

Published in final edited form as:

*Cancer Epidemiol Biomarkers Prev.* 2009 January ; 18(1): 255–259. doi:  
10.1158/1055-9965.EPI-08-0704.

## Common Genetic Variation in Candidate Genes and Susceptibility to Sub-types of Breast Cancer

Nasim Mavaddat<sup>1</sup>, Alison M Dunning<sup>2</sup>, Bruce AJ Ponder<sup>2</sup>, Douglas F Easton<sup>1</sup>, and Paul D Pharoah<sup>2</sup>

<sup>1</sup>Department of Public Health and Primary Care, University of Cambridge

<sup>2</sup>Department of Oncology, University of Cambridge

### Abstract

Association studies have been widely used to search for common low penetrance susceptibility alleles to breast cancer in general. However, breast cancer is a heterogeneous disease and it has been suggested that it may be possible to identify additional susceptibility alleles by restricting analyses to particular subtypes. We used data on 710 SNPs in 120 candidate genes from a large candidate-gene association study of up to 4470 cases and 4560 controls to compare the results of analyses of “overall” breast cancer with sub-group analyses based on the major clinico-pathological characteristics of breast cancer (stage, grade, morphology and hormone receptor status). No single nucleotide polymorphism (SNP) was highly significant in overall-effects analysis. Subgroup analysis resulted in substantial reordering of ranks of SNPs, as assessed by the magnitude of the test statistics and some associations that were not significant for an overall effect were detected in sub-groups at a nominal 5% level adjusted for multiple testing. The most significant association, of *CCND1* SNP rs3212879 with estrogen receptor negative tumour types ( $p = 0.001$ ), did not reach genome-wide significance levels. These results demonstrate that it may be possible to detect associations using subgroup analysis that are missed in overall-effects analysis. If the associations we found can be replicated in independent studies they may provide important insights into disease mechanisms in breast cancer.

### Keywords

breast; cancer; sub-types; genetic; variation

### Introduction

Breast cancer tends to cluster in families, the disease being approximately twice as common in first-degree relatives of cases, than in the general population (1). Some of this clustering occurs as part of specific familial breast cancer syndromes where disease results from single alleles conferring a high risk. However such alleles are rare in the population and the majority of multiple case breast cancer families do not segregate mutations in these genes (2). The model that best describes aggregation of breast cancer in these families is a polygenic model in which susceptibility to breast cancer is conferred by a large number of genetic variants, each of which has a modest effect (3, 4, 5, 6).

Despite large research efforts over the past ten years the number of common susceptibility alleles identified has been small. Loci implicated thus far include SNP alleles located in the

*CASP8*, *FGFR2*, *TNRC9*, *MAP3K1* and *LSP1* genes, and in regions in the DNA devoid of known genes at chromosome positions 8q24 and 2q35 (7, 8). One of the reasons for this lack of success may be disease heterogeneity. Invasive breast cancers can be divided into several pathologic subtypes with different histological appearances of the malignant cells, and different clinical presentations and outcomes. Novel subtypes have also emerged as a result of gene expression-profiling (9). It is plausible that the aetiology of the sub-types is different. The association of the 'basal' phenotype in breast cancers with rare, deleterious mutations in *BRCA1* (10, 11, 12, 13) demonstrates the principle that different genetic determinants can underlie different subtypes of the disease. In addition, the association with SNPs at the *FGFR2*, 2q35 and *MAP3K1* loci have been shown to be largely restricted to estrogen receptor positive disease (14, 15).

Few genetic association studies have systematically evaluated association between putative common susceptibility alleles and specific sub-types of disease. However, under some models of disease susceptibility, sub-group analysis may identify associations missed by analysis of overall effects. For example, if the genetic effects in the subgroups are sufficiently heterogeneous, the power of sub-group analysis may exceed that of simple overall-effects analysis. This is supported by the findings that, under some models, power to detect epistasis between genes is greater than power to detect overall effects (16). In that report the sub-group was defined by genotype at a second locus. The largest study to-date of common variants in multiple candidate genes for breast cancer susceptibility has been published recently (17). Over 700 single nucleotide polymorphisms (SNPs) in 120 genes in approximately 4,400 cases and 4,400 controls were tested. None of the SNPs reached genome-wide significance levels after adjusting for population stratification (17). However when the admixture maximum likelihood (AML) experiment-wise test for association was applied, there was evidence for an excess of positive associations over the proportion expected by chance, suggesting that some SNPs in these candidate genes are associated with breast cancer risk (17).

The purpose of this study was to reanalyse this data set in order to identify highly significant associations with specific subgroups of breast cancer in the absence of overall effects. Subgroups were based on the major clinico-pathological features of the tumours, namely morphology type, stage, grade, estrogen and progesterone receptor status.

## Methods

### Study participants

Study participants (up to 4,470 cases and 4,560 controls) were selected, as described by Pharoah et al. (17). In brief, cases were drawn from Studies of Epidemiology and Risk factors in Cancer Heredity (SEARCH), an ongoing population-based study ascertained through the Eastern Cancer Registration and Information Centre. All patients diagnosed with invasive breast cancer below age 55 years since 1991 and still alive in 1996 (prevalent cases, median age 48 years), together with all those diagnosed below age 70 years between 1996 and the present (incident cases, median age 54 years) are eligible to take part. Of 12,767 eligible patients, 2,284 were not contacted because their general practitioner did not respond or thought that it would be inappropriate to contact the patient. Of the 10,583 patients who were contacted, 67% have returned a questionnaire, and 64% provided a blood sample for DNA analysis. Eligible patients who did not take part in the study were similar to participants except, as might be expected, the proportion of clinical stage III/IV cases was somewhat higher in nonparticipants (10% versus 5%). Female controls were randomly selected from the Norfolk component of the European Prospective Investigation of Cancer (EPIC) a prospective study of diet and cancer being carried out in nine European countries (18). The EPIC-Norfolk cohort comprises 25,000 individuals resident in Norfolk, East

Anglia—the same region from which the cases have been recruited. Controls are not matched to cases, but are broadly similar in age (42-81 years). The ethnic background of both cases and controls as reported on the questionnaires is similar, with >98% being white. Staging and phenotyping of breast cancer cases were obtained through the cancer registry from routine pathology and clinical records. The study is approved by the Eastern Region Multi-centre Research Ethics Committee, and all patients gave written informed consent.

The samples were split into two sets in order to save DNA and reduce genotyping costs: the first set ( $n = 2,270$  cases and 2,280 controls) was genotyped for all SNPs, and the second set ( $n = 2,200$  cases and 2,280 controls) were then tested for those SNPs that showed marginally significant associations in overall effects analysis for set 1 ( $p$  - heterogeneity or  $p$  - trend < 0.1). This staged approach substantially reduces genotyping costs without significantly affecting statistical power of the overall effects analysis. Results of sub-group analyses for set 1 data were however not used to select SNPs for analysis in set 2.

### Candidate gene, SNP selection and Genotyping methods

Data on 710 SNPs in 120 candidate genes were available for analysis<sup>17</sup> (Supplementary Table 1). Genes that encode proteins in cellular pathways that are likely to be involved in breast carcinogenesis were chosen as candidates. The major pathways studied were steroid hormone metabolism and signalling, double strand break DNA repair, oxidative damage repair, epigenetic modifiers, and cell-cycle control. Genes in the 17q21 region commonly amplified in a variety of animal models of cancer, and some carcinogen metabolism genes were also tested. For some pathways, only a small subset of genes was selected for study. Genes evaluated by pathway and number of SNPs assayed for each are described in Pharaoh et al (17). Common variation in most genes was captured using a minimal set of tagging SNPs (17, 19). Genotyping methods are as described (17). Concordance for duplicate samples was 98% for all assays. Failed genotypes were not repeated (the rate for failed genotypes did not exceed 8.3% for any of the SNPs under study). Hardy Weinberg Equilibrium was tested as part of genotyping quality assurance and SNPs with serious deviations excluded.

### Statistical methods

The aim of this study was to test for statistical association between each of 710 individual SNPs and breast cancer sub-types, and to compare the results of these with the overall-effects analyses. The sub-types were categorized according to clinical stage at diagnosis (I, II, III/ IV), histopathological grade (1, 2, and 3), estrogen receptor (ER) status and progesterone receptor (PR) status (negative or positive) and histopathological morphology. Only the most common morphological types - lobular and ductal were analysed. Pair-wise correlation coefficient was calculated to assess correlation structure between sub-groups (Supplementary Table 2). Sample sizes for most joint phenotypes were too small for sub-grouping to be based on phenotypic correlations. Association between disease and genotype for each SNP within each subtype category for ER status, PR status and morphology was assessed using the one degree of freedom Cochran-Armitage trend test with a single parameter for allele dose. The analyses were conducted with each sub-group being compared with all of the available controls. Grade and stage were assessed as ordered categories using ordinal polytomous logistic regression.

Results for all tests were summarised using standard quantile-quantile (Q-Q) plots constructed by ranking the set of values for the test statistic from smallest to largest and plotting them against their expected values. Per-allele odds ratios and confidence intervals were estimated using logistic regression. In order to compare the previously reported overall effects analysis with the sub-group analyses reported here a Bonferroni correction was

applied to correct for the number of sub-group analyses (eight). A nominal significance level of  $p < 0.05$  was chosen for overall effects and an equivalent  $p < 0.00625$  ( $=0.05/8$ ) for subgroup analyses. Note that this is not a correction for the number of SNPs tested as such a correction would be the same for each sub-group analysis and would make no difference to the comparison between overall effects and sub-group analyses.

## Results

The number of cases by sub-group is shown in Table 1. This is a maximum sample size as not all SNPs were genotyped for both set 1 and set 2. Figure 1(A) shows the Q-Q plot for the univariate trend test for association between SNPs and breast cancer (overall effect). For chi-squared values less than three, the observed values lie close to the line expected under the null hypothesis of no association, providing no evidence of inflation of the test statistic that would suggest population stratification or other systematic bias. The deviation of the higher observed values from those expected is suggestive of multiple weak associations. One SNP showed a much higher chi-squared statistic than the others. This SNP - rs3020314 in the estrogen receptor  $\alpha$  gene ( $p = 8 \times 10^{-5}$ ) – did not reach genome-wide significance, but did reach the  $p < 10^{-4}$  threshold that has been suggested for candidate gene studies (20). Figures 1 (B)-(I) show Q-Q plots for the univariate trend test for each subgroup scan. None of the associations reached the level of significance for the most significant association in the overall effects analysis.

In the overall effects analyses, 52 SNPs (7.5%) were significant at the  $p < 0.05$  level. In subgroup analysis, at the equivalent threshold significance of  $p < 0.00625$ , 7 SNPs were significantly associated with increasing cancer grade, 16 with increasing cancer stage, 7 with lobular cancer, 7 with ductal cancer, 7 with ER positive disease, 14 with ER negative disease, 6 with PR positive cancer and 5 with PR negative cancer (data not shown). Most of the SNPs detected at the  $p < 0.00625$  level in subgroup analysis achieve at least  $p < 0.05$  significance in overall effects analysis, but 18 SNPs found to be significant in subgroup analysis did not. For these SNPs per allele OR and the corresponding 95% confidence intervals, in subgroup and overall effects analysis are shown in Table 2. Sample sizes of cases and controls for each SNP tested is also indicated. The strongest association observed was for *CCND1* rs3212879 in ER negative disease ( $P=0.0001$ , OR = 1.40, 95% CI = 1.20-1.70).

Thus, sub-group analysis has not identified any highly significant associations missed by overall effects analysis. Nevertheless, sub-group analysis may still be useful when selecting SNPs of borderline significance for further replication. In general, the number of SNPs selected for replication studies is limited by the cost of genotyping. For example, assume funding is available to attempt to replicate 50 SNPs in further studies in order to provide definitive evidence of association. One strategy would be to simply select the top 50 'hits' ie SNPs that are significant at some pre-defined level – for example  $p < 0.05$ , from the overall effects analysis. However, better candidates for replication may be identified from the sub-group analyses. A possible strategy would be to include SNPs for replication that achieve the same  $p$ -value as overall effects analysis after Bonferroni correction. Here we applied a correction of eight in order to compare directly the results of the subgroup and overall-effects analysis, although this correction is overly conservative as sub-groups (and therefore tests) are correlated (Supplementary Table 2). A ranking of corrected  $p$ -values representing the top 50 'hits' ( $p < 0.05$ ), from both overall effects and sub-group analyses is shown in Table 3. With this strategy, twelve SNPs achieved higher ranks in subgroup analysis than in overall-effects analysis. Thus 50 SNPs chosen for replication would include 38 SNPs from the overall effects analysis and 12 SNPs from the sub-group analyses.

## Discussion

Complex diseases such as breast cancer are phenotypically heterogeneous, and this heterogeneity may obscure genetic associations. If alleles at different genetic loci are responsible for different subtypes of disease, genetic associations may be best detected by subgroup analysis. However there is a trade off between the increased specificity that may be obtained by sub-group analysis, and the loss of statistical power from reduction in sample size and increase in number of hypotheses being evaluated. The most powerful way to test for and detect associations is not known because the true underlying biological/genetic models for the data are not known.

Our results support the notion that sub-group analysis may be worthwhile because some associations that were not detected using an overall-effects approach were detected using a subgroup approach at a nominal level of  $p < 0.05$  adjusted for multiple sub-group testing. Nevertheless, the findings are only illustrative of the potential for sub-group analysis, because none of the associations detected could be regarded as definitive, and we cannot state with certainty that sub-group analyses have identified true associations missed by the overall effects analysis. The most strongly associated SNP was rs3212879 in *CCND1*, which was associated with risk of estrogen receptor negative tumours ( $p = 0.001$  adjusted for multiple sub-group testing). This does not reach the threshold for genome-wide significance ( $p < 10^{-8}$ ), which is necessarily stringent due to the low prior probability of any individual SNP being associated with disease. The prior probability for SNPs in candidate gene studies may be somewhat higher but there is still no biological *a priori* hypothesis for association between any particular SNP with a particular subtype of cancer. Despite good total sample size, the sub-group sizes are modest and power to detect sub-group effects at very stringent significance levels may be small. Statistical power may be further limited by sub-group classification error. Data for sub-group categorisation were obtained from clinical records and there is likely to be some degree of misclassification of phenotype. Nevertheless, the effect size for *GPX4* rs4087542 in lobular carcinoma, and the SNPs in the *TBXAS1* family in ER positive tumours and those in the *CCND1* family in ER negative tumours are of sufficient magnitude to warrant replication in larger studies of patients with these subtypes of cancer.

There are other published examples of associations between genetic variants and breast cancer restricted to specific subtypes. The associations of common variants at *FGFR2*, *MAP3K1* and *2q35* have been reported to be confined to estrogen receptor positive cancers (14, 15). These loci were not evaluated as part of our candidate gene study. In these examples, the overall effect analyses between the variants and breast cancer had reached genome-wide significance levels prior to subgroup analysis.

Our results also demonstrate that sub-group analyses may be incorporated into a strategy for selection of SNPs for replication in independent data sets. Staged study designs are commonly used in genome-wide association studies in order to reduce costs (21). The most appropriate selection of SNPs for the second and subsequent stages is critical to maximise power. To date most GWAS have based this selection on the results of overall effects analyses, but SNP selection based on overall effects and sub-group analysis with an appropriate correction for multiple testing may prove more efficient.

The potential advantages of reducing phenotypic heterogeneity by restricting analysis to specific sub-types of disease are clear. Further evaluation of such a strategy is required to provide definitive evidence of its value.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

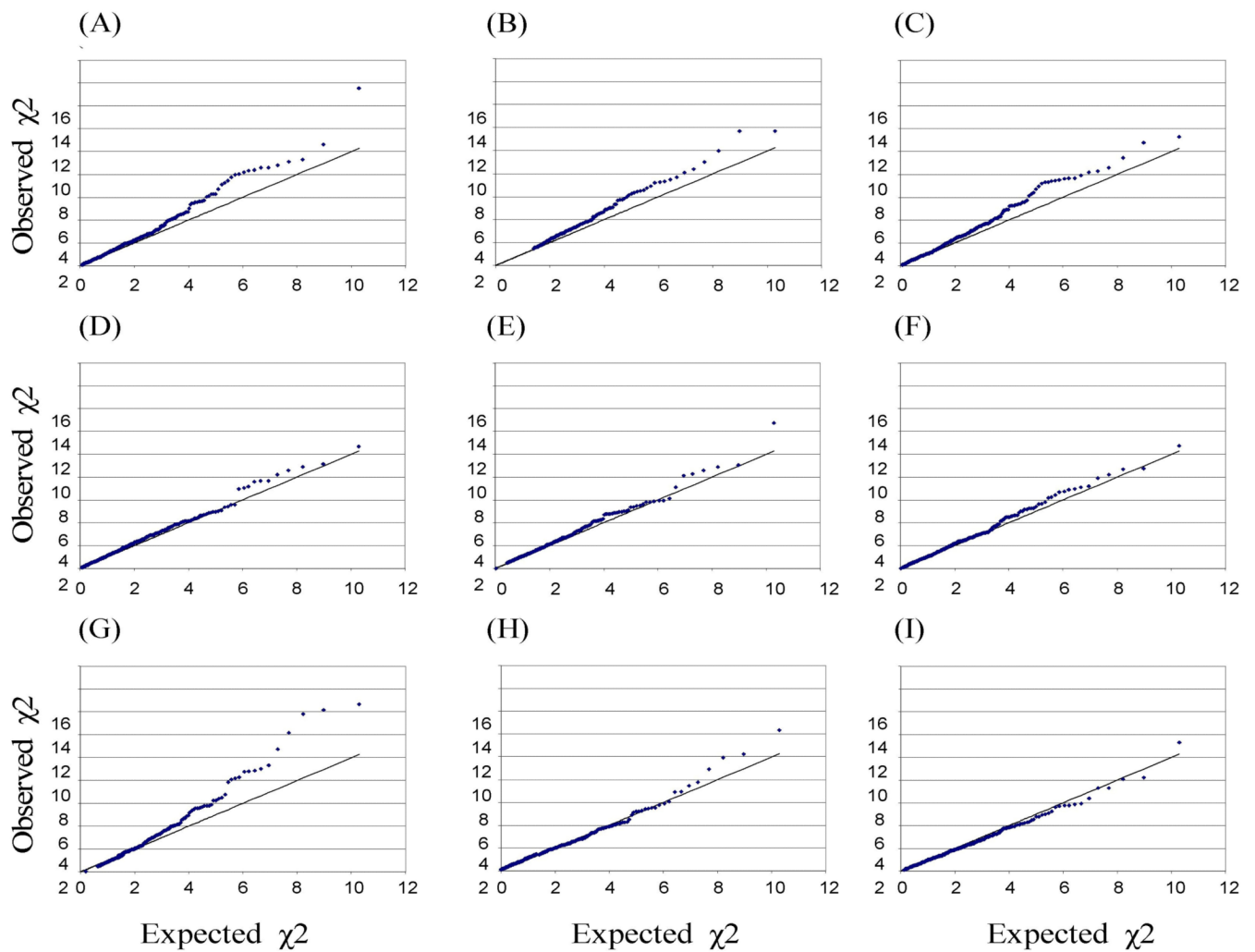
We thank the SEARCH team, the EPIC collaborators and the Eastern Cancer Registration and Information Centre (patient recruitment and phenotype data). Genotyping was carried out by many individuals from the Department Oncology at Strangeways Research Laboratory and funded by Cancer Research UK NM was funded by scholarships from Cancer Research UK and the Medical Research Council, PP is a Senior Clinical Research Fellow and DFE a Principal Fellow of Cancer Research UK.

Funding: NM was funded by scholarships from Cancer Research UK and the Medical Research Council, PP is a Senior Clinical Research Fellow and DFE a Principal Fellow of Cancer Research UK.

## References

1. Familial breast cancer: collaborative reanalysis of individual data from 52 epidemiological studies including 58,209 women with breast cancer and 101,986 women without the disease. *Lancet*. 2001; 358:1389–1399. [PubMed: 11705483]
2. Easton DF. How many more breast cancer predisposition genes are there? *Breast Cancer Res*. 1999; 1:14–17. [PubMed: 11250676]
3. Antoniou AC, Pharoah PD, McMullan G, et al. A comprehensive model for familial breast cancer incorporating BRCA1, BRCA2 and other genes. *Br J Cancer*. 2002; 86:76–83. [PubMed: 11857015]
4. Antoniou AC, Pharoah PP, Smith P, et al. The BOADICEA model of genetic susceptibility to breast and ovarian cancer. *Br J Cancer*. 2004; 91:1580–1590. [PubMed: 15381934]
5. Cui J, Antoniou AC, Dite GS, et al. After BRCA1 and BRCA2-what next? Multifactorial segregation analyses of three-generation, population-based Australian families affected by female breast cancer. *Am J Hum Genet*. 2001; 68:420–431. [PubMed: 11133358]
6. Pharoah PD, Antoniou A, Bobrow M, Zimmern RL, Easton DF, Ponder BA. Polygenic susceptibility to breast cancer and implications for prevention. *Nat Genet*. 2002; 31:33–36. [PubMed: 11984562]
7. Easton DF, Pooley KA, Dunning AM, et al. Genome-wide association study identifies novel breast cancer susceptibility loci. *Nature*. 2007; 447(7148):1087–93. [PubMed: 17529967]
8. Cox A, Dunning AM, Garcia-Closas M, et al. Corrigendum: A common coding variant in CASP8 is associated with breast cancer risk. *Nat Genet*. 2007; 39:688.
9. Sorlie T, Perou CM, Tibshirani R, et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci USA*. 2001; 98:10869–10874. [PubMed: 11553815]
10. Hedenfalk I, Duggan D, Chen Y, et al. Gene-expression profiles in hereditary breast cancer. *N Engl J Med*. 2001; 344:539–548. [PubMed: 11207349]
11. Hedenfalk IA, Ringner M, Trent JM, Borg A. Gene expression in inherited breast cancer. *Adv Cancer Res*. 2002; 84:1–34. [PubMed: 11883525]
12. Sorlie T, Tibshirani R, Parker J, et al. Repeated observation of breast tumor subtypes in independent gene expression data sets. *Proc Natl Acad Sci USA*. 2003; 100:8418–8423. [PubMed: 12829800]
13. van't Veer LJ, Dai H, van de Vijver MJ, et al. Gene expression profiling predicts clinical outcome of breast cancer. *Nature*. 2002; 415:530–536. [PubMed: 11823860]
14. Stacey SN, Manolescu A, Sulem P, et al. Common variants on chromosomes 2q35 and 16q12 confer susceptibility to estrogen receptor-positive breast cancer. *Nat Genet*. 2007; 39:865–869. [PubMed: 17529974]
15. Garcia-Closas M, Hall P, Nevanlinna H, et al. Heterogeneity of breast cancer associations with five susceptibility loci by clinical and pathological characteristics. *PLoS Genet*. Apr 25.2008 4(4):e1000054. [PubMed: 18437204]

16. Evans DM, Marchini J, Morris AP, Cardon LR. Two-stage two-locus models in genome-wide association. *PLoS Genet.* 2006; 2:e157. [PubMed: 17002500]
17. Pharoah PD, Tyrer J, Dunning AM, Easton DF, Ponder BA. Association between common variation in 120 candidate genes and breast cancer risk. *PLoS Genet.* 2007; 3:e42. [PubMed: 17367212]
18. Day N, Oakes S, Luben R, et al. EPIC-Norfolk: study design and characteristics of the cohort. European Prospective Investigation of Cancer. *Br J Cancer.* 1999; 80(Suppl 1):95–103. [PubMed: 10466767]
19. de Bakker PI, Yelensky R, Pe'er I, Gabriel SB, Daly MJ, Altshuler D. Efficiency and power in genetic association studies. *Nat Genet.* 2005; 37:1217–1223. [PubMed: 16244653]
20. Pharoah PD, Dunning AM, Ponder BA, Easton DF. Association studies for finding cancer-susceptibility genetic variants. *Nat Rev Cancer.* 2004; 4:850–860. [PubMed: 15516958]
21. McCarthy MI, Abecasis GR, Cardon LR, et al. Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nat Rev Genet.* 2008; 9:356–369. [PubMed: 18398418]



**Figure 1.**

Q-Q plot for association of SNPs with breast cancer, observed vs expected Chi-squared  $\chi^2$ : overall effects (A), grade (B), stage (C), lobular morphology (D), and ductal morphology (E), ER positive (F), ER negative (G), PR positive (H), and PR negative breast cancers (I). The line shows the plot of expected values according to the null hypothesis.



**Table 1**  
**Number of breast cancers included in analyses by sub-type\***

Clinical phenotype / sub-group	set 1	set 2	Total
Controls *	2280	2280	4560
Stage *			
I	1114	1077	2191
II	987	996	1983
III/IV	110	84	194
Missing	59	43	102
Grade *			
1	437	437	874
2	788	899	1687
3	505	515	1020
Missing	540	349	889
Morphology-type			
Lobular	351	308	659
Ductal	1674	1642	3316
Other	222	233	455
Missing	23	17	40
ER status			
Positive	1054	983	2037
Negative	278	192	470
Missing	938	1025	1963
PR status			
Positive	536	240	776
Negative	233	116	349
Missing	1501	1844	3345

Maximum number of cases and controls are listed.

\* Controls are coded as zero for ordered logistic regression. Otherwise all other analyses were conducted with all of the controls.

**Table 2**  
**Odds ratio of SNPs showing evidence of association with breast cancer in subgroup but not overall effect analysis\***

Subgroup	GENE	dBSNP ref	Cases/controls	Per allele OR <sup>†</sup> (95% CI)	Per allele OR <sup>‡</sup> (95% CI)
Lobular	GPX4	rs4807542	340/2259	1.38 (1.14-1.69)	1.10 (0.97-1.2)
ER positive	TBXAS1	rs6971207	269/2278	1.85 (1.23-2.70)	1.30 (0.90-1.9)
ER negative	CCND1	rs3212879	268/2271	1.40 (1.20-1.70)	1.04 (0.96-1.13)
	CCND1	rs678653	447/4471	1.30 (1.13-1.50)	1.06 (1.00-1.13)
	CCND1	rs3212891	268/2275	1.40 (1.15-1.66)	1.04 (0.96-1.13)
	CCND1	rs602652	267/2268	0.74 (0.61-0.88)	0.98 (0.93-1.07)
	CCND1	rs603965	257/2194	0.74 (0.62-0.90)	0.94 (0.87-1.03)
	DNNMT3A	rs11677670	269/2275	0.68 (0.52-0.87)	0.91 (0.82-1.01)
	PTGIS	rs693649	268/2272	0.69 (0.53-0.88)	0.92 (0.83-1.02)
	GPX1	rs3448	460/4557	1.24 (1.07-1.44)	1.03 (0.97-1.10)
	STE	rs1881668	449/4473	1.24 (1.07-1.44)	1.05 (0.99-1.13)
PR positive	CDK6	rs3757823	339/4540	1.36 (1.14-1.60)	1.10 (0.99-1.22)
	PTGIS	rs6095541	222/2270	1.32 (1.11-1.56)	1.07 (0.96-1.20)
	PTGIS	rs6095543	223/2264	1.32 (1.11-1.56)	1.08 (0.97-1.21)
	CCND3	rs2479717	224/2267	0.80 (0.68-0.94)	0.94 (0.86-1.03)
PR negative	IGF1R	rs1546713	209/2149	0.74 (0.60-0.90)	0.94 (0.86-1.02)
	NBS1	rs1805787	215/2243	1.34 (1.10-1.65)	1.04 (0.95-1.14)
	BRCA2	rs4942485	217/2251	0.60 (0.42-0.87)	0.95 (0.83-1.08)

\* SNPs at significance level  $p < 0.00625$  in subgroup analysis, but not  $p < 0.05$  in overall effects analysis are shown.

<sup>†</sup> per allele OR for subgroup

<sup>‡</sup> per allele OR for overall effects analysis

**Table 3**  
**SNPs showing evidence of association with breast cancer in rank order of significance:**  
**results from overall effects and subgroup analysis**

Gene	dBSNP ref	p-value	p-value*	subgroup
ESR1	rs3020314	$8 \times 10^{-5}$	$8 \times 10^{-5}$	overall effect
CCND1	rs3212879	0.0001	0.0010	ER negative
MCS5A6	rs2182317	0.0011	0.0011	overall effect
CCND1	rs678653	0.0003	0.0023	ER negative
DNMT3B	rs406193	0.0026	0.0026	overall effect
CDKN1A	rs3176336	0.0030	0.0030	overall effect
ESR1	rs3020317	0.0030	0.0030	overall effect
ESR1	rs3020394	0.0034	0.0034	overall effect
CCNE1	rs997669	0.0034	0.0034	overall effect
CCND1	rs3212891	0.0005	0.0038	ER negative
ESR1	rs3020396	0.0038	0.0038	overall effect
ESR1	rs3020401	0.0039	0.0039	overall effect
CDK6	rs3757823	0.0005	0.0036	ER positive
ESR1	rs3020377	0.0043	0.0043	overall effect
ESR1	rs3020390	0.0046	0.0046	overall effect
ESR1	rs3020400	0.0047	0.0047	overall effect
STE	rs3736599	0.0053	0.0053	overall effect
ESR1	rs3020407	0.0062	0.0062	overall effect
RB1	rs2854344	0.0070	0.0070	overall effect
ESR1	rs3020405	0.0077	0.0077	overall effect
MMP8	rs1892886	0.0097	0.0097	overall effect
GPX4	rs4807542	0.0001	0.0086	lobular
CCND1	rs602652	0.0001	0.0086	ER negative
PTGIS	rs6095541	0.0001	0.0106	ER positive
CDKN1B	rs34330	0.0120	0.0120	overall effect
IGFBP3	rs2132572	0.0122	0.0122	overall effect
CDKN2A	rs3731239	0.0122	0.0122	overall effect
PTGIS	rs6095543	0.0016	0.0130	ER negative
BAT8	rs535586	0.0138	0.0138	overall effect
ERBB2	rs1801201	0.0139	0.0139	overall effect
ESR1	rs726282	0.0167	0.0167	overall effect
RB1	rs198580	0.0174	0.0174	overall effect
CHEK2	rs9608698	0.0177	0.0177	overall effect
CCND1	rs603965	0.0023	0.0180	ER negative
PGR	rs1042838	0.0180	0.0180	overall effect
STE	rs3775775	0.0182	0.0182	overall effect
STK15	rs732417	0.0182	0.0182	overall effect
EHMT1	rs4634736	0.0194	0.0194	overall effect

Gene	dBSNP ref	p-value	p-value*	subgroup
ADH1C	rs698	0.0196	0.0196	overall effect
RIZ1	rs2235515	0.0204	0.0204	overall effect
DNMT3A	rs11677670	0.0030	0.0240	ER negative
PTGIS	rs693649	0.0031	0.0250	ER negative
TBXAS1	rs6971207	0.0032	0.0254	ER positive
ADH1B	rs1042026	0.0253	0.0253	overall effect
ESR1	rs1884051	0.0299	0.0299	overall effect
XRCC2	rs3218536	0.0302	0.0302	overall effect
IGF1R	rs2229765	0.0308	0.0308	overall effect
IGF1R	rs1546713	0.0039	0.0310	ER negative
ESR1	rs1884054	0.0328	0.0328	overall effect
SHBG	rs858524	0.0329	0.0329	overall effect

\* Bonferroni correction factor of eight is used to adjust tests in subgroup analysis