Published in final edited form as:
*Breast Cancer Res Treat*. 2011 January ; 125(1): 221–228. doi:10.1007/s10549-010-0950-8.

# Centrosome-related genes, genetic variation, and risk of breast cancer

**J. E. Olson**, **X. Wang**, **V. S. Pankratz**, **Z. S. Fredericksen**, **C. M. Vachon**, **R. A. Vierkant**, **J. R. Cerhan**, and **F. J. Couch**
Mayo Clinic College of Medicine, Rochester, MN, USA

J. E. Olson: olsonj@mayo.edu

## Abstract

Centrosome amplification has been detected in premalignant lesions and in situ tumors in the breast and in over 70% of invasive breast tumors, and has been associated with aneuploidy and tumor development. Based on these observations, the contribution of commonly inherited genetic variation in candidate genes related to centrosome structure and function to breast cancer risk was evaluated in an association study. Seven-hundred and 82 single nucleotide polymorphisms (SNPs) from 101 centrosomal genes were analyzed in 798 breast cancer cases and 843 controls from the Mayo Clinic Breast Cancer Study to assess the association between these SNPs (both individually and combined) and risk of breast cancer in this population. Eleven SNPs out of 782 from six genes displayed associations with breast cancer risk ($P < 0.01$). Haplotypes in five genes also displayed significant associations with risk. A two SNP combination of rs10145182 in *NIN* and rs2134808 in the *TUBG1* locus (*P*-interaction = 0.00001), suggested SNPs in mediators of microtubule nucleation from the centrosome contribute to breast cancer. Evaluation of the simultaneous significance of all SNPs in the centrosome pathway suggested that the centrosome pathway is highly enriched ($P = 4.76 \times 10^{-50}$) for SNPs that are associated with breast cancer risk. Collections of weakly associated genetic variants in the centrosome pathway, rather than individual highly significantly associated SNPs, may account for a putative role for the centrosome pathway in predisposition to breast cancer.

### Keywords

Centrosome; Mitosis; Single nucleotide polymorphism (SNP); Haplotype; Breast cancer risk

## Introduction

The centrosome is the primary microtubule organizing center of the cell. It is responsible for maintenance of cellular polarity, is required for entry into S-phase of the cell cycle [1], and mediates the process of chromosome segregation during mitosis. Centrosomes duplicate during late G1 phase, separate in G2, and establish two spindle poles during mitosis to



Correspondence to: J. E. Olson, olsonj@mayo.edu.

**Electronic supplementary material** The online version of this article (doi:10.1007/s10549-010-0950-8) contains supplementary material, which is available to authorized users.

**Conflicts of interest statement** None of the authors have a financial relationship with the organization that sponsored the research (NIH). We further state that we have full control of all primary data and that we agree to allow the journal to review our data if requested.

facilitate chromosome segregation [2,3]. Numerical or functional defects of the centrosome can result in improper chromosome segregation, leading to aneuploidy and polyploidy [3].

As many cancer cells have aberrant numbers of chromosomes, a link between centrosomes and cancer has been suggested. Centrosomal defects, including the presence of extra centrioles, and increased ability to nucleate microtubules, are common in many cancers. Centrosomal aberrations have been detected in premalignant lesions and in situ tumors in the breast and in over 70% of invasive breast tumors [4–6]. Similarly, cells with monopolar spindles resulting from a failure to duplicate or separate centrosomes are often observed in tumors. While centrosomal defects in cancer may arise as a result of malignant processes, early centrosomal aberrations may also lead to increased malignancy [6]. Owing to the frequent involvement of centrosome defects in breast cancer, we conducted an association study to determine whether common genetic variations in 101 genes involved in centrosome structure and function contribute to breast cancer risk.

## Methods

### Study Subjects

This study was reviewed and approved by the Human Subjects Institutional Review Board at the Mayo Clinic, and all participants provided informed consent. Full details of the ongoing clinic-based Mayo Clinic Breast Cancer Case–Control Study have been previously reported [7]. Briefly, cases included white women newly diagnosed (within 6 months of first diagnosis) with invasive cancer of the breast. Controls were selected from women visiting Mayo Clinic for a pre-scheduled full medical exam in the Department of Internal Medicine and were frequency matched to cases on region of residence, race, and 5-year age group. Case participation was 69%, and control participation was 71%. Eligible women were asked to provide risk factor information via a self-administered questionnaire, and a sample of blood as a source of DNA. This analysis is based on 798 cases and 843 controls enrolled from February 1, 2001 through June 30, 2005. Estrogen receptor (ER) status and HER2 status of tumors was available for 788 (99%) and 498 (62%) of cases, respectively. Progesterone receptor data were also available but were strongly correlated with ER status and were not included here.

### Candidate Gene and SNP Selection

Candidate genes were compiled primarily based on the protein profiling work of the human interphase centrosome by Andersen and colleagues [8]. Only genes encoding true centrosome proteins as characterized in this article were selected for this pathway-based analysis. By searching NCBI databases of Entrez Gene information and PubMed abstracts, the list was further expanded to a total of 101 genes to include newly identified centrosome components during all phases of the cell cycle. The SNP selection process has been previously described [7]. Briefly, candidate functional SNPs and SNPs in the genomic region from 5 kb upstream to 5 kb downstream of the largest cDNA isoform (NCBI35) of each gene with MAF > 0.05 in Caucasian populations were selected from publicly available databases. TagSNPs representing SNPs with pair-wise correlation of $r^2 \geq 0.8$ were chosen by ldSelect [9] (See Supplemental Table 1). A total of 66 centrosomal genes had gene coverage (the proportion of the SNP variability accounted for by the tagSNPs) between 90 and 100%, and another 28 had coverage between 50 and 89%. *CSNK1E* had 25% gene coverage. Six genes included non-synonymous protein coding SNPs only (Supplemental Table 2).

### Genotyping

A total of 1,741 samples (798 cases, 843 controls, 100 duplicates) were assayed on an Illumina GoldenGate genotyping platform as previously described [10,11]. Only samples and SNPs with call rates >95% were included in analyses. Concordance between 100 duplicate samples was >99.99%.

### Statistical Analysis

Allele frequencies were estimated from both cases and controls, and departures from Hardy–Weinberg equilibrium among controls were assessed using standard goodness-of-fit tests or exact tests [12]. Primary tests of individual SNP associations with breast cancer status were performed using unconditional logistic regression assuming an ordinal (log-additive) genotypic relationship. All models were adjusted for age and region of residence; multivariate models also included age at menarche, oral contraceptive use, age at first childbirth, pack-years of cigarettes smoked, HRT use, and menopausal status. Exploratory analyses were also conducted in subgroups of women defined by histological subtype of tumors based on ER and HER2 status, as reported in pathology records.

Estimates of pair-wise linkage disequilibrium, both D′ and $r^2$, were obtained using genotype data from the controls. We determined haplotype blocks within and across genes using the method of Gabriel et al. [13]. Overall differences in breast cancer risk among gene-specific haplotypes (with estimated frequencies greater than 0.01) were assessed using the global score test in the Haplo.stat software [14].

Bonferroni correction was conducted by multiplying the estimated *P*-values by the number of SNPs in the analysis ($N = 782$).

All analyses described above were specified *a priori*. We also conducted exploratory analyses to identify potential combinations of SNPs that might contribute jointly to the risk of breast cancer. We explored interactions between all possible pairs of SNPs by including in logistic regression models the genotype count variables (i.e., 0, 1, 2 copies of minor allele) for each pair of SNPs, along with the product of these two count variables. The significance of each multiplicative interaction was assessed using a likelihood ratio test. To further test the simultaneous significance of all SNPs in the centrosomal genes, we ran 500 permutations of GLOSSI (Gene-loci Set Analysis), an algorithm designed to determine if the distribution of *P*-values in a pathway deviates from what is expected when no significant associations are present [15]. Subsequent to this assessment, two stepwise logistic regression procedures were executed, with *P*-value thresholds of 0.05 and 0.01 for inclusion, to identify the SNPs most likely to explain associations suggested by the gene-set analyses.

## Results

### Individual SNPs and haplotypes associated with breast cancer risk

The primary purpose of this analysis was to look for evidence of associations between 782 predominantly tagSNPs ($r^2 > 0.8$) in 101 candidate genes encoding proteins implicated in the structure and/or function of the centrosome and risk of breast cancer. Forty-eight SNPs from 29 genes (out of 782 SNPs examined) showed evidence of significant associations with breast cancer risk in our population in the log-additive model (*P*-trend < 0.05) (Table 1), whereas 11 SNPs exhibited significant associations at $P < 0.01$ (Table 1). These results suggest a slight enrichment for SNPs associated with breast cancer in the centrosome pathway. One SNP, rs1374468 SNP in *TACC3*, displayed the most significant association with risk in the overall analyses (*P*-trend = 0.001) (Table 1). Two SNPs from each of seven genes (*JUB*, *CHUK*, *MCPH1*, *NEK7*, *PAK1*, *PIK3CB*, and *GPSM2*) were significantly

associated with risk of breast cancer (*P*-trend < 0.05; Table 1). Three genes, *AXIN2, NIN, NUMA1*, had four or more SNPs that were significantly associated with breast cancer risk (*P*-trend < 0.05; Table 1). All of the SNPs within each of these four genes were in strong linkage disequilibrium ($r^2 > 0.6$). All *P*-values for the above associations exceeded 0.05 after Bonferroni adjustment for multiple testing.

Recent studies have shown substantial differences in the strength and significance of associations between established genetic risk factors from GWAS and histological subtypes of breast cancer defined by estrogen receptor status [17]. To explore the effect of disease heterogeneity on our results, we stratified our cases based on the expression of estrogen receptor (ER-positive) and the absence of HER2 (HER2-negative). We did not conduct analyses for the ER-negative or HER2-positive subgroups because of limited sample size within these groups. Twenty-seven SNPs from 18 genes that had not been identified when examined among all breast cancers had *P*-values less than 0.05 when cases were restricted to the ER-positive and/or HER2-negative tumors (Table 2). In HER2-negative cases, rs6693750 in the *RAPGAP1L* locus displayed a substantially strengthened inverse association with risk (OR = 0.70, 95% CI 0.51–0.97, *P*-trend = 0.03) when compared to the overall study population (OR = 0.90, 95% CI 0.70–1.15, *P*-trend = 0.39). Likewise, rs153867 in *KIF2A* and rs3804443 in *SKP2* also exhibited more extreme associations with risk (>20% change) when restricting to HER2-negative cases (Table 2). Among ER-positive cases, the association of rs3013512 in *NUF2* with risk of breast cancer became strengthened (OR = 0.64, 95% CI 0.42–0.96, *P*-trend = 0.03) compared to the overall population (OR = 0.80, 95% CI 0.51–1.25, *P*-trend = 0.33).

We also examined haplotype associations with risk of breast cancer (Supplemental Table 3). When considering all haplotype blocks from the candidate loci, a total of 20 haplotypes displayed significant associations ($P < 0.05$) with breast cancer risk. This included specific haplotypes from 13 genes in which specific SNPs also displayed associations with risk (Table 1 and Supplementary Table 3). In particular, a specific haplotype in the *TACC3* locus was highly significantly associated with risk ($P = 0.0008$), as was the global haplotype accounting for all haplotypes in this gene ($P < 0.02$) (Supplementary Table 3). Specific haplotypes in six genes (*YWHAE, CDC16, CKAP5, KIF2A, NEK9, NPM2*) were associated with risk of breast cancer, although none of the individual typed SNPs in these genes reached significance.

## Multi-SNP and Pathway Assessments

We assessed breast cancer risk associations with multiplicative interactions for all pairs of SNPs. The two SNP combination with the greatest significance was that of rs10145182 in intron 2 of *NIN* and rs2134808 located between *TUBG1* (Gamma Tubulin 1) and *TUBG2* (Gamma Tubulin 2), *P*-interaction = 0.00001 (data not shown), both of which also independently display significant associations with risk. We also conducted a gene-set analysis using GLOSSI [15] to evaluate the simultaneous significance of all SNPs in the centrosome pathway and risk of breast cancer. The highly significant result ($P = 4.76 \times 10^{-50}$) obtained with this method suggested that the centrosome pathway is enriched for SNPs that are associated with breast cancer risk. In an effort to identify the SNPs in this pathway most likely simultaneously associated with breast cancer, we conducted stepwise logistic regression analyses. Forty SNPs were identified when the threshold for SNPs to enter and remain in the model was set at $P < 0.05$ (Supplementary Table 4). Seventeen of these SNPs were not individually associated with breast cancer risk. However, only five SNPs from five genes (*GPSM2, TACC3, CDC25C, NIN, AXIN2*) remained in the model when a threshold of $P < 0.01$ was used (Table 3). Each of these five SNPs individually displayed associations with breast cancer (Table 1).

## Discussion

Centrosome abnormalities are a common feature of breast cancers [18,19] that have also been detected in premalignant lesions in the mammary gland. It has long been postulated that centrosome amplification may result in multipolar mitoses, unequal segregation of chromosomes, and aneuploidy, but clear evidence in support of this model has been absent. However, recent studies have shown that centrosome amplification can lead to inappropriate merotelic attachment of spindle fibers nucleating from multiple spindle poles to kinetochores, resulting in aberrant chromosome segregation and aneuploidy [20], which is itself a hallmark of cancer. In addition, recent studies in animal models have established that aberrant expression of mitotic checkpoint proteins leading to aneuploidy can enhance tumor formation [21], suggesting that aneuploidy has a direct role in cancer. Moreover, the centrosome functions as a licensing body for the progression of the cell cycle from G1 to S-phase and the G2 to M phase, suggesting that disruption of centrosome signaling can influence cellular proliferation. Based on these findings, we proposed that inherited genetic variation in genes involved in centrosome structure and function may contribute to the development of breast cancer.

In our studies, we observed that risk of breast cancer in the Mayo Clinic population was associated with individual SNPs in 29 genes ($P < 0.05$). Consistent with recent evidence suggesting that known genetic risk factors for breast cancer identified through genome-wide association studies often display specific associations with subtypes of breast cancer [17], we found that some genetic variants were more strongly associated with risk of a particular pathologic subtype of breast cancer. Eight SNPs in six genes that were not associated with risk of overall disease displayed significant associations when cases were restricted to those with ER-positive disease. Similarly 17 SNPs in 12 genes were only associated with risk when analyses were restricted to HER2-negative cases. The results of these exploratory analyses need further investigation in independent data sets.

We further evaluated associations between haplotypes in the candidate genes and breast cancer and identified specific haplotypes exhibiting significant associations in 20 genes, 13 of which harbored individual SNPs that displayed significant associations. In addition, specific haplotypes in six genes were associated with risk of breast cancer despite the fact that none of the individual SNPs in these genes displayed significant associations with breast cancer risk. Overall, we noted that a single SNP, a haplotype block, and the global haplotypes within the *TACC3* locus displayed the most significant associations with breast cancer risk in each of these categories. TACC3 is localized to the centrosome in an Aurora A dependent manner and has been implicated in regulating the stability of microtubules in the mitotic spindle [22]. Mislocalization of TACC3 from the centrosome or reduced levels of TACC3 are associated with chromosome congression and segregation defects and onset of aneuploidy. Similarly, a single SNP and a haplotype block in the *PINS* locus, which is also involved in microtubule formation, displayed further associations with risk. Despite a lack of significance when adjusting for multiple testing, these consistent associations are interesting and warrant follow up in other populations.

We also assessed the possibility of interactions between SNPs and found that rs10145182 in intron 2 of *NIN* and rs2134808 located between *TUBG1* (Gamma Tubulin 1) and *TUBG2* (Gamma Tubulin 2), showed the most significant evidence of interaction (*P*-interaction = 0.00001). This is particularly interesting because functional studies have shown that ninein is important for positioning and anchoring the ends of the microtubules in epithelial cells [23] and that ninein binds to gamma-tubulin. Elevated levels of ninein cause mislocalization of gamma-tubulin, recruiting it to ectopic (non-centrosomal) ninein-containing sites which are not active in nucleating microtubules during mitosis [24]. This can result in failure to

fully develop mitotic spindles leading to mitotic checkpoint arrest and/or chromosome segregation defects. Importantly, individual SNPs and a haplotype in *NIN* also displayed some of the most significant associations with breast cancer in this study.

A multi-SNP gene-set analysis [15] of all SNPs in the centrosome structure and function pathway strongly suggested that a collection of SNPs in the pathway were associated with risk of breast cancer. To identify those most likely to be playing a part, we ran a stepwise logistic regression model incorporating our standard adjustment factors and allowing the model to select the best combination of SNPs. The results (presented in Table 3 and Supplemental Table 4) yielded intriguing evidence that as many as 40 SNPs in the centrosomal pathway could be associated with breast cancer risk. A more stringent analysis identified a combination of five SNPs, including the four SNPs showing the most significant individual associations, which was highly significantly associated with breast cancer. These findings suggest that combinations of weakly associated genetic variants in the centrosome pathway, rather than individual highly significantly associated SNPs, may account for a putative role for the centrosome pathway in predisposition to breast cancer. These results also warrant replication in other studies.

Several limitations should be considered when interpreting our results. One hundred and one genes were included in this analysis, each with numerous SNPs. When statistical significance was adjusted for the many statistical tests conducted, none of these associations remained statistically significant. In addition, none of these SNPs were significant within the Cancer Genetic Markers of Susceptibility (CGEMS) [25] data. However, the strong biological rationale by which these genes were selected strengthens the evidence in support of a real biological connection between these genes and development of breast cancer. Another limitation is the ethnic makeup of our population, which was 100% Caucasian from the upper Midwest portion of the United States. Although this may reduce generalizability, the homogeneous nature of our population limits the effects of population stratification on the association with risk.

In summary, in this first epidemiological study to focus on the centrosome structure and function pathway, we examined both individual and multi-SNP associations between genetic variation in genes known to be related with structure and/or function of the centrosome and breast cancer risk in a breast cancer case control study at the Mayo Clinic. Of the 101 genes evaluated, several had interesting associations with risk of breast cancer in our population. In addition, our multi-SNP analyses suggested that many SNPs in this pathway may need to be examined simultaneously in order to truly understand the relevance of genetic variation in this pathway on risk of breast cancer. This opens up a new area for investigation that is worthy of follow-up in other populations and in other cancer types.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

This study was funded in part by a grant 5R01CA122340-02 and a Breast Cancer Specialized Program of Research Excellence (SPORE) grant P50CA166201 from the National Cancer Institute.

## References

1. Sluder G. Two-way traffic: centrosomes and the cell cycle. Nat Rev Mol Cell Biol 2005;6(9):743–748. [PubMed: 16231423]

2. Lange BM, Gull K. Structure and function of the centriole in animal cells: progress and questions. Trends Cell Biol 1996;6(9):348–352. [PubMed: 15157432]
3. Fukasawa K. Oncogenes and tumour suppressors take on centrosomes. Nat Rev Cancer 2007;7(12): 911–924. [PubMed: 18004399]
4. D'Assoro AB, Lingle WL, Salisbury JL. Centrosome amplification and the development of cancer. Oncogene 2002;21(40):6146–6153. [PubMed: 12214243]
5. Brinkley BR. Managing the centrosome numbers game: from chaos to stability in cancer cell division. Trends Cell Biol 2001;11(1):18–21. [PubMed: 11146294]
6. Marx J. Cell biology. Do centrosome abnormalities lead to cancer? Science 2001;292(5516):426–429. [PubMed: 11330289]
7. Wang X, et al. Association of genetic variation in genes implicated in the beta-catenin destruction complex with risk of breast cancer. Cancer Epidemiol Biomarkers Prev 2008;17(8):2101–2108. [PubMed: 18708403]
8. Andersen JS, et al. Proteomic characterization of the human centrosome by protein correlation profiling. Nature 2003;426(6966):570–574. [PubMed: 14654843]
9. Carlson CS, et al. Selecting a maximally informative set of single-nucleotide polymorphisms for association analyses using linkage disequilibrium. Am J Hum Genet 2004;74(1):106–120. [PubMed: 14681826]
10. Oliphant A, et al. BeadArray technology: enabling an accurate, cost-effective approach to high-throughput genotyping. Biotechniques 2002 Suppl:56–58. 60–61. [PubMed: 12083399]
11. Steemers FJ, Gunderson KL. Illumina, Inc. Pharmacogenomics 2005;6(7):777–782. [PubMed: 16207153]
12. Weir, BS. Genetic data analysis II: methods for discrete population genetic data. Sunderland MA: Sinauer Associates, Inc.; 1996. p. 98-99.
13. Gabriel SB, et al. The structure of haplotype blocks in the human genome. Science 2002;296(5576):2225–2229. [PubMed: 12029063]
14. Schaid DJ, et al. Score tests for association between traits and haplotypes when linkage phase is ambiguous. Am J Hum Genet 2002;70(2):425–434. [PubMed: 11791212]
15. Chai HS, et al. GLOSSI: a method to assess the association of genetic loci-sets with complex diseases. BMC Bioinformatics 2009;10:102. [PubMed: 19344520]
16. Kelemen LE, et al. Genetic variation in the chromosome 17q23 amplicon and breast cancer risk. Cancer Epidemiol Biomarkers Prev 2009;18(6):1864–1868. [PubMed: 19454617]
17. Garcia-Closas M, et al. Heterogeneity of breast cancer associations with five susceptibility loci by clinical and pathological characteristics. PLoS Genet 2008;4(4):e1000054. [PubMed: 18437204]
18. Guo HQ, et al. Analysis of the cellular centrosome in fine-needle aspirations of the breast. Breast Cancer Res 2007;9(4):R48. [PubMed: 17662154]
19. Schneeweiss A, et al. Centrosomal aberrations in primary invasive breast cancer are associated with nodal status and hormone receptor expression. Int J Cancer 2003;107(3):346–352. [PubMed: 14506732]
20. Ganem NJ, Godinho SA, Pellman D. A mechanism linking extra centrosomes to chromosomal instability. Nature 2009;460(7252):278–282. [PubMed: 19506557]
21. Baker DJ, et al. Whole chromosome instability caused by Bub1 insufficiency drives tumorigenesis through tumor suppressor gene loss of heterozygosity. Cancer Cell 2009;16(6):475–486. [PubMed: 19962666]
22. LeRoy PJ, et al. Localization of human TACC3 to mitotic spindles is mediated by phosphorylation on Ser558 by Aurora A: a novel pharmacodynamic method for measuring Aurora A activity. Cancer Res 2007;67(11):5362–5370. [PubMed: 17545617]
23. Delgehyr N, Sillibourne J, Bornens M. Microtubule nucleation and anchoring at the centrosome are independent processes linked by ninein function. J Cell Sci 2005;118(Pt 8):1565–1575. [PubMed: 15784680]
24. Stillwell EE, Zhou J, Joshi HC. Human ninein is a centrosomal autoantigen recognized by CREST patient sera and plays a regulatory role in microtubule nucleation. Cell Cycle 2004;3(7):923–930. [PubMed: 15190203]

25. Thomas G, 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]

**Table 1**

Multivariate adjusted odds ratios and 95% confidence intervals on all variants with a p-value less than 0.05 in the 2 degree of freedom test or ordinal (1 df). Mayo Clinic Breast Cancer Case–Control Study, Rochester, MN

| Chm | Chromosome position (bp) | Gene name | Entrez Gene ID | SNP ID | Location | Minor allele frequency | | Multivariate-adjusted[a] OR (95% Confidence Interval) | *P*-value log-additive model |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | | Cases | Controls | Log-additive model OR (95% CI) | |
| 4 | 1691810 | TACC3 | 10460 | rs1374468 | tag-SNP | 0.1679 | 0.2114 | 0.74 (0.61, 0.89) | 0.00127 |
| 1 | 109128267 | GPSM2 | 29899 | rs12090453 | tag-SNP | 0.3452 | 0.39905 | 0.79 (0.68, 0.92) | 0.00186 |
| 5 | 137653040 | CDC25C | 995 | rs11567998 | tag-SNP | 0.0082 | 0.02017 | 0.36 (0.18, 0.69) | 0.00237 |
| 14 | 50340017 | NIN | 51199 | rs9788504 | tag-SNP | 0.4228 | 0.36817 | 1.25 (1.08, 1.44) | 0.00259 |
| 14 | 50345759 | NIN | 51199 | rs10145182 | tag-SNP | 0.3758 | 0.43535 | 0.81 (0.70, 0.93) | 0.00334 |
| 14 | 50350422 | NIN | 51199 | rs7153720 | tag-SNP | 0.5226 | 0.46318 | 1.23 (1.07, 1.42) | 0.00392 |
| 1 | 194900676 | NEK7 | 140609 | rs2884765 | tag-SNP | 0.0714 | 0.09668 | 0.70 (0.54, 0.91) | 0.00687 |
| 1 | 195024460 | NEK7 | 140609 | rs12403821 | tag-SNP | 0.0763 | 0.10369 | 0.71 (0.55, 0.91) | 0.00729 |
| 8 | 6488159 | MCPH1 | 79648 | rs2433149 | 3utr | 0.3302 | 0.28537 | 1.23 (1.06, 1.44) | 0.00786 |
| 14 | 50343524 | NIN | 51199 | rs6650505 | tag-SNP | 0.4956 | 0.44187 | 1.21 (1.05, 1.39) | 0.00854 |
| 8 | 6487952 | MCPH1 | 79648 | rs1057091 | non_synon | 0.3264 | 0.28325 | 1.23 (1.05, 1.44) | 0.00875 |
| 17 | 60979143 | AXIN2 | 8313 | rs11079571 | tag-SNP | 0.1861 | 0.15006 | 1.28 (1.05, 1.54) | 0.01229 |
| 10 | 101953289 | CHUK | 1147 | CHUK-028086 | tag-SNP | 0.4718 | 0.51485 | 0.83 (0.72, 0.96) | 0.01279 |
| 17 | 54799891 | YPEL2 | 388403 | rs16943468[b] | tag-SNP | 0.0783 | 0.05991 | 1.42 (1.07, 1.89) | 0.01476 |
| 8 | 31144196 | WRN | 7486 | rs1346044 | non_synon | 0.2863 | 0.25297 | 1.21 (1.03, 1.42) | 0.01941 |
| 6 | 52255268 | MCM3 | 4172 | rs7774976 | tag-SNP | 0.1159 | 0.09739 | 1.32 (1.04, 1.67) | 0.01993 |
| 5 | 36193087 | SKP2 | 6502 | rs4440390 | tag-SNP | 0.1485 | 0.121 | 1.28 (1.04, 1.57) | 0.02053 |
| 15 | 41488814 | TUBGCP4 | 27229 | rs17725343 | tag-SNP | 0.0389 | 0.02435 | 1.65 (1.08, 2.52) | 0.02093 |
| 22 | 20410927 | YPEL1 | 29799 | rs4821217 | tag-SNP | 0.057 | 0.03686 | 1.50 (1.06, 2.12) | 0.02306 |
| 14 | 50335203 | NIN | 51199 | rs12893300 | tag-SNP | 0.1372 | 0.11032 | 1.29 (1.04, 1.60) | 0.02311 |
| 3 | 180432999 | PIK3CA | 5290 | rs1607237 | tag-SNP | 0.3813 | 0.42221 | 0.85 (0.73, 0.98) | 0.02346 |
| 3 | 139909126 | PIK3CB | 5291 | rs361084 | tag-SNP | 0.4323 | 0.47805 | 0.85 (0.74, 0.98) | 0.02556 |
| 11 | 71425076 | NUMA1 | 4926 | rs1573502 | 5upstream | 0.0639 | 0.04626 | 1.43 (1.04, 1.96) | 0.02574 |
| 11 | 71425429 | NUMA1 | 4926 | rs7127865 | 5upstream | 0.0639 | 0.04626 | 1.43 (1.04, 1.96) | 0.02574 |
| 3 | 139961242 | PIK3CB | 5291 | rs361072 | 5upstream | 0.4298 | 0.47565 | 0.85 (0.74, 0.98) | 0.02597 |

| Chm | Chromosome position (bp) | Gene name | Entrez Gene ID | SNP ID | Location | Minor allele frequency | | Multivariate-adjusted[a] OR (95% Confidence Interval) | *P*-value log-additive model |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | | Cases | Controls | Log-additive model OR (95% CI) | |
| 17 | 60979950 | AXIN2 | 8313 | rs3923086 | tag-SNP | 0.4536 | 0.4139 | 1.18 (1.02, 1.36) | 0.02663 |
| 10 | 101967873 | CHUK | 1147 | rs7903344 | non_synon | 0.5094 | 0.47331 | 1.18 (1.02, 1.36) | 0.02695 |
| 17 | 60971959 | AXIN2 | 8313 | rs4791171[c] | tag-SNP | 0.312 | 0.27936 | 1.19 (1.02, 1.39) | 0.02789 |
| 14 | 22510244 | JUB | 84962 | rs2180834 | tag-SNP | 0.2293 | 0.19988 | 1.22 (1.02, 1.45) | 0.0282 |
| 17 | 38035609 | TUBG2 | 27175 | rs2134808 | tag-SNP | 0.2274 | 0.26868 | 0.83 (0.71, 0.98) | 0.02949 |
| 11 | 71425794 | NUMA1 | 4926 | rs4945426 | 5upstream | 0.0634 | 0.04632 | 1.41 (1.03, 1.94) | 0.03068 |
| 11 | 71386920 | NUMA1 | 4926 | rs3814721 | tag-SNP | 0.064 | 0.04686 | 1.41 (1.03, 1.92) | 0.03188 |
| 1 | 171201562 | RAPGAP1L | 9910 | rs12071794 | tag-SNP | 0.2802 | 0.31384 | 0.85 (0.74, 0.99) | 0.03317 |
| 14 | 22517181 | JUB | 84962 | rs6572891 | tag-SNP | 0.4147 | 0.3796 | 1.17 (1.01, 1.35) | 0.03361 |
| 17 | 60959412 | AXIN2 | 8313 | rs7210356 | tag-SNP | 0.1191 | 0.10154 | 1.28 (1.02, 1.60) | 0.03502 |
| 17 | 42616954 | CDC27 | 996 | rs16941635 | tag-SNP | 0.0802 | 0.1038 | 0.77 (0.60, 0.98) | 0.03536 |
| 20 | 33542614 | CEP250 | 11190 | rs224373 | tag-SNP | 0.1598 | 0.19039 | 0.82 (0.68, 0.99) | 0.03645 |
| 10 | 102760072 | LZTS2 | 84445 | rs807023 | tag-SNP | 0.1299 | 0.15065 | 0.8 (0.65, 0.99) | 0.03698 |
| 14 | 50294408 | NIN | 51199 | rs2073348 | tag-SNP | 0.2901 | 0.31807 | 0.85 (0.72, 0.99) | 0.03709 |
| 14 | 101617977 | HSP90AA1 | 3320 | rs2298877 | tag-SNP | 0.1773 | 0.15065 | 1.23 (1.01, 1.49) | 0.03752 |
| 20 | 29787706 | TPX2 | 22974 | rs6089055 | tag-SNP | 0.2068 | 0.23428 | 0.83 (0.70, 0.99) | 0.03769 |
| 11 | 76711347 | PAK1 | 5058 | rs2729762 | tag-SNP | 0.3127 | 0.34875 | 0.85 (0.73, 0.99) | 0.03818 |
| 1 | 109182227 | GPSM2 | 29899 | rs839859 | tag-SNP | 0.3062 | 0.27106 | 1.18 (1.01, 1.38) | 0.03869 |
| 20 | 30906597 | MAPRE1 | 22919 | rs17297652 | tag-SNP | 0.0345 | 0.04804 | 0.68 (0.47, 0.98) | 0.0392 |
| 1 | 100682367 | CDC14A | 8556 | rs10875295 | tag-SNP | 0.4906 | 0.44537 | 1.16 (1.01, 1.34) | 0.04033 |
| 17 | 60979723 | AXIN2 | 8313 | rs3923087 | tag-SNP | 0.2422 | 0.2114 | 1.19 (1.00, 1.41) | 0.04378 |
| 16 | 56337904 | KATNB1 | 10300 | rs12708992 | tag-SNP | 0.0846 | 0.06762 | 1.32 (1.00, 1.74) | 0.04719 |
| 11 | 76776127 | PAK1 | 5058 | rs2729747 | tag-SNP | 0.312 | 0.34679 | 0.86 (0.74, 1.00) | 0.04758 |

[a] Adjusted for age, region of residence, age at menarche, oral contraceptive use, age at first childbirth, pack-years of cigarettes smoked, HRT use, and menopausal status

[b] Previously reported by Kelemen et al. [16]

[c] Previously reported by Wang et al. [7]

**Table 2**

Associations between SNPs and breast tumor subtypes

| CHROM | Gene | SNP | All tumors | | Her-2/Neu negative tumors | | ER-positive tumors | |
|---|---|---|---|---|---|---|---|---|
| | | | Log-Additive OR (95% CI) (798 cases) | *P*-value | Multivariate adjusted[a] log-additive OR (95% CI) (419 cases) | *P*-value | Multivariate adjusted[a] log-additive OR (95% CI) (663 cases) | *P*-value |
| Group 1[b] | | | | | | | | |
| 1 | RAPGAP1L | rs6693750 | 0.90 (0.70, 1.15) | 0.39 | **0.70 (0.51, 0.97)** | **0.031** | 0.89 (0.69, 1.14) | 0.35 |
| 1 | NEK7 | rs12058769 | 1.11 (0.94, 1.30) | 0.22 | **1.25 (1.04, 1.52)** | **0.020** | 1.10 (0.93, 1.30) | 0.24 |
| 1 | NEK7 | rs716784 | 1.17 (0.95, 1.43) | 0.14 | **1.31 (1.04, 1.67)** | **0.024** | 1.13 (0.91, 1.39) | 0.26 |
| 3 | GSK3B | rs1154597 | 1.23 (0.94, 1.60) | 0.13 | **1.37 (1.01, 1.87)** | **0.045** | 1.29 (0.99, 1.69) | 0.06 |
| 3 | GSK3B | rs17811013 | 0.83 (0.65, 1.07) | 0.16 | **0.72 (0.53, 0.99)** | **0.043** | 0.90 (0.70, 1.16) | 0.40 |
| 5 | KIF2A | rs153867 | 1.37 (0.98, 1.92) | 0.07 | **1.73 (1.19, 2.53)** | **0.004** | 1.37 (0.97, 1.93) | 0.07 |
| 5 | SKP2 | rs33678 | 1.21 (0.97, 1.50) | 0.10 | **1.30 (1.00, 1.67)** | **0.048** | 1.20 (0.96, 1.51) | 0.10 |
| 7 | YWHAG | rs11763069 | 1.37 (0.94, 2.01) | 0.10 | **1.64 (1.06, 2.54)** | **0.026** | 1.42 (0.97, 2.08) | 0.07 |
| 10 | LZTS2 | rs11190790 | 0.87 (0.75, 1.01) | 0.06 | **0.81 (0.68, 0.97)** | **0.020** | 0.86 (0.74, 1.00) | 0.05 |
| 14 | HSP90AA1 | rs7155973 | 1.20 (0.91, 1.57) | 0.19 | **1.38 (1.02, 1.86)** | **0.038** | 1.17 (0.89, 1.53) | 0.26 |
| 14 | NIN | rs1004832 | 1.23 (0.97, 1.56) | 0.08 | **1.42 (1.08, 1.86)** | **0.012** | 1.19 (0.93, 1.52) | 0.16 |
| 14 | NIN | rs6572697 | 1.15 (0.96, 1.37) | 0.13 | **1.24 (1.01, 1.53)** | **0.038** | 1.17 (0.97, 1.40) | 0.09 |
| 17 | CDC27 | rs11570579 | 0.85 (0.67, 1.06) | 0.15 | **0.74 (0.56, 0.98)** | **0.034** | 0.83 (0.66, 1.05) | 0.12 |
| 17 | CDC27 | rs701982 | 1.10 (0.95, 1.27) | 0.19 | **1.19 (1.00, 1.41)** | **0.045** | 1.07 (0.92, 1.24) | 0.36 |
| 17 | CDC27 | rs764792 | 1.09 (0.94, 1.26) | 0.26 | **1.25 (1.05, 1.48)** | **0.014** | 1.06 (0.91, 1.23) | 0.44 |
| 17 | TUBD1 | rs12150500 | 1.12 (0.93, 1.36) | 0.23 | **1.29 (1.04, 1.61)** | **0.021** | 1.13 (0.93, 1.37) | 0.21 |
| 19 | DNM2 | rs11666111 | 0.77 (0.57, 1.05) | 0.09 | **0.67 (0.46, 0.96)** | **0.032** | 0.78 (0.58, 1.06) | 0.11 |
| Group 2[c] | | | | | | | | |
| 1 | CDC14A | rs12096135 | 0.86 (0.71, 1.04) | 0.11 | 0.91 (0.73, 1.14) | 0.42 | **0.82 (0.68, 1.00)** | **0.049** |
| 1 | NUF2 | rs3013512 | 0.78 (0.53, 1.13) | 0.18 | 0.80 (0.51, 1.25) | 0.33 | **0.64 (0.43, 0.96)** | **0.032** |
| 5 | SKP2 | rs7715070 | 0.75 (0.54, 1.05) | 0.09 | 0.66 (0.44, 1.00) | 0.05 | **0.68 (0.48, 0.97)** | **0.031** |
| 6 | MCM3 | rs3765447 | 1.29 (0.98, 1.71) | 0.07 | 1.31 (0.95, 1.80) | 0.10 | **1.33 (1.00, 1.76)** | **0.046** |
| 17 | TUBG1 | rs2089114 | 0.88 (0.76, 1.01) | 0.06 | 0.87 (0.73, 1.02) | 0.09 | **0.81 (0.70, 0.94)** | **0.004** |
| 17 | TUBG1 | rs9911799 | 0.88 (0.76, 1.01) | 0.07 | 0.87 (0.73, 1.02) | 0.09 | **0.81 (0.70, 0.93)** | **0.004** |

| CHROM | Gene | SNP | All tumors | | Her-2/Neu negative tumors | | ER-positive tumors | |
|---|---|---|---|---|---|---|---|---|
| | | | Log-Additive OR (95% CI) (798 cases) | *P*-value | Multivariate adjusted[a] log-additive OR (95% CI) (419 cases) | *P*-value | Multivariate adjusted[a] log-additive OR (95% CI) (663 cases) | *P*-value |
| 22 | YPEL1 | rs8135758 | 0.85 (0.70, 1.03) | 0.10 | 0.94 (0.75, 1.17) | 0.57 | **0.78 (0.64, 0.95)** | **0.015** |
| 22 | YPEL1 | rs861818 | 0.84 (0.69, 1.03) | 0.10 | 0.85 (0.67, 1.08) | 0.18 | **0.75 (0.61, 0.93)** | **0.008** |
| Group 3[d] | | | | | | | | |
| 3 | PIK3CB | rs12493155 | 1.15 (1.00, 1.33) | 0.06 | **1.29 (1.09, 1.52)** | **0.004** | **1.22 (1.06, 1.42)** | **0.007** |
| 5 | SKP2 | rs3804443 | 0.84 (0.65, 1.08) | 0.17 | **0.69 (0.50, 0.95)** | **0.02** | **0.75 (0.57, 0.97)** | **0.029** |

[a] Adjusted for age, region of residence, age at menarche, oral contraceptive use, age at first childbirth, pack-years of cigarettes smoked, HRT use, and menopausal status

[b] SNPs in this group were significant only within HER-2/Neu Negative pathological subtype

[c] SNPs in this group were significant only within ER-positive pathological subtype

[d] SNPs in this group were significant within ER-positive and HER-2/Neu negative pathological subtype

**Table 3**

Stepwise Regression Analysis results using $P < 0.01$ criterion for SNP inclusion and retention in the final regression model. Results are multivariate adjusted

| Chm | Gene | Geneid | SNP | MAF | Multivariate adjusted[a] log-additive OR (95% CI) | Change |
|---|---|---|---|---|---|---|
| 1 | GPSM2 | 29899 | rs12090453 | 0.35 | 0.77 (0.66, 0.89) | A–>G |
| 4 | TACC3 | 10460 | rs1374468 | 0.12 | 0.74 (0.62, 0.90) | G–>A |
| 5 | CDC25C | 995 | rs11567998 | 0.06 | 0.34 (0.18, 0.67) | C–>G |
| 14 | NIN | 51199 | rs9788504 | 0.45 | 1.28 (1.10, 1.48) | C–>G |
| 17 | AXIN2 | 8313 | rs11079571 | 0.11 | 1.29 (1.06, 1.56) | G–>A |

[a]Adjusted for age, region of residence, age at menarche, oral contraceptive use, age at first childbirth, pack-years of cigarettes smoked, HRT use, and menopausal status