Author Manuscript

HHS Public Access

Author manuscript

Breast Cancer Res Treat. Author manuscript; available in PMC 2018 January 01.

Published in final edited form as:

Breast Cancer Res Treat. 2017 January ; 161(2): 333–344. doi:10.1007/s10549-016-4053-z.

Previous GWAS Hits in Relation to Young-onset Breast Cancer

Min Shi[1,*], **Katie M. O'Brien**[1,*], **Dale P. Sandler**[2], **Jack A. Taylor**[2], **Dmitri V. Zaykin**[1], and **Clarice R. Weinberg**[1]

[1]Biostatistics and Computational Biology Branch, National Institute of Environmental Health Sciences, Research Triangle Park, NC 27709

[2]Epidemiology Branch, National Institute of Environmental Health Sciences, Research Triangle Park, NC 27709

Abstract

Purpose—Genome-wide association studies (GWAS) have identified dozens of single nucleotide polymorphisms (SNPs) associated with breast cancer. Few studies focused on young-onset breast cancer, which exhibits etiologic and tumor-type differences from older-onset disease. Possible confounding by prenatal effects of the maternal genome has also not been considered.

Methods—Using a family-based design for breast cancer before age 50, we assessed the relationship between breast cancer and 77 GWAS-identified breast cancer risk SNPs. We estimated relative risks (RR) for inherited and maternally-mediated genetic effects. We also used published RR estimates to calculate genetic risk scores and model joint effects.

Results—Seventeen of the candidate SNPs were nominally associated with young-onset breast cancer in our 1,296 non-Hispanic white affected families (uncorrected p-value<0.05). Top-ranked SNPs included rs3803662-A (*TOX3*, RR=1.39; p=7.0×10^{-6}), rs12662670-G (*ESR1*, RR=1.56; p=5.7×10^{-4}), rs2981579-A (*FGFR2*, RR=1.24; p=0.002), and rs999737-G (*RAD51B*, RR=1.37; p=0.003). No maternally-mediated effects were found. A risk score based on all 77 SNPs indicated that their overall relationship to young-onset breast cancer risk was more than additive (additive-fit p=2.2×10^{-7}) and consistent with a multiplicative joint effect (multiplicative-fit p=0.27). With the multiplicative formulation, the case sister's genetic risk score exceeded that of her unaffected sister in 59% of families.

Conclusions—The results of this family-based study indicate that no effects of previously-identified risk SNPs were explained by prenatal effects of maternal variants. Many of the known breast cancer risk variants were associated with young-onset breast cancer, with evidence that *TOX3*, *ESR1*, *FGFR2*, and *RAD51B* are important for young-onset disease.

Corresponding author: Clarice R. Weinberg, 111 TW Alexander Dr, Research Triangle Park, NC 27709; (919) 541-4927; weinber2@niehs.nih.gov.
*These authors contributed equally.

Conflicts of interest: The authors have no conflicts of interest to disclose.

Ethics statement: All participants provided written or verbal consent and the study was approved by the National Institute of Environmental Health Sciences and the Copernicus Group Institutional Review Boards. All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable clinical standards.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Keywords

young-onset breast cancer; genome-wide association study; single nucleotide polymorphism; family-based genetic methods; log-linear models; genetic risk score

Introduction

In the search for genetic risk variants associated with breast cancer, dozens of genome-wide association studies (GWAS) have identified more than a hundred single nucleotide polymorphisms (SNPs) [1–33]. Many of these susceptibility loci have been confirmed in other study samples and across racial groups [34–39], but relatively few studies were designed to examine whether these GWAS 'hits' are associated with young-onset breast cancer [1, 16, 23, 40–43]. Given evidence that age modifies the association between breast cancer and some non-genetic risk factors [44–47], and that younger cases are more likely to have strong family histories of the disease [48], additional studies of the genetic determinants of young-onset breast cancer are warranted.

Young-onset breast cancer is often defined as breast cancer before the age of 50, as this age is both a pro×y for menopausal status and an inflection point for incidence trends in US women [44]. Young-onset disease tends to be relatively more aggressive and difficult to treat [49, 50], and includes a higher fraction of triple-negative breast cancers [51], which may have distinct genetic risk factors [52–56].

An early GWAS of young-onset breast cancer identified one risk-associated SNP in *GLG1* [16]. In a larger and more recent GWAS of young-onset breast cancer, Ahsan et al. [1] identified 96 SNPs with genome-wide significant p-values, all of which were located in six regions previously linked to breast cancer risk unrestricted by age. In total, the authors found that 32 of 83 previously identified GWAS hits were associated with young-onset disease ($p<0.05$).

Family-based designs are feasible for diseases that occur at young ages and offer certain advantages over case-control studies, including robustness to bias due to population stratification and the ability to assess maternally-mediated or imprinting effects [57, 58]. We previously conducted a family-based GWAS of young-onset breast cancer, in which we identified 9 SNPs with unadjusted p-values $<10^{-5}$, including several novel loci [59].

Here, we further investigate the role of 77 known risk variants by reporting their individual associations with young-onset breast cancer and exploring whether their joint effects follow additive or multiplicative risk models. The selected risk variants were previously identified by Mavaddat et al. [60] as having an association with breast cancer at $p<5\times10^{-8}$. We also examined maternally-mediated effects [58] for these same 77 SNPs to assess the influence of the mother's genotype acting prenatally on her daughter, controlling for the daughter's genotype. Because mother's and daughter's genotypes are correlated, maternally-mediated genetic effects could confound inherited-gene effects and may have biased some of the previously-observed GWAS associations.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Methods

Study participants

Young-onset breast cancer cases and their families were recruited as part of the Sister Study and Two Sister Study. The Sister Study is a prospective cohort of women who had one or more sisters diagnosed with breast cancer, but had never had breast cancer themselves at enrollment (2003–2009). It includes 50,884 US and Puerto Rican women aged 35–74. Because they all have a first-degree family history of breast cancer, participants have, on average, approximately twice the risk of developing breast cancer compared to women with no family history. Sister Study participants who developed breast cancer before age 50 (n=235) were included as cases for this analysis.

When we enrolled women in the Sister Study, we asked them the age and date of diagnosis of their affected sister. Proband sisters whose breast cancer was diagnosed before age 50 and within the previous four years were eligible for inclusion in the Two Sister Study. We developed a family-based genetic study by asking an unaffected sister to forward a study invitation to her eligible affected sister and then asking all of the participating young-onset cases to forward a letter from us to any living parents, asking them to provide saliva samples. If one or both parents were unavailable, we genotyped DNA from the blood or saliva from unaffected sister(s) participating in the Sister Study. In total, 3,331 individuals from 1,477 families were genotyped.

All participants provided written or verbal consent and the study was approved by the National Institute of Environmental Health Sciences and the Copernicus Group Institutional Review Boards. The GWAS data is publically accessible (http://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000678.v1.p1). None of the Sister Study or Two Sister Study participants were included in any previously published pooled or meta-analyses.

In addition to providing DNA samples, participants completed computer-assisted telephone interviews about various health-related and lifestyle factors. All cases were asked about their breast cancer diagnosis and treatment and to authorize release of their medical records. Tumor information, including invasiveness and estrogen receptor (ER) status was extracted from the medical records for most cases (85–90%). We relied on self-report for the remainder after observing that Sister Study participants could recall their status with reasonable accuracy (positive predictive values of 99%, 64%, 99% and 84% for invasive, *in situ*, ER+ and ER− breast cancer, respectively). We also reviewed medical records for *BRCA1/2* findings and asked participants to report the results of any *BRCA1/2* mutation tests. A case was assumed to be *BRCA1/2* positive if she had a positive test or if her sister had a positive test, but the case had not been tested.

Genotyping analysis

DNA samples were collected from saliva (80%), whole blood (19%) or blood clot (1%) and shipped to the Johns Hopkins University Center for Inherited Disease Research for genotyping, with subsequent quality control carried out by the Genetics Coordinating Center at the University of Washington. After extraction, DNA samples were processed using 96-well plates, with family members assigned to the same plate. We included 76 HapMap

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

controls and 74 duplicate samples, balanced across plates. All samples were analyzed using the Illumina OmniExpress plus HumanExome-8v1-2 array, which included 964,193 SNPs.

Genotyping revealed that 11 of the sister pairs were half-sisters. The 11 unaffected half-sisters were excluded. Because numbers were inadequate for sufficiently well-parameterized analyses of minority categories, we limited analyses to the majority category of 1,296 non-Hispanic white families (Table 1). No additional participants were excluded, as individuals' missing call rates were all <2%, and duplicate discordance and Mendelian inconsistency rates were low (7.9×10^{-6} per SNP per duplicate pair, and 0.0003 per SNP, respectively). Individuals with chromosomal anomalies (n=6) were assigned missing values for affected regions.

Forty of the 77 candidate SNPs were directly genotyped on our arrays. All met the inclusion criteria in non-Hispanic whites: call rates 97%, 1 discordant call in 74 study duplicates, 5 Mendelian errors, Hardy-Weinberg equilibrium p-values 1×10^{-6} among founders. Two SNPs with a minor allele frequency <1% were excluded from our assessment of individual SNP effects, but were retained in the genetic risk score analysis.

Imputation analysis

Imputation analyses were conducted at the University of Washington's Genetics Coordinating Center using data from the 1000 Genomes Project [61]. Haplotypes were pre-phased using SHAPEIT2 [62] to improve efficiency and make use of known family structures. Imputation was then conducted using IMPUTE (v2.3.0) [63]. The imputation was highly accurate, with good concordance between measured and masked but imputed as most-probable genotypes (99.7% and 98.0% when minor allele frequencies were <5% and 5%, respectively).

In total, 22 million loci were imputed, including 37 of the non-genotyped SNPs from our list of 77. We imputed based on the most probable imputed genotype. If none of an individual's estimated genotype probabilities exceeded 90%, the genotype for that locus was considered missing.

Statistical analysis

Individual SNP effects—For each of the 75 typed candidate SNPs (38 genotyped, 37 imputed), we recorded the odds ratio (OR) reported in Mavaddat et al. [60] and noted which allele was associated with increased breast cancer risk. We then tested the association between each of the candidate SNPs and young-onset breast cancer using a likelihood-based log-linear model to assess transmission distortion within families [58]. Briefly, this conditional model examines whether the relative frequencies of particular offspring genotypes at a di-allelic locus are consistent with Mendelian inheritance. Expected frequencies are modeled using a multinomial distribution with six mating type parameters, which impose conditioning on parental genotypes, and fifteen possible case-parent outcomes. Using likelihood ratio tests (χ^2 distribution), we tested the association between each SNP and young-onset breast cancer by comparing models with and without a term for the relative risk of the offspring's genetic effect. We estimated the relative risks (RR) using

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Poisson regression. If either or both parents were unavailable, an expectation-maximization algorithm was applied to maximize the likelihood [64, 65]. Because we selected these SNPs based on evidence that they were associated with breast cancer, we did not correct p-values for multiple comparisons, but used α=0.05 to evaluate statistical significance.

To facilitate comparisons with Mavaddat et al. [60] and other prior studies, we estimated RRs for each SNP in relation to the previously-established risk allele. We assumed a log-additive model, coding genotype as the number of copies of the risk allele carried by the affected offspring. All analyses used LEM (http://members.home.nl/jeroenvermunt/#Software) and R (v3.2.1). We performed t-tests to compare the risk ratio observed in our study (representative of the young-onset breast cancer effect estimate) and the pooled odds ratio reported by Mavaddat et al. [60] (representative of the overall breast cancer effect estimate).

We also assessed the effects of each SNP in family subsets limited to invasive, ER+, or premenopausal breast cancer and in the subset of families not known to carry BRCA1/2 mutations. Small numbers precluded assessments for the complementary categories. We also estimated RRs for maternally-mediated effects [58].

As secondary analyses, we considered 89 other SNPs known to be associated with breast cancer risk. Additional details of the SNP selection process and results can be found in the Supplementary Materials.

Genetic Risk Scores—In addition to assessing their main effects, we also assessed the efficacy of our 77-SNP set by calculating both multiplicative and additive scores and comparing their model fits and predictive properties. The multiplicative genetic risk score was calculated as: $\Sigma_{i=1}^{k} x_i \ln(OR_i)$ where OR^i is the previously reported OR for one copy of the risk variant at SNP i and x^i is the number of copies of that risk variant carried by the individual. The summation is over k=77 candidate SNPs. Therefore, if 1) the effect size for each SNP in our study is equal to the previously-reported effect size, 2) the joint effect of more than one copy is multiplicative, and 3) the joint effect of multiple SNPs is multiplicative, then the risk score will equal the ln-OR for the 77-SNP set. Thus, the coefficient of $\Sigma_{i=1}^{k} x_i \ln(OR_i)$ is 1.0 under a multiplicative joint effects model. To the extent that the multiplicative genetic risk score and the ln-odds of developing young-onset breast cancer differ (as when the β coefficient differs from 1), there is a departure from multiplicity. The multiplicative joint effects model fit can be tested by comparing β to 1.0 using a Wald test (β-1 divided by standard error). One can also test whether the multiplicative genetic risk score is associated with young-onset breast cancer by comparing β to 0, again using a Wald test.

The additive genetic risk score was calculated as: $\ln\left(\left(\Sigma_{i=1}^{k} x_i(\mathrm{OR}_{xi} - 1)\right) + 1\right)$ Here, if 1) the effect size for each SNP in our study is equal to the previously-reported effect size, 2) the effect of having more than one copy of the risk allele is additive, and 3) the joint effect of multiple SNPs is additive, then that risk score equals the ln-OR for young-onset breast cancer. Thus, the true coefficient of $\ln\left(\left(\Sigma_{i=1}^{k} x_i(\mathrm{OR}_{xi} - 1)\right) + 1\right)$ is 1.0 under a fully

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

additive joint effects model. As with the multiplicative model, the additive model fit can be tested by comparing β to 1.0 using a Wald test, and the association between the additive risk score and young-onset breast cancer can be tested by comparing β to 0.

We calculated the effect estimate (β) per unit increase in risk score using a conditional logistic regression model to compare cases to unaffected sister controls. When both sisters were genotyped, we compared them directly. If the unaffected sister was not genotyped but both parents were, we compared cases to an equally-likely, complementary pseudo-sister whose genotype was defined by her parents' non-transmitted alleles. Altogether, we included 850 case-sister or case-pseudo-sister pairs in the risk scores analyses. For sporadic missing genotypes, we filled in the expected risk allele counts based on the allele frequencies in the parents. As both sets of risk score models were able to accommodate genotypes for imputed SNPs that were not whole numbers, we assigned individual genotypes for imputed SNPs based on expected allele counts.

The relative risk estimates for the 77 SNPs included in the risk score were taken from Mavaddat et al. [60]. These estimates were calculated using data from a pooled analysis with 33,673 cases. None of the 77 selected SNPs was in high linkage disequilibrium (LD) with another included SNP (all $r^2<0.80$) [60].

To assess model fit and risk score utility, we calculated four p-values for each SNP set: an additive score testing β=0; an additive-fit score testing β=1; a multiplicative score testing β=0; and a multiplicative-fit score testing β=1. Additionally, we compared scores between sister pairs to assess their ability to predict risk. We also tested for maternally-mediated genetic effects by comparing scores of mothers to scores of fathers in families where both parents were genotyped (n=418).

As a natural extension and a means to further examine risk score utility, we also used logistic regression to examine whether the mothers' scores were associated with their own breast cancer risk. Cases included 119 mothers with breast cancer at any age (cases) versus 599 mothers with no history of breast cancer (controls).

Results—At diagnosis, most of the 1,279 cases from the 1,296 non-Hispanic white families included in our analysis were aged 40–49 (89%) and premenopausal (93%) (Table 2). The proportions of invasive, ER+ and recognized *BRCA1/2* mutation positive cases were 84%, 81% and 8%, respectively.

Seventeen of the candidate SNPs were associated with increased risk of young-onset breast cancer (p-value<0.05; Table 3). As demonstrated by the quantile-quantile plot (Figure 1), the distribution of p-values for the candidate SNPs was markedly shifted relative to the uniform distribution expected under a global null for the set. The SNP with the smallest p-value was rs3803662-A, which is located upstream of *TOX3* (RR=1.39, 95% CI:1.20–1.60; $p=7.0\times10^{-6}$). The next smallest p-values were for rs12662670-G in *ESR1* (RR=1.56, 95% CI: 1.20–2.03; $p=5.7\times10^{-4}$), rs2981579-A in *FGFR2* (RR=1.24, 95% CI: 1.08–1.42; p=0.002), and rs999737-G in *RAD51B* (RR=1.27, 95% CI: 1.09–1.48; p=0.003).

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

The remaining 13 SNPs with statistically-significant associations are located in the following loci: *ESR1*, 2q14.2, *TERT*, ARHGEF5, 10q26.12, *CCND1*, *MKL1*, 9q31.2, *MRPS30*, 8q24.21, *PEX14* and *ADAM29*. The *ESR1* and 8q24.21 regions both had two statistically-significant SNPs which were not in LD with one another. When we compared our observed effect estimates for young-onset breast cancer to those observed for overall breast cancer [60], the distribution of p-values for those tests did not deviate from the expected distribution under the null of no difference (Supplementary Figure S1).

Analyses restricted to families with invasive, ER+, premenopausal cancer or not known to carry risk-related *BRCA1/2* mutations showed similar patterns, with small p-values again seen for rs3803662, rs12662670, rs2981579, rs999737, and other top-ranked SNPs from the overall analysis (Supplementary Tables S1–S4). For maternally-mediated effects, we found only five SNPs associated ($p<0.05$) with young-onset breast cancer, a number compatible with random chance (73 included SNPs, Table 4 and Figure 2). None of those SNPs were also identified as significant in our primary analysis.

For the polygenic risk score, both the multiplicative and additive genetic risk scores were associated with young-onset breast cancer ($p=2.7\times10^{-10}$ and $p=1.3\times10^{-10}$, respectively, testing $\beta=0$). The coefficient for the additive risk score was different from 1.0 ($\beta= 5.1$, additive-fit-$p=2.2\times10^{-7}$), but the coefficient for the multiplicative risk score was not ($\beta=0.85$, multiplicative-fit-$p=0.27$). This indicates that the overall joint effect of these risk SNPs was more than additive and consistent with multiplicative and that each one-unit increase in score corresponds to a relative risk increase of 2.34 ($e^{0.85}$). For both scores, the case sister had a higher score than her control/pseudo-control sister in 59% of pairs, suggesting there is within-family predictive power.

There was no evidence that these genetic effects act prenatally through the mother's phenotype, as there was no difference between mothers' and fathers' genetic scores on either the multiplicative or additive scale ($p=0.19$ and $p=0.18$, for paired testing, respectively). When we tested the association between the polygenic risk score and breast cancer in mothers, the score preformed similarly to our original analysis, with associations seen for both multiplicative ($p=4.6\times10^{-7}$) and additive ($p=4.0\times10^{-7}$) models. The effects were again more than additive and consistent with multiplicative (multiplicative $\beta=0.88$, multiplicative-fit $p=0.50$; additive $\beta=5.4$, additive-fit $p=3.5\times10^{-5}$), with an estimated area under the receiver operating characteristics curve of 0.65 for both formulations.

Discussion—The purpose of this family-based genetic study was to examine whether previously-identified breast cancer risk variants were associated with young-onset disease through inherited or maternally-mediated effects. We focused on 77 common polymorphisms identified in published GWAS [60]. Seventeen of those candidate SNPs were nominally associated with young-onset breast cancer and a genetic risk score consisting of all 77 SNPs was related to the risk of young-onset breast cancer in daughters and to a history of breast cancer at any age in their mothers. There was no evidence that the independent or combined effects of these candidate SNPs was due to a prenatal effect mediated by the maternal genome.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

The SNPs with the lowest p-values were located in well-studied susceptibility genes – *TOX3*, *ESR1*, *FGFR2,* and *RAD51B*. The results were similar when we restricted analyses to include only families with invasive, ER+, or premenopausal young-onset breast cancer, or to families without known deleterious *BRCA1/2* variants.

TOX3 encodes a nuclear protein that regulates calcium-dependent transcription in neurons [66]. Its link to breast cancer is unclear, though there is some evidence that it is a tumor suppressor [66, 67]. Our top hit, rs3803662, was first identified by Easton et al., [7], with several subsequent GWAS confirming the region's importance [11, 18, 22, 23, 30–32]. Presence of this variant allele has been linked to lower *TOX3* expression levels in breast tumor tissue [68]. Several candidate gene studies also reported risks of similar magnitude for rs3803662-T and young-onset breast cancer [1, 38, 40, 43, 69]. In Ahsan et al.'s [1] young-onset GWAS, *TOX3* was one of the identified susceptibility regions, and rs3803662 had the smallest p-value for that region. For a more direct comparison of our results to those of Ahsan et al., see Supplementary Table S5 (a comparison of p-values for the 83 previously-established susceptibility loci selected by Ahsan et al. and our results for the same SNPs) and Supplementary Table S6 (an assessment of other established susceptibility loci not included in Mavaddat et al. [60], including the novel GWAS hits from Ahsan et al.).

The SNPs with the second and fifth smallest p-values - rs12662670 and rs2046210 - are located in the 6q25 region just upstream of *ESR1*. *ESR1* encodes Estrogen Receptor α, a ligand-activated transcription factor crucial to sexual development and reproduction [70]. rs12662670 was previously linked to increased risk of triple-negative breast cancer, a subtype more common in younger women [71]. We did not have enough cases of triple-negative cancer for a separate analysis. Carriers of the other SNP, rs2046210-A, have lower *ESR1* levels in both tumor and normal breast tissue [72]. The 6q25 region was not one of the regions with genome-wide significant p-values in Ahsan et al. [1], though rs2046210-A was positively associated with young-onset breast cancer at $p<0.05$.

FGFR2 encodes a fibroblast growth factor receptor, the overexpression of which may contribute to carcinogenesis through increased cell proliferation, migration, and resistance to apoptosis [73]. Our *FGFR2* SNP, rs2981579, was first identified by Thomas et al. [31] and independently replicated in two other GWAS [23, 32]. Other GWAS found stronger associations with other SNPs in the same gene [1, 7, 8, 10, 12, 15, 19, 22, 26]. Ahsan et al. [1], for example, reported genome-wide significant p-values for 4 *FGFR2* SNPs and young-onset breast cancer, with the smallest p-value seen for rs2981579. Other studies have also observed associations between *FGFR2* SNPs and young-onset breast cancer [41, 42, 69, 74].

RAD51B is a homologous recombination repair gene related to *BRCA1* and *BRCA2* [75]. Thomas et al. [31] were the first to link rs999737-G to breast cancer risk. This was confirmed in a second GWAS [23] and subsequent meta-analysis [76]. Although *RAD51B* was not one of the genome-wide significant regions identified in Ahsan et al. [1], rs999737 was associated with young-onset breast cancer at $p<0.05$.

In addition to *TOX3*, *ESR1*, *FGFR2*, and *RAD51,* we corroborated Ahsan et al.'s findings that SNPs in 8q24.21 and 11q13.3 (*CCND1*) may be important for young-onset breast

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

cancer. However, we did not see statistically significant effects for SNPs in *SLC4A7* or *MAP3K1*. It is not clear whether these differences are due to chance or heterogeneity across studies. We had limited power to detect small effects, with approximately 80% power to detect RRs of 1.24 or 1.20 for allele frequencies of 20% or 40%, respectively for $\alpha=0.05$. This modest power limits our ability to detect small effects and age-at-onset interaction effects. Given this caveat, the results of the statistical tests comparing our effect estimates to those of Mavaddat et al. [60] showed no evidence that effect sizes differed by age at breast cancer onset.

Dite et al. [77] also applied the 77-SNP risk score developed by Mavaddat et al. [60] to evaluate young-onset breast cancer. They reported an area under the curve of 0.61 for the risk score alone, which is consistent with the predictive capability we report here. We additionally found that the 77-SNP risk score was predictive of breast cancer in the mothers of cases (which were predominantly older-onset), and that similar coefficients were seen for the two generations.

Our use of previously reported effect measures helped us avoid over-fitting the prediction model to our data [78] and enabled independent replication. Several previous studies have used similar approaches, but most have considered only models with multiplicative joint effects [18, 22, 60, 77, 79]. Like us, Joshi et al. [80] constructed scores for both multiplicative and additive polygenic effects using externally-reported effect estimates. We both found evidence that the joint effects were super-additive, but our results showed consistency with a multiplicative model, while Joshi et al. found that the observed associations were sub-multiplicative. Future studies should consider both additive and multiplicative models to see which mathematical form of the risk score best fits the data and to determine if the same model is appropriate for different subgroups (e.g. younger versus older women). Ideally, investigators will reach a consensus regarding what SNPs to include in the score and how to correctly specify their combined effects. Ultimately, we hope a risk score can be utilized as a tool to classify women's individual breast cancer risk and identify who should be selected for more or less frequent screening.

A major strength of this study is its family-based design, which is robust to population stratification and amenable to missing-data imputation methods [64, 65]. However, we restricted our analysis to non-Hispanic whites, which somewhat limits the generalizability of our findings. There is also the potential for survival bias if some SNPs are related to survival after breast cancer. However, short-term survival rates for breast cancer are high (91% 5-year relative survival for invasive breast cancer before age 50 [81]), suggesting the effects of such attrition would be minimal.

In this family-based genetic study, we saw little evidence of maternally-mediated genetic effects, but found that many of the known breast cancer risk variants were also associated with young-onset breast cancer. Our analyses provide further evidence that certain loci, including *TOX3*, *ESR1*, *FGFR2*, and *RAD51B*, are important for the development of breast cancer at any age.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Funding support: This work was supported by the Intramural Research Program of the NIH, the National Institute of Environmental Health Sciences (project Z01-ES044005 to DPS and project Z01-ES102245 to CRW) and Susan G. Komen for the Cure (grant FAS0703856 to CRW).

The authors would like to thank Stephanie London and Sophia Harlid for their comments on an early draft of this paper. Genotyping, quality control analysis and imputation were completed at the Center for Inherited Disease Research, the Genetics Coordinating Center at the University of Washington (Cecelia Laurie, Quenna Wong, Sarah Nelson, and Cathy Laurie).

References

1. Ahsan H, Halpern J, Kibriya MG, Pierce BL, Tong L, Gamazon E, et al. A genome-wide association study of early-onset breast cancer identifies PFKM as a novel breast cancer gene and supports a common genetic spectrum for breast cancer at any age. Cancer Epidemiol Biomarkers Prev. 2014; 23(4):658–669. [PubMed: 24493630]
2. Antoniou AC, Wang X, Fredericksen ZS, McGuffog L, Tarrell R, Sinilnikova OM, 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(10):885–892. [PubMed: 20852631]
3. Cai Q, Long J, Lu W, Qu S, Wen W, Kang D, et al. Genome-wide association study identifies breast cancer risk variant at 10q21.2: results from the Asia Breast Cancer Consortium. Hum Mol Genet. 2011; 20(24):4991–4999. [PubMed: 21908515]
4. Cai Q, Zhang B, Sung H, Low SK, Kweon SS, Lu W, et al. Genome-wide association analysis in East Asians identifies breast cancer susceptibility loci at 1q32.1, 5q14.3 and 15q26.1. Nat Genet. 2014; 46(8):886–890. [PubMed: 25038754]
5. Chen F, Chen GK, Stram DO, Millikan RC, Ambrosone CB, John EM, et al. A genome-wide association study of breast cancer in women of African ancestry. Hum Genet. 2013; 132(1):39–48. [PubMed: 22923054]
6. Couch FJ, Wang X, McGuffog L, Lee A, Olswold C, Kuchenbaecker KB, et al. Genome-wide association study in BRCA1 mutation carriers identifies novel loci associated with breast and ovarian cancer risk. PLoS Genet. 2013; 9(3):e1003212. [PubMed: 23544013]
7. Easton DF, Pooley KA, Dunning AM, Pharoah PD, Thompson D, Ballinger DG, et al. Genome-wide association study identifies novel breast cancer susceptibility loci. Nature. 2007; 447(7148):1087–1093. [PubMed: 17529967]
8. Elgazzar S, Zembutsu H, Takahashi A, Kubo M, Aki F, Hirata K, 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(12):766–771. [PubMed: 22951594]
9. Fejerman L, Ahmadiyeh N, Hu D, Huntsman S, Beckman KB, Caswell JL, et al. Genome-wide association study of breast cancer in Latinas identifies novel protective variants on 6q25. Nat Commun. 2014; 5:5260. [PubMed: 25327703]
10. Fletcher O, Johnson N, Orr N, Hosking FJ, Gibson LJ, Walker K, et al. Novel breast cancer susceptibility locus at 9q31.2: results of a genome-wide association study. J Natl Cancer Inst. 2011; 103(5):425–435. [PubMed: 21263130]
11. Garcia-Closas M, Couch FJ, Lindstrom S, Michailidou K, Schmidt MK, Brook MN, et al. Genome-wide association studies identify four ER negative-specific breast cancer risk loci. Nat Genet. 2013; 45(4):392–398. [PubMed: 23535733]
12. Gaudet MM, Kirchhoff T, Green T, Vijai J, Korn JM, Guiducci C, et al. Common genetic variants and modification of penetrance of BRCA2-associated breast cancer. PLoS Genet. 2010; 6(10):e1001183. [PubMed: 21060860]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

13. 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(11):4340–4345. [PubMed: 18326623]
14. Haiman CA, Chen GK, Vachon CM, Canzian F, Dunning A, Millikan RC, et al. A common variant at the TERT-CLPTM1L locus is associated with estrogen receptor-negative breast cancer. Nat Genet. 2011; 43(12):1210–1214. [PubMed: 22037553]
15. Hunter DJ, Kraft P, Jacobs KB, Cox DG, Yeager M, Hankinson SE, et al. A genome-wide association study identifies alleles in FGFR2 associated with risk of sporadic postmenopausal breast cancer. Nat Genet. 2007; 39(7):870–874. [PubMed: 17529973]
16. Kibriya MG, Jasmine F, Argos M, Andrulis I, John EM, Chang-Claude J, et al. A pilot genome-wide association study of early-onset breast cancer. Breast Cancer Res Treat. 2009; 114(3):463–467. [PubMed: 18463975]
17. Kim HC, Lee JY, Sung H, Choi JY, Park SK, Lee KM, 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(2):R56. [PubMed: 22452962]
18. Li J, Humphreys K, Darabi H, Rosin G, Hannelius U, Heikkinen T, et al. A genome-wide association scan on estrogen receptor-negative breast cancer. Breast Cancer Res. 2010; 12(6):R93. [PubMed: 21062454]
19. Li J, Humphreys K, Heikkinen T, Aittomaki K, Blomqvist C, Pharoah PD, et al. A combined analysis of genome-wide association studies in breast cancer. Breast Cancer Res Treat. 2011; 126(3):717–727. [PubMed: 20872241]
20. Long J, Cai Q, Shu XO, Qu S, Li C, Zheng Y, 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. [PubMed: 20585626]
21. Long J, Cai Q, Sung H, Shi J, Zhang B, Choi JY, et al. Genome-wide association study in east Asians identifies novel susceptibility loci for breast cancer. PLoS Genet. 2012; 8(2):e1002532. [PubMed: 22383897]
22. Low SK, Takahashi A, Ashikawa K, Inazawa J, Miki Y, Kubo M, et al. Genome-wide association study of breast cancer in the Japanese population. PLoS One. 2013; 8(10):e76463. [PubMed: 24143190]
23. Michailidou K, Hall P, Gonzalez-Neira A, Ghoussaini M, Dennis J, Milne RL, et al. Large-scale genotyping identifies 41 new loci associated with breast cancer risk. Nat Genet. 2013; 45(4):353–361. [PubMed: 23535729]
24. Murabito JM, Rosenberg CL, Finger D, Kreger BE, Levy D, Splansky GL, 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. [PubMed: 17903305]
25. Purrington KS, Slager S, Eccles D, Yannoukakos D, Fasching PA, Miron P, et al. Genome-wide association study identifies 25 known breast cancer susceptibility loci as risk factors for triple-negative breast cancer. Carcinogenesis. 2014; 35(5):1012–1019. [PubMed: 24325915]
26. Rinella ES, Shao Y, Yackowski L, Pramanik S, Oratz R, Schnabel F, et al. Genetic variants associated with breast cancer risk for Ashkenazi Jewish women with strong family histories but no identifiable BRCA1/2 mutation. Hum Genet. 2013; 132(5):523–536. [PubMed: 23354978]
27. Sehrawat B, Sridharan M, Ghosh S, Robson P, Cass CE, Mackey JR, et al. Potential novel candidate polymorphisms identified in genome-wide association study for breast cancer susceptibility. Hum Genet. 2011; 130(4):529–537. [PubMed: 21424380]
28. Siddiq A, Couch FJ, Chen GK, Lindstrom S, Eccles D, Millikan RC, 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(24):5373–5384. [PubMed: 22976474]
29. Song C, Chen GK, Millikan RC, Ambrosone CB, John EM, Bernstein L, et al. A genome-wide scan for breast cancer risk haplotypes among African American women. PLoS One. 2013; 8(2):e57298. [PubMed: 23468962]
30. Stacey SN, Manolescu A, Sulem P, Rafnar T, Gudmundsson J, Gudjonsson SA, et al. Common variants on chromosomes 2q35 and 16q12 confer susceptibility to estrogen receptor-positive breast cancer. Nat Genet. 2007; 39(7):865–869. [PubMed: 17529974]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

31. Thomas G, Jacobs KB, Kraft P, Yeager M, Wacholder S, Cox DG, et al. A multistage genome-wide association study in breast cancer identifies two new risk alleles at 1p11.2 and 14q24.1 (RAD51L1). Nat Genet. 2009; 41(5):579–584. [PubMed: 19330030]
32. 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(6):504–507. [PubMed: 20453838]
33. Zheng W, Long J, Gao YT, Li C, Zheng Y, Xiang YB, et al. Genome-wide association study identifies a new breast cancer susceptibility locus at 6q25.1. Nat Genet. 2009; 41(3):324–328. [PubMed: 19219042]
34. Campa D, Kaaks R, Le Marchand L, Haiman CA, Travis RC, Berg CD, 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(16):1252–1263. [PubMed: 21791674]
35. Hutter CM, Young AM, Ochs-Balcom HM, Carty CL, Wang T, Chen CT, 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(9):1950–1959. [PubMed: 21795501]
36. O'Brien KM, Cole SR, Poole C, Bensen JT, Herring AH, Engel LS, et al. Replication of breast cancer susceptibility loci in whites and African Americans using a Bayesian approach. Am J Epidemiol. 2014; 179(3):382–394. [PubMed: 24218030]
37. Reeves SG, Meldrum C, Groombridge C, Spigelman A, Suchy J, Kurzawski G, et al. DNA repair gene polymorphisms and risk of early onset colorectal cancer in Lynch syndrome. Cancer Epidemiol. 2012; 36(2):183–189. [PubMed: 21974800]
38. 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(2):531–539. [PubMed: 21475998]
39. Zheng W, Cai Q, Signorello LB, Long J, Hargreaves MK, Deming SL, et al. Evaluation of 11 breast cancer susceptibility loci in African-American women. Cancer Epidemiol Biomarkers Prev. 2009; 18(10):2761–2764. [PubMed: 19789366]
40. Elematore I, Gonzalez-Hormazabal P, Reyes JM, Blanco R, Bravo T, Peralta O, et al. Association of genetic variants at TOX3, 2q35 and 8q24 with the risk of familial and early-onset breast cancer in a South-American population. Mol Biol Rep. 2014; 41(6):3715–3722. [PubMed: 24532140]
41. Fu F, Wang C, Huang M, Song C, Lin S, Huang H. Polymorphisms in Second Intron of the FGFR2 Gene Are Associated with the Risk of Early-Onset Breast Cancer in Chinese Han Women. The Tohoku Journal of Experimental Medicine. 2012; 226(3):221–229. [PubMed: 22374580]
42. Jara L, Gonzalez-Hormazabal P, Cerceno K, Di Capua GA, Reyes JM, Blanco R, et al. Genetic variants in FGFR2 and MAP3K1 are associated with the risk of familial and early-onset breast cancer in a South-American population. Breast Cancer Res Treat. 2013; 137(2):559–569. [PubMed: 23225170]
43. Tapper W, Hammond V, Gerty S, Ennis S, Simmonds P, Collins A, et al. The influence of genetic variation in 30 selected genes on the clinical characteristics of early onset breast cancer. Breast Cancer Res. 2008; 10(6):R108. [PubMed: 19094228]
44. Anderson WF, Matsuno RK, Sherman ME, Lissowska J, Gail MH, Brinton LA, et al. Estimating age-specific breast cancer risks: a descriptive tool to identify age interactions. Cancer Causes Control. 2007; 18(4):439–447. [PubMed: 17216325]
45. Michels KB, Xue F. Role of birthweight in the etiology of breast cancer. Int J Cancer. 2006; 119(9):2007–2025. [PubMed: 16823839]
46. Trentham-Dietz A, Sprague BL, Hampton JM, Miglioretti DL, Nelson HD, Titus LJ, et al. Modification of breast cancer risk according to age and menopausal status: a combined analysis of five population-based case-control studies. Breast Cancer Res Treat. 2014; 145(1):165–175. [PubMed: 24647890]
47. Van den Brandt PA, Spiegelman D, Yaun SS, Adami HO, Beeson L, Folsom AR, et al. Pooled analysis of prospective cohort studies on height, weight, and breast cancer risk. Am J Epidemiol. 2000; 152:514–527. [PubMed: 10997541]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

48. Collaborative Group on Hormonal Factors in Breast Cancer. Familial breast cancer: collaborative reanalysis of individual data from 52 epidemiological studies including 58 209 women with breast cancer and 101 986 women without the disease. The Lancet. 2001; 358(9291):1389–1399.
49. Althuis MD, Brogan DD, Coates RJ, Daling JR, Gammon MD, Malone K, et al. Breast cancers among very young premenopausal women (United States). Cancer Causes & Control. 2003; 14:151–160. [PubMed: 12749720]
50. Anders CK, Hsu DS, Broadwater G, Acharya CR, Foekens JA, Zhang Y, et al. Young age at diagnosis correlates with worse prognosis and defines a subset of breast cancers with shared patterns of gene expression. J Clin Oncol. 2008; 26(20):3324–3330. [PubMed: 18612148]
51. Carey LA, Perou CM, Livasy CA, Dressler LG, Cowan D, Conway K, et al. Race, breast cancer subtypes, and survival in the Carolina Breast Cancer Study. JAMA. 2006; 295(21):2492–2502. [PubMed: 16757721]
52. Broeks A, Schmidt MK, Sherman ME, Couch FJ, Hopper JL, Dite GS, 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(16):3289–3303. [PubMed: 21596841]
53. Chan M, Ji SM, Liaw CS, Yap YS, Law HY, Yoon CS, et al. Association of common genetic variants with breast cancer risk and clinicopathological characteristics in a Chinese population. Breast Cancer Res Treat. 2012; 136(1):209–220. [PubMed: 22965832]
54. Han W, Woo JH, Yu JH, Lee MJ, Moon HG, Kang D, 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(5):793–798. [PubMed: 21415360]
55. O'Brien KM, Cole SR, Engel LS, Bensen JT, Poole C, Herring AH, et al. Breast cancer subtypes and previously established genetic risk factors: a bayesian approach. Cancer Epidemiol Biomarkers Prev. 2014; 23(1):84–97. [PubMed: 24177593]
56. Rebbeck TR, DeMichele A, Tran TV, Panossian S, Bunin GR, Troxel AB, 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(2):269–274. [PubMed: 19028704]
57. Laird NM, Lange C. Family-based designs in the age of large-scale gene-association studies. Nat Rev Genet. 2006; 7(5):385–394. [PubMed: 16619052]
58. Weinberg CR, Wilcox AJ, Lie RT. A log-linear approach to case-parent-triad data: Assessing effects of disease genes that act either directly or through maternal effects and that may be subject to parental imprinting. Am J Hum Genet. 1998; 62:969–978. [PubMed: 9529360]
59. O'Brien KM, Shi M, Sandler DP, Taylor JA, Zaykin DV, Keller J, et al. A family-based, genome-wide association study of young-onset breast cancer: inherited variants and maternally mediated effects. Eur J Hum Genet. 2016
60. Mavaddat N, Pharoah PD, Michailidou K, Tyrer J, Brook MN, Bolla MK, et al. Prediction of breast cancer risk based on profiling with common genetic variants. J Natl Cancer Inst. 2015; 107(5)
61. The 1000 Genomes Project Consortium. An integrated map of genetic variation from 1,092 human genomes. Nature. 2012; 491(7422):56–65. [PubMed: 23128226]
62. Delaneau O, Marchini J, Zagury JF. A linear complexity phasing method for thousands of genomes. Nat Methods. 2012; 9(2):179–181.
63. Howie BN, Donnelly P, Marchini J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. PLoS Genet. 2009; 5(6):e1000529. [PubMed: 19543373]
64. Shi M, Umbach DM, Weinberg CR. Case-sibling studies that acknowledge unstudied parents and permit the inclusion of unmatched individuals. Int J Epidemiol. 2013; 42(1):298–307. [PubMed: 23248214]
65. Weinberg CR. Allowing for missing parents in genetic studies of case-parent triads. Am J Hum Genet. 1999; 64:1186–1193. [PubMed: 10090904]
66. Dittmer S, Kovacs Z, Yuan SH, Siszler G, Kogl M, Summer H, et al. TOX3 is a neuronal survival factor that induces transcription depending on the presence of CITED1 or phosphorylated CREB in the transcriptionally active complex. J Cell Sci. 2011; 124(Pt 2):252–260. [PubMed: 21172805]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

67. Cowper-Sal lari R, Zhang X, Wright JB, Bailey SD, Cole MD, Eeckhoute J, et al. Breast cancer risk-associated SNPs modulate the affinity of chromatin for FOXA1 and alter gene expression. Nat Genet. 2012; 44(11):1191–1198. [PubMed: 23001124]
68. Riaz M, Berns EM, Sieuwerts AM, Ruigrok-Ritstier K, de Weerd V, Groenewoud A, et al. Correlation of breast cancer susceptibility loci with patient characteristics, metastasis-free survival, and mRNA expression of the nearest genes. Breast Cancer Res Treat. 2012; 133(3):843–851. [PubMed: 21748294]
69. Barzan D, Veldwijk MR, Herskind C, Li Y, Zhang B, Sperk E, et al. Comparison of genetic variation of breast cancer susceptibility genes in Chinese and German populations. Eur J Hum Genet. 2013; 21(11):1286–1292. [PubMed: 23486537]
70. ESR1 estrogen receptor 1 [Homo sapiens (human)] [Internet]. [cited December 22, 2015]
71. Stevens KN, Vachon CM, Lee AM, Slager S, Lesnick T, Olswold C, et al. Common breast cancer susceptibility loci are associated with triple-negative breast cancer. Cancer Res. 2011; 71(19): 6240–6249. [PubMed: 21844186]
72. Sun Y, Ye C, Guo X, Wen W, Long J, Gao YT, et al. Evaluation of potential regulatory function of breast cancer risk locus at 6q25.1. Carcinogenesis. 2015
73. Jain VK, Turner NC. Challenges and opportunities in the targeting of fibroblast growth factor receptors in breast cancer. Breast Cancer Res. 2012; 14:208. [PubMed: 22731805]
74. Kawase T, Matsuo K, Suzuki T, Hiraki A, Watanabe M, Iwata H, 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(8):1946–1952. [PubMed: 19582883]
75. Roy R, Chun J, Powell SN. BRCA1 and BRCA2: different roles in a common pathway of genome protection. Nat Rev Cancer. 2012; 12(1):68–78.
76. Dong H, Gao Z, Li C, Wang J, Jin M, Rong H, et al. Analyzing 395,793 samples shows significant association between rs999737 polymorphism and breast cancer. Tumour Biol. 2014; 35(6):6083–6087. [PubMed: 24729084]
77. Dite GS, MacInnis RJ, Bickerstaffe A, Dowty JG, Allman R, Apicella C, et al. Breast Cancer Risk Prediction Using Clinical Models and 77 Independent Risk-Associated SNPs for Women Aged Under 50 Years: Australian Breast Cancer Family Registry. Cancer Epidemiol Biomarkers Prev. 2015
78. Kraft P. Curses--winner's and otherwise--in genetic epidemiology. Epidemiology. 2008; 19(5):649–651. discussion 57-8. [PubMed: 18703928]
79. Machiela MJ, Chen CY, Chen C, Chanock SJ, Hunter DJ, Kraft P. Evaluation of polygenic risk scores for predicting breast and prostate cancer risk. Genet Epidemiol. 2011; 35(6):506–514. [PubMed: 21618606]
80. Joshi AD, Lindstrom S, Husing A, Barrdahl M, VanderWeele TJ, Campa D, et al. Additive interactions between susceptibility single-nucleotide polymorphisms identified in genome-wide association studies and breast cancer risk factors in the Breast and Prostate Cancer Cohort Consortium. Am J Epidemiol. 2014; 180(10):1018–1027. [PubMed: 25255808]
81. National Cancer Institute, Surveillance, Epidemiology, and End Results. [January 24, 2016] http://seercancergov/ Accessed

Abbreviations

CI	confidence interval
ER	estrogen receptor
GWAS	genome-wide association study
LD	linkage disequilibrium
OR	odds ratio

RR	relative risk
SNP	single nucleotide polymorphism

Author Manuscript
Author Manuscript
Author Manuscript
Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

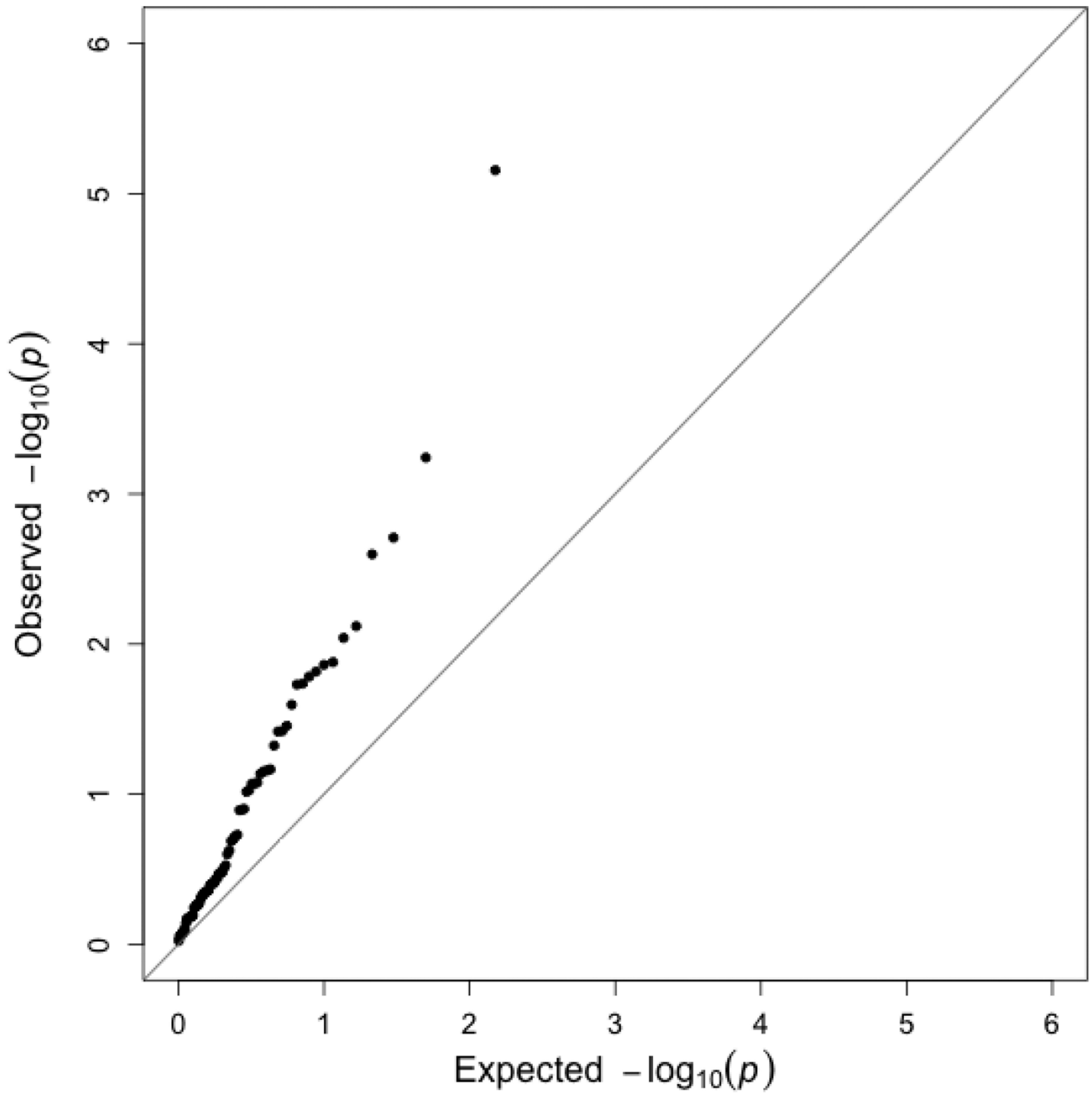

Figure 1.

Quantile-quantile plot for the association between 75 candidate SNPs and young-onset breast cancer in the Two Sister Study (2008–2012).

Author Manuscript
Author Manuscript
Author Manuscript
Author Manuscript

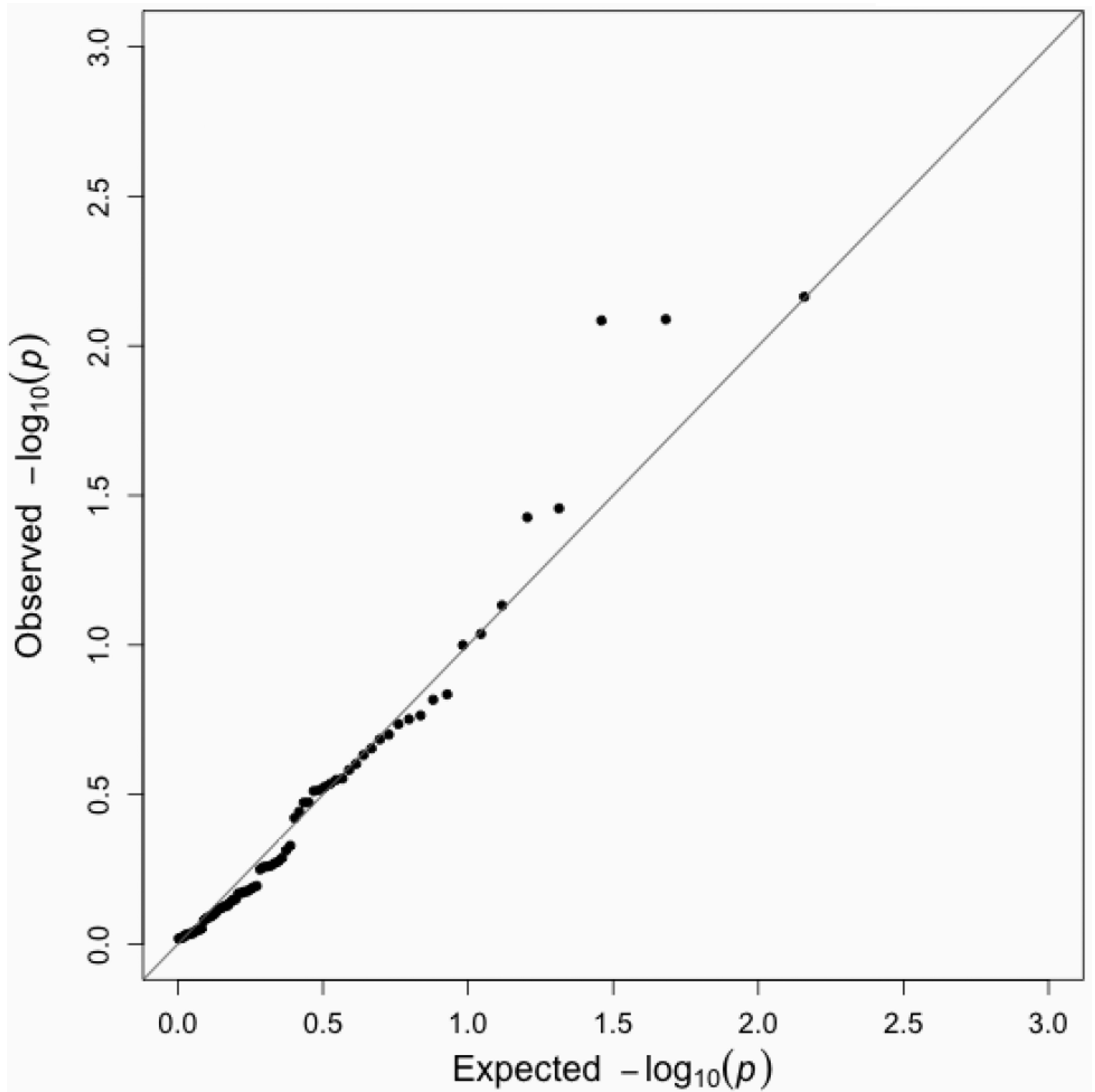

Figure 2.
Quantile-quantile plot for maternally-mediated genetic effects of 73 candidate SNPs and young-onset breast cancer in the Two Sister Study (2008–2012).

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Table 1

Non-Hispanic white participants included in the young-onset breast cancer genotyping analysis.

Group Description	Number of Families	Number of Cases
Trios (affected sister, both parents)	416	416
Sister-pairs (1 affected, 1 unaffected), father	81	81
Sister-pairs	353	353
Affected sister and 1 parent	321	321
Unaffected sister and father	2	0
Parents only	2	0
Affected sister only	108	108
Unaffected sister only	11	0
Mother only	2	0
TOTAL	**1296**	**1279**

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Table 2

Characteristics of young-onset breast cancer casesincluded in the genotyping analysis (n=1279).

	N (%)
Age at diagnosis	
<40	136 (11)
40–49	1143 (89)
Menopausal Status at Diagnosis	
Premenopausal	1186 (93)
Postmenopausal	86 (7)
Missing	7
Invasive Status	
Ductal carcinoma in situ	206 (16)
Invasive	1057 (84)
Missing	16
Estrogen Receptor Status	
Positive	1013 (81)
Negative	239 (19)
Missing	27
BRCA1/2 status[a]	
Case carries BRCA1/2 mutation	100 (8)
Case not known to have BRCA1/2 mutation	1179 (92)

[a]Families were categorized as *BRCA1* or *BRCA2* mutation positive if (1) the case sister reported that she had had a positive test or (2) the case sister was not tested but the unaffected sister reported that she had had a positive test.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Table 3

Relative risks and 95% confidence intervals for the association between candidate SNPs and young-onset breast cancer

Rank	SNP[a]	Gene/ Region	Risk Allele	RAF	RR (95% CI)	p-value	Reported OR[b]
1	rs3803662	TOX3	A	0.30	1.39 (1.20, 1.60)	7.0×10^{-6}	1.23
2	rs12662670	ESR1	G	0.08	1.56 (1.20, 2.03)	5.7×10^{-4}	1.14
3	rs2981579	FGFR2	A	0.44	1.24 (1.08, 1.42)	0.002	1.25
4	rs999737	RAD51B	G	0.77	1.27 (1.09, 1.48)	0.003	1.09
5	rs2046210	ESR1	A	0.35	1.21 (1.05, 1.39)	0.008	1.05
6	*rs4849887*	2q14.2	C	0.89	1.31 (1.07, 1.60)	0.009	1.09
7	*rs7726159*	TERT	A	0.32	1.28 (1.05, 1.55)	0.01	1.04
8	rs720475	ARHGEF5	G	0.75	1.20 (1.04, 1.39)	0.01	1.06
9	*rs11199914*	10q26.12	C	0.69	1.20 (1.04, 1.38)	0.02	1.06
10	*rs554219*	CCND1	G	0.13	1.27 (1.04, 1.55)	0.02	1.12
11	rs6001930	MKL1	G	0.10	1.29 (1.04, 1.59)	0.02	1.13
12	*rs10759243*	9q31.2	A	0.27	1.21 (1.03, 1.41)	0.02	1.05
13	*rs10941679*	MRPS30	G	0.25	1.20 (1.02, 1.40)	0.03	1.12
14	rs13281615	8q24.21	G	0.42	1.16 (1.01, 1.32)	0.04	1.10
15	rs616488	PEX14	A	0.66	1.16 (1.01, 1.33)	0.04	1.06
16	*rs11780156*	8q24.21	T	0.18	1.19 (1.01, 1.41)	0.04	1.07
17	*rs6828523*	ADAM29	C	0.90	1.24 (1.00, 1.54)	0.05	1.10
18	rs2943559	HNF4G	G	0.08	1.24 (0.98, 1.57)	0.07	1.13
19	*rs12493607*	TGFBR2	C	0.34	1.14 (0.99, 1.32)	0.07	1.05
20	rs4973768	SLC4A7	A	0.50	1.13 (0.99, 1.28)	0.07	1.09
21	*rs75915166*	FGF3	A	0.06	1.29 (0.97, 1.72)	0.07	1.02
22	*rs17356907*	12q22	A	0.73	1.14 (0.98, 1.32)	0.08	1.10
23	*rs2736108*	TERT	T	0.73	1.17 (0.98, 1.39)	0.09	1.07
24	rs7072776	MLLT10	A	0.31	0.88 (0.77, 1.02)	0.09	1.06
25	*rs3903072*	11q13.1	G	0.53	1.12 (0.98, 1.27)	0.09	1.06
26	rs941764	CCDC88C	G	0.35	1.12 (0.98, 1.29)	0.10	1.06
27	rs10472076	RPL5P15	G	0.36	1.11 (0.97, 1.27)	0.13	1.04
28	rs1292011	12q24.21	A	0.60	1.11 (0.97, 1.26)	0.13	1.08

Rank	SNP[a]	Gene/ Region	Risk Allele	RAF	RR (95% CI)	p-value	Reported OR[b]
29	rs1011970	CDKN2B	A	0.17	1.14 (0.96, 1.36)	0.13	1.05
30	rs11242675	FOXQ1	A	0.63	1.10 (0.96, 1.26)	0.19	1.06
31	*rs78540526*	CCND1	T	0.08	1.19 (0.92, 1.55)	0.19	1.18
32	rs1045485	CASP8	C	0.88	1.14 (0.93, 1.39)	0.20	1.04
33	*rs12710696*	LINC01376	T	0.37	1.09 (0.95, 1.24)	0.20	1.04
34	*rs9693444*	8p12	A	0.44	1.09 (0.95, 1.25)	0.24	1.07
35	*rs11075995*	FTO	A	0.24	0.92 (0.79, 1.06)	0.25	1.04
36	*rs6762644*	ITPR1	G	0.39	0.93 (0.82, 1.06)	0.30	1.07
37	*rs527616*	18q11.2	G	0.68	1.09 (0.92, 1.29)	0.31	1.04
38	rs17817449	FTO	A	0.61	1.07 (0.94, 1.22)	0.33	1.08
39	rs4245739	MDM4	C	0.27	1.08 (0.93, 1.25)	0.33	1.03
40	rs889312	MAP3K1	C	0.31	1.07 (0.93, 1.23)	0.34	1.12
41	rs1436904	CHST9	A	0.60	1.06 (0.93, 1.21)	0.36	1.06
42	rs11249433	EMBP1	G	0.42	1.06 (0.93, 1.22)	0.37	1.10
43	rs704010	ZMIZ1	A	0.41	1.06 (0.93, 1.21)	0.39	1.07
44	*rs7904519*	TCF7L2	G	0.47	0.94 (0.82, 1.08)	0.39	1.06
45	*rs2236007*	PAX9	G	0.88	1.07 (0.92, 1.25)	0.40	1.09
46	rs6472903	8q21.11	A	0.83	1.08 (0.91, 1.28)	0.40	1.10
47	*rs11820646*	11q24.3	C	0.61	0.95 (0.82, 1.09)	0.44	1.05
48	rs2363956	ANKLE1	A	0.51	0.95 (0.83, 1.08)	0.44	1.03
49	rs3817198	LSP1	G	0.34	1.05 (0.92, 1.21)	0.45	1.07
50	rs2016394	DLX2	G	0.53	1.05 (0.92, 1.20)	0.45	1.05
51	rs865686	RPL31P43	A	0.63	1.05 (0.92, 1.21)	0.46	1.11
52	*rs17529111*	6q14.1	G	0.22	1.06 (0.90, 1.26)	0.48	1.05
53	rs2823093	CYYR1	G	0.73	1.05 (0.91, 1.21)	0.49	1.08
54	*rs132390*	EMID1	C	0.02	1.16 (0.74, 1.81)	0.52	1.11
55	*rs10771399*	12p11.22	G	0.89	1.07 (0.87, 1.31)	0.54	1.16
56	rs9790517	TET2	A	0.24	1.05 (0.90, 1.22)	0.55	1.05
57	*rs6504950*	STXBP4	G	0.73	1.05 (0.90, 1.22)	0.55	1.07
58	rs3760982	KCNN4	A	0.47	1.04 (0.91, 1.19)	0.57	1.06
59	rs13387042	TNP1	G	0.54	1.04 (0.91, 1.18)	0.57	1.14

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Rank	SNP[a]	Gene/ Region	Risk Allele	RAF	RR (95% CI)	p-value	Reported OR[b]
60	*rs12422552*	12p13.1	C	0.28	0.97 (0.83, 1.12)	0.64	1.03
61	*rs6678914*	LGR6	G	0.59	0.97 (0.85, 1.11)	0.65	1.01
62	rs11552449	AP4B1	A	0.15	0.96 (0.81, 1.14)	0.65	1.08
63	*rs1550623*	CDCA7	A	0.85	1.04 (0.87, 1.24)	0.66	1.06
64	rs10069690	TERT	A	0.27	1.03 (0.89, 1.19)	0.67	1.02
65	*rs11814448*	10p12.31	C	0.02	0.92 (0.60, 1.39)	0.68	1.22
66	*rs204247*	RANBP9	G	0.45	1.03 (0.90, 1.17)	0.68	1.05
67	*rs13329835*	CDYL2	G	0.23	0.97 (0.83, 1.14)	0.72	1.08
68	*rs4808801*	ELL	A	0.68	1.02 (0.88, 1.17)	0.80	1.07
69	*rs2588809*	RAD51B	T	0.17	0.98 (0.82, 1.16)	0.80	1.07
70	*rs1432679*	EBF1	C	0.46	0.99 (0.86, 1.13)	0.84	1.07
71	rs10995190	ZNF365	G	0.86	1.02 (0.85, 1.23)	0.84	1.17
72	*rs16857609*	DIRC3	T	0.28	0.99 (0.85, 1.15)	0.87	1.07
73	rs2380205	GDI2	G	0.58	0.99 (0.87, 1.13)	0.88	1.02
74	rs8170	BIBAM1	A	0.17	0.99 (0.83, 1.17)	0.88	1.03
75	*rs1353747*	PDE4D	T	0.91	1.01 (0.81, 1.25)	0.95	1.09

Abbreviations: CI= Confidence Interval; RR = risk ratio; NR = not reported; RAF= Risk allele frequency

[a] Imputed SNPs are italicized

[b] Odds ratio for the log-additive model reported in Mavaddat et al. [60] Original ORs<1.00 were transformed to reflect the measured effect of the risk allele.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Table 4

Effect estimates for maternally-mediated genetic effects of previous breast cancer genome-wide association study risk variants on the risk of young-onset breast cancer

Rank	SNP[a]	Gene	RR (95% CI)	p-value
1	rs9790517	TET2	0.78 (0.66, 0.94)	0.007
2	rs10995190	ZNF365	0.75 (0.61, 0.93)	0.008
3	*rs9693444*	8p12	0.81 (0.70, 0.93)	0.008
4	*rs2236007*	PAX9	1.21 (1.01, 1.45)	0.03
5	rs2016394	DLX2	0.85 (0.73, 0.99)	0.04
6	*rs2588809*	RAD51B	0.84 (0.69, 1.02)	0.07
7	rs2943559	HNF4G	0.80 (0.62, 1.04)	0.09
8	*rs6504950*	STXBP4	1.15 (0.97, 1.35)	0.10
9	*rs6828523*	ADAM29	1.19 (0.94, 1.49)	0.15
10	rs2981579	FGFR2	0.90 (0.78, 1.04)	0.15
11	*rs10771399*	12p11.22	0.84 (0.66, 1.08)	0.17
12	rs720475	ARHGEF5	0.89 (0.75, 1.05)	0.18
13	rs8170	BIBAM1	0.87 (0.71, 1.07)	0.18
14	rs2046210	ESR1	1.10 (0.95, 1.28)	0.20
15	*rs12422552*	12p13.1	0.90 (0.76, 1.06)	0.21
16	rs4245739	MDM4	1.11 (0.94, 1.32)	0.22
17	rs1045485	CASP8	0.87 (0.70, 1.09)	0.23
18	*rs7904519*	TCF7L2	0.92 (0.79, 1.06)	0.25
19	rs3817198	LSP1	0.91 (0.77, 1.07)	0.26
20	rs865686	RPL31P43	1.09 (0.93, 1.27)	0.28
21	rs13281615	8q24.21	1.09 (0.93, 1.27)	0.28
22	*rs3903072*	11q13.1	1.09 (0.93, 1.27)	0.29
23	rs11242675	FOXQ1	1.09 (0.93, 1.27)	0.30
24	*rs10759243*	9q31.2	0.91 (0.76, 1.09)	0.31
25	*rs11820646*	11q24.3	0.93 (0.80, 1.07)	0.31
26	rs12662670	ESR1	0.88 (0.67, 1.15)	0.34
27	rs2380205	GDI2	0.93 (0.79, 1.08)	0.34
28	rs13387042	TNP1	1.07 (0.92, 1.25)	0.36
29	rs999737	RAD51B	0.93 (0.78, 1.10)	0.38
30	*rs4849887*	2q14.2	0.92 (0.72, 1.16)	0.47
31	*rs17356907*	12q22	0.94 (0.80, 1.12)	0.49
32	*rs13329835*	CDYL2	1.06 (0.88, 1.28)	0.52
33	*rs6762644*	ITPR1	0.95 (0.82, 1.11)	0.53
34	*rs204247*	RANBP9	0.95 (0.82, 1.11)	0.54
35	rs1292011	12q24.21	0.95 (0.81, 1.12)	0.55
36	rs616488	PEX14	0.95 (0.82, 1.11)	0.55
37	rs4973768	SLC4A7	0.95 (0.82, 1.11)	0.55
38	rs10069690	TERT	0.95 (0.80, 1.13)	0.56

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Rank	SNP[a]	Gene	RR (95% CI)	p-value
39	*rs11814448*	10p12.31	0.90 (0.57, 1.41)	0.64
40	*rs10941679*	MRPS30	0.96 (0.80, 1.15)	0.65
41	*rs16857609*	DIRC3	0.96 (0.81, 1.14)	0.66
42	*rs1550623*	CDCA7	0.96 (0.78, 1.18)	0.67
43	*rs2736108*	TERT	0.96 (0.79, 1.17)	0.67
44	*rs6678914*	LGR6	1.03 (0.88, 1.21)	0.68
45	*rs527616*	18q11.2	1.04 (0.87, 1.25)	0.68
46	*rs12493607*	TGFBR2	0.97 (0.83, 1.13)	0.71
47	rs704010	ZMIZ1	0.97 (0.83, 1.14)	0.71
48	*rs78540526*	CCND1	1.05 (0.78, 1.42)	0.73
49	rs3803662	TOX3	0.97 (0.82, 1.15)	0.74
50	rs889312	MAP3K1	0.97 (0.83, 1.15)	0.75
51	*rs11075995*	FTO	0.97 (0.82, 1.15)	0.76
52	*rs75915166*	FGF3	1.05 (0.77, 1.42)	0.76
53	rs1436904	CHST9	1.02 (0.87, 1.20)	0.77
54	*rs1353747*	PDE4D	0.97 (0.75, 1.24)	0.79
55	rs1011970	CDKN2B	0.97 (0.79, 1.19)	0.80
56	*rs11199914*	10q26.12	0.98 (0.83, 1.16)	0.81
57	rs2363956	ANKLE1	0.98 (0.85, 1.14)	0.81
58	*rs4808801*	ELL	0.98 (0.84, 1.15)	0.82
59	rs941764	CCDC88C	0.98 (0.84, 1.16)	0.83
60	*rs1432679*	EBF1	1.01 (0.87, 1.17)	0.89
61	rs11249433	EMBP1	0.99 (0.85, 1.15)	0.89
62	*rs12710696*	LINC01376	1.01 (0.86, 1.19)	0.90
63	rs17817449	FTO	0.99 (0.85, 1.16)	0.90
64	rs11552449	AP4B1	0.99 (0.81, 1.21)	0.92
65	*rs7726159*	TERT	0.99 (0.82, 1.20)	0.92
66	*rs554219*	CCND1	1.01 (0.82, 1.25)	0.92
67	rs10472076	RPL5P15	1.01 (0.87, 1.17)	0.93
68	*rs132390*	EMID1	1.02 (0.62, 1.69)	0.93
69	*rs11780156*	8q24.21	0.99 (0.81, 1.22)	0.95
70	rs2823093	CYRR1	1.00 (0.85, 1.17)	0.95
71	rs6001930	MKL1	1.01 (0.79, 1.29)	0.95
72	rs7072776	MLLT10	1.00 (0.85, 1.16)	0.96
73	rs3760982	KCNN4	1.00 (0.86, 1.16)	1.00

Abbreviations: CI= Confidence Interval; RR = risk ratio

[a]Imputed SNPs are italicized