



Published in final edited form as:

Cancer Res. 2011 June 1; 71(11): 3881–3888. doi:10.1158/0008-5472.CAN-10-2653.

A multi-stage association study identifies a breast cancer genetic locus at *NCOA7*

Kathryn S. P. Higginbotham¹, Joan P. Breyer¹, Kevin M. Bradley¹, Peggy A. Schuyler², W. Dale Plummer Jr.², Marcia E. Freudenthal², Amy Trentham-Dietz³, Polly A. Newcomb⁴, Melinda E. Sanders⁵, David L. Page⁵, Fritz F. Parl⁵, Kathleen M. Egan⁶, William D. Dupont², and Jeffrey R. Smith^{1,7}

¹Department of Medicine, Vanderbilt-Ingram Cancer Center, Vanderbilt University School of Medicine, Nashville, TN

²Department of Biostatistics, Vanderbilt-Ingram Cancer Center, Vanderbilt University School of Medicine, Nashville, TN

³Department of Population Health Sciences, Paul P. Carbone Comprehensive Cancer Center, University of Wisconsin-Madison, Madison, WI

⁴Cancer Prevention Program, Fred Hutchinson Cancer Research Center, Seattle, WA

⁵Department of Pathology, Vanderbilt-Ingram Cancer Center, Vanderbilt University School of Medicine, Nashville, TN

⁶Division of Cancer Prevention and Control, H. Lee Moffitt Cancer Center and Research Institute, Tampa, FL

⁷Medical Research Service, Veteran's Affairs Tennessee Valley Healthcare System, Nashville, TN

Abstract

Estrogen metabolism and growth factor signaling pathway genes play key roles in breast cancer development. We evaluated associations between breast cancer and tagging SNPs of 107 candidate genes of these pathways using single allele- and haplotype-based tests. We first sought concordance of associations between two study populations: the Nashville Breast Cohort (510 cases, 988 controls), and the Cancer Genetic Markers of Susceptibility breast cancer study (1,145 cases, 1,142 controls). Findings across the two study populations were concordant at tagging SNPs of six genes, and at previously published SNPs of *FGFR2*. We sought further replication of results for *EGFR*, *NCOA7*, and *FGFR2* in the independent Collaborative Breast Cancer Study (1,552 cases, 1,185 controls). Associations at *NCOA7* and *FGFR2* replicated across all three studies. The association at *NCOA7* on 6q22.32, detected by a haplotype spanning the initial protein-coding exon (5' - rs9375411, rs11967627, rs549438, rs529858, rs490361, rs17708107 - 3'), has not been previously reported. The haplotype had a significant inverse association with breast cancer in each study (OR_{Het} 0.69 (NBC), 0.76 (CGEMS), 0.79 (CBCS)), and a meta-analysis OR_{Het} of 0.75 (95% CI 0.65-0.87, $P = 1.4 \times 10^{-4}$) in the combined study populations. The haplotype frequency was 0.07 among cases, and 0.09 among controls; homozygotes were infrequent and each OR_{Hom} was not significant. *NCOA7* encodes a nuclear receptor co-activator that interacts with estrogen

Correspondence: William D. Dupont, Department of Biostatistics, Vanderbilt University School of Medicine, S-2323 Medical Center North, 1161 21st Ave. S., Nashville, TN 37232-2158, william.dupont@vanderbilt.edu; and Jeffrey R. Smith, Department of Medicine, Vanderbilt University School of Medicine, 529 Light Hall, 2215 Garland Avenue, Nashville, TN, 37232-0275, jeffrey.smith@vanderbilt.edu.

There are no financial disclosures.

receptor α to modulate its activity. These observations provide consistent evidence that genetic variants at the *NCOA7* locus may confer a reduced risk of breast cancer.

Keywords

Breast cancer; estrogen; genetic susceptibility; polymorphism; association; growth factor; benign breast disease

Introduction

Breast cancer is one of the most common malignancies among women worldwide. Epidemiological studies show that first-degree female relatives of women with breast cancer are at approximately two-fold increased risk of developing the disease, and twin studies suggest that 27% of risk for breast cancer is heritable (1). The comprehensive identification of both risk and protective alleles may yield important risk prediction models (2, 3). Mendelian, high- and moderate-penetrance breast cancer susceptibility genes together account for approximately one fifth of the observed heritable risk for breast cancer (4-13). Susceptibility alleles identified in genome-wide association studies (GWAS) have smaller effects on relative breast cancer risk (1.2 to 1.7 fold relative risks), but have a relatively high population prevalence with allele frequencies of 5% to 50% (12). At present, these account for approximately 8% of observed heritable risk (13-22). Several variants identified by GWAS of breast cancer are in genes of growth factor signaling and estrogen metabolism pathways, pathways previously known to play key roles in the development of breast cancer.

We describe an ongoing multi-stage association study of breast cancer risk investigating candidate genes of growth factor signaling and estrogen metabolism pathways. We employed a tagging SNP approach to efficiently detect genetic variation associated with breast cancer based upon known patterns of linkage disequilibrium among regional SNPs. In contrast to prior studies using a single SNP allele approach, we additionally employed a systematic haplotype-based analytic strategy to better capture ancestral genetic diversity for tests of association. We inferred the extant haplotypes of the study population at a given region of the genome, representing the ancestral versions of the region, any of which could harbor a disease-associated genetic variant. Haplotypes can more directly mark a disease variant than individual SNP alleles, and enable more powerful tests of association. Statistical multiple comparisons were addressed through sequential replication of observations across independent study populations. Our principal study population is a cohort of women with pathologically defined forms of benign breast disease, among them women with benign proliferative breast disease at increased risk of subsequently developing invasive breast cancer. This enabled correlation of identified genetic risk factors for breast cancer with presence of potential histological precursor lesions.

Materials and Methods

Overview

We conducted a nested case-control investigation of women within the Nashville Breast Cohort (NBC) (23-26), a longitudinal study of women diagnosed with benign breast disease, a subset of whom later developed invasive breast cancer. We evaluated associations between breast cancer and common SNPs in 107 candidate genes in the NBC study population, and compared the results to parallel analyses from the Cancer Genetic Markers of Susceptibility (CGEMS) breast cancer study data (16), seeking concordant observations. We further replicated the concordant association of variants at three genes with breast cancer risk in the independent Collaborative Breast Cancer Study (CBCS) (27). Finally, we investigated

potential interaction of the genetic and benign proliferative breast disease risk factors among women of the NBC.

Study Subjects

The Nashville Breast Cohort (NBC) is an ongoing, retrospective cohort study of 17,030 predominantly Caucasian women who underwent a breast biopsy that revealed benign parenchyma or fibroadenoma at Vanderbilt, St. Thomas, and Baptist Hospitals in Nashville, Tennessee from 1954 to the present. Benign proliferative breast lesions carry an increased risk for subsequent invasive breast cancer, and are believed to be a non-obligate precursor lesion (28-30). Roughly a third of the subjects of the cohort had this form of benign breast disease. Extensive histopathological data and benign breast disease biopsies accompany each of the subjects of the NBC. The source of germline DNA for NBC subjects is the archival (FFPE) benign tissue biopsy. Additional details on the NBC have been published elsewhere (24, 26). Entry biopsy FFPE blocks and complete follow up were available for 8,897 women of the cohort, of whom 575 developed incident breast cancer and were Caucasian. The mean age of cases of the NBC at benign breast biopsy was 45 (range 16 to 75) and at breast cancer diagnosis was 61 (range 32 to 96); 80% were postmenopausal at breast cancer diagnosis. We performed a nested case-control study, as previously described (23). Briefly, for each case we selected two controls from the risk set of women who had not been diagnosed with breast cancer in a similar period of observation. These controls were selected without replacement. Controls were matched to cases by age, race, and year of entry biopsy. DNA extraction was done by standard paraffin removal, proteinase K digestion, phenol/chloroform extraction, and ethanol precipitation. Successful DNA extractions from benign archival entry biopsy specimens were performed for 549 of the 575 Caucasian cases (95.5%). A total of 510 of the 549 cases were matched to controls for whom DNA extractions were also successful. Successful DNA extraction was achieved for 988 of the 1020 selected controls (96.9%). The study included a total of 478 trios (1 case matched to 2 controls), and 32 pairs (1 case matched to 1 control). The original archival FFPE biopsy blocks (benign tissue) that were used for DNA extraction generally were quite old: 13.3% date to the 1950s, 28.3% to the 1960s, 39.3% to the 1970s, 15.5% to the 1980s, 3.5% to the 1990s, and 0.1% to the 2000s (2002 was the most recent).

Two independent Caucasian study populations were used for replication of significant associations identified in the NBC, both well described within the published literature. The first is the CGEMS GWAS with 1,145 postmenopausal breast cancer cases and 1,142 controls from the Nurses' Health Study (16). This was an existing data set available for our analysis. Age at diagnosis for CGEMS cases was provided in five-year intervals, with a median interval of 65-69. The second study population is the Collaborative Breast Cancer Study, for which buccal cell DNA samples were available for our genotyping for 1,552 cases and 1,185 controls (27). The mean age at diagnosis of cases of the CBCS was 54 (range 28-73); 59% were postmenopausal at breast cancer diagnosis.

Selection of Candidate Genes and Variants

We selected genes in estrogen metabolism pathways and growth factor signaling pathways using databases of gene function, of biochemical pathways, and of protein-protein interaction (BioCarta, KEGG, OMIM, Entrez Gene, BIND, and HPRD). These searches yielded a set of 870 candidate genes in the targeted pathways. Among these, 107 were selected for this first-round investigation (others remain under ongoing study). We selected tagging SNPs based upon HapMap data for Caucasian (CEU) subjects (31). We identified HapMap SNPs within a gene and 20 kb flanking intervals, omitting SNPs with a minor allele frequency < 0.05 or within non-unique sequences. SNPs were assigned to linkage disequilibrium (LD) bins using LDSelect ($r^2 \geq 0.8$) (32). A SNP categorized as a candidate

functional SNP or represented on Illumina GWAS chips was preferentially selected as the tag for a given LD bin. Tagging SNPs for each gene are provided in Supplementary Table 1.

Genotyping of NBC and CBCS Subjects

Genotypes of CGEMS subjects became available as an *in silico* dataset from the National Institutes of Health during the course of this study, while genotypes of NBC and CBCS subjects were directly assayed. Genotyping was accomplished using the Illumina GoldenGate assay (Illumina, San Diego, CA). GoldenGate chemistry combines allele-specific primer extension, adapter ligation, and universal amplification of a highly multiplexed assay on a bead-based microarray system (33, 34). Genotyping of 1,613 SNPs was attempted in NBC subjects, encompassing the set of 107 genes; 1,469 SNPs (91%) successfully converted for assay (given in Supplementary Table 1). Saliva DNA samples were genotyped for a subset of 56 living NBC study subjects for comparison to the genotype of the original FFPE benign breast disease tissue block (dating from 1963 to 1994) of the same subjects. FFPE DNA genotyping fidelity was estimated to be 99.5%: 81,692 of 82,096 saliva DNA / FFPE DNA genotype pairs were concordant. On average, each 96-well plate of DNA that was genotyped contained 3.3 duplicate DNA samples of these pairs. For CBCS study DNA samples, each 96 well plate contained an average of 5.4 duplicate DNA samples (231 duplicate pairs in total), and 20,534 of 20,585 genotype pairs (99.8%) were concordant. We successfully obtained 2,628,086 of 2,636,946 (99.7%) attempted genotypes of the NBC and CBCS study subjects.

Statistical Analysis

We implemented BEAGLE version 3.1.0 (35, 36) within custom software to impute missing genotypes and determine haplotypes for each sample, across each gene. Systematic tests of association were then performed for sliding windows of SNPs across each gene. These windows varied in width from 1 to 10 SNPs; “sliding” involved scanning a given window consisting of a fixed number of SNPs along a gene one SNP at a time. At each step a test of association with breast cancer was performed. Thus both single allele- (one SNP) and haplotype-based analyses (two to ten SNPs) were conducted. Haplotype-based analytic strategies can be more powerful when the disease variant is not directly tagged (37-40). Overlapping windows are typically observed to redundantly detect a given association with this range of window sizes. Where a given genotype assay failed for a NBC or CBCS study subject, we imputed the missing data if it could be assigned with probability of 1.0. After imputation, only 0.16% of genotypes were missing in these subjects. Subjects with missing genotypes for a given haplotype were excluded from analysis. Genotypes of SNPs assayed in the NBC and CBCS studies, but not directly assayed in the CGEMS study, were imputed for CGEMS subjects by BEAGLE with a mean probability of 0.94 (35, 36). To accomplish this, BEAGLE used the known genotypes for NBC, CGEMS and reference HapMap CEU trios for a given gene, as well as HapMap genotypes of 500 SNPs to each flank as additional input.

Conditional logistic regression analyses of the NBC data were used to estimate breast cancer odds ratios (ORs), adjusted for age at entry biopsy and year of entry biopsy. Unconditional logistic regression was used to calculate these ORs in the CGEMS and CBCS studies, adjusted for age. The ORs, 95% confidence intervals (CIs), and *P*-values were derived under a model that included two β parameters to concurrently assess the risk in those heterozygous or homozygous for the allele or haplotype under consideration. Associations between genotype and breast cancer were considered to be nominally significant if the associated two-sided *P* value was less than 0.05. The associations of particular interest in this paper were those that were significant in all three study populations. Note that haplotypes consist of alleles of multiple SNPs on the same chromosome. Subjects who are homozygous for a

haplotype are homozygous at each SNP of the haplotype. Subjects who are heterozygous have one copy of the haplotype and one of a distinct haplotype; the subject must be heterozygous at a least one of the haplotype's SNPs. The combined analyses of all three data sets were performed in two ways. First, we combined all three studies together for each haplotype using a fixed-effects meta-analysis (41). An accompanying test of inter-study heterogeneity for all odds ratios of *FGFR2* and of *NCOA7* was not significant, but was significant for six of eighteen odds ratios of *EGFR* (Table 2). These tests of heterogeneity used a χ^2 statistic derived by the *metan* program (42). Second, we analyzed data of all three studies combined using unconditional logistic regression, adjusting for age and study site. For NBC women in this combined analyses, individual matching of cases to controls was ignored. This second analysis was highly consistent with the first analysis, with somewhat more significant *P* values.

A variant that concordantly replicated as a breast cancer risk factor within the NBC, CGEMS, and CBCS studies was then further analyzed for evidence of statistical interaction with a history of benign proliferative breast disease. Benign breast disease histology was uniquely available for subjects of the NBC, all of whom had forms of biopsy-proven benign breast disease. We derived the combined effects on breast cancer risk of haplotype and proliferative disease (PD) in the patient's entry biopsy. These models contained two parameters for the heterozygous or homozygous haplotype under consideration, a parameter for PD, two interaction parameters for the joint effects of heterozygous and homozygous haplotypes with PD, age at entry biopsy, and year of entry biopsy.

Results

We employed sliding window single allele- and haplotype-based tests to systematically seek concordant associations with breast cancer risk in both the NBC and CGEMS breast cancer studies. We evaluated tagging SNPs of 107 candidate genes, as well as previously published GWAS SNPs in *FGFR2* and *ESR1*. Figure 1 presents the log-transformed significance of *all* sliding window tests conducted within the NBC and CGEMS studies. These included all single allele- and haplotype-based tests at each gene. Genetic variants of six genes (*EGFR*, *ESRRG*, *NCOA2*, *NCOA7*, *PRMT8*, and *VDR*), as well as the GWAS SNPs in *FGFR2*, were associated with breast cancer risk in both studies (Figure 1, red highlight for CGEMS, green highlight for NBC). *ESR1* GWAS SNPs (rs10872676, rs2046210, and rs3020314) were not significant in either study population. All concordant results of NBC and CGEMS study subjects among these 107 candidate genes are detailed in Table 1 and in Supplementary Table 2.

We completed analysis of tagging SNPs of two of the six genes, *EGFR* and *NCOA7*, as well as the GWAS SNPs of *FGFR2*, in the independent CBCS population. The remaining four genes shown in Supplementary Table 2 are pending further investigation. Table 1 presents nominally significant tests of association for *EGFR*, *NCOA7*, and *FGFR2* in the NBC and CGEMS studies; Table 2 presents results for the CBCS study, and for a meta-analysis of all three studies combined. Only *FGFR2* and *NCOA7* were associated with breast cancer in the independent CBCS study population. *FGFR2* carries known susceptibility variants confirmed in prior GWAS (rs11200014, rs2981579, rs1219648, and rs2420946), whereas the association at *NCOA7* on 6q22.32 has not been previously reported. Six common *NCOA7* haplotypes are detected by rs9375411, rs11967627, rs549438, rs529858, rs490361, and rs17708107; one haplotype was consistently more common among controls than among cases (Tables 1, 2, and 3). This haplotype spans the initial protein-coding exon and its flanking introns, in a region of intermediate LD that extends to the neighboring gene (*HEY2* and *HINT3*, see Supplementary Figure 1). This *NCOA7* haplotype had a significant inverse association with breast cancer in each study (OR_{Het} of 0.69 (NBC), 0.76 (CGEMS), 0.79

(CBCS)); meta-analysis combining all three studies yielded an OR_{Het} of 0.75 ($P = 1.4 \times 10^{-4}$). The test of heterogeneity of these study odds ratios studies was not significant ($P = 0.77$). An alternative logistic regression analysis of the combined data of these study populations that adjusted for age and for study site provided similar results (for example, the *NCOA7* haplotype GATGCA yielded an OR_{Het} of 0.75 (95% CI 0.65-0.87, $P = 1.4 \times 10^{-4}$). A test for trend based on a simple additive logistic regression model of the combined data of the studies, adjusted for age, yielded an OR of 0.77 (95% CI 0.68-0.88, $P = 1.4 \times 10^{-4}$). With a control frequency of 0.09, homozygotes for the *NCOA7* haplotype were infrequent and our power was limited to accurately estimate the effect of the homozygous state.

We investigated the hypothesis that either of two replicating genetic variants (*NCOA7* GATGCA haplotype heterozygote or the *FGFR2* rs1219648 G allele homozygote) might alter risk of breast cancer associated with a history of PD in NBC subjects. Women with a history of PD had a 66% excess risk of breast cancer (OR = 1.66; 95% CI, 1.32 – 2.10; $P = 1.7 \times 10^{-5}$). Among women without PD, those who were heterozygous for the *NCOA7* haplotype had an OR of 0.53 (95% CI 0.32 – 0.89, $P = 0.017$) compared to women who did not inherit the haplotype. Women with both PD and the heterozygous *NCOA7* haplotype had an OR of 0.83 (95% CI, 0.45-1.53; $P = 0.54$) compared to women without PD and without the haplotype. This suggests that the haplotype attenuated the excess risk associated with PD. Among women without PD, those who were homozygous for the risk allele of *FGFR2* had an OR of 1.42 (95% CI 0.90–2.25, $P = 0.13$) compared to women homozygous for the alternative allele. Women with PD who were homozygous for the risk allele of *FGFR2* had an OR of 2.40 (95% CI, 1.20 – 4.81; $P = 0.013$) compared to women without PD and who were homozygous for the alternative allele. The tests of interaction (i.e. departure from the multiplicative model) between PD and *NCOA7*, and between PD and *FGFR2* were not significant.

Discussion

This multi-stage study investigated the potential role of candidate genes of estrogen metabolism pathways and growth factor signaling pathways in breast cancer risk within three well-characterized cohorts. Our current results identify a novel haplotype at *NCOA7* that is associated with a reduction in breast cancer risk. The *NCOA7* haplotype had a consistent significant inverse association with breast cancer in each of three independent study populations. Each of the individual SNPs that comprise the haplotype failed to detect the association, and illustrate the potential for improved detection of genetic associations through haplotype-based analysis (see Table 3). Our results also replicated the known risk association of SNPs of the second intron of *FGFR2*. At *FGFR2*, the four genotyped SNPs (each previously published) distinguish only two common haplotypes of equally significant opposite effect. However, CGEMS data of this region of *FGFR2* includes additional SNPs (seven, from rs11200014 to rs2420946) that discern four common haplotypes. Among them, a haplotype that is inversely associated with breast cancer is the most significant (data not shown). Thus, at both *NCOA7* and *FGFR2*, prominent haplotypes were observed to be inversely associated with breast cancer.

Among the set of genes explored, a number of additional genes were also associated with breast cancer in both the NBC and CGEMS (detailed in Supplementary Table 2). Each of these genes is plausibly linked to breast cancer risk, including: *ESRRG* (encoding an estrogen-related receptor), *NCOA2* (encoding a nuclear receptor co-activator), *PRMT8* (a member of a gene family, some known to act as steroid receptor co-activators), and *VDR* (a gene with a long investigational history in epidemiological studies of cancer risk). Further study in the CBCS and additional populations is needed to discern true from false positive observations among them. *NCOA7* encodes a nuclear receptor co-activator that directly

interacts with estrogen receptor α to modulate its activity. NCOA7 is recruited to estrogen receptor α target gene promoters following estradiol treatment in a dynamic fashion, similar to other co-activators. But NCOA7 is not similar to other co-activators in sequence or structure, and may represent a novel co-activator class. The ability of NCOA7 to bind ER β , TR β , PPAR γ , and RAR α has also been demonstrated (43). Variants affecting NCOA7 co-activator functions could impact hormone-induced gene transcription. Alternatively, NCOA7 may function in the prevention of oxidative DNA damage (44) by estrogen quinones as a mechanism for carcinogenesis (reviewed in (45)).

The associated *NCOA7* haplotype encompassed the initial protein-encoding exon and was not efficiently tagged by any of the individual SNPs evaluated within the gene and its flanking sequence (20 kb upstream or downstream). A maximum r^2 of 0.20 was observed between the haplotype and tagging SNP rs490361 among NBC subjects. In HapMap data of Caucasians, the 140 kb gene resides within a larger region of moderate linkage disequilibrium, spanning approximately 250 kb (Supplementary Figure 1). The *NCOA7* haplotype is not in notable linkage disequilibrium with other individual HapMap SNPs of the 250 kb interval (maximum r^2 = 0.16 with rs674859). It is possible that the association signal that we detected at *NCOA7* may be attributable to variation anywhere within the larger region of linkage disequilibrium. The identification of candidate functional variants will require additional investigation of this interval, including the genes *HEY2*, *NCOA7*, and *HINT3*.

The first stage of this study was conducted using samples from the Nashville Breast Cohort, whose participants all have a history of benign breast disease. Studies have shown that those with benign proliferative breast disease have an approximately two-fold increase in breast cancer risk (24-26, 28, 29, 46). Epithelial hyperplastic breast lesions are thought to represent an initial step in a non-obligate path that can result in invasive breast cancer (28, 30). Given this, one might anticipate that this histologic risk factor could interact with additional genetic risk factors for the development of breast cancer. We observed that the joint effects of the genetic (*FGFR2* or *NCOA7*) and histologic risk factors on breast cancer risk appear to be multiplicative. It is possible that proliferative breast disease is on a causal pathway between genotype and breast cancer, or that genotype directly affects the risk of both proliferative breast disease and breast cancer. Our nested case-control study was not designed to distinguish between these two hypotheses.

We initially conducted an exploratory analysis of the NBC of 1,469 tagging SNPs from 107 genes associated with estrogen metabolism and growth factor signaling pathways. The source of DNA from the NBC was from archival FFPE benign breast tissue. DNA from this source was not amenable to current chip-based GWAS techniques, but was amenable to alternative moderate-throughput methods. In our sliding window analytic approach we evaluated 199,937 overlapping single allele and haplotype windows in the NBC. The extent of this overlap meant that many of these tests were highly correlated and were effectively evaluating the same variation in allele frequencies between cases and controls. Nevertheless, the risk of spurious findings in our original analysis due to multiple comparisons artifacts was very high. To address this, we replicated our results in the CGEMS study population, and then again in the CBCS study population. The fact that an *NCOA7* haplotype was significantly associated with breast cancer risk in all three of these studies markedly increases the likelihood that this association is real. However, further investigation in independent study populations is warranted.

An association at *NCOA7* has not previously been described by GWAS of breast cancer. Five GWAS investigations of breast cancer have identified twenty-three susceptibility loci to date. Each of these studies have conducted single-allele analyses as the principal analytic

strategy, and have sought sequential replication of observations to discern true from false positive results. One of the discovered loci resides at 6q22.33, initially detected by the SNP rs2180341 (20, 47, 48). This locus is 1.4 Mb telomeric to *NCOA7*; the two loci are not in linkage disequilibrium. Single-allele analysis of tagging SNPs at *NCOA7* failed to detect an association in the NBC, CGEMS, or CBCS breast cancer study populations. Only haplotype-based analysis detected the association within these study populations. Given this observation, single allele analysis within each of the GWAS studies might not be likely to detect the association at *NCOA7*. CGEMS GWAS data was accessible to us for haplotype-based investigation, and was concordant.

The goal of this project has been to identify genetic factors that permit the identification of women who are at particular risk for breast cancer and for whom informed clinical decisions regarding prevention are valuable. The risks conferred by individual common variants identified to date are small. Risk profiles generated by these alleles may be useful in stratification of population risk, but they do not yet have sufficient positive predictive value for individual assessment (2, 3). The portfolio of genetic variants inherited by an individual includes both risk and protective alleles. More comprehensive knowledge of risk-modifying genetic variants is needed for development of clinically useful risk-stratification tests. Such tests may in the future improve clinical care by identifying women who would most benefit from more intensive screening, those who may be at particularly increased risk if they were to take hormone replacement therapy, or those who may benefit from prophylactic treatment.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We wish to acknowledge the contributions of the participating study subjects, as well as that of investigators of the Cancer Genetic Markers of Susceptibility Study, and additional investigators of the Collaborative Breast Cancer Study, Drs. Montserrat Garcia-Closas and Linda T. Titus-Ernstoff. This study was supported by NIH grants CA098131, CA050468, CA068485, 1UL1RR024975, by a MERIT grant from the US Department of Veterans Affairs, and by an award from the V Foundation. The CBCS was supported by NIH grants CA105197, CA47147, CA47305, and CA69664.

References

1. Lichtenstein P, Holm NV, Verkasalo PK, Iliadou A, Kaprio J, Koskenvuo M, et al. Environmental and heritable factors in the causation of cancer--analyses of cohorts of twins from Sweden, Denmark, and Finland. *N Engl J Med.* 2000; 343:78–85. [PubMed: 10891514]
2. Pharoah PD, Antoniou AC, Easton DF, Ponder BA. Polygenes, risk prediction, and targeted prevention of breast cancer. *N Engl J Med.* 2008; 358:2796–803. [PubMed: 18579814]
3. Wacholder S, Hartge P, Prentice R, Garcia-Closas M, Feigelson HS, Diver WR, et al. Performance of common genetic variants in breast-cancer risk models. *N Engl J Med.* 2010; 362:986–93. [PubMed: 20237344]
4. Garber JE, Goldstein AM, Kantor AF, Dreyfus MG, Fraumeni JF Jr, Li FP. Follow-up study of twenty-four families with Li-Fraumeni syndrome. *Cancer Res.* 1991; 51:6094–7. [PubMed: 1933872]
5. Hearle N, Schumacher V, Menko FH, Olschwang S, Boardman LA, Gille JJ, et al. Frequency and spectrum of cancers in the Peutz-Jeghers syndrome. *Clin Cancer Res.* 2006; 12:3209–15. [PubMed: 16707622]
6. Masciari S, Larsson N, Senz J, Boyd N, Kaurah P, Kandel MJ, et al. Germline E-cadherin mutations in familial lobular breast cancer. *J Med Genet.* 2007; 44:726–31. [PubMed: 17660459]

7. Nevanlinna H, Bartek J. The CHEK2 gene and inherited breast cancer susceptibility. *Oncogene*. 2006; 25:5912–9. [PubMed: 16998506]
8. Rahman N, Seal S, Thompson D, Kelly P, Renwick A, Elliott A, et al. PALB2, which encodes a BRCA2-interacting protein, is a breast cancer susceptibility gene. *Nat Genet*. 2007; 39:165–7. [PubMed: 17200668]
9. Renwick A, Thompson D, Seal S, Kelly P, Chagtai T, Ahmed M, et al. ATM mutations that cause ataxia-telangiectasia are breast cancer susceptibility alleles. *Nat Genet*. 2006; 38:873–5. [PubMed: 16832357]
10. Saal LH, Gruvberger-Saal SK, Persson C, Lovgren K, Jumppanen M, Staaf J, et al. Recurrent gross mutations of the PTEN tumor suppressor gene in breast cancers with deficient DSB repair. *Nat Genet*. 2008; 40:102–7. [PubMed: 18066063]
11. Seal S, Thompson D, Renwick A, Elliott A, Kelly P, Barfoot R, et al. Truncating mutations in the Fanconi anemia J gene BRIP1 are low-penetrance breast cancer susceptibility alleles. *Nat Genet*. 2006; 38:1239–41. [PubMed: 17033622]
12. Stratton MR, Rahman N. The emerging landscape of breast cancer susceptibility. *Nat Genet*. 2008; 40:17–22. [PubMed: 18163131]
13. Turnbull C, Ahmed S, Morrison J, Pernet D, Renwick A, Maranian M, et al. Genome-wide association study identifies five new breast cancer susceptibility loci. *Nat Genet*. 2010; 42:504–7. [PubMed: 20453838]
14. Cox A, Dunning AM, Garcia-Closas M, Balasubramanian S, Reed MW, Pooley KA, et al. A common coding variant in CASP8 is associated with breast cancer risk. *Nat Genet*. 2007; 39:352–8. [PubMed: 17293864]
15. Easton DF, Pooley KA, Dunning AM, Pharoah PD, Thompson D, Ballinger DG, et al. Genome-wide association study identifies novel breast cancer susceptibility loci. *Nature*. 2007; 447:1087–93. [PubMed: 17529967]
16. Hunter DJ, Kraft P, Jacobs KB, Cox DG, Yeager M, Hankinson SE, et al. A genome-wide association study identifies alleles in FGFR2 associated with risk of sporadic postmenopausal breast cancer. *Nat Genet*. 2007; 39:870–4. [PubMed: 17529973]
17. Stacey SN, Manolescu A, Sulem P, Rafnar T, Gudmundsson J, Gudjonsson SA, et al. Common variants on chromosomes 2q35 and 16q12 confer susceptibility to estrogen receptor-positive breast cancer. *Nat Genet*. 2007; 39:865–9. [PubMed: 17529974]
18. Zheng W, Long J, Gao YT, Li C, Zheng Y, Xiang YB, et al. Genome-wide association study identifies a new breast cancer susceptibility locus at 6q25.1. *Nat Genet*. 2009; 41:324–8. [PubMed: 19219042]
19. Ahmed S, Thomas G, Ghoussaini M, Healey CS, Humphreys MK, Platte R, et al. Newly discovered breast cancer susceptibility loci on 3p24 and 17q23.2. *Nat Genet*. 2009; 41:585–90. [PubMed: 19330027]
20. Gold B, Kirchhoff T, Stefanov S, Lautenberger J, Viale A, Garber J, 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; 105:4340–5. [PubMed: 18326623]
21. Stacey SN, Manolescu A, Sulem P, Thorlacius S, Gudjonsson SA, Jonsson GF, et al. Common variants on chromosome 5p12 confer susceptibility to estrogen receptor-positive breast cancer. *Nat Genet*. 2008; 40:703–6. [PubMed: 18438407]
22. Thomas G, Jacobs KB, Kraft P, Yeager M, Wacholder S, Cox DG, et al. A multistage genome-wide association study in breast cancer identifies two new risk alleles at 1p11.2 and 14q24.1 (RAD51L1). *Nat Genet*. 2009; 41:579–84. [PubMed: 19330030]
23. Dupont WD, Breyer JP, Bradley KM, Schuyler PA, Plummer WD, Sanders ME, et al. Protein phosphatase 2A subunit gene haplotypes and proliferative breast disease modify breast cancer risk. *Cancer*. 2010; 116:8–19. [PubMed: 19890961]
24. Dupont WD, Page DL. Risk factors for breast cancer in women with proliferative breast disease. *N Engl J Med*. 1985; 312:146–51. [PubMed: 3965932]
25. Dupont WD, Parl FF, Hartmann WH, Brinton LA, Winfield AC, Worrell JA, et al. Breast cancer risk associated with proliferative breast disease and atypical hyperplasia. *Cancer*. 1993; 71:1258–65. [PubMed: 8435803]

26. Page DL, Dupont WD, Rogers LW, Rados MS. Atypical hyperplastic lesions of the female breast. A long-term follow-up study. *Cancer*. 1985; 55:2698–708. [PubMed: 2986821]
27. Zhang Y, Newcomb PA, Egan KM, Titus-Ernstoff L, Chanock S, Welch R, et al. Genetic polymorphisms in base-excision repair pathway genes and risk of breast cancer. *Cancer Epidemiol Biomarkers Prev*. 2006; 15:353–8. [PubMed: 16492928]
28. Hartmann LC, Sellers TA, Frost MH, Lingle WL, Degnim AC, Ghosh K, et al. Benign breast disease and the risk of breast cancer. *N Engl J Med*. 2005; 353:229–37. [PubMed: 16034008]
29. London SJ, Connolly JL, Schnitt SJ, Colditz GA. A prospective study of benign breast disease and the risk of breast cancer. *JAMA*. 1992; 267:941–4. [PubMed: 1734106]
30. Lakhani SR. The transition from hyperplasia to invasive carcinoma of the breast. *J Pathol*. 1999; 187:272–8. [PubMed: 10398078]
31. Altshuler D, Brooks LD, Chakravarti A, Collins FS, Daly MJ, Donnelly P. A haplotype map of the human genome. *Nature*. 2005; 437:1299–320. [PubMed: 16255080]
32. Carlson CS, Eberle MA, Rieder MJ, Yi Q, Kruglyak L, Nickerson DA. Selecting a maximally informative set of single-nucleotide polymorphisms for association analyses using linkage disequilibrium. *Am J Hum Genet*. 2004; 74:106–20. [PubMed: 14681826]
33. Shen R, Fan JB, Campbell D, Chang W, Chen J, Doucet D, et al. High-throughput SNP genotyping on universal bead arrays. *Mutation Research*. 2005; 573:70–82. [PubMed: 15829238]
34. Fan JB, Oliphant A, Shen R, Kermani BG, Garcia F, Gunderson KL, et al. Highly parallel SNP genotyping. *Cold Spring Harbor Symposia on Quantitative Biology*. 2003; 68:69–78.
35. Browning BL, Browning SR. A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals. *Am J Hum Genet*. 2009; 84:210–23. [PubMed: 19200528]
36. Browning SR, Browning BL. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *Am J Hum Genet*. 2007; 81:1084–97. [PubMed: 17924348]
37. de Bakker PI, Yelensky R, Pe'er I, Gabriel SB, Daly MJ, Altshuler D. Efficiency and power in genetic association studies. *Nat Genet*. 2005; 37:1217–23. [PubMed: 16244653]
38. Breyer JP, Sanders ME, Airey DC, Cai Q, Yaspan BL, Schuyler PA, et al. Heritable variation of ERBB2 and breast cancer risk. *Cancer Epidemiol Biomarkers Prev*. 2009; 18:1252–8. [PubMed: 19336545]
39. Fallin D, Schork NJ. Accuracy of haplotype frequency estimation for biallelic loci, via the expectation-maximization algorithm for unphased diploid genotype data. *Am J Hum Genet*. 2000; 67:947–59. [PubMed: 10954684]
40. Yaspan BL, Breyer JP, Cai Q, Dai Q, Elmore JB, Amundson I, et al. Haplotype analysis of CYP11A1 identifies promoter variants associated with breast cancer risk. *Cancer Res*. 2007; 67:5673–82. [PubMed: 17575134]
41. Sterne, JA. *Meta-analysis in Stata: an updated collection from the Stata Journal*. College Station TX: Stata Press; 2009.
42. Harris R, Bradburn M, Deeks J, Harbord R, Altman D, Sterne J. *Metan: fixed- and random-effects meta-analysis*. *Stata J*. 2008; 8:3–28.
43. Shao W, Halachmi S, Brown M. ERAP140, a conserved tissue-specific nuclear receptor coactivator. *Mol Cell Biol*. 2002; 22:3358–72. [PubMed: 11971969]
44. Durand M, Kolpak A, Farrell T, Elliott NA, Shao W, Brown M, et al. The OXR domain defines a conserved family of eukaryotic oxidation resistance proteins. *BMC Cell Biol*. 2007; 8:13. [PubMed: 17391516]
45. Bolton JL, Thatcher GR. Potential mechanisms of estrogen quinone carcinogenesis. *Chem Res Toxicol*. 2008; 21:93–101. [PubMed: 18052105]
46. Palli D, Rosselli del Turco M, Simoncini R, Bianchi S. Benign breast disease and breast cancer: a case-control study in a cohort in Italy. *Int J Cancer*. 1991; 47:703–6. [PubMed: 2004851]
47. Hemminki K, Muller-Myhsok B, Lichtner P, Engel C, Chen B, Burwinkel B, et al. Low-risk variants FGFR2, TNRC9 and LSP1 in German familial breast cancer patients. *Int J Cancer*. 2010; 126:2858–62. [PubMed: 19856316]

48. Kirchoff T, Chen ZQ, Gold B, Pal P, Gaudet MM, Kosarin K, et al. The 6q22.33 locus and breast cancer susceptibility. *Cancer Epidemiol Biomarkers Prev.* 2009; 18:2468–75. [PubMed: 19690183]

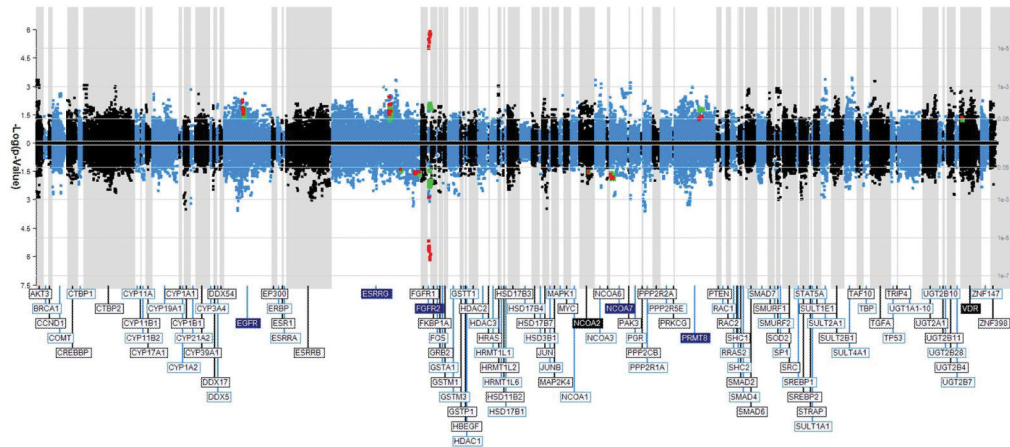


Figure 1.

Association of breast cancer with evaluated estrogen and growth factor signaling pathway genes. Data of the Nashville Breast Cohort (NBC) and of the CGEMS breast cancer study are jointly illustrated. A Manhattan plot shows $-\log_{10}(P\text{-values})$ from sliding-window analysis of tagging SNPs of 107 genes, and of published GWAS SNPs in *ESR1* and *FGFR2*. Windows ranged in width from 1 SNP (single allele analysis) to 10 SNPs (haplotype analysis). All statistical tests conducted in NBC and CGEMS study data are plotted. A given data point depicted at a SNP may represent either a single allele test, or a test of a haplotype beginning at the SNP and that extends to the right (corresponding to the window width). Data points include both those for the heterozygous and for the homozygous state for each variant. Data points plotted above the horizontal midline correspond to a test with an OR >1 , while those plotted below it correspond to a test with an OR <1 . Highlighted data points in 7 genes (named in solid boxes) each designate a genetic variant that was significant within NBC data (green) and also within the corresponding analysis of CGEMS data (red). Black/blue shading distinguishes adjacent genes of the figure.

Table 1
Associations between variants of *FGFR2*, *NCOA7*, and *EGFR* and breast cancer risk: NBC, and CGEMS

Gene	Haplotype/Allele	NBC discovery set			CGEMS replication set				
		OR _{Het} (95% CI)	P	OR _{Het} (95% CI)	OR _{Het} (95% CI)	P	OR _{Het} (95% CI)		
<i>FGFR2</i>	A	1.16 (0.91-1.47)	0.24	1.54 (1.12-2.12)	0.007	1.25 (1.04-1.50)	0.02	1.71 (1.35-2.17)	8.2 x 10 ⁻⁶
	A T	1.16 (0.91-1.47)	0.24	1.54 (1.12-2.12)	0.007	1.26 (1.04-1.51)	0.02	1.71 (1.35-2.17)	8.8 x 10 ⁻⁶
	A T G	1.18 (0.93-1.49)	0.18	1.56 (1.12-2.15)	0.008	1.26 (1.05-1.52)	0.01	1.78 (1.40-2.26)	2.9 x 10 ⁻⁶
	A T G T	1.18 (0.93-1.49)	0.18	1.57 (1.13-2.20)	0.008	1.28 (1.06-1.53)	0.01	1.75 (1.37-2.23)	6.9 x 10 ⁻⁶
	T	1.19 (0.93-1.51)	0.17	1.54 (1.12-2.10)	0.007	1.25 (1.04-1.51)	0.02	1.74 (1.38-2.21)	4.3 x 10 ⁻⁶
	T G	1.20 (0.95-1.53)	0.13	1.57 (1.14-2.17)	0.006	1.26 (1.05-1.51)	0.01	1.81 (1.42-2.30)	1.6 x 10 ⁻⁶
	T G T	1.21 (0.95-1.53)	0.12	1.60 (1.15-2.24)	0.006	1.27 (1.06-1.53)	0.01	1.78 (1.39-2.27)	3.8 x 10 ⁻⁶
	G	1.20 (0.94-1.52)	0.14	1.55 (1.12-2.14)	0.008	1.25 (1.04-1.50)	0.02	1.82 (1.43-2.31)	1.2 x 10 ⁻⁶
	G T	1.19 (0.94-1.50)	0.16	1.57 (1.13-2.18)	0.008	1.26 (1.05-1.51)	0.01	1.80 (1.41-2.29)	2.2 x 10 ⁻⁶
	T	1.18 (0.93-1.49)	0.18	1.54 (1.11-2.13)	0.010	1.26 (1.05-1.51)	0.02	1.81 (1.42-2.31)	1.6 x 10 ⁻⁶
<i>NCOA7</i>	G A T G C	0.69 (0.49-0.96)	0.03	0.92 (0.35-2.44)	0.87	0.76 (0.60-0.96)	0.02	0.64 (0.18-2.27)	0.49
	G A T G C A	0.69 (0.49-0.96)	0.03	0.92 (0.35-2.44)	0.87	0.76 (0.60-0.96)	0.02	0.64 (0.18-2.27)	0.49
	G A T G C C	0.69 (0.49-0.96)	0.03	0.92 (0.35-2.44)	0.87	0.76 (0.60-0.96)	0.02	0.64 (0.18-2.27)	0.49
	G A T G C A C	0.69 (0.49-0.96)	0.03	0.92 (0.35-2.44)	0.87	0.76 (0.60-0.96)	0.02	0.64 (0.18-2.27)	0.49
<i>EGFR</i>	G C G G	1.41 (1.05-1.90)	0.02	0.91 (0.23-3.61)	0.89	1.34 (1.09-1.66)	0.01	1.47 (0.71-3.01)	0.30
	G C G C	1.40 (1.04-1.89)	0.03	0.91 (0.23-3.61)	0.89	1.36 (1.10-1.68)	0.005	1.39 (0.67-2.87)	0.38
	G C G C G	1.46 (1.08-1.97)	0.01	0.63 (0.13-3.19)	0.58	1.31 (1.05-1.62)	0.01	1.53 (0.71-3.32)	0.28
	G C G C G C	1.46 (1.08-1.97)	0.01	0.63 (0.13-3.19)	0.58	1.31 (1.05-1.62)	0.01	1.53 (0.71-3.32)	0.28
	C G	1.30 (1.03-1.64)	0.03	1.06 (0.64-1.75)	0.83	1.22 (1.02-1.45)	0.03	1.32 (0.94-1.85)	0.11
	C G C G	1.40 (1.06-1.86)	0.02	0.79 (0.24-2.53)	0.69	1.28 (1.04-1.58)	0.02	1.58 (0.76-3.30)	0.22
	G	1.34 (1.06-1.68)	0.01	1.03 (0.62-1.69)	0.91	1.22 (1.03-1.45)	0.02	1.34 (0.80-2.51)	0.09
	G G	1.32 (1.02-1.72)	0.04	1.38 (0.66-2.86)	0.39	1.26 (1.04-1.53)	0.02	1.42 (0.80-2.51)	0.24
	G G C	1.32 (1.01-1.72)	0.04	1.38 (0.66-2.85)	0.39	1.27 (1.05-1.55)	0.01	1.37 (0.77-2.44)	0.29
	G G C C	1.32 (1.01-1.72)	0.04	1.38 (0.66-2.85)	0.39	1.27 (1.05-1.55)	0.01	1.37 (0.77-2.44)	0.29

Each table row for a given gene presents a haplotype/allele for which tests of association were concordantly significant in the NBC and CGEMS study populations. The underlined SNP of *EGFR* was not genotyped in the CGEMS study, but was imputed with probability 0.79.

Each table row for a given gene presents a haplotype/allele for which tests of association were concordantly significant in the NBC and CGEMS study populations. The underlined SNP of *EGFR* was not genotyped in the CGEMS study, but was imputed with probability 0.79.

Table 2
Associations between variants of *FGFR2*, *NCOA7*, and *EGFR* and breast cancer risk: CBCS, and combined studies (NBC, CGEMS, CBCS)

Gene	Haplotype/Allele	CBCS replication set			Variant Frequency			Combined study sets			
		OR _{hwt} (95% CI)	P	OR _{hwt} (95% CI)	P	Case	Control	OR _{hwt} (95% CI)	P	OR _{hwt} (95% CI)	P
<i>FGFR2</i>	A	1.12 (0.93-1.35)	0.24	1.59 (1.21-2.07)	7.0 x 10 ⁻⁴	0.45	0.39	1.17 (1.04-1.32)	0.007	1.62 (1.39-1.89)	9.3 x 10 ⁻¹⁰
	A T	1.14 (0.94-1.37)	0.18	1.61 (1.23-2.11)	5.5 x 10 ⁻⁴	0.45	0.39	1.18 (1.05-1.33)	0.005	1.63 (1.39-1.90)	7.7 x 10 ⁻¹⁰
	A T G	1.15 (0.95-1.38)	0.15	1.64 (1.24-2.17)	4.6 x 10 ⁻⁴	0.43	0.37	1.19 (1.07-1.34)	0.002	1.67 (1.43-1.96)	2.5 x 10 ⁻¹⁰
	A T G T	1.14 (0.95-1.38)	0.16	1.57 (1.18-2.09)	0.002	0.42	0.36	1.20 (1.07-1.34)	0.002	1.64 (1.40-1.93)	2.9 x 10 ⁻⁹
	T	1.19 (1.01-1.42)	0.04	1.43 (1.14-1.80)	0.002	0.45	0.40	1.21 (1.08-1.35)	0.001	1.56 (1.35-1.80)	2.8 x 10 ⁻⁹
	T G	1.19 (1.01-1.42)	0.04	1.43 (1.14-1.80)	0.002	0.44	0.38	1.22 (1.09-1.36)	0.001	1.63 (1.40-1.89)	2.5 x 10 ⁻¹⁰
	T G T	1.18 (1.00-1.41)	0.05	1.42 (1.11-1.81)	0.005	0.43	0.37	1.22 (1.09-1.36)	0.001	1.59 (1.36-1.85)	3.8 x 10 ⁻⁹
	G	1.21 (1.02-1.43)	0.03	1.52 (1.20-1.92)	4.6 x 10 ⁻⁴	0.44	0.38	1.22 (1.09-1.36)	0.001	1.63 (1.40-1.89)	1.6 x 10 ⁻¹⁰
	G T	1.19 (1.00-1.41)	0.05	1.42 (1.12-1.81)	0.004	0.43	0.38	1.21 (1.08-1.35)	0.001	1.59 (1.36-1.85)	2.9 x 10 ⁻⁹
	T	1.19 (1.00-1.40)	0.05	1.42 (1.12-1.80)	0.003	0.44	0.38	1.20 (1.08-1.34)	0.001	1.58 (1.36-1.84)	2.3 x 10 ⁻⁹
<i>NCOA7</i>	A	0.80 (0.64-0.99)	0.04	0.73 (0.24-2.29)	0.60	0.07	0.09	0.76 (0.66-0.88)	1.9 x 10 ⁻⁴	0.78 (0.41-1.48)	0.45
	A T	0.79 (0.63-0.96)	0.03	0.73 (0.23-2.27)	0.59	0.07	0.09	0.75 (0.65-0.87)	1.4 x 10 ⁻⁴	0.78 (0.41-1.48)	0.44
	G										
	G T										
<i>EGFR</i>	A	0.96 (0.78-1.18)	0.68	1.67 (0.72-3.86)	0.23	0.11	0.09	1.18 (1.04-1.35)	0.01	1.44 (0.87-2.40)	0.16
	A T	0.96 (0.78-1.18)	0.70	1.68 (0.73-3.89)	0.22	0.11	0.09	1.19 (1.04-1.39)	0.01	1.41 (0.84-2.34)	0.19
	A T G	0.97 (0.79-1.20)	0.79	1.59 (0.68-3.71)	0.28	0.10	0.09	1.19 (1.04-1.36)	0.01	1.41 (0.83-2.42)	0.21
	A T G C	0.98 (0.79-1.21)	0.85	1.60 (0.69-3.72)	0.28	0.09	0.09	1.19 (1.04-1.36)	0.01	1.49 (0.83-2.42)	0.21
	C	0.91 (0.77-1.07)	0.26	1.21 (0.87-1.68)	0.25	0.26	0.23	1.09 (0.98-1.22)	0.10	1.23 (0.99-1.52)	0.06
	C G	1.00 (0.82-1.22)	0.99	1.24 (0.63-2.42)	0.53	0.12	0.10	1.18 (1.04-1.34)	0.01	1.29 (0.81-2.01)	0.30
	C G C	0.93 (0.79-1.09)	0.36	1.22 (0.89-1.67)	0.22	0.26	0.24	1.11 (1.00-1.23)	0.06	1.23 (1.00-1.52)	0.05
	C G C C	0.97 (0.81-1.17)	0.77	1.35 (0.78-2.35)	0.29	0.15	0.13	1.14 (1.02-1.29)	0.03	1.38 (0.97-1.96)	0.07
	G	0.97 (0.81-1.16)	0.73	1.35 (0.78-2.35)	0.29	0.15	0.13	1.15 (1.02-1.29)	0.03	1.37 (0.96-1.94)	0.08
	G C										

Each table row for a given gene presents a haplotype/allele for which tests of association were concordantly significant in the NBC and CGEMS study populations. The underlined SNP of *EGFR* was not genotyped in the CGEMS study, but was imputed with probability 0.79. Results of the combined studies are derived from meta-analysis.

Table 3

NCOA7 haplotype and single SNP analyses.

Haplotype		Variant Frequency		OR _{Het} (95% CI)	P	OR _{Hom} (95% CI)	P	
		Case	Control					
A	G A T G C A	0.07	0.09	0.75 (0.65-0.87)	1.5 x 10 ⁻⁴	0.78 (0.41-1.48)	0.44	
B	G A T G T A	0.37	0.36	1.00 (0.90-1.11)	1.00	1.12 (0.95-1.31)	0.17	
C	A A T G C A	0.21	0.20	1.00 (0.90-1.11)	1.00	1.21 (0.93-1.58)	0.16	
D	A G C G T A	0.09	0.09	1.00 (0.87-1.15)	0.98	0.94 (0.52-1.72)	0.85	
E	A A C G T G	0.09	0.09	1.00 (0.87-1.15)	0.97	1.21 (0.61-2.40)	0.59	
F	A A C G T A	0.07	0.06	1.15 (0.99-1.34)	0.07	1.38 (0.59-3.23)	0.46	
Single SNP Allele								
	rs9375411	G	0.49	0.50	0.90 (0.79-1.01)	0.08	0.95 (0.82-1.09)	0.43
	rs11967627	A	0.90	0.90	1.06 (0.62-1.82)	0.83	1.09 (0.64-1.84)	0.76
	rs549438	T	0.70	0.71	1.00 (0.83-1.20)	1.00	0.94 (0.78-1.12)	0.49
	rs529858	G	0.98	0.98	0.97 (0.67-1.40)	0.87	0.86 (0.12-6.09)	0.88
	rs490361	C	0.30	0.31	0.96 (0.87-1.07)	0.49	0.83 (0.70-0.42)	0.04
	rs17708107	A	0.90	0.90	0.85 (0.48-1.49)	0.57	0.86 (0.49-1.49)	0.59

Results are those of a meta-analysis of the combined NBC, CGEMS and CBCS study populations.