association with a particular breast cancer subtype was rs8051542. This SNP was positively associated with unclassified disease in CBCS, but was not associated with ER- disease in any of three previous studies [205, 266, 349]. In contrast, all three reported positive associations between rs8051542 and ER+ breast cancer.

I found a positive association between rs2046210 (*ESR1*) and triple-negative breast cancer. This association was previously reported by Han et al [226]. I also replicated Han et al.'s finding that rs889312 (*MAP3K1*) was associated with luminal A breast cancer, but not their findings that rs2046210 and rs4973768 (*SLC4A7*) were associated with luminal A disease. Similarly, I did not replicate their finding that rs2046210 was associated with HER2+/ER- disease.

Broeks et al. [168] found that rs3817198 (*LSP1*) was associated with triple-negative breast cancer, but the same SNP was not associated with any subtype in the CBCS population. Further, where Broeks et al. found that rs13387042 (2q35) and rs889312 (*MAP3K1*) were associated with triple-negative breast cancer and that rs13387042 was associated with basal-like breast cancer, I found that these two SNPs were associated with luminal A breast cancer only. None of the susceptibility loci for triple-negative breast cancer highlighted by Stevens et al. were genotyped in CBCS [167].

According to my results, SNPs in 4p, *TLR1, ANKRD16*, and *ZM1Z1* may also be related to subtype-specific etiology. However, as with the majority of *FGFR2* and *TNRC9/TOX3* SNPs, it is difficult to compare my results with previous reports, as most prior investigations of this topic were limited to single hormone receptor comparisons [189, 192,

193, 198, 214, 216, 227, 247]. I have attempted to summarize the findings for all included SNPs in Tables 23 and 24. In general, those associated with luminal A breast cancer in CBCS were associated with ER+ disease in previous investigations. Few previous studies found SNPs associated with triple-negative or ER- disease, and those that did were not replicated in this investigation.

In the race-stratified subtype analyses, I found evidence that rs10757278 (*CDKN2A/B*) was differentially related to basal-like disease by race and that a few other SNP-subtype associations were much stronger when the analysis was limited to whites. This included rs1562430 (8q24) and rs3112562 and rs12443621 (*TNRC9/TOX3*). These findings support the hypothesis that there are both race and subtype differences in breast cancer susceptibility variants. Additional studies are needed to further clarify the role of these SNPs and the other potentially important genes identified in my investigation.

6.2 Strengths and limitations

The CBCS has some inherent limitations. These include selection bias, exposure and outcome misclassification, and informative missing data. Inappropriately specified genetic models and the inclusion of *in situ* breast cancer cases may also limit the validity of these particular analyses.

Differences in response rates by case-control status, race, and age may have introduced selection bias. Further, non African Americans were more likely than African Americans to provide blood for genotyping, but were less likely to have tumor tissue available for IHC analysis. Women with early stage disease were also less likely to provide tumor tissue.

Including race and age as model covariates should limit the bias resulting from this potentially informative missing data. However, given that women with more advanced disease were more likely to provide tumor tissue, the subtype-specific analyses may still be biased if one or more of the included SNPs is related to disease aggressiveness or medical care utilization. Though not included here, the extent of this bias could be explored further using inverse-probability of selection weighting or Bayesian imputation methods.

Despite quality control checks, some genotyping errors may be present. My decision to retain SNPs that violated HWE may have exacerbated this issue, but I only did so after verifying genotype clustering images and considering other potential threats to validity. This process led me to drop rs614367 (*MYEOV*). This particular SNP was not in HWE in white controls (p=0.05) and closer inspection of the TaqMan genotype results plots revealed disparate clusters for the homozygous rare genotype. I also discovered that the primer used for the analysis included a second polymorphic site. As this could result in failed amplification and allelic drop-out, I compared the CBCS MAFs for rs614367 to the MAFs in HapMap CEU and ASW populations. After finding these incompatible, I decided to exclude the SNP from further analysis. The genotype clusters for the remaining six SNPs that failed HWE tests formed three distinct clusters with no overlap and were therefore retained.

Age and self-identified race were collected and validated for all CBCS participants. I accounted for additional variation within self-identified racial groups by adjusting for proportion of African ancestry.

Classification of incident and *in situ* breast cancer cases is likely accurate, as disease status was first identified using a cancer registry and then confirmed by the participant. However, breast cancer subtypes are much more difficult to classify accurately. First of all,

as each tumor was stained and interpreted by a pathologist, there may have been inaccuracies due to human error. Additionally, even though CBCS investigators decided 81% concordance between medical record reports and UNC-run assays of ER and PR status was sufficiently high [273], additional misclassification due to inter-laboratory discrepancies in staining techniques or positivity cut-points was clearly present. Lastly, there is some disagreement in the field as to how best to classify breast cancer subtypes. The immunhistochemical markers are themselves proxies for gene expression patterns [39, 40] and while the majority of investigators support the subtype definitions described here, many authors do not differentiate between basal-like and unclassified breast cancer [8, 79, 85, 117], while others note that additional markers, such as claudin or Ki-67, may also be important [43- 46, 57, 350].

I included *in situ* cases to increase sample size and improve precision. Because most studies of genetic risk factors for breast cancer are limited to invasive breast cancer cases it is unclear whether the findings are generalizable to *in situ* cases. The inclusion of *in situ* cases could be particularly problematic if any of the included SNPs are associated with disease aggressiveness or if genotype affects the probability that an *in situ* case progresses into an invasive case. However, there is strong evidence that *in situ* tumors have similar subtype [4, 89, 346] and risk profiles [4, 341, 345] to invasive cases, and can therefore be included in the proposed analysis as early stage cases without substantially biasing the results.

Although I included MLE ORs estimated using co-dominant modeling assumptions, I assumed additive models for all Bayesian analyses and drew inferences from additive models only. This approach is consistent with the existing literature, as most of the GWAS and replication studies cited in this report provided effects for additive models. If the true effect

of the causal variant did not follow an additive pattern, the reported OR estimates would be biased and my inferences would be incorrect. While it is possible to do statistical tests to compare model fit under different assumptions (e.g. Akaike Information Criteria [AIC]), such global tests are underpowered to detect subtle differences in fit [291] and are rarely used. Alternatively, I could have incorporated model uncertainty using Bayesian methods, as proposed by Stephens et al [15]. However, because the quality and quantity of *a priori* information varied widely among the selected SNPs, it would have been difficult to construct priors for many of the included SNPs.

The SNPs selected for this analysis were both a strength and a limitation. I was able to evaluate the role of a number of important GWAS and candidate gene hits, but the panel was incomplete and I was not able to include any new discoveries. Additionally, while I believe that the inclusion of more finely mapped regions augmented theses analyses, only a few of the selected regions included tag SNPs in addition to the variant of interest. In particular, this limited my ability to fully explore the performance of the hierarchical methods within high LD regions.

The diverse composition of CBCS is a major strength of this study. Study investigators originally chose to include large numbers of African American women so they could explore racial differences in breast cancer risk factors and breast cancer related outcomes. Because subtype distributions vary by race, this recruitment scheme inadvertently ensured that all five intrinsic subtypes were well represented in the CBCS population. In this way, the study is uniquely suited to answer questions about both race and subtype differences in etiology and prognosis. Studying the disease in racial groups with unique LD patterns also provided information that may help pinpoint causal variants. To date, this is one of the

largest population-based studies of African Americans and also one of the largest studies to evaluate breast cancer subtypes using a five-marker IHC panel.

A strength of these particular analyses was the use of Bayesian methods. Breast cancer susceptibility loci are a commonly studied topic and Bayesian methods offered a way to make use of the plethora of prior information to generate more precise estimates. I believe the priors specified here were reasonable given our existing knowledge of breast cancer susceptibility variants and linkage disequilibrium patterns. Therefore, the results presented here should be more accurate, on average, than those produced using traditional frequentist methods. Further, by selecting only null-centered, informative priors for the full Bayes analysis, any bias resulting from these methods should be towards the null [17]. In this way, this application of Bayesian methods also reduces the probability of observing false positive associations. Similarly, hierarchical models shrink everything towards the group-level mean, which should also help control type I error.

Lastly, while I do believe that my prior assumptions were reasonable, I provided a sensitivity analysis that included the MLE estimates as well as Bayesian estimates given more informative priors. These supplementary analyses demonstrated the relative influence of the prior assumptions.

6.3 Public health implications

While I did not identify any completely new susceptibility loci, I replicated several previously established risk variants and explored some of the potentially causal regions more thoroughly. For some SNPs, I reported novel race- or subtype-specific effect estimates. As most previous attempts to isolate effects within African Americans or for individual subtypes

have produced mostly imprecise and inconsistent findings, my work may help advance our understanding of breast cancer etiology.

Additionally, since LD patterns vary by race, these race-stratified analyses may help to narrow down the specific region or regions of the candidate genes that are most strongly associated with the disease of interest. Improved understanding of genetic risk factors and disease etiology can ultimately improve breast cancer control and prevention by identifying potential targets for directed therapies and for locating high-risk individuals. Subsequent research on environmental or treatment-related factors that interact with these genes could contribute further public health benefit.

This research also has potential methodological impacts. For both manuscripts, I provided a simple, yet descriptive discussion of the merits of Bayesian analysis and how to implement the selected methods. This included SAS code for hierarchical models and Bayesian polytomous logistic regression, both of which are sparsely documented. As these methods are well accepted but rarely used in the applied epidemiology literature, I hope that my work will encourage others to consider them for similar investigations.

6.4 Future research

The breast cancer genetic epidemiology field is large and evolves quickly. The 21 GWAS hits selected for this analysis were from the first 8 GWAS. Since then, another 15 GWAS have been published, bringing the total count to 58 hits [183]. It would be interesting to see how these more recently identified SNPs replicate, especially within different racial groups and disease subtypes. There were also several other candidate genes in Zhang et al. [20] that merit further study.

This ever-expanding body of literature could also be used to formulate more informative, SNP-specific priors than were selected for the analysis presented here. For example, one could conduct meta-analyses of each of the selected SNPs and use these results to choose tailored prior distributions. Such an approach would further improve the precision and overall accuracy of the estimated effects, as long as the chosen priors more accurately reflected the true effect and the selected variances were at least as informative as those specified in the previously described analyses. However, such an approach would require extensive literature reviews of each of the candidate SNPs, with careful consideration of all potential biases, including publication bias, selection bias, and population stratification.

Our understanding of breast cancer subtypes is quickly evolving. While my use of the five IHC marker definition is more refined than most previous investigations, it does not capture the true heterogeneity of the disease [47]. As these subtype definitions improve, I hope that we will start to see even more convincing evidence of etiologic differences between tumor types.

In terms of methodological pursuits, I can see myself applying Bayesian methods in several other genetic epidemiology scenarios, including further replication studies, gene-environment or gene-gene interaction studies, and fine-mapping investigations. While it might be more difficult to define and justify priors for environmental factors, which may have larger or less predictable effects than SNPs, the breast cancer literature includes copious investigations and meta-analyses from which to draw informative priors. As both gene-environment and gene-gene interaction analyses are often vastly underpowered, the field would benefit from the improved precision of Bayesian approaches. Similarly, I think that

fine-mapping investigations would benefit from hierarchical approaches, as standard MLE methods cannot accurately accommodate high between-SNP correlations.

The ultimate goal of any of these proposed directions is to learn more about breast cancer etiology so that we may better prevent, detect, and treat the disease. My main aspiration is to do what I can to contribute to this worthy endeavor.

Table 23: Summary of replication results and subtype-specific findings for GWAS-identified and candidate gene SNP hits

Chrom-osome	Gene or gene region	SNP	EA		AA		Previous subtype findings	CBCS subtype findings
			Replication?*	OR>1.0?	Replication?*	OR>1.0?		
1	1p12	rs11249433	No	Yes	No	Yes	strongest association in Luminal A and B cancers	null
2	CASP8	rs1045485	No	Yes	No	Yes	Equally associated with all subtypes	null
2	2p	rs4666451	No	Yes	No	Yes	evidence of association with ER+ and ER- disease	null
2	2q35	rs13387042	No	Yes	No	Yes	some association with all subtypes	positively associated with luminal A
3	SLC4A7	rs4973768	No	Yes	No	No	strongest association in Luminal A cancers; no association in triple-negative	null
4	4p	rs12505080	No	Yes	No	Yes	NE	positively associated with luminal A
4	TLR1	rs7696175	No	Yes	No	Yes	Associated with ER+, ER- and HER2- disease in Chinese	positively associated with luminal A and basal-like, inversely associated with HER2+/ER-
5	MRPS30	rs4415084	Yes	Yes	No	Yes	stronger associations with ER+ or PR+ disease; mostly null associations with ER- or PR- disease	null
5	MRPS30	rs10941679	Yes	Yes	No	Yes	strongest association in Luminal A and B cancers; null association in triple-negative cancers	positively associated with basal-like
5	5p12	rs981782	No	No	No	Yes	strongest associations with ER- and HER2- disease	null

5	5q	rs30099	No	Yes	No	Yes	possible associations with ER+, ER- and HER2+ disease	null
5	MAP3K1	rs889312	Yes	Yes	No	No	evidence of association with all subtypes	positively associated with luminal A
6	ECHDC1	rs2180341	No	Yes	No	No	predominantly null associations with ER and PR status	null
6	ESR1	rs2046210	No	Yes	Yes	Yes	some association with all subtypes	positively associated with basal-like and HER2+/ER-
6	ESR1	rs1801132	No	No	No	Yes	NE	null
6	ESR1	rs3020314	No	Yes	No	Yes	NE	null
7	RELN	rs17157903	No	No	No	Yes	NE	null
8	8q24	rs13281615	No	Yes	No	Yes	strongest association in Luminal A and B; null association in triple-negative	null
8	8q24	rs1562430	No	Yes	No	Null	Associated with ER+, ER-, PR+, and PR- disease in Chinese (one study)	Positively associated with luminal A in whites only
9	CDKN2A/B	rs1011970	No	Yes	No	No	some evidence of association with ER+ and triple-negative	null
10	ANKRD16	rs2380205	No	Yes	No	No	possible inverse association with ER+ disease, null association with triple negative	negatively associated with unclassified
10	ZNF365	rs10995190	No	Yes	No	Yes	inversely associated with ER+ disease	null
10	ZMIZ1	rs704010	Yes	Yes	No	Yes	positively associated with all ER/PR subtypes, but not triple-negative disease	positively associated with basal-like, luminal B and HER2+/ER-
10	FGFR2	rs10736303	Yes	Yes	Yes	Yes	evidence of associations	positively associated

							with ER+ and PR+ disease	with luminal A, luminal B and unclassified
10	FGFR2	rs2981579	Yes	Yes	Yes	Yes	associated with ER+/PR+ in AA,and most subtypes in Chinese study	positively associated with luminal A, HER2+/ER- and unclassified
10	FGFR2	rs1078806	Yes	Yes	No	Yes	NE	positively associated with luminal A and unclassified
10	FGFR2	rs2981578	Yes	Yes	Yes	Yes	evidence of association with ER+, ER- and PR+ disease	positively associated with luminal A, luminal B, HER2+/ER- and unclassified
10	FGFR2	rs1219648	Yes	Yes	Yes	Yes	evidence of association with ER+, PR+ and HER2-	positively associated with luminal A and unclassified
10	FGFR2	rs2912774	Yes	Yes	Yes	Yes	evidence of association with ER+ and ER- disease	positively associated with luminal A, HER2+/ER- and unclassified
10	FGFR2	rs2936870	Yes	Yes	Yes	Yes	NE	positively associated with luminal A, HER2+/ER- and unclassified
10	FGFR2	rs2420946	Yes	Yes	Yes	Yes	NE	positively associated with luminal A, HER2+/ER- and unclassified
10	FGFR2	rs2981582	Yes	Yes	Yes	Yes	strongest association in Luminal A and B; null association in HER2+/ER- and triple-negative	positively associated with luminal A, HER2+/ER- and unclassified
10	FGFR2	rs3135718	Yes	Yes	Yes	Yes	NE	positively associated with luminal A, HER2+/ER- and

								unclassified
10	10q	rs1051012	No	Yes	No	Yes	NE	null
10	ATM	rs1800057	No	Yes	NE	NE	NE	null
11	LSP1	rs3817198	No	Yes	No	Yes	Generally imprecise and inconsistent findings, possible association with unclassified disease	null
11	LSP1	rs909116	No	Yes	No	Yes	NE	null
11	H19	rs2107425	Yes	Yes	No	No	NE	null
11	MYEOV	rs614367	NE	NE	NE	NE	possible association with ER+ and PR+ disease	NE
16	TNRC9	rs8051542	No	Yes	No	Yes	evidence of association with ER+ disease	positively associated with unclassified
16	TNRC9	rs1244362	Yes	Yes	No	No	no evidence of association	Positively associated with HER2+/ER- in whites
16	TNRC9	rs3803662	Yes	Yes	No	Yes	evidence of association with all subtypes	positively associated with luminal A
16	TNRC9	rs4784227	Yes	Yes	No	Yes	positively associated with all ER and PR subtypes in one Korean study	positively associated with luminal A and unclassified
17	TP53	rs12951053	No	Yes	No	Yes	NE	null
17	COX11	rs7222197	No	No	No	Yes	NE	null
17	COX11	rs6504950	No	No	No	Yes	inverse association in Luminal A and B; null association in HER2+/ER- and triple-negative	null

*Based on Bayesian posterior mean and 95% PI

Table 24: Comparison of previous findings and CBCS results for less established (non-GWAS, non-candidate) SNPs

Chrom-osome	Gene or gene region	SNP	Range of estimates in previous studies		CBCS effect estimates		Summary of previous subtype findings	CBCS subtype findings
			EA	AA	EA	AA		
2	CASP8	rs17468277	1.05-1.06	NE	1.11	1.07	NE	null
6	ESR1	rs851974	1.09	0.88	0.91	0.94	NE	inversely associated with luminal A
6	ESR1	rs2077647	1.03*	NE	0.97	1.07	NE	null
6	ESR1	rs2234693	1.03*	0.90#	0.95	0.97	NE	null
6	ESR1	rs3798577	1.02*	NE	1.03	1.02	evidence of association with ER- and PR- disease	inversely associated with HER2+/ER-
9	CDKN2A/B	rs3731257	1.1#	NE	0.94	0.91	NE	null
9	CDKN2A/B	rs3731249	1.5&	NE	0.94	NE	NE	null
9	CDKN2A/B	rs518394	NE	NE	1.03	1.02	NE	Inversely associated with HER2+/ER-
9	CDKN2A/B	rs564398	NE	NE	1.04	1.01	NE	null
9	CDKN2A/B	rs10757278	NE	NE	1.16	0.93	NE	Positively associated with unclassified in whites, inversely associated in African Americans
9	CDKN2A/B	rs10811661	NE	NE	1.01	1.01	NE	null
10	FGFR2	rs1896395	NE	NE	NE	1.02	NE	null
10	FGFR2	rs3750817	1.16-1.28	NE	1.22	1.61	GWAS hit in Japanese study of hormone receptor positive disease (inverse association)	positively associated with luminal A
10	FGFR2	rs11200014	1.24- 1.31	NE	1.29	1.04	Associated with ER+, PR+,	Positively associated with

							HER2+ and HER2- in Chinese study	luminal A and unclassified
10	FGFR2	rs2162540	NE	NE	1.29	1.21	NE	Positively associated with luminal A, HER2+/ER- and unclassified
10	ATM	rs1800054	1.16*&	2.02	1.03	NE	associated with PR+ tumors	null
10	ATM	rs4986761	1.08*&	NE	NE	NE	not associated with any subtype	NE
10	ATM	rs1800056	1.27*&	NE	NE	NE	NE	NE
10	ATM	rs1800058	1.25*&	NE	0.90	NE	not associate with ER or PR status	null
10	ATM	rs1801516	1.09*	1.14-1.27	0.99	1.14	ER+ inversely associated, relative to ER-	null
10	ATM	rs3092992	NE	NE	1.16	NE	NE	null
10	ATM	rs664143	1.02*	NE	1.02	0.96	associated with PR+	null
10	ATM	rs170548	0.94	NE	0.98	0.90	not associated with ER+ or PR+	null
10	ATM	rs3092993	NE	NE	0.99	1.14	NE	null
16	TNRC9	rs8049149	NE	NE	NE	0.98	NE	NE
16	TNRC9	rs16951186	NE	NE	1.10	0.92	NE	null
16	TNRC9	rs3104746	NE	1.17-1.23	1.42	1.49	evidence of association with ER+ and ER- disease	positively associated with luminal A and basal-like
16	TNRC9	rs3112562	NE	1.17	0.99	1.26	NE	positively associated with HER2+/ER- and basal-like, inversely associated with unclassified; inversely associated with luminal B in whites
16	TNRC9	rs9940048	NE	NE	1.03	1.10	NE	positively associated with luminal A
17	TP53	rs9894946	1.07*	NE	0.86	0.93	NE	inversely associated with luminal A
17	TP53	rs1614984	1.01*	NE	1.03	1.07	NE	null

17	TP53	rs4968187	NE	NE	NE	NE	NE	null
17	TP53	rs17880604	NE	NE	0.92	NE	NE	null
17	TP53	rs1800372	NE	NE	0.95	NE	NE	null
17	TP53	rs2909430	1.00*	NE	1.10	1.06	NE	null
17	TP53	rs1042522	1.02*	NE	0.99	0.98	associated with ER+	null
17	TP53	rs8079544	1.11*	NE	1.19	0.92	NE	null

Appendix 1: SAS code

Bayesian model, dichotomous outcome

```
PROC MCMC DATA=DATA NMC=30000 NBI=1000 THIN=5
MONITOR=(OR1 beta0 beta1 beta2 beta3 var) SEED =6544;
WHERE RACE=1;
PARMS beta0 beta1 beta2 beta3 var; *default starting value of 0;
HYPERPRIOR var~igamma(shape=3,scale=0.2); *distribution with mode at 0.05;
PRIOR beta0~normal(0, var=1e6);
PRIOR beta1~normal(0, var=var); *var is the hyperprior defined above;
PRIOR beta2 beta3~normal(0, var=0.68);
OR1=exp(beta1);
mu1=(beta0+beta1*snp+beta2*age+beta3*ances+offset);
logl= ((case=0)*(0-log(1+exp(mu1))) + (case=1)*(mu1-log(1+exp(mu1))));
MODEL case~general(logl);
RUN;
QUIT;
```

Hierarchical model with identity covariance matrix, 3 SNPs in LD block

```
DATA covar3;
INPUT covp1-covp3;
DATALINES;
0.05 0.05 0.05
;
RUN;

DATA data;
SET data;
LD= snp1+snp2+snp3;
RUN;

PROC GLIMMIX data=data;
MODEL CASE= age ancestry LD /
SOLUTION DIST=binomial LINK=logit OFFSET=OFF;
PARMS / pdata=covar3 hold= 1,2,3;
RANDOM snp1 snp2 snp3 / TYPE=VC G SOLUTION;
ESTIMATE "snp1" intercept 0 age 0 ancestry 0 LD 1 | snp1 1 / cl exp;
ESTIMATE "snp2" intercept 0 age 0 ancestry 0 LD 1 | snp2 1 / cl exp;
ESTIMATE "snp3" intercept 0 age 0 ancestry 0 LD 1 | snp3 1 / cl exp;
RUN;
```

Hierarchical model with exponential decay matrix

For 3 SNPs at positions 7518935, 7519370, and 7520197, such that

$|d_{12}| = 435$

$|d_{13}| = 1262$

$|d_{12}| = 827$

$$t_{ij} = \exp\left[-\left(\frac{d_{ij}}{1000}\right)\right]$$

$$T=\begin{bmatrix} 1 & 0.6473 & 0.2831 \\ 0.6473 & 1 & 0.4374 \\ 0.2831 & 0.4374 & 1 \end{bmatrix}$$

$$\sigma^2 T=\begin{bmatrix} 0.05 & 0.03236 & 0.01415 \\ 0.03236 & 0.05 & 0.02187 \\ 0.01415 & 0.02187 & 0.05 \end{bmatrix}$$

*Input covariances left of and including the diagonal, reading right to left across rows
DATA covar_LDdist;
INPUT covp1-covp6;
DATALINES;
0.05 0.03236 0.05 0.01415 0.02187 0.05
;
RUN;

PROC GLIMMIX data=data;
MODEL CASE= age ancestry LD /
SOLUTION DIST=binomial LINK=logit OFFSET=OFF;
PARMS / pdata=covar_LDdist hold= 1,2,3,4,5,6;
RANDOM snp1 snp2 snp3 / TYPE=UN G SOLUTION;
ESTIMATE "snp1" intercept 0 age 0 ancestry 0 LD 1 | snp1 1 / cl exp;
ESTIMATE "snp2" intercept 0 age 0 ancestry 0 LD 1 | snp2 1 / cl exp;
ESTIMATE "snp3" intercept 0 age 0 ancestry 0 LD 1 | snp3 1 / cl exp;
RUN;

Bayesian model, polytomous outcome

In the code below, subtype is a variable with 6 levels, where subtype=0 corresponds to controls (the referent group). SNP is a 3 level ordinal variable, age and ancestry are mean-centered continuous variables, race is a 2 level categorical variable and offset is the offset term.

PROC MCMC DATA=data NMC=50000 NBI=1000 THIN=10 SEED=3342
MONITOR=(var OR1 OR2 OR3 OR4 OR5 b01 b02 b03 b04 b05 b11 b12 b13 b14 b15 b21 b22 b23 b24 b25 b31 b32 b33 b34 b35 b41 b42 b43 b44 b45);

PARMS var b01 b02 b03 b04 b05 b11 b12 b13 b14 b15 b21 b22 b23 b24 b25 b31 b32 b33 b34 b35 b41 b42 b43 b44 b45 0; *this starting value should be arbitrary

HYPERPRIOR var~igamma(shape=4,scale=0.5); *distribution with mode at 0.1;
PRIOR b01 b02 b03 b04 b05~normal(0, var=1e6);
PRIOR b11 b12 b13 b14 b15~normal(0, var=var); *var is the hyperprior defined above;
PRIOR b21 b22 b23 b24 b25 b31 b32 b33 b34 b35~normal(0, var=0.68);
PRIOR b41 b42 b43 b44 b45~normal(0,var=1);

mu1=b01+b11*snp+b21*age+b31*ances+b41*race+offset;
mu2=b02+b12*snp+b22*age+b32*ances+b42*race+offset
mu3=b03+b13*snp+b23*age+b33*ances+b43*race+offset;
mu4=b04+b14*snp+b24*age+b34*ances+b44*race+offset;

```
        mu5=b05+b15*snp+b25*age+b35*ances+b45*race+offset;

        logl= (subtype=0)*(0-log(1+exp(mu1)+exp(mu2)+exp(mu3)+exp(mu4)+exp(mu5)))
        + (subtype=1)*(mu1-log(1+exp(mu1)+ exp(mu2)+exp(mu3)+ exp(mu4)+exp(mu5)))
        + (subtype=2)*(mu2-log(1+exp(mu1)+ exp(mu2)+exp(mu3)+ exp(mu4)+exp(mu5)))
        + (subtype=3)*(mu3-log(1+exp(mu1)+ exp(mu2)+exp(mu3)+ exp(mu4)+exp(mu5)))
        + (subtype=4)*(mu4-log(1+exp(mu1)+ exp(mu2)+exp(mu3)+ exp(mu4)+exp(mu5)))
        + (subtype=5)*(mu5-log(1+exp(mu1)+exp(mu2)+exp(mu3)+ exp(mu4)+exp(mu5)));

        MODEL subtype~general(logl);

        OR1=exp(b11);
        OR2=exp(b12);
        OR3=exp(b13);
        OR4=exp(b14);
        OR5=exp(b15);

RUN;
```

Appendix 2: Extra figures, overall breast cancer analysis

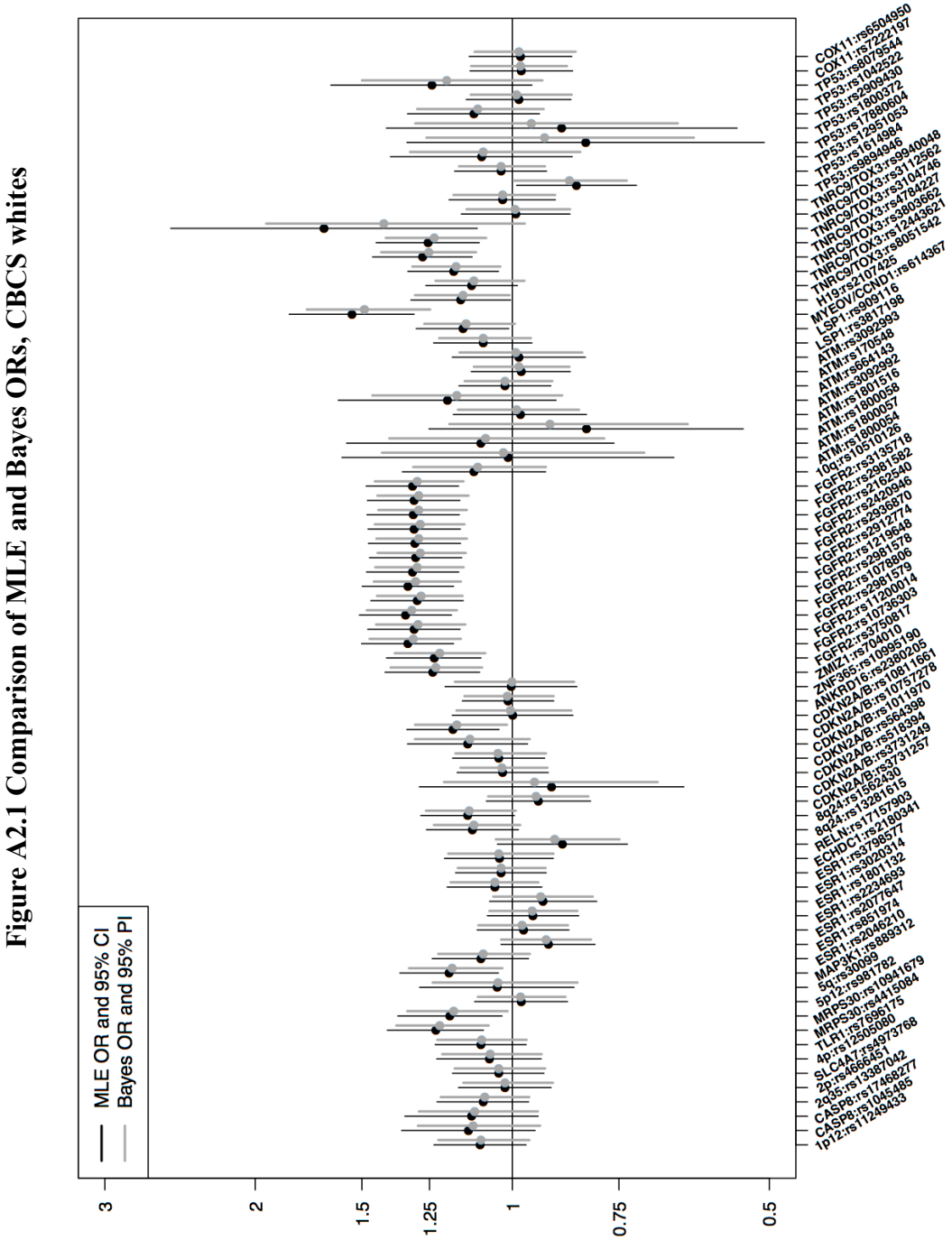


Figure A2.1 Comparison of MLE and Bayes ORs, CBCS whites

Figure A2.2 Comparison of MLE and Bayes ORs, CBCS African Americans

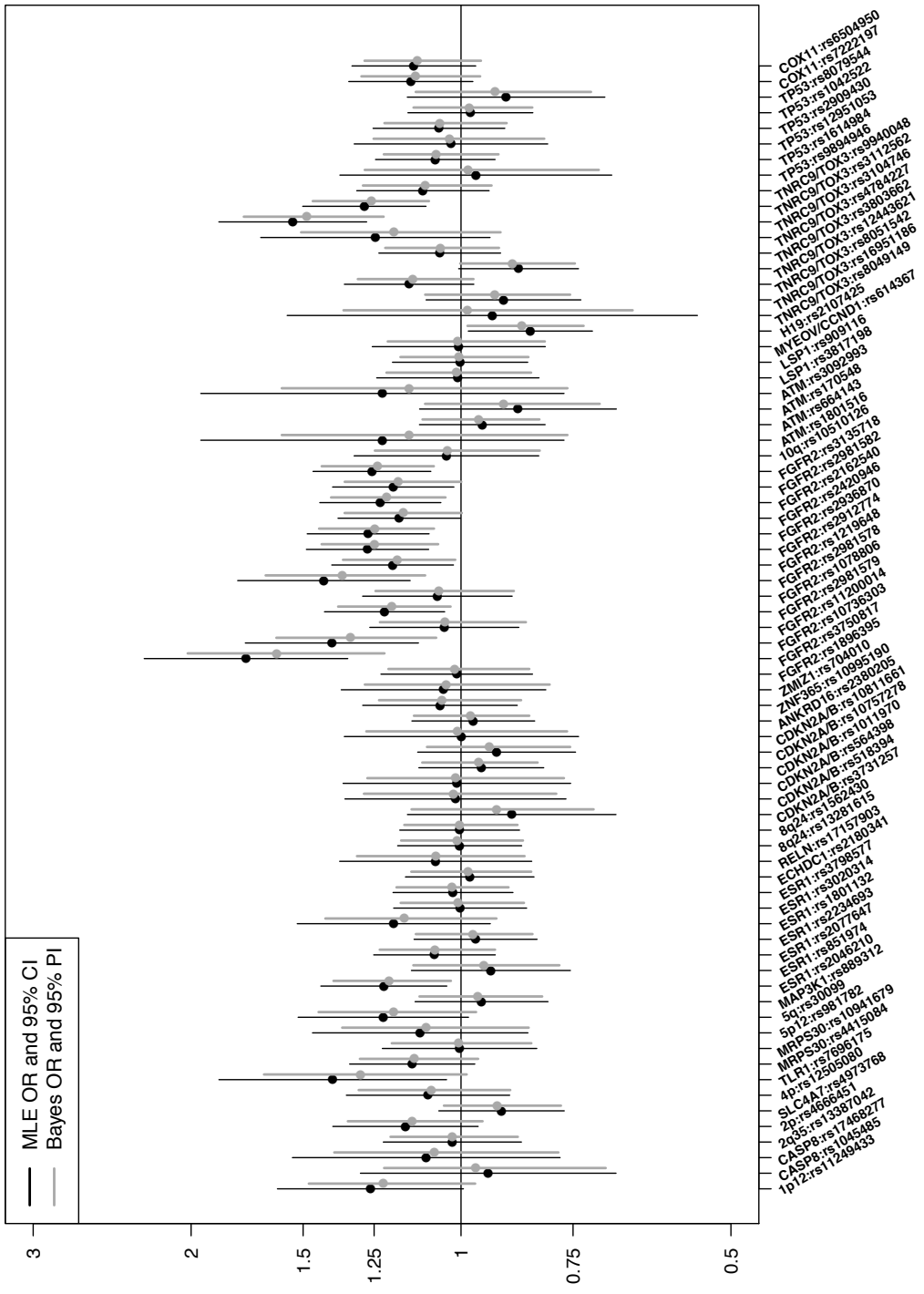


Figure A2.3 Comparison of MLE and hierarchical ORs, CBCS Whites

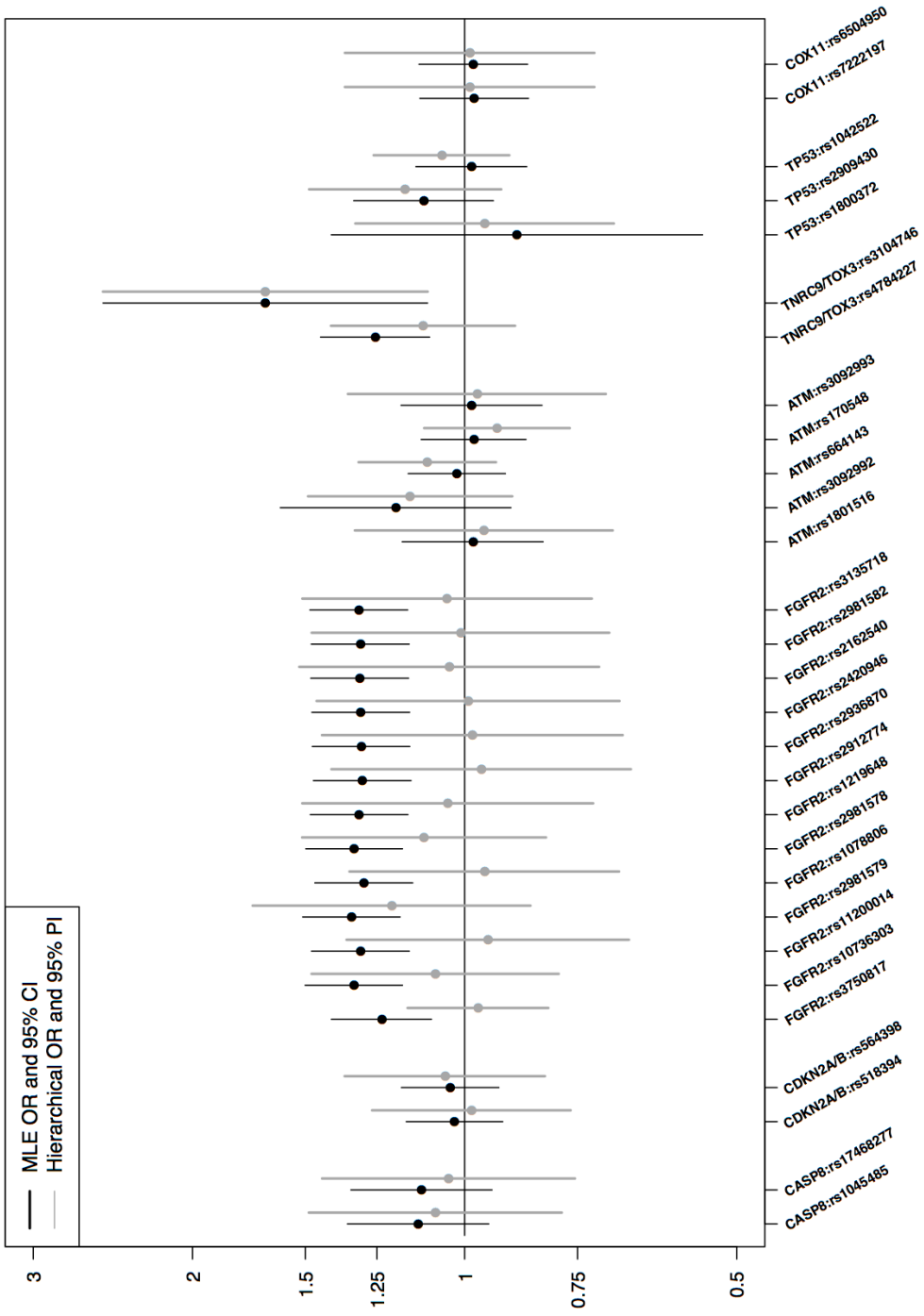
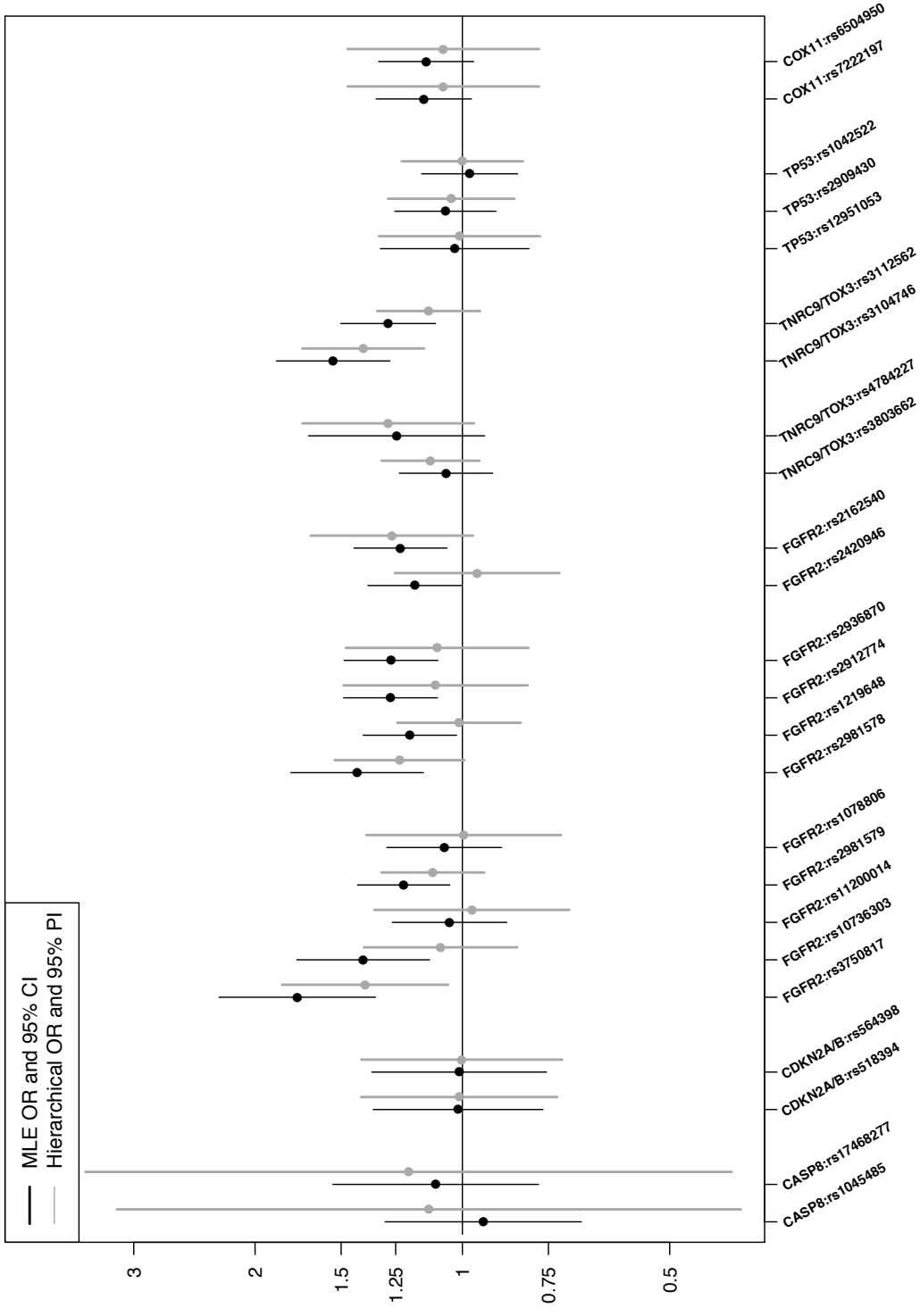


Figure A2.4 Comparison of MLE and hierarchical ORs, CBCS African Americans



References

1. SEER Incidence Statistics - SEER Cancer Query Systems. http://seer.cancer.gov/canques/incidence.html. Accessed January 9, 2013.

2. US Cancer Morality Statistics. http://seer.cancer.gov/canques/mortality.html. Accessed January 9, 2013.

3. Carey LA, Perou CM, Livasy CA, et al. Race, breast cancer subtypes, and survival in the Carolina Breast Cancer Study. *JAMA.* 2006;295:2492-2502.

4. Millikan RC, Newman B, Tse C, et al. Epidemiology of basal-like breast cancer. *Breast Cancer Res Treat.* 2008;109:123-139, 2008.

5. Cheang MCU, Voduc D, Bajdik C, et al. Basal-like breast cancer defined by five biomarkers has superior prognostic value than triple-negative phenotype. *Clin Cancer Res.* 2008;14:1368-1376.

6. Shin BK, Lee Y, Lee JB, et al. Breast carcinomas expressing basal markers have poor clinical outcome regardless of estrogen receptor status. *Oncol Rep.* 2008;19:617-625.

7. Kurian AW, Fish K, Shema SJ, Clarke CA. Lifetime risks of specific breast cancer subtypes among women in four racial/ethnic groups. *Breast Cancer Res,* 2010;12:R99. doi:10.1186/bcr2780.

8. Lund MJ, Trivers KF, Porter PL, et al. Race and triple negative threats to breast cancer survival: a population-based study in Atlanta, GA. *Breast Cancer Res Treat.* 2009;113:357-370.

9. International HapMap Consortium, Frazer KA, Ballinger DG, et al. A second generation human haplotype map of over 3.1 million SNPs. *Nature.* 2007;449:851-861.

10. Haffty BG, Silber A, Matloff E, Chung J, Lannin D. Racial differences in the incidence of BRCA1 and BRCA2 mutations in a cohort of early onset breast cancer patients: African American compared to white women. *J Med Genet.* 2006;43:133-137.

11. Hutter CM, Young AM, Ochs-Balcom HM, et al. Replication of Breast Cancer GWAS Susceptibility Loci in the Women's Health Initiative African American SHARe Study. *Cancer Epidemiol Biomarkers Prev.* 2011;20:1950-1959.

12. Zheng W, Cai Q, Signorello LB, et al. Evaluation of 11 breast cancer susceptibility loci in African-American women. *Cancer Epidemiol Biomarkers Prev.* 2009;18:2761-2764.

13. Greenland S. Principles of multilevel modelling. *Int J Epidemiol.* 2000;29:158-167.

14. Greenland S. Bayesian perspectives for epidemiological research: I. Foundations and basic methods. *Int J Epidemiol.* 2006;35:765-775.

15. Stephens M, Balding DJ. Bayesian statistical methods for genetic association studies. *Nat Rev Genet.* 2009;10:681-690.

16. Hung RJ, Brennan P, Malaveille C, et al. Using hierarchical modeling in genetic association studies with multiple markers: application to a case-control study of bladder cancer. *Cancer Epidemiol Biomarkers Prev.* 2004;13:1013-1021.

17. Hamra GB, Maclehose RF, Cole SR. Sensitivity Analyses for Sparse-Data Problems-Using Weakly Informative Bayesian Priors. *Epidemiology.* 2013;24:233-239.

18. Hindorff LA, MacArthur J (European Bioinformatics Institute), Morales J (European Bioinformatics Institute), et al. A Catalog of Published Genome-Wide Association Studies. Available at: www.genome.gov/gwastudies. Accessed January 13, 2013.

19. Varghese JS, Easton DF. Genome-wide association studies in common cancers-what have we learnt? *Curr Opin Genet Dev.* 2010;20:201-209.

20. Zhang B, Beeghly-Fadiel A, Long J, Zheng W. Genetic variants associated with breast-cancer risk: comprehensive research synopsis, meta-analysis, and epidemiological evidence. *Lancet Oncol.* 2011;12:477-488.

21. GLOBOCAN: Country Fast Stat. http://globocan.iarc.fr/factsheets/populations/factsheet.asp?uno=900. Accessed January 9, 2013.

22. Breast Cancer Home Page - National Cancer Institute. http://www.cancer.gov/cancertopics/types/breast. Accessed January 9, 2013.

23. Cancer of the Breast - SEER Stat Fact Sheets. http://www.seer.cancer.gov/statfacts/html/breast.html. Accessed January 9, 2013.

24. Cancer - State Cancer Facts - North Carolina vs. United States Comparisons. http://apps.nccd.cdc.gov/statecancerfacts/Table.aspx?Group=5f&TableType=INCI&SelectedState=North%20Carolina. Accessed January 9, 2013.

25. Breast Equivalent Terms, Definitions, Tables and Illustration. http://seer.cancer.gov/manuals/2010/AppendixC/breast/terms_defs.pdf. Accessed July 25, 2011.

26. Anderson WF, Chu KC, Chang S, Sherman ME. Comparison of age-specific incidence rate patterns for different histopathologic types of breast carcinoma. *Cancer Epidemiol Biomarkers Prev.* 2004;13:1128-1135.

27. Li CI. Risk of mortality by histologic type of breast cancer in the United States. *Horm Cancer.* 2010;1:156-165.

28. Dontu G, El-Ashry D, Wicha MS. Breast cancer, stem/progenitor cells and the estrogen receptor. *Trends Endocrinol Metab.* 2004;15:193-197.

29. Melchor L, Benítez J. An integrative hypothesis about the origin and development of sporadic and familial breast cancer subtypes. *Carcinogenesis.* 2008;29:1475-1482.

30. Stingl J. Estrogen and progesterone in normal mammary gland development and in cancer. *Horm Cancer.* 2011;2:85-90.

31. Jatoi I, Chen BE, Anderson WF, Rosenberg PS. Breast cancer mortality trends in the United States according to estrogen receptor status and age at diagnosis. *J Clin Oncol.* 2007;25:1683-1690.

32. Li CI, Uribe DJ, Daling JR. Clinical characteristics of different histologic types of breast cancer. *Br J Cancer.* 2005;93:1046-1052.

33. Anderson WF, Luo S, Chatterjee N, et al. Human epidermal growth factor receptor-2 and estrogen receptor expression, a demonstration project using the residual tissue repository of the Surveillance, Epidemiology, and End Results (SEER) program. *Breast Cancer Res Treat.* 2009;113:189-196.

34. Li CI, Moe RE, Daling JR. Risk of mortality by histologic type of breast cancer among women aged 50 to 79 years. *Arch Intern Med.* 2003;163:2149-2153.

35. Ma H, Wang Y, Sullivan-Halley J, et al. Breast cancer receptor status: do results from a centralized pathology laboratory agree with SEER registry reports? *Cancer Epidemiol Biomarkers Prev.* 2009;18:2214-2220.

36. Grann VR, Troxel AB, Zojwalla NJ, Jacobson JS, Hershman D, Neugut AI. Hormone receptor status and survival in a population-based cohort of patients with breast carcinoma. *Cancer.* 2005;103:2241-2251.

37. Joslyn SA. Hormone receptors in breast cancer: racial differences in distribution and survival. *Breast Cancer Res Treat.* 2002;73:45-59.

38. Al-Abbadi MA, Washington TA, Saleh HA, Tekyi-Mensah SE, Lucas DR, Briston CA. Differential expression of HER-2/NEU receptor of invasive mammary carcinoma between Caucasian and African American patients in the Detroit metropolitan area. Correlation with overall survival and other prognostic factors. *Breast Cancer Res Treat.* 2006;97:3-8.

39. Perou CM, Sørlie T, Eisen MB, et al. Molecular portraits of human breast tumours. *Nature.* 2000;406:747-752.

40. Sørlie T, Perou CM, Tibshirani R, et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci U S A.* 2001;98:10869-10874.

41. Sorlie T, Tibshirani R, Parker J, et al. Repeated observation of breast tumor subtypes in independent gene expression data sets. *Proc Natl Acad Sci U S A.* 2003;100:8418-8423.

42. Nielsen TO, Hsu FD, Jensen K, et al. Immunohistochemical and clinical characterization of the basal-like subtype of invasive breast carcinoma. *Clin Cancer Res.* 2004;10:5367-5374.

43. Kreike B, van Kouwenhove M, Horlings H, et al. Gene expression profiling and histopathological characterization of triple-negative/basal-like breast carcinomas. *Breast Cancer Res*, 2007;9:R65. doi:10.1186/bcr1771.

44. Ma CX, Luo J, Ellis MJ. Molecular Profiling of Triple Negative Breast Cancer. *Breast Dis.* 2011;32:73-84.

45. Perou CM. Molecular stratification of triple-negative breast cancers. *Oncologist.* 2011;16 Suppl 1:61-70.

46. Rody A, Karn T, Liedtke C, et al. A clinically relevant gene signature in triple negative and basal-like breast cancer. *Breast Cancer Res*, 2011;13:R97. doi:10.1186/bcr3035.

47. Cancer Genome Atlas Network. Comprehensive molecular portraits of human breast tumours. *Nature.* 2012;490:61-70.

48. Calza S, Hall P, Auer G, et al. Intrinsic molecular signature of breast cancer in a population-based cohort of 412 patients. *Breast Cancer Res*, 2006;8:R34. doi:10.1186/bcr1517.

49. Hines LM, Risendal B, Byers T, Mengshol S, Lowery J, Singh M. Ethnic Disparities in Breast Tumor Phenotypic Subtypes in Hispanic and Non-Hispanic White Women. *J Womens Health (Larchmt).* 2011;20:1543-1550.

50. Huo D, Ikpatt F, Khramtsov A, et al. Population differences in breast cancer: survey in indigenous African women reveals over-representation of triple-negative breast cancer. *J Clin Oncol.* 2009;27:4515-4521.

51. Kennecke H, Yerushalmi R, Woods R, et al. Metastatic behavior of breast cancer subtypes. *J Clin Oncol.* 2010;28:3271-3277.

52. Kim M, Ro JY, Ahn S, Kim HH, Kim S, 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;37:1217-1226.

53. Kurebayashi J, Moriya T, Ishida T, et al. The prevalence of intrinsic subtypes and prognosis in breast cancer patients of different races. *Breast.* 2007;16 Suppl 2:S72-77.

54. Liu H, Fan Q, Zhang Z, Li X, Yu H, Meng F. Basal-HER2 phenotype shows poorer survival than basal-like phenotype in hormone receptor-negative invasive breast cancers. *Hum Pathol.* 2008;39:167-174.

55. Munirah MA, Siti-Aishah MA, Reena MZ, et al. Identification of different subtypes of breast cancer using tissue microarray. *Rom J Morphol Embryol.* 2011;52:669-677.

56. Sihto H, Lundin J, Lehtimäki T, et al. Molecular subtypes of breast cancers detected in mammography screening and outside of screening. *Clin Cancer Res.* 2008;14:4103-4110.

57. Voduc KD, Cheang MCU, Tyldesley S, Gelmon K, Nielsen TO, Kennecke H. Breast cancer subtypes and the risk of local and regional relapse. *J Clin Oncol.* 2010;28:1684-1691.

58. Yamamoto Y, Ibusuki M, Nakano M, Kawasoe T, Hiki R, Iwase H. Clinical significance of basal-like subtype in triple-negative breast cancer. *Breast Cancer.* 2009;16:260-267.

59. Zaha DC, Lazăr E, Lăzureanu C. Clinicopathologic features and five years survival analysis in molecular subtypes of breast cancer. *Rom J Morphol Embryol.* 2010;51:85-89.

60. Bauer KR, Brown M, Cress RD, Parise CA, Caggiano V. Descriptive analysis of estrogen receptor (ER)-negative, progesterone receptor (PR)-negative, and HER2-negative invasive breast cancer, the so-called triple-negative phenotype: a population-based study from the California cancer Registry. *Cancer.* 2007;109:1721-1728.

61. Ihemelandu CU, Leffall LD Jr, Dewitty RL, et al. Molecular breast cancer subtypes in premenopausal African-American women, tumor biologic factors and clinical outcome. *Ann Surg Oncol.* 2007;14:2994-3003.

62. Kwan ML, Kushi LH, Weltzien E, et al. Epidemiology of breast cancer subtypes in two prospective cohort studies of breast cancer survivors. *Breast Cancer Res*, 2009;11:R31. doi:10.1186/bcr2261.

63. Parise CA, Bauer KR, Caggiano V. Variation in breast cancer subtypes with age and race/ethnicity. *Crit Rev Oncol Hematol.* 2010;76:44-52.

64. Stark A, Kleer CG, Martin I, et al. African ancestry and higher prevalence of triple-negative breast cancer: findings from an international study. *Cancer.* 2010;116:4926-4932.

65. Stead LA, Lash TL, Sobieraj JE, et al. Triple-negative breast cancers are increased in black women regardless of age or body mass index. *Breast Cancer Res* 2009;11:R18. doi:10.1186/bcr2242.

66. Trivers KF, Lund MJ, Porter PL, et al. The epidemiology of triple-negative breast cancer, including race. *Cancer Causes Control.* 2009;20:1071-1082.

67. Brown M, Tsodikov A, Bauer KR, Parise CA, Caggiano V. The role of human epidermal growth factor receptor 2 in the survival of women with estrogen and progesterone receptor-negative, invasive breast cancer: the California Cancer Registry, 1999-2004. *Cancer.* 2008;112:737-747.

68. Nalwoga H, Arnes JB, Wabinga H, Akslen LA. Frequency of the basal-like phenotype in African breast cancer. *APMIS.* 2007;115:1391-1399.

69. Agboola AJ, Musa AA, Wanangwa N, et al. Molecular characteristics and prognostic features of breast cancer in Nigerian compared with UK women. *Breast Cancer Res Treat.* 2012;135:555-569.

70. Caldarella A, Crocetti E, Bianchi S, Vet al. Female Breast Cancer Status According to ER, PR and HER2 Expression: A Population Based Analysis. *Pathol Oncol Res.* 2011;17:753-758.

71. Muñoz M, Fernández-Aceñero MJ, Martín S, Schneider J. Prognostic significance of molecular classification of breast invasive ductal carcinoma. *Arch Gynecol Obstet.* 2009;280:43-48.

72. Spitale A, Mazzola P, Soldini D, Mazzucchelli L, Bordoni A. Breast cancer classification according to immunohistochemical markers: clinicopathologic features and short-term survival analysis in a population-based study from the South of Switzerland. *Ann Oncol.* 2009;20:628-635.

73. Zarcone M, Amodio R, Campisi I, et al. Application of a new classification to a breast tumor series from a population-based cancer registry: demographic, clinical, and prognostic features of incident cases, Palermo Province, 2002-2004. *Ann N Y Acad Sci.* 2009;1155:222-226.

74. Lin C, Liau J, Lu Y, et al. Molecular subtypes of breast cancer emerging in young women in Taiwan: evidence for more than just westernization as a reason for the disease in Asia. *Cancer Epidemiol Biomarkers Prev.* 2009;18:1807-1814.

75. Kim E, Noh WC, Han W, Noh D. Prognostic significance of young age (<35 years) by subtype based on ER, PR, and HER2 status in breast cancer: a nationwide registry-based study. *World J Surg.* 2011;35:1244-1253.

76. Nakajima H, Fujiwara I, Mizuta N, et al. Prognosis of Japanese breast cancer based on hormone receptor and HER2 expression determined by immunohistochemical staining. *World J Surg.* 2008;32:2477-2482.

77. Su Y, Zheng Y, Zheng W, et al. Distinct distribution and prognostic significance of molecular subtypes of breast cancer in Chinese women: a population-based cohort study. *BMC Cancer,* 2011. 11:292. doi:10.1186/1471-2407-11-292.

78. Telli ML, Chang ET, Kurian AW, et al. Asian ethnicity and breast cancer subtypes: a study from the California Cancer Registry. *Breast Cancer Res Treat.* 2011;127:471-478.

79. Xing P, Li J, Jin F. A case-control study of reproductive factors associated with subtypes of breast cancer in Northeast China. *Med Oncol.* 2010;27:926-931.

80. Kwong A, Mang OWK, Wong CHN, Chau WW, The Hong Kong Breast Cancer Research Group, Law SCK. Breast Cancer in Hong Kong, Southern China: The First Population-Based Analysis of Epidemiological Characteristics, Stage-Specific, Cancer-Specific, and Disease-Free Survival in Breast Cancer Patients: 1997-2001. *Ann Surg Oncol.* 2011;18:3072-3078.

81. Zhao J, Liu H, Wang M, et al. Characteristics and prognosis for molecular breast cancer subtypes in Chinese women. *J Surg Oncol.* 2009;100:89-94.

82. Yang XR, Sherman ME, Rimm DL, et al. Differences in risk factors for breast cancer molecular subtypes in a population-based study. *Cancer Epidemiol Biomarkers Prev.* 2007;16:439-443.

83. Tamimi RM, Colditz GA, Hazra A, et al. Traditional breast cancer risk factors in relation to molecular subtypes of breast cancer. *Breast Cancer Res Treat.* 2012;131:159-167.

84. Lara-Medina F, Pérez-Sánchez V, Saavedra-Pérez D, et al. Triple-negative breast cancer in Hispanic patients: High prevalence, poor prognosis, and association with menopausal status, body mass index, and parity. *Cancer.* 2011;117:3658-3669.

85. Parise CA, Bauer KR, Brown MM, Caggiano V. Breast cancer subtypes as defined by the estrogen receptor (ER), progesterone receptor (PR), and the human epidermal growth factor receptor 2 (HER2) among women with invasive breast cancer in California, 1999-2004. *Breast J.* 2009;15:593-602.

86. O'Brien KM, Cole SR, Tse C, et al. Intrinsic breast tumor subtypes, race, and long-term survival in the Carolina Breast Cancer Study. 2010;*Clin Cancer Res.* 16:6100-110.

87. Liu Z, Wu J, Ping B, et al. Basal cytokeratin expression in relation to immunohistochemical and clinical characterization in breast cancer patients with triple negative phenotype. *Tumori.* 2009;95:53-62.

88. Malorni L, Shetty PB, De Angelis C, et al. Clinical and biologic features of triple-negative breast cancers in a large cohort of patients with long-term follow-up. *Breast Cancer Res Treat.* 2012;136:795-804.

89. Livasy CA, Perou CM, Karaca G, et al. Identification of a basal-like subtype of breast ductal carcinoma in situ. *Hum Pathol.* 2007;38:197-204.

90. de Ruijter TC, Veeck J, de Hoon JPJ, van Engeland M, Tjan-Heijnen VC. Characteristics of triple-negative breast cancer. *J Cancer Res Clin Oncol.* 2011;137:183-192.

91. Seal MD, Chia SK. What is the difference between triple-negative and basal breast cancers? *Cancer J.* 2010;16:12-16.

92. Tamoxifen - National Cancer Institute. http://www.cancer.gov/cancertopics/factsheet/Therapy/tamoxifen. Accessed August 31, 2011.

93. Dictionary of Cancer Terms. http://www.cancer.gov/common/popUps/popDefinition.aspx?term=letrozole. Accessed August 31, 2011.

94. Herceptin® (Trastuzumab): Questions and Answers - National Cancer Institute. http://www.cancer.gov/cancertopics/factsheet/Therapy/herceptin. Accessed August 31, 2011.

95. Perez EA, Romond EH, Suman VJ, et al. Four-year follow-up of trastuzumab plus adjuvant chemotherapy for operable human epidermal growth factor receptor 2-positive breast cancer: joint analysis of data from NCCTG N9831 and NSABP B-31. *J Clin Oncol.* 2011;29:3366-3373.

96. Morin PJ. Claudin proteins in human cancer: promising new targets for diagnosis and therapy. *Cancer Res.* 2005;65:9603-9606.

97. What are the risk factors for breast cancer?. http://www.cancer.org/Cancer/BreastCancer/DetailedGuide/breast-cancer-risk-factors. Accessed September 29, 2011.

98. NCI Breast Cancer Prevention. http://www.cancer.gov/cancertopics/pdq/prevention/breast/HealthProfessional. Accessed January 10, 2013.

99. Hankinson SE, Colditz GA, Willett WC. Towards an integrated model for breast cancer etiology: the lifelong interplay of genes, lifestyle, and hormones. *Breast Cancer Res.* 2004;6:213-218.

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

101. 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;13:1558-1568.

102. Ma H, Bernstein L, Pike MC, Ursin G. Reproductive factors and breast cancer risk according to joint estrogen and progesterone receptor status: a meta-analysis of epidemiological studies. *Breast Cancer Res*, 2006;8:R43. doi:10.1186/bcr1525.

103. Yang XR, Chang-Claude J, Goode EL, 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;103:250-263.

104. Suzuki R, Orsini N, Saji S, Key TJ, Wolk A. Body weight and incidence of breast cancer defined by estrogen and progesterone receptor status-a meta-analysis. *Int J Cancer.* 2009;124:698-712.

105. Morabia A. Smoking (active and passive) and breast cancer: epidemiologic evidence up to June 2001. *Environ Mol Mutagen.* 2002;39:89-95.

106. Suzuki R, Orsini N, Mignone L, Saji S, Wolk A. Alcohol intake and risk of breast cancer defined by estrogen and progesterone receptor status-a meta-analysis of epidemiological studies. *Int J Cancer.* 2008;122:1832-1841.

107. Stark A, Schultz D, Kapke A, et al. Obesity and risk of the less commonly diagnosed subtypes of breast cancer. *Eur J Surg Oncol.* 2009;35:928-935.

108. Shinde SS, Forman MR, Kuerer HM, et al. Higher parity and shorter breastfeeding duration: association with triple-negative phenotype of breast cancer. *Cancer.* 2010;116:4933-4943.

109. Dolle JM, Daling JR, White E, et al. Risk factors for triple-negative breast cancer in women under the age of 45 years. *Cancer Epidemiol Biomarkers Prev.* 2009;18:1157-1166.

110. Gaudet MM, Press MF, Haile RW, et al. Risk factors by molecular subtypes of breast cancer across a population-based study of women 56 years or younger. *Breast Cancer Res Treat.* 2011;130:587-597.

111. Phipps AI, Buist DSM, Malone KE, et al. Family history of breast cancer in first-degree relatives and triple-negative breast cancer risk. *Breast Cancer Res Treat.* 2011;126:671-678.

112. Welsh ML, Buist DSM, Aiello Bowles EJ, Anderson ML, Elmore JG, Li CI. Population-based estimates of the relation between breast cancer risk, tumor subtype, and family history. *Breast Cancer Res Treat.* 2009;114:549-558.

113. Islam T, Matsuo K, Ito H, et al. Reproductive and hormonal risk factors for luminal, HER2-overexpressing, and triple-negative breast cancer in Japanese women. *Ann Oncol.* 2012;23:2435-2441.

114. Phipps AI, Malone KE, Porter PL, Daling JR, Li CI. Reproductive and hormonal risk factors for postmenopausal luminal, HER-2-overexpressing, and triple-negative breast cancer. *Cancer.* 2008;113:1521-1526.

115. Phipps AI, Chlebowski RT, Prentice R, et al. Reproductive history and oral contraceptive use in relation to risk of triple-negative breast cancer. *J Natl Cancer Inst.* 2011;103:470-477.

116. Li CI, Beaber EF, Tang MC, Porter PL, Daling JR, Malone KE. Reproductive factors and risk of estrogen receptor positive, triple-negative, and HER2-neu overexpressing breast cancer among women 20-44 years of age. *Breast Cancer Res Treat.* 2013;137:579-587.

117. Ma H, Wang Y, Sullivan-Halley J, et al. Use of four biomarkers to evaluate the risk of breast cancer subtypes in the women's contraceptive and reproductive experiences study. *Cancer Res.* 2010;70:575-587.

118. Phipps AI, Buist DSM, Malone KE, et al. Reproductive history and risk of three breast cancer subtypes defined by three biomarkers. *Cancer Causes Control.* 2011;22:399-405.

119. Saxena T, Lee E, Henderson KD, et al. Menopausal hormone therapy and subsequent risk of specific invasive breast cancer subtypes in the California Teachers Study. *Cancer Epidemiol Biomarkers Prev.* 2010;19:2366-2378.

120. Cruz GI, Martínez ME, Natarajan L, et al. Hypothesized role of pregnancy hormones on HER2+ breast tumor development. *Breast Cancer Res Treat.* 2013;137:237-246.

121. WHO: Global Database on Body Mass Index. http://apps.who.int/bmi/index.jsp?introPage=intro_3.html. Accessed October 30, 2011.

122. Phipps AI, Chlebowski RT, Prentice R, et al. Body size, physical activity, and risk of triple-negative and estrogen receptor-positive breast cancer. *Cancer Epidemiol Biomarkers Prev.* 2011;20:454-463.

123. Phipps AI, Malone KE, Porter PL, Daling JR, Li CI. Body size and risk of luminal, HER2-overexpressing, and triple-negative breast cancer in postmenopausal women. *Cancer Epidemiol Biomarkers Prev.* 2008;17:2078-2086.

124. Pierobon M, Frankenfeld CL. Obesity as a risk factor for triple-negative breast cancers: a systematic review and meta-analysis. *Breast Cancer Res Treat.* 2013;137:307-314.

125. Kabat GC, Kim M, Phipps AI, et al. Smoking and alcohol consumption in relation to risk of triple-negative breast cancer in a cohort of postmenopausal women. *Cancer Causes Control.* 2011;22:775-783.

126. Ma H, Luo J, Press MF, Wang Y, Bernstein L, Ursin G. Is there a difference in the association between percent mammographic density and subtypes of breast cancer? Luminal

A and triple-negative breast cancer. *Cancer Epidemiol Biomarkers Prev.* 2009;18:479-485.

127. Antoniou A, Pharoah PDP, Narod S, et al. Average risks of breast and ovarian cancer associated with BRCA1 or BRCA2 mutations detected in case series unselected for family history: a combined analysis of 22 studies. *Am J Hum Genet.* 2003;72:1117-1130.

128. Machiela MJ, Chen C, Chen C, Chanock SJ, Hunter DJ, Kraft P. Evaluation of polygenic risk scores for predicting breast and prostate cancer risk. *Genet Epidemiol.* 2011;35:506-514.

129. Jostins L, Barrett JC. Genetic risk prediction in complex disease. *Hum Mol Genet*, 2011;20:R182-8. doi:10.1093/hmg/ddr378.

130. Hall JM, Lee MK, Newman B, et al. Linkage of early-onset familial breast cancer to chromosome 17q21. *Science.* 1990;250:1684-1689.

131. Narod SA, Feunteun J, Lynch HT, et al. Familial breast-ovarian cancer locus on chromosome 17q12-q23. *Lancet.* 1991;338:82-83.

132. Wooster R, Neuhausen SL, Mangion J, et al. Localization of a breast cancer susceptibility gene, BRCA2, to chromosome 13q12-13. *Science.* 1994;265:2088-2090.

133. Welcsh PL, King MC. BRCA1 and BRCA2 and the genetics of breast and ovarian cancer. *Hum Mol Genet.* 2001;10:705-713.

134. Yoshida K, Miki Y. Role of BRCA1 and BRCA2 as regulators of DNA repair, transcription, and cell cycle in response to DNA damage. *Cancer Sci.* 2004;95:866-871.

135. Malone KE, Daling JR, Doody DR, et al. Prevalence and predictors of BRCA1 and BRCA2 mutations in a population-based study of breast cancer in white and black American women ages 35 to 64 years. *Cancer Res.* 2006;66:8297-8308.

136. Whittemore AS, Gong G, John EM, et al. Prevalence of BRCA1 mutation carriers among U.S. non-Hispanic Whites. *Cancer Epidemiol Biomarkers Prev.* 2004;13:2078-2083.

137. John EM, Miron A, Gong G, et al. Prevalence of pathogenic BRCA1 mutation carriers in 5 US racial/ethnic groups. *JAMA.* 2007;298:2869-2876.

138. Chen S, Parmigiani G. Meta-analysis of BRCA1 and BRCA2 penetrance. *J Clin Oncol.* 2007;25:1329-1333.

139. Comen E, Davids M, Kirchhoff T, Hudis C, Offit K, Robson M. Relative contributions of BRCA1 and BRCA2 mutations to "triple-negative" breast cancer in Ashkenazi Women. *Breast Cancer Res Treat.* 2011;129:185-190.

140. Evans DG, Howell A, Ward D, Lalloo F, Jones JL, Eccles DM. Prevalence of BRCA1 and BRCA2 mutations in triple negative breast cancer. *J Med Genet.* 2011;48:520-522.

141. Haffty BG, Yang Q, Reiss M, et al. Locoregional relapse and distant metastasis in conservatively managed triple negative early-stage breast cancer. *J Clin Oncol.* 2006;24:5652-5657.

142. Kwong A, Wong LP, Wong HN, et al. Clinical and pathological characteristics of Chinese patients with BRCA related breast cancer. *Hugo J.* 2009;3:63-76.

143. Lee E, McKean-Cowdin R, Ma H, et al. Characteristics of Triple-Negative Breast Cancer in Patients With a BRCA1 Mutation: Results From a Population-Based Study of Young Women. *J Clin Oncol.* 2011;29:4373-4380.

144. Xu J, Wang B, Zhang Y, Li R, Wang Y, Zhang S. Clinical implications for BRCA gene mutation in breast cancer. *Mol Biol Rep.* 2012;39:3097-3012.

145. Young SR, Pilarski RT, Donenberg T, et al. The prevalence of BRCA1 mutations among young women with triple-negative breast cancer. *BMC Cancer*, 2008;9:86, doi:10.1186/1471-2407-9-86.

146. Zhang J, Pei R, Pang Z, et al. Prevalence and characterization of BRCA1 and BRCA2 germline mutations in Chinese women with familial breast cancer. *Breast Cancer Res Treat.* 2011;132:421-428.

147. Mavaddat N, Barrowdale D, Andrulis IL, et al. Pathology of breast and ovarian cancers among BRCA1 and BRCA2 mutation carriers: results from the Consortium of Investigators of Modifiers of BRCA1/2 (CIMBA). *Cancer Epidemiol Biomarkers Prev.* 2012;21:134-147.

148. Foulkes WD, Stefansson IM, Chappuis PO, et al. Germline BRCA1 mutations and a basal epithelial phenotype in breast cancer. *J Natl Cancer Inst.* 2003;95:1482-1485.

149. Antoniou AC, Easton DF. Models of genetic susceptibility to breast cancer. *Oncogene.* 2006;25:5898-5905.

150. Freisinger F, Domchek SM. Clinical implications of low-penetrance breast cancer susceptibility alleles. *Curr Oncol Rep.* 2009;11:8-14.

151. Hirshfield KM, Rebbeck TR, Levine AJ. Germline mutations and polymorphisms in the origins of cancers in women. *J Oncol* 2010. doi:10.1155/2010/297671.

152. de Bock GH, Mourits MJE, Schutte M, et al. Association between the CHEK2*1100delC germ line mutation and estrogen receptor status. *Int J Gynecol Cancer.* 16 Suppl 2006;2:552-555.

153. Cybulski C, Huzarski T, Byrski T, et al. Estrogen receptor status in CHEK2-positive breast cancers: implications for chemoprevention. *Clin Genet.* 2009;75:72-78.

154. Kilpivaara O, Bartkova J, Eerola H, et al. Correlation of CHEK2 protein expression and c.1100delC mutation status with tumor characteristics among unselected breast cancer patients. *Int J Cancer.* 2005;113:575-580.

155. Schmidt MK, Tollenaar RAEM, de Kemp SR, et al. Breast cancer survival and tumor characteristics in premenopausal women carrying the CHEK2*1100delC germline mutation. *J Clin Oncol.* 2007;25:64-69.

156. Meyer A, Dörk T, Sohn C, Karstens JH, Bremer M. Breast cancer in patients carrying a germ-line CHEK2 mutation: Outcome after breast conserving surgery and adjuvant radiotherapy. *Radiother Oncol.* 2007;82:349-353.

157. Weischer M, Bojesen SE, Tybjaerg-Hansen A, Axelsson CK, Nordestgaard BG. Increased risk of breast cancer associated with CHEK2*1100delC. *J Clin Oncol.* 2007;25:57-63.

158. Nagel JHA, Peeters JK, Smid M, et al. Gene expression profiling assigns CHEK2 1100delC breast cancers to the luminal intrinsic subtypes. *Breast Cancer Res Treat.* 2012;132:439-448.

159. Domagala P, Wokolorczyk D, Cybulski C, Huzarski T, Lubinski J, Domagala W. Different CHEK2 germline mutations are associated with distinct immunophenotypic molecular subtypes of breast cancer. *Breast Cancer Res Treat.* 2012;132:937-945.

160. Cox A, Dunning AM, Garcia-Closas M, et al. A common coding variant in CASP8 is associated with breast cancer risk. *Nat Genet.* 2007;39:352-358.

161. Barroso E, Pita G, Arias JI, et al. The Fanconi anemia family of genes and its correlation with breast cancer susceptibility and breast cancer features. *Breast Cancer Res Treat.* 2009;118:655-660.

162. Reeves GK, Travis RC, Green J, et al. Incidence of breast cancer and its subtypes in relation to individual and multiple low-penetrance genetic susceptibility loci. *JAMA.* 2010;304:426-434.

163. Tapper W, Hammond V, Gerty S, et al. The influence of genetic variation in 30 selected genes on the clinical characteristics of early onset breast cancer. *Breast Cancer Res,* 2008;10:R108. doi:10.1186/bcr2213.

164. Barrett JC. Haploview: Visualization and analysis of SNP genotype data. *Cold Spring Harb Protoc* 2009. doi:10.1101/pdb.ip71.

165. Gabriel SB, Schaffner SF, Nguyen H, et al. The structure of haplotype blocks in the human genome. *Science.* 2002;296:2225-2229.

166. Haiman CA, Chen GK, Vachon CM, et al. A common variant at the TERT-CLPTM1L locus is associated with estrogen receptor-negative breast cancer. *Nat Genet.* 2011;43:1210-1214.

167. Stevens KN, Vachon CM, Lee AM, et al. Common breast cancer susceptibility loci are associated with triple-negative breast cancer. *Cancer Res.* 2011;71:6240-6249.

168. Broeks A, Schmidt MK, Sherman ME, et al. Low penetrance breast cancer susceptibility loci are associated with specific breast tumor subtypes: findings from the Breast Cancer Association Consortium. *Hum Mol Genet.* 2011;20:3289-3303.

169. Campa D, Kaaks R, Le Marchand L, et al. Interactions between genetic variants and breast cancer risk factors in the breast and prostate cancer cohort consortium. *J Natl Cancer Inst.* 2011;103:1252-1263.

170. Han S, Lee K, Choi J, et al. CASP8 polymorphisms, estrogen and progesterone receptor status, and breast cancer risk. *Breast Cancer Res Treat.* 2008;110:387-393.

171. Hamaguchi M, Nishio M, Toyama T, et al. Possible difference in frequencies of genetic polymorphisms of estrogen receptor alpha, estrogen metabolism and P53 genes between estrogen receptor-positive and -negative breast cancers. *Jpn J Clin Oncol.* 2008;38:734-742.

172. Han W, Kang D, Park IA, et al. Associations between breast cancer susceptibility gene polymorphisms and clinicopathological features. *Clin Cancer Res.* 2004;10:124-130.

173. Noma C, Miyoshi Y, Taguchi T, Tamaki Y, Noguchi S. Association of p53 genetic polymorphism (Arg72Pro) with estrogen receptor positive breast cancer risk in Japanese women. *Cancer Lett.* 2004;210:197-203.

174. Yoshimoto N, Nishiyama T, Toyama T, et al. Genetic and environmental predictors, endogenous hormones and growth factors and risk of estrogen receptor-positive breast cancer in Japanese women. *Cancer Sci.* 2011;102:2065-2072.

175. Dufloth RM, Arruda A, Heinrich JKR, Schmitt F, Zeferino LC. The investigation of DNA repair polymorphisms with histopathological characteristics and hormone receptors in a group of Brazilian women with breast cancer. *Genet Mol Res.* 2008;7:574-582.

176. Erfani N, Razmkhah M, Talei AR, et al. Cytotoxic T lymphocyte antigen-4 promoter variants in breast cancer. *Cancer Genet Cytogenet.* 2006;165:114-120.

177. Fasching PA, Loehberg CR, Strissel PL, et al. Single nucleotide polymorphisms of the aromatase gene (CYP19A1), HER2/neu status, and prognosis in breast cancer patients.

Breast Cancer Res Treat. 2008;112:89-98.

178. Heikkinen T, Kärkkäinen H, Aaltonen K, et al. The breast cancer susceptibility mutation PALB2 1592delT is associated with an aggressive tumor phenotype. *Clin Cancer Res.* 2009;15:3214-3222.

179. Krupa R, Synowiec E, Pawlowska E, et al. Polymorphism of the homologous recombination repair genes RAD51 and XRCC3 in breast cancer. *Exp Mol Pathol.* 2009;87:32-35.

180. Pooley KA, Baynes C, Driver KE, et al. Common single-nucleotide polymorphisms in DNA double-strand break repair genes and breast cancer risk. *Cancer Epidemiol Biomarkers Prev.* 2008;17:3482-3489.

181. Synowiec E, Stefanska J, Morawiec Z, Blasiak J, Wozniak K. Association between DNA damage, DNA repair genes variability and clinical characteristics in breast cancer patients. *Mutat Res.* 2008;648:65-72.

182. Zhang L, Gu L, Qian B, et al. Association of genetic polymorphisms of ER-alpha and the estradiol-synthesizing enzyme genes CYP17 and CYP19 with breast cancer risk in Chinese women. *Breast Cancer Res Treat.* 2009;114:327-338.

183. Definition of genome-wide association study - NCI Dictionary of Cancer Terms - National Cancer Institute. http://www.cancer.gov/dictionary?cdrid=636779. Accessed December 7, 2011.

184. Genome.gov | GWAS: Full Description of Methods. http://www.genome.gov/27529028. Accessed December 7, 2011.

185. Genome.gov | A Catalog of Published Genome-Wide Association Studies. http://www.genome.gov/26525384. Accessed, January 10, 2013.

186. Thomas G, Jacobs KB, Kraft P, et al. A multistage genome-wide association study in breast cancer identifies two new risk alleles at 1p11.2 and 14q24.1 (RAD51L1). *Nat Genet.* 2009;41:579-584.

187. Stacey SN, Manolescu A, Sulem P, et al. Common variants on chromosomes 2q35 and 16q12 confer susceptibility to estrogen receptor-positive breast cancer. *Nat Genet.* 2007;39:865-869.

188. Murabito JM, Rosenberg CL, Finger D, et al. A genome-wide association study of breast and prostate cancer in the NHLBI's Framingham Heart Study. *BMC Med Genet*, 2007;8 Suppl 1:S6. doi:10.1186/1471-2350-8-S1-S6.

189. Turnbull C, Ahmed S, Morrison J, et al. Genome-wide association study identifies five new breast cancer susceptibility loci. *Nat Genet.* 2010;42:504-507.

190. Fletcher O, Johnson N, Orr N, et al. Novel breast cancer susceptibility locus at 9q31.2: results of a genome-wide association study. *J Natl Cancer Inst.* 2011;103:425-435.

191. Li J, Humphreys K, Heikkinen T, et al. A combined analysis of genome-wide association studies in breast cancer. *Breast Cancer Res Treat.* 2011;126:717-727

192. Kim H, Lee J, Sung H, et al. A genome-wide association study identifies a breast cancer risk variant in ERBB4 at 2q34: results from the Seoul Breast Cancer Study. *Breast Cancer Res,* 2012;14:R56. doi:10.1186/bcr3158.

193. Elgazzar S, Zembutsu H, Takahashi A, et al. A genome-wide association study identifies a genetic variant in the SIAH2 locus associated with hormonal receptor-positive breast cancer in Japanese. *J Hum Genet.* 2012;57:766-771.

194. Easton DF, Pooley KA, Dunning AM, et al. Genome-wide association study identifies novel breast cancer susceptibility loci. *Nature.* 2007;447:1087-1093.

195. Sehrawat B, Sridharan M, Ghosh S, et al. Potential novel candidate polymorphisms identified in genome-wide association study for breast cancer susceptibility. *Hum Genet.* 2011;130:529-537.

196. Gold B, Kirchhoff T, Stefanov S, et al. Genome-wide association study provides evidence for a breast cancer risk locus at 6q22.33. *Proc Natl Acad Sci U S A.* 2008;105:4340-4345.

197. Zheng W, Long J, Gao Y, et al. Genome-wide association study identifies a new breast cancer susceptibility locus at 6q25.1. *Nat Genet.* 2009;41:324-328.

198. Siddiq A, Couch FJ, Chen GK, et al. A meta-analysis of genome-wide association studies of breast cancer identifies two novel susceptibility loci at 6q14 and 20q11. *Hum Mol Genet.* 2012;21:5373-5384.

199. Long J, Cai Q, Sung H, et al. Genome-wide association study in east Asians identifies novel susceptibility loci for breast cancer. *PLoS Genet,* 2012 8(2):e1002532, doi:10.1371/journal.pgen.1002532.

200. Cai Q, Long J, Lu W, et al. Genome-wide association study identifies breast cancer risk variant at 10q21.2: results from the Asia Breast Cancer Consortium. *Hum Mol Genet.* 2012;20:4991-4999.

201. Hunter DJ, Kraft P, Jacobs KB, et al. A genome-wide association study identifies alleles in FGFR2 associated with risk of sporadic postmenopausal breast cancer. *Nat Genet.* 2007;39:870-874.

202. Gaudet MM, Kirchhoff T, Green T, et al. Common genetic variants and modification of penetrance of BRCA2-associated breast cancer. *PLoS Genet,* 2010;6(10):e1001183,

doi:10.1371/journal.pgen.1001183.

203. Chen F, Chen GK, Stram DO, et al. A genome-wide association study of breast cancer in women of African ancestry. *Hum Genet.* 2013;132:39-48.

204. Kibriya MG, Jasmine F, Argos M, et al. A pilot genome-wide association study of early-onset breast cancer. *Breast Cancer Res Treat.* 2009;114:463-477.

205. Long J, Cai Q, Shu X, et al. Identification of a functional genetic variant at 16q12.1 for breast cancer risk: results from the Asia Breast Cancer Consortium. *PLoS Genet*, 2010;6(6):e1001002. doi:10.1371/journal.pgen.1001002.

206. Antoniou AC, Wang X, Fredericksen ZS, et al. A locus on 19p13 modifies risk of breast cancer in BRCA1 mutation carriers and is associated with hormone receptor-negative breast cancer in the general population. *Nat Genet.* 2010;42:885-892.

207. Li J, Humphreys K, Darabi H, et al. A genome-wide association scan on estrogen receptor-negative breast cancer. *Breast Cancer Res,* 2010;12:R93. 10.1186/bcr2772.

208. Stacey SN, Manolescu A, Sulem P, et al. Common variants on chromosome 5p12 confer susceptibility to estrogen receptor-positive breast cancer. *Nat Genet.* 2008;40:703-706.

209. Ahmed S, Thomas G, Ghoussaini M, et al. Newly discovered breast cancer susceptibility loci on 3p24 and 17q23.2. *Nat Genet.* 2009;41:585-590.

210. Barnholtz-Sloan JS, Shetty PB, Guan X, et al. FGFR2 and other loci identified in genome-wide association studies are associated with breast cancer in African-American and younger women. *Carcinogenesis.* 2010;31:1417-1423.

211. Bhatti P, Doody MM, Rajaraman P, et al. Novel breast cancer risk alleles and interaction with ionizing radiation among U.S. radiologic technologists. *Radiat Res.* 2010;173:214-224.

212. Figueroa JD, Garcia-Closas M, Humphreys M, et al. Associations of common variants at 1p11.2 and 14q24.1 (RAD51L1) with breast cancer risk and heterogeneity by tumor subtype: findings from the Breast Cancer Association Consortium. *Hum Mol Genet.* 2011;20:4693-4706.

213. Loizidou MA, Hadjisavvas A, Ioannidis JPA, Kyriacou K. Replication of genome-wide discovered breast cancer risk loci in the Cypriot population. *Breast Cancer Res Treat.* 2011;128:267-272.

214. Chen F, Chen GK, Millikan RC, et al. Fine-mapping of breast cancer susceptibility loci characterizes genetic risk in African Americans. *Hum Mol Genet.* 2011;20:4491-4503.

215. Huo D, Zheng Y, Ogundiran TO, et al. Evaluation of 19 susceptibility loci of breast cancer in women of African ancestry. *Carcinogenesis.* 2012;33:835-840.

216. Palmer JR, Ruiz-Narvaez EA, Rotimi CN, et al. Genetic susceptibility Loci for subtypes of breast cancer in an african american population. *Cancer Epidemiol Biomarkers Prev.* 2013;22:127-134.

217. Long J, Shu X, Cai Q, et al. Evaluation of breast cancer susceptibility loci in Chinese women. *Cancer Epidemiol Biomarkers Prev.* 2010;19:2357-2365.

218. Harlid S, Ivarsson MIL, Butt S, et al. Combined effect of low-penetrant SNPs on breast cancer risk. *Br J Cancer.* 2012;106:389-396.

219. Hemminki K, Müller-Myhsok B, Lichtner P, et al. Low-risk variants FGFR2, TNRC9 and LSP1 in German familial breast cancer patients. *Int J Cancer.* 2010;126:2858-2862.

220. Milne RL, Gaudet MM, Spurdle AB, et al. Assessing interactions between the associations of common genetic susceptibility variants, reproductive history and body mass index with breast cancer risk in the breast cancer association consortium: a combined case-control study. *Breast Cancer Res*, 2010;12:R110. doi:10.1186/bcr2797.

221. Milne RL, Benítez J, Nevanlinna H, et al. Risk of estrogen receptor-positive and -negative breast cancer and single-nucleotide polymorphism 2q35-rs13387042. *J Natl Cancer Inst.* 2009;101:1012-1018.

222. Slattery ML, Baumgartner KB, Giuliano AR, Byers T, Herrick JS, Wolff RK. Replication of five GWAS-identified loci and breast cancer risk among Hispanic and non-Hispanic white women living in the Southwestern United States. *Breast Cancer Res Treat.* 2011;129:531-539.

223. Higginbotham KS, Breyer JP, McReynolds KM, et al. A multistage genetic association study identifies breast cancer risk Loci at 10q25 and 16q24. *Cancer Epidemiol Biomarkers Prev.* 2012;21:1565-1573.

224. Jiang Y, Han J, Liu J, et al. Risk of genome-wide association study newly identified genetic variants for breast cancer in Chinese women of Heilongjiang Province. *Breast Cancer Res Treat.* 2011;128:251-257.

225. Chen W, Zhong R, Ming J, et al. The SLC4A7 variant rs4973768 is associated with breast cancer risk: evidence from a case-control study and a meta-analysis. *Breast Cancer Res Treat.* 2012;136:847-857.

226. Han W, Woo JH, Yu J, et al. Common genetic variants associated with breast cancer in Korean women and differential susceptibility according to intrinsic subtype. *Cancer Epidemiol Biomarkers Prev.* 2011;20:793-798.

227. Chan M, Ji S, Liaw C, et al. Association of common genetic variants with breast cancer risk and clinicopathological characteristics in a Chinese population. *Breast Cancer Res Treat.* 2012;136:209-220.

228. Huang Y, Ballinger DG, Dai JY, et al. Genetic variants in the MRPS30 region and postmenopausal breast cancer risk. *Genome Med.* 2011;3:42.

229. Mcinerney N, Colleran G, Rowan A, et al. Low penetrance breast cancer predisposition SNPs are site specific. *Breast Cancer Res Treat.* 2009;117:151-159.

230. Milne RL, Goode EL, García-Closas M, et al. Confirmation of 5p12 As a susceptibility locus for progesterone-receptor-positive, lower grade breast cancer. *Cancer Epidemiol Biomarkers Prev.* 2011;20:2222-2231.

231. Ruiz-Narvaez EA, Rosenberg L, Rotimi CN, et al. Genetic variants on chromosome 5p12 are associated with risk of breast cancer in African American women: the Black Women's Health Study. *Breast Cancer Res Treat.* 2010;123:525-530.

232. Garcia-Closas M, Hall P, Nevanlinna H, et al. Heterogeneity of breast cancer associations with five susceptibility loci by clinical and pathological characteristics. *PLoS Genet*, 2008;4(4):e1000054. doi:10.1371/journal.pgen.1000054.

233. Gorodnova TV, Kuligina ES, Yanus GA, et al. Distribution of FGFR2, TNRC9, MAP3K1, LSP1, and 8q24 alleles in genetically enriched breast cancer patients versus elderly tumor-free women. *Cancer Genet Cytogenet.* 2010;199:69-72.

234. Huijts PEA, Vreeswijk MPG, Kroeze-Jansema KHG, et al. Clinical correlates of low-risk variants in FGFR2, TNRC9, MAP3K1, LSP1 and 8q24 in a Dutch cohort of incident breast cancer cases. *Breast Cancer Res*, 2007;9:R78. doi:10.1186/bcr1793.

235. Latif A, Hadfield KD, Roberts SA, et al. Breast cancer susceptibility variants alter risks in familial disease. *J Med Genet.* 2010;47:126-131.

236. Rebbeck TR, DeMichele A, Tran TV, et al. Hormone-dependent effects of FGFR2 and MAP3K1 in breast cancer susceptibility in a population-based sample of post-menopausal African-American and European-American women. *Carcinogenesis.* 2009;30:269-274.

237. Tamimi RM, Lagiou P, Czene K, et al. Birth weight, breast cancer susceptibility loci, and breast cancer risk. *Cancer Causes Control.* 2010;21:689-696.

238. Zheng W, Wen W, Gao Y, et al. Genetic and clinical predictors for breast cancer risk assessment and stratification among Chinese women. *J Natl Cancer Inst.* 2010;102:972-981.

239. Lu P, Yang J, Li C, et al. Association between mitogen-activated protein kinase kinase kinase 1 rs889312 polymorphism and breast cancer risk: evidence from 59,977 subjects.

Breast Cancer Res Treat. 2011;126:663-670.

240. Sueta A, Ito H, Kawase T, et al. A genetic risk predictor for breast cancer using a combination of low-penetrance polymorphisms in a Japanese population. *Breast Cancer Res Treat.* 2012;132:711-721.

241. Kirchhoff T, Chen Z, Gold B, et al. The 6q22.33 locus and breast cancer susceptibility. *Cancer Epidemiol Biomarkers Prev.* 2009;18:2468-2475.

242. Cai Q, Wen W, Qu S, et al. Replication and functional genomic analyses of the breast cancer susceptibility locus at 6q25.1 generalize its importance in women of chinese, Japanese, and European ancestry. *Cancer Res.* 2011;71:1344-1355.

243. Stacey SN, Sulem P, Zanon C, et al. Ancestry-shift refinement mapping of the C6orf97-ESR1 breast cancer susceptibility locus. *PLoS Genet*, 2010;6(7):e1001029, doi:10.1371/journal.pgen.1001029.

244. Dai J, Hu Z, Jiang Y, et al. Breast cancer risk assessment with five independent genetic variants and two risk factors in Chinese women. *Breast Cancer Res,* 2012;14:R17. doi:10.1186/bcr3101.

245. Fletcher O, Johnson N, Gibson L, et al. Association of genetic variants at 8q24 with breast cancer risk. *Cancer Epidemiol Biomarkers Prev.* 2008;17:702-705.

246. Jiang Y, Shen H, Liu X, et al. Genetic variants at 1p11.2 and breast cancer risk: a two-stage study in Chinese women. *PLoS One*, 2011;6(6):e21563. doi:10.1371/journal.pone.0021563.

247. Lambrechts D, Truong T, Justenhoven C, et al. 11q13 is a susceptibility locus for hormone receptor positive breast cancer. *Hum Mutat.* 2012;33:1123-1132.

248. Debniak T, Górski B, Huzarski T, et al. A common variant of CDKN2A (p16) predisposes to breast cancer. *J Med Genet.* 2005;42:763-765.

249. Driver KE, Song H, Lesueur F, et al. Association of single-nucleotide polymorphisms in the cell cycle genes with breast cancer in the British population. *Carcinogenesis.* 2008;29:333-341.

250. Lindström S, Vachon CM, Li J, et al. Common variants in ZNF365 are associated with both mammographic density and breast cancer risk. *Nat Genet.* 2011;43:185-187.

251. Boyarskikh UA, Zarubina NA, Biltueva JA, et al. Association of FGFR2 gene polymorphisms with the risk of breast cancer in population of West Siberia. *Eur J Hum Genet.* 2009;17:1688-1691.

252. Prentice RL, Huang Y, Hinds DA, et al. Variation in the FGFR2 gene and the effects of postmenopausal hormone therapy on invasive breast cancer. *Cancer Epidemiol Biomarkers Prev.* 2009;18:3079-3085.

253. Cherdyntseva NV, Denisov EV, Litviakov NV, et al. Crosstalk Between the FGFR2 and TP53 Genes in breast cancer: Data from an association study and epistatic interaction analysis. *DNA Cell Biol.* 2012;31:305-315.

254. Higginbotham KSP, Breyer JP, Bradley KM, et al. A multistage association study identifies a breast cancer genetic locus at NCOA7. *Cancer Res.* 2011;71:3881-3888.

255. Marian C, Ochs-Balcom HM, Nie J, et al. FGFR2 intronic SNPs and breast cancer risk: Associations with tumor characteristics and interactions with exogenous exposures and other known breast cancer risk factors. *Int J Cancer.* 2011;129:702-712.

256. Raskin L, Pinchev M, Arad C, et al. FGFR2 is a breast cancer susceptibility gene in Jewish and Arab Israeli populations. *Cancer Epidemiol Biomarkers Prev.* 2008;17:1060-1065.

257. Wang H, Yang Z, Zhang H. Assessing interactions between the associations of fibroblast growth factor receptor 2 common genetic variants and hormone receptor status with breast cancer risk. *Breast Cancer Res Treat.* 2013;137:511-522.

258. Zhou L, Yao F, Luan H, et al. Three novel functional polymorphisms in the promoter of FGFR2 gene and breast cancer risk: a HuGE review and meta-analysis. *Breast Cancer Res Treat.* 2012;136:885-897.

259. Udler MS, Meyer KB, Pooley KA, et al. FGFR2 variants and breast cancer risk: fine-scale mapping using African American studies and analysis of chromatin conformation. *Hum Mol Genet.* 2009;18:1692-1703.

260. Chen M, Li C, Shen W, Guo Y, Shen W, Lu P. Association of a LSP1 gene rs3817198T>C polymorphism with breast cancer risk: evidence from 33,920 cases and 35,671 controls. *Mol Biol Rep.* 2011;38:4687-4695.

261. Teraoka SN, Bernstein JL, Reiner AS, et al. Single nucleotide polymorphisms associated with risk for contralateral breast cancer in the Women's Environment, Cancer and Radiation Epidemiology (WECARE) study. *Breast Cancer Res*, 2011;13:R114. doi:10.1186/bcr3057.

262. Udler MS, Ahmed S, Healey CS, et al. Fine scale mapping of the breast cancer 16q12 locus. *Hum Mol Genet.* 2010;19:2507-2515.

263. Ruiz-Narváez EA, Rosenberg L, Cozier YC, Cupples LA, Adams-Campbell LL, Palmer JR. Polymorphisms in the TOX3/LOC643714 locus and risk of breast cancer in African-American women. *Cancer Epidemiol Biomarkers Prev.* 2010;19:1320-1327.

264. Chen M, Wu X, Shen W, et al. Association between polymorphisms of trinucleotide repeat containing 9 gene and breast cancer risk: evidence from 62,005 subjects. *Breast Cancer Res Treat.* 2011;126:177-183.

265. Li L, Zhou X, Huang Z, Liu Z, Song M, Guo Z. TNRC9/LOC643714 polymorphisms are not associated with breast cancer risk in Chinese women. *Eur J Cancer Prev.* 2009;18:285-290.

266. Liang J, Chen P, Hu Z, et al. Genetic variants in trinucleotide repeat-containing 9 (TNRC9) are associated with risk of estrogen receptor positive breast cancer in a Chinese population. *Breast Cancer Res Treat.* 2010;124:237-241.

267. Tang L, Xu J, Wei F, et al. Association of STXBP4/COX11 rs6504950 (G>A) polymorphism with breast cancer risk: evidence from 17,960 cases and 22,713 controls. *Arch Med Res.* 2012;43:383-388.

268. Bensen JT, Xu Z, Smith GJ, Mohler JL, Fontham ETH, Taylor JA. Genetic polymorphism and prostate cancer aggressiveness: A case-only study of 1,536 GWAS and candidate SNPs in African-Americans and European-Americans. *Prostate.* 2013;73:11-22.

269. Haiman CA, Stram DO. Exploring genetic susceptibility to cancer in diverse populations. *Curr Opin Genet Dev.* 2010;20:330-335.

270. Hinch AG, Tandon A, Patterson N, et al. The landscape of recombination in African Americans. *Nature.* 2011;476:170-175.

271. Aldrich TE, Vann D, Moorman PG, Newman B. Rapid reporting of cancer incidence in a population-based study of breast cancer: one constructive use of a central cancer registry. *Breast Cancer Res Treat.* 1995; 35:61-64.

272. Newman B, Moorman PG, Millikan R, et al. The Carolina Breast Cancer Study: integrating population-based epidemiology and molecular biology. *Breast Cancer Res Treat.* 1995; 35:51-60.

273.Nyante, SJ. *Single nucleotide polymorphisms and the etiology of basal-like and luminal A breast cancer: a pathway-based approach.* Ph.D. dissertation, 2009;University of North Carolina at Chapel Hill.

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

275. Barnholtz-Sloan JS, McEvoy B, Shriver MD, Rebbeck TR. Ancestry estimation and correction for population stratification in molecular epidemiologic association studies. *Cancer Epidemiol Biomarkers Prev.* 2008;17:471-477.

276. 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;79:640-649.

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

278. dbSNP Home Page. http://www.ncbi.nlm.nih.gov/projects/SNP. Accessed February 20, 2012.

279. Sherry ST, Ward MH, Kholodov M, et al. dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res.* 2001;29:308-311.

280. Shen R, Fan J, Campbell D, et al. High-throughput SNP genotyping on universal bead arrays. *Mutat Res.* 2005;573:70-82.

281. TaqMan SNP Genotyping Assays. http://www3.appliedbiosystems.com/cms/groups/mcb_support/documents/generaldocuments/cms_042998.pdf. Accessed February 20, 2012.

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

283. 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;151:703-714.

284. Millikan R, Eaton A, Worley K, et al. HER2 codon 655 polymorphism and risk of breast cancer in African Americans and whites. *Breast Cancer Res Treat.* 2003;79:355-364.

285. AJCC Cancer Staging Manual, 7th edition. http://www.cancerstaging.org/staging/index.html. Accessed December 9, 2012.

286. 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;14:1545-1551.

287. Nyante SJ, Gammon MD, Kaufman JS, et al. Common genetic variation in adiponectin, leptin, and leptin receptor and association with breast cancer subtypes. *Breast Cancer Res Treat.* 2011;129:593-606.

288. Ziegler, A and Konig, IR. *A statistical approach to genetic epidemiology,* 1st edition. Weinheim, Germany: Wiley-VCH Verlag GmbH & Co. KGaA; 2006.

289. Souren NYP, Zeegers MP. Is Hardy-Weinberg on its retreat? *J Clin Epidemiol.* 2011;64:819-820.

290. Weale, ME. Quality Control for Genome-Wide Association Studies. In: Barnes MR and Breen G, eds, *Genetic Variation: Methods and Protocols*. New York: Humana Press; 2010: 341-372.

291. Rothman, KJ, Greenland, S, Lash, TL. *Modern Epidemiology*. 3rd edition. Philadelphia: Lippincott, Williams, and Wilkins; 2008.

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

293. Weinberg CR, Sandler DP. Randomized recruitment in case-control studies. *Am J Epidemiol.* 1991;134:421-432.

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

295. Anghel A, Raica M, Narita D, et al. Estrogen receptor alpha polymorphisms: correlation with clinicopathological parameters in breast cancer. *Neoplasma.* 2010;57:306-315.

296. Ding S, Yu J, Chen S, et al. Diverse associations between ESR1 polymorphism and breast cancer development and progression. *Clin Cancer Res.* 2010;16:3473-3484.

297. Dunning AM, Healey CS, Baynes C, et al. Association of ESR1 gene tagging SNPs with breast cancer risk. *Hum Mol Genet.* 2009;18:1131-1139.

298. Fernández LP, Milne RL, Barroso E, et al. Estrogen and progesterone receptor gene polymorphisms and sporadic breast cancer risk: a Spanish case-control study. *Int J Cancer.* 2006;119:467-471.

299. Ghoussaini M, Fletcher O, Michailidou K, et al. Genome-wide association analysis identifies three new breast cancer susceptibility loci. *Nat Genet.* 2012;44:312-318.

300. Sonestedt E, Ivarsson MIL, Harlid S, et al. The protective association of high plasma enterolactone with breast cancer is reasonably robust in women with polymorphisms in the estrogen receptor alpha and beta genes. *J Nutr.* 2009;139:993-1001.

301. Wang J, Higuchi R, Modugno F, et al. Estrogen receptor alpha haplotypes and breast cancer risk in older Caucasian women. *Breast Cancer Res Treat.* 2007;106:273-280.

302. Bretsky P, Haiman CA, Gilad S, et al. The relationship between twenty missense ATM variants and breast cancer risk: the Multiethnic Cohort. *Cancer Epidemiol Biomarkers Prev.*

2003;12:733-738.

303. Buchholz TA, Weil MM, Ashorn CL, et al. A Ser49Cys variant in the ataxia telangiectasia, mutated, gene that is more common in patients with breast carcinoma compared with population controls. *Cancer.* 2004;100:1345-1351.

304. Dombernowsky SL, Weischer M, Allin KH, Bojesen SE, Tybjaerg-Hansen A, Nordestgaard BG. Risk of cancer by ATM missense mutations in the general population. *J Clin Oncol.* 2008;26:3057-3062.

305. Edvardsen H, Tefre T, Jansen L, et al. Linkage disequilibrium pattern of the ATM gene in breast cancer patients and controls; association of SNPs and haplotypes to radio-sensitivity and post-lumpectomy local recurrence. *Radiat Oncol.* 2007;2:25.

306. Hirsch AE, Atencio DP, Rosenstein BS. Screening for ATM sequence alterations in African-American women diagnosed with breast cancer. *Breast Cancer Res Treat.* 2008;107:139-144.

307. Shen L, Yin Z, Wan Y, Zhang Y, Li K, Zhou B. Association between ATM polymorphisms and cancer risk: a meta-analysis. *Mol Biol Rep.* 2012;39:5719-5725.

308. Spurdle AB, Hopper JL, Chen X, et al. No evidence for association of ataxia-telangiectasia mutated gene T2119C and C3161G amino acid substitution variants with risk of breast cancer. *Breast Cancer Res*, 2002;4:R15. doi:10.1186/bcr534.

309. Ye C, Dai Q, Lu W, et al. Two-stage case-control study of common ATM gene variants in relation to breast cancer risk. *Breast Cancer Res Treat.* 2007;106:121-126.

310. Gao L, Pan X, Sun H, et al. The association between ATM D1853N polymorphism and breast cancer susceptibility: a meta-analysis. *J Exp Clin Cancer Res.* 2010;29:117.

311. Mao C, Chung VCH, He B, Luo R, Tang J. Association between ATM 5557G>A polymorphism and breast cancer risk: a meta-analysis. *Mol Biol Rep.* 2012;39:1113-1118.

312. Agresti, A. Logit Models for Multinomial Responses. In: Agresti, A. *Categorical Data Analysis*, 2nd ed. Hoboken, New York: John Wiley & Sons, Inc.; 2002:267-314.

313. Greenland S. Bayesian interpretation and analysis of research results. *Semin Hematol.* 2008;45:141-9, 2008.

314. Greenland S. Bayesian perspectives for epidemiological research. II. Regression analysis. *Int J Epidemiol.* 2007;36:195-202.

315. Gill, J. *Bayesian methods: A social and behavioral sciences approach*, 2nd edition. Washington, DC, USA: CRC press; 2002.

316. Yu X, Xun P, Hu Z, Liu P, Shen H, Chen F. Combining previously published studies with current data in Bayesian logistic regression model: an example for identifying susceptibility genes related to lung cancer in humans. *J Toxicol Environ Health A.* 2009;72:683-689.

317. Newcombe PJ, Reck BH, Sun J, et al. A comparison of Bayesian and frequentist approaches to incorporating external information for the prediction of prostate cancer risk. *Genet Epidemiol.* 2012;36:71-83.

318. Hunter DJ. Lessons from genome-wide association studies for epidemiology. *Epidemiology.* 2012;23:363-367.

319. Greenland S, Robins JM. Empirical-Bayes adjustments for multiple comparisons are sometimes useful. *Epidemiology.* 1991;2:244-251.

320. Ferguson JP, Cho JH, Yang C, Zhao H. Empirical Bayes Correction for the Winner's Curse in Genetic Association Studies. *Genet Epidemiol.* 2013;37:60-68.

321. Corbin M, Richiardi L, Vermeulen R, et al. Hierarchical regression for multiple comparisons in a case-control study of occupational risks for lung cancer. *PLoS One,* 2012;7(6):e38944. doi:10.1371/journal.pone.0038944.

322. Greenland S, Poole C. Empirical-Bayes and semi-Bayes approaches to occupational and environmental hazard surveillance. *Arch Environ Health.* 1994;49:9-16.

323. The MCMC Procedure: The MCMC Procedure :: SAS/STAT(R) 9.2 User's Guide, Second Edition. http://support.sas.com/documentation/cdl/en/statug/63033/HTML/default/viewer.htm#mcmc_toc.htm. Accessed March 5, 2012.

324. Capanu M, Orlow I, Berwick M, Hummer AJ, Thomas DC, Begg CB. The use of hierarchical models for estimating relative risks of individual genetic variants: an application to a study of melanoma. *Stat Med.* 2008;27:1973-1992.

325. Carmichael SL, Witte JS, Shaw GM. Nutrient pathways and neural tube defects: a semi-Bayesian hierarchical analysis. *Epidemiology.* 2009;20:67-73.

326. Chen GK, Witte JS. Enriching the analysis of genomewide association studies with hierarchical modeling. *Am J Hum Genet.* 2007;81:397-404.

327. Conti DV, Gauderman WJ. SNPs, haplotypes, and model selection in a candidate gene region: the SIMPle analysis for multilocus data. *Genet Epidemiol.* 2004;27:429-441.

328. Conti DV, Witte JS. Hierarchical modeling of linkage disequilibrium: genetic structure and spatial relations. *Am J Hum Genet.* 2003;72:351-363.

329. Fridley BL, Jenkins GD. Localizing putative markers in genetic association studies by incorporating linkage disequilibrium into bayesian hierarchical models. *Hum Hered.* 2010;70:63-73.

330. MacLehose RF, Dunson DB, Herring AH, Hoppin JA. Bayesian methods for highly correlated exposure data. *Epidemiology.* 2007;18:199-207.

331. Thomas DC, Stram DO, Conti D, Molitor J, Marjoram P. Bayesian spatial modeling of haplotype associations. *Hum Hered.* 2003;56:32-40.

332. Witte JS, Greenland S, Haile RW, Bird CL. Hierarchical regression analysis applied to a study of multiple dietary exposures and breast cancer. *Epidemiology.* 1994;5:612-621.

333. Hung RJ, Baragatti M, Thomas D, et al. Inherited predisposition of lung cancer: a hierarchical modeling approach to DNA repair and cell cycle control pathways. *Cancer Epidemiol Biomarkers Prev.* 2007;16:2736-2744.

334. Witte JS. Genetic analysis with hierarchical models. *Genet Epidemiol.* 1997; 14:1137-1142.

335. Reding KW, Chen C, Lowe K, et al. Estrogen-related genes and their contribution to racial differences in breast cancer risk. *Cancer Causes Control.* 2012;23:671-681.

336. Chen J, Jiang Y, Liu X, et al. Genetic variants at chromosome 9p21, 10p15 and 10q22 and breast cancer susceptibility in a Chinese population. *Breast Cancer Res Treat.* 2012;132:741-746.

337. Kawase T, Matsuo K, Suzuki T, et al. FGFR2 intronic polymorphisms interact with reproductive risk factors of breast cancer: results of a case control study in Japan. *Int J Cancer.* 2009;125:1946-1952.

338. Liang J, Chen P, Hu Z, et al. Genetic variants in fibroblast growth factor receptor 2 (FGFR2) contribute to susceptibility of breast cancer in Chinese women. *Carcinogenesis.* 2008;29:2341-2346.

339. Xu W, Shu X, Long J, et al. Relation of FGFR2 genetic polymorphisms to the association between oral contraceptive use and the risk of breast cancer in Chinese women. *Am J Epidemiol.* 2011;173:923-931.

340. Bortsov AV, Millikan RC, Belfer I, Boortz-Marx RL, Arora H, McLean SA. μ-Opioid receptor gene A118G polymorphism predicts survival in patients with breast cancer. *Anesthesiology.* 2012;116:896-902.

341. Kerlikowske K. Epidemiology of ductal carcinoma in situ. *J Natl Cancer Inst Monogr.* 2010;2010:139-141.

342. Gierach G, Burke A, Anderson WF. Epidemiology of triple negative breast cancers. *Breast Dis.* 2010;32:5-24.

343. Cole SR, Chu H, Greenland S, Hamra G, Richardson DB. Bayesian posterior distributions without Markov chains. *Am J Epidemiol.* 2012;175:368-375.

344. Stevens KN, Fredericksen Z, Vachon CM, et al. 19p13.1 is a triple-negative-specific breast cancer susceptibility locus. *Cancer Res.* 2012;72:1795-1803.

345. Phillips LS, Millikan RC, Schroeder JC, Barnholtz-Sloan JS, Levine BJ. Reproductive and hormonal risk factors for ductal carcinoma in situ of the breast. *Cancer Epidemiol Biomarkers Prev.* 2009;18:1507-1514.

346. Zhou W, Jirström K, Johansson C, et al. Long-term survival of women with basal-like ductal carcinoma in situ of the breast: a population-based cohort study. *BMC Cancer.* 2010;10:653.

347. Jia C, Jia C, Cai Y, Ma Y, Fu D. Quantitative assessment of the effect of FGFR2 gene polymorphism on the risk of breast cancer. *Breast Cancer Res Treat.* 2010;124:521-528.

348. Zhang J, Qiu L, Wang Z, et al. Current evidence on the relationship between three polymorphisms in the FGFR2 gene and breast cancer risk: a meta-analysis. *Breast Cancer Res Treat.* 2010;124:419-424.

349. Shan J, Mahfoudh W, Dsouza SP, et al. Genome-Wide Association Studies (GWAS) breast cancer susceptibility loci in Arabs: susceptibility and prognostic implications in Tunisians. *Breast Cancer Res Treat.* 2012;135:715-724.

350. Cheang MCU, Chia SK, Voduc D, et al. Ki67 index, HER2 status, and prognosis of patients with luminal B breast cancer. *J Natl Cancer Inst.* 2009;101:736-750.