AD_____

Award Number: W81XWH-05-1-0204

TITLE: Identification, Characterisation and Clinical Development of the New General of Breast Cancer Susceptibility Alleles

PRINCIPAL INVESTIGATOR: Nazneen Rahman, M.D., Ph.D.

CONTRACTING ORGANIZATION: The Institute of Cancer Research
London SW7 3RP; United Kingdom

REPORT DATE: March 2010

TYPE OF REPORT: Annual

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
31-03-2010	Annual	1 MAR 2009- 28 FEB 2010

4. TITLE AND SUBTITLE
Identification, characterisation and clinical development of the new generation of breast cancer susceptibility alleles

5a. CONTRACT NUMBER

5b. GRANT NUMBER
W81XWH-05-1-0204

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Nazneen Rahman M.D. Ph.D.

Email: nazneen.rahman@icr.ac.uk

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

The Institute of Cancer Research
123 Old Brompton Road
London SW7 3RP, United Kingdom

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT
There is considerable evidence that genetic factors play an important role in causing breast cancer, but the genes involved in the majority of breast cancers are currently unknown. Our aim is to identify genetic factors that increase the risk of breast cancer occurring. We have collected samples and clinical information from over 4000 breast cancer families. We compare the frequency of genetic factors in these cases with control individuals. Over the last year we have been engaged in two complementary strategies. 1) Undertaking genome-wide association analyses to identify common, low-penetrance variants that increase breast cancer risk by a modest amount. Our collaborative endeavours in this area have already led to the identification of several variants and during the last year we completed the largest breast cancer genome-wide association study to date identifying five new breast cancer susceptibility loci. 2) Undertaking mutational analyses to identify rare, low or intermediate-penetrance variants. This previously led to our identification of 4 breast cancer genes, *CHEK2*, *ATM*, *PALB2* and *BRIP1*. In the last year we have begun exome sequencing and we aim to sequence 20 exomes in breast cancer patients and to follow-up promising genes within the final year of the grant.

15. SUBJECT TERMS
None provided.

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			USAMRMC
U	U	U	UU	20	19b. TELEPHONE NUMBER *(include area code)*

Table of Contents

Introduction

Breast cancer is a common disease in women but the causes are still largely unknown. There is considerable evidence to suggest that genetic factors play an important role in causing breast cancer. In recent years, our understanding of genetic predisposition to breast cancer has advanced significantly. Three classes of predisposition factor, categorized by their associated risks of breast cancer, are currently known (1,2) . *BRCA1* and *BRCA2* are high-penetrance breast cancer predisposition genes identified by genome-wide linkage analysis and positional cloning (3,4). Our group undertook mutational screening of genes functionally related to *BRCA1* and/or *BRCA2* which revealed four genes, *CHEK2, ATM, BRIP1* and *PALB2*, mutations of which are rare and confer an intermediate risk of breast cancer (5-8). Recent genome-wide association studies have identified common variants at twelve loci that are associated with an increased risk of breast cancer, while an additional locus, *CASP8* D302H, was identified through a candidate gene association study (9-16). However, despite these discoveries, most of the familial risk of breast cancer remains unexplained.

Our aim, therefore, remains the identification and characterisation of genetic factors that increase the risk of breast cancer occurring. We have collected clinical information and samples from over 4000 breast cancer families to facilitate these aims . Over the reporting period we have primarily been engaged in completing the largest genome-wide association study (GWAS) in breast cancer to date, which includes 3,659 cases enriched for a family history and 4,897 controls in the first phase with replication in 12,576 cases and 12,223 controls. This is now completed and we identified five novel susceptibility loci, on chromosomes 9, 10 and 11 (P=4.6x10^{-7} to P=3.2x10^{-15}). We also identified SNPs in the 6q25.1, 8q24 and *LSP1* regions that were more strongly associated with risk than those reported previously.

The other primary area that we have been engaged in over the reporting period is optimizing of next-generation sequencing analysis to allow us to expand our gene sequencing experiments to directly identify rare breast cancer susceptibility genes (which are not detectable by GWAS) from candidate to genome-wide level. In collaboration with the Wellcome Trust Sanger Institute we have completed exome resequencing in two individuals with breast cancer and we will extend this to 20 exomes in the final year of the grant. We will follow-up truncating variants with Sanger resequencing in cases and controls to identify rare, low-intermediate penetrance genes.

Body

As part of the programme of work we defined five tasks. The progress towards the tasks is outlined in detail below.

Task 1: Evaluate the contribution of BRCA1 and BRCA2 exonic deletions and duplications to breast cancer susceptibility.

This task is complete – see previous report.

Task 2. Perform familial case-control analyses of non-synonymous coding single nucleotide polymorphisms (SNPs) in DNA repair genes in familial breast cancer cases.

This task is complete – see previous report and paper (17)

Task 3. Characterise the histopathology and immunohistochemistry of familial breast cancer.

We have been primarily focussing on trying to clarify the contribution of *BRCA1* mutations to triple-negative tumors *(ER, PR* and *HER2* negative*)* as this is a contentious and clinically important area. Genetic and biological data indicate that triple-negative, basal-like tumors are a distinctive sub-phenotype of breast cancers that may have different underlying causes (18). There is a known, strong association of triple-negative tumor phenotype and *BRCA1* mutations. However, the contribution of *BRCA1* to triple-negative breast cancer in the absence of a strong family history remains unclear and is a source of considerable confusion diagnostically. We aim to address this question by screening substantial numbers of individuals with triple-negative tumors for *BRCA1* mutations. We have now collected 212 cases and we will complete the mutational analysis and submit the data for publication by the end of the grant.

Task 4. Perform genome-wide familial case-control analyses of non-synonymous coding SNPs,

This task is completed (as originally proposed) and published – see previous report and paper (17). We initially undertook a direct analysis of the known (at that time) 15,000 non-synonymous coding SNPs. Our genome-wide association study in 3,659 cases and 4,897 controls additionally allowed us to tag almost all non-synonymous coding SNPs of 5% or greater frequency (see below) and therefore allowed us to interrogate the remaining non-synonymous coding SNPs.

Task 5. Identify and characterize low-penetrance breast cancer susceptibility alleles

a) *Undertake a second-generation genome-wide association study to identify common, low-penetrance breast cancer susceptibility alleles.*

In 2007, we completed the first such study in breast cancer in collaboration with Professor Doug Easton. This was based on 400 genetically enriched breast cancer cases and 400 controls typed for over 220,000 SNPs. These SNPs were correlated with ~71% of known common SNPs, at r^2>0.5.

Putative associations were followed up in ~26,000 cases and 26,000 controls. This study provided clear evidence for five novel breast cancer susceptibility loci (9). Further studies of this initial scan have recently led to the identification of two further variants on 3p24 and 17q23 which we outlined in last year's report (13). Several other loci have been identified (10-12, 14-15). However, since the risks associated with these variants are modest (per-allele odds ratios, ORs, <1.3), they explain only a small fraction of the estimated two-fold familial relative risk of breast cancer in first degree relatives of affected women. Moreover, the GWAS conducted to date have been relatively small, and it is likely that many susceptibility variants have been missed due to lack of power.

In an attempt to identify further breast cancer loci, within this reporting period we completed a GWAS that was substantially larger than those conducted to date. We studied 3,960 cases of breast cancer from the United Kingdom, selected for a positive family history of breast cancer. We selected cases with a positive family history since, under a polygenic model of susceptibility, this is expected to increase the effect size and hence improve power (19). DNA samples from these women were genotyped using an Illumina Infinium 660k array. Case genotypes were compared with those from 5,069 controls, drawn from two UK population-based studies; 2,930 controls were drawn from the 1958 Birth Cohort (1958BC), a population based study in the UK of individuals born in one week in 1958. The remaining 2,737 controls were identified through the UK National Blood Service (NBS). The control samples were genotyped as part of the Wellcome Trust Case Control Consortium (WTCCC2; http://www.wtccc.org.uk/). Our analyses utilized 2,482 1958BC and 2,587 NBS controls for which genotype data were available at the time of analysis.

Genotypes for stage 1 cases were generated using a custom Illumina Infinium 670k array, while controls were genotyped using an Illumina Infinium 1.2M array, at the Wellcome Trust Sanger Institute. We analysed data on 594,375 SNPs that were successfully genotyped on both arrays. Genotypes for both arrays were called using the Illumnus algorithm. We utilised genotypes for which

Illumnus generated a posterior probability of >0.95. Cluster plots were inspected manually for all SNPs considered for stage 2.

After quality control exclusions, we utilised data on 582,886 SNPs in 3,659 cases and 4,897 controls. Genotype frequencies in cases and controls were compared using a 1 degree of freedom (df) Cochran-Armitage trend test. There was modest evidence for inflation in the test statistic (λ=1.12, equivalent to λ_{1000}=1.03 for a study of 1,000 cases and 1,000 controls). Adjustment for differential population structure using the first two components based on a principal components analysis of uncorrelated SNPs reduced the inflation to λ=1.06.

We observed evidence of association for all twelve of the susceptibility loci identified through previous GWAS, with the same SNP as that previously identified or a strongly correlated SNP (*P*=.02 to *P*=3.6x10^{-31}; Table 1). Seven of these reached *P*<10^{-4}, five of which have previously been evaluated in large collaborative analyses of case-control studies by the Breast Cancer Association Consortium (BCAC). The BCAC analyses involved more than 20,000 cases and 20,000 controls, providing a reliable estimate of the per-allele OR (9, 13).

For three loci (6q25.1, *LSP1* and 8q24) we identified a SNP that was more strongly associated than the SNP originally reported. The most significant SNP at 6q25.1 (rs3757318) lies ~200kb upstream of *ESR1,* in an intron of C6orf97. rs3757318 is only weakly correlated, in Europeans, with SNP rs2046210 identified as a susceptibility SNP by Zheng et al (15) in a study from Shanghai (r^2=0.088), though these SNPs are more strongly correlated in East Asians (r^2=0.48 in HapMap CHB). Both rs3757318 and rs6900157 (a surrogate for rs2046210, r^2=0.96) remain significantly associated with breast cancer in multiple logistic regression analysis (*P*=.0003 and *P*=.002, respectively). These results suggest either the presence of a single causal variant that is more strongly correlated with rs3757318 than rs2046210 in Europeans, or the presence of two causal variants. The more strongly

associated SNPs that we identified in the 8q24 and *LSP1* regions lie within the same LD blocks as the original SNP, and in each case the original SNP was not significantly associated with risk after adjusting for the new SNP. Thus, these may reflect the same underlying association, and these results should assist in narrowing the search for the causal variants. A more strongly associated variant, rs10931936, was also identified at the *CASP8* locus (P=.0014; r^2=0.74).

After eliminating SNPs in previously identified susceptibility regions, we identified 28 SNPs in 13 regions of linkage disequilibrium that were significant at P<.00001. After eliminating SNPs that were strongly correlated, we attempted to replicate these associations by genotyping 15 SNPs in a second stage involving 11,431 cases and 11,081 controls from three studies in the UK and the Netherlands. We also incorporated available data from 1,145 cases and 1,142 controls from the CGEMS study (http://cgems.cancer.gov/). As part of the Era of Hope award we genotyped 3,992 cases and 3,450 controls in our lab by 5'exonuclease assay (Taqman™) using the ABI Prism 7900HT sequence detection system. All our Taqman assays included at least two negative controls and 2-5% duplicates per plate.

Six SNPs from five regions on chromosomes 9,10 and 11 showed clear evidence of replication in stage 2 (P=.0017 or better, in the same direction as stage 1) and reached significance levels over both stages combined of P=4.6x10^{-7} to P=3.2x10^{-15}; Table 2). SNPs rs614367 and rs624797, which both showed strong evidence of association, were correlated and rs624797 showed no independent association after adjustment for rs614367. There was no evidence for departure from a log-additive model for any SNP (that is, the OR for rare homozygot es did not differ significantly from the square of the OR for heterozygotes).

SNP rs1011970 lies in a 180kb block on 9p21 that includes the *CDKN2A* and *CDKN2B* genes. These genes encode cyclin-dependent kinase inhibitors and are frequently mutated or deleted in a wide

variety of human tumours (20). Germline mutations in *CDKN2A* predispose to malignant melanoma and pancreatic cancer (21), while recent GWAS also identified rs1011970 to be associated with melanoma risk (22), and SNPs within the same region are associated with naevus density and melanoma (23), basal cell carcinoma (25), glioma (26,27), diabetes (27) and coronary heart disease (28). This is the first example of the same common variant predisposing to breast cancer and another cancer type. SNP rs10757278, which is correlated with rs1011970 (r^2=0.7), is associated with levels of expression in lymphocytes of *CDKN2A*, *CDKN2B* and a non-coding RNA in the same block, *ANR1L* (29).

SNP rs614367 on 11q13 lies in an LD block of ~166kb that contains no annotated genes. This region is frequently amplified in human tumours including br east cancers (30). Plausible genes flanking this block include: proximally - *MYEOV*, a gene overexpressed in myeloma; and distally - *CCND1*, encoding cyclin D1, a protein critical for cell-cycle c ontrol that is somatically altered in many tumour types; *ORAOV1*, a gene overexpressed in oral cancer, and three fibroblast growth factors: *FGF19*, *FGF4* and *FGF3*. FGF3 and FGF4 are oncogenic growth factors that bind distinct FGFR2 isoforms, providing a possible link with the FGFR2 susceptibility locus (31).

rs10995190 on chromosome 10 lies within intron 4 of *ZNF365* (encoding zinc finger protein 365). An amino acid substitution in this gene has been associated with uric acid nephrolithiasis (32). Recent GWAS have identified another variant within this gene, rs10995271, located 159 kb downstream of rs10995190, to be associated with Crohn's disease (33). rs2380205 lies in a 105kb block on chromosome 10 containing the genes *ANKRD16* (ankyrin repeat domain 16) and *FBXO18* (F-box protein, helicase 18). rs704010 on chromosome 10 lies in a 20kb block, 90kb upstream of *ZMIZ1* (zinc finger MIZ-type containing 1).

Based on the estimated per-allele ORs from stage 2 of our study, the newly identified loci explain approximately 1.2% of the familial risk of breast cancer, though the overall contribution may be larger, since the true causal variants may be more strongly associated with disease than the SNPs tagging them in this study. Taken together with estimates from previous studies, the 18 confirmed breast cancer susceptibility loci explain approximately 8% of the familial risk of breast cancer, while rarer mutations in the known high risk (principally *BRCA1* and *BRCA2*) and moderate risk loci explain a further ~20%. The residual familial risk is likely to be due to a combination of further common variants with smaller effects, together with rarer variants not testable with GWAS arrays. Identification of rarer variants will require other strategies, such as exome sequencing described below.

b) *Undertake case-control resequencing of genes to identify further rare, low-penetrance genes.*

In our original application we aimed to undertake candidate gene case-control resequencing of genes, focusing on DNA repair genes. The original strategy involved sequencing 96 (1 tray) familial breast cancer cases through the full gene and undertaking additional sequencing of genes in which we identified truncating variants in larger series of cases and controls (typically 1000 cases and 1000 controls). This strategy led to the identification of four rare, low-intermediate penetrance genes (5-8). However, as detailed in previous reports, analysis of a further 30 DNA repair genes has not led to the identification of additional genes. The availability of new sequencing technologies together with pull down arrays targeting the exons of 16,000 genes (known as exome arrays) has made it feasible to progress from a candidate to a genome-wide gene re sequencing strategy. The applicability of exome arrays to the identification of classical Mendelian disease genes has recently been published (34). As articulated in our report last year our aim is to apply the technology to complex diseases. We will analyse the exomes of breast cancer cases to identify genes with truncating variants which we will then follow-up in larger series by Sanger sequencing. We are undertaking the exome arrays in collaboration with the Wellcome Trust Sanger Institute. We have received the data for the first two exomes and will receive a further 18 shortly. We are analyzing the data using NextGENe software

(Biogene). Our analysis is in the early stages but we have identified truncating mutations in two genes that are highly plausible breast cancer susceptibility genes. We are currently mutationally analyzing these two genes in 400 cases and 400 controls in the first instance. We aim to complete exome sequencing in 20 familial breast cancer cases and will follow-up as many truncating mutations as possible by the end of the grant. We are hopeful that this strategy will lead to the discovery of further breast cancer predisposition genes.

Key Research Accomplishments

In this reporting period we have achieved the following:

- We have completed the largest genome-wide association study in breast cancer analysing 582,886 SNPs in 3,659 cases enriched for a family history of the disease and 4,897 controls. We evaluated promising associations in a second stage, comprising 12,576 cases and 12,223 controls. This has led to the identification of five novel susceptibility loci, on chromosomes 9, 10 and 11 (P=4.6x10^{-7} to P=3.2x10^{-15}). We will publish these data before the end of the grant

- We have set-up next-generation exome sequenci ng and have completed two exomes. We are analyzing the data currently and we are following-up two promising genes. We will complete 20 exomes in total and we will follow-up truncating variants by case-control resequencing in the final year of the grant.

Reportable Outcomes

Based in part on the work supported by this award we successfully applied to the Wellcome Trust to fund the first phase of our second generation genome-wide association study and received £655,500.

Conclusion

We have had another productive year. The primary work of this reporting period has been completing the GWAS and undertaking replication to identify five novel breast cancer susceptibility loci. In our final year we will complete the analysis of *BRCA1* in triple-negative-tumors and we will complete exome sequencing in 20 familial breast cancer ca ses and will follow-up promising genes by case-control resequencing.

References

1. Stratton, M.R. and Rahman, N. The emerging landscape of breast cancer susceptibility. *Nat Genet* **40**, 17-22 (2008).

2. Turnbull, C. andRahman, N. Genetic predisposition to breast cancer, past, present, and future. *Annu.Rev Genomics Hum Genet* **9**, 321-45 (2008).

3. Miki Y. *et al.* A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. *Science* **266**, 66-71 (1994)

4. Wooster, R. *et al*., Identification of the breast cancer susceptibility gene BRCA2 *Nature* **378**, 789-92 (1995)

5. Meijers-Heijboer, H. *et al.,* Low-penetrance susceptibility to breast cancer due to CHEK2*1100delC in noncarriers of BRCA1 or BRCA2 mutations. *Nat Genet* **31**, 55-59 (2002).

6. Seal, S. *et al.* Truncating mutations in BRIP1 are low penetrance breast cancer susceptibility alleles. *Nat Genet* **38**, 1239-1241 (2006).

7. Renwick, A. *et al.* ATM mutations that cause ataxia-telangiectasia are breast cancer susceptibility alleles. *Nat Genet* **38**, 873-875 (2006).

8. Rahman, N. *et al.* PALB2, which encodes a BRCA2 interacting protein, is a breast cancer susceptibility gene. *Nat Genet* 39, 165-167 (2007).

9. Easton, D.F. *et al.* Genome-wide association study identifies novel breast cancer susceptibility loci. *Nature* **447**, 1087-1093 (2007).

10. Hunter,D.J. *et al.* A genome-wide association study identifies alleles in FGFR2 associated with risk of sporadic postmenopausal breast cancer. *Nat Genet* **39**, 870-874 (2007).

11. Stacey,S.N. *et al.* Common variants on chromosomes 2q35 and 16q12 confer susceptibility to estrogen receptor-positive breast cancer. *Nat Genet* **39**, 865-869 (2007).

12. Stacey,S.N. *et al.* Common variants on chromosome 5p12 confer susceptibility to estrogen receptor-positive breast cancer. *Nat Genet* **40**, 703-706 (2008).

13. Ahmed S *et al.* Newly discovered breast cancer susceptibility loci on 3p24 and 17q23.2. *Nat Genet* **41**, 585-590 (2009).

14. 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* **41**, 579-584 (2009).

15. Zheng,W. *et al.* Genome-wide association study identifies a new breast cancer susceptibility locus at 6q25.1. *Nat Genet* **41**, 324-328 (2009).

16. Cox A *et al.* A common coding variant in CASP8 is associated with breast cancer risk. *Nat Genet* **39**, 352-358 (2007).

17. Wellcome Trust Case Control Consortium and The Australo-Angle-American Spondylitis Consortium Association scan of 14,500 nonsynonymous SNPs in four diseases identifies autoimmunity variants. *Nat Genet* **39**, 1329-1338 (2007).

18. Garcia-Closas, M. *et al.* Genetic susceptibility loci for breast cancer by estrogen receptor status. *Clin Cancer Res* **14**,8000-9 (2008).

19. Antoniou AC & Easton DF Polygenic inheritance of breast cancer: implications for design of association studies. *Genet Epidemiol* **25**, 190-202 (2003).

20. Kamb,A. *et al.* A cell cycle regulator potentially involved in genesis of many tumor types. *Science* **264**, 436-440 (1994).

21. Kamb,A. *et al.* Analysis of the p16 gene (CDKN2) as a candidate for the chromosome 9p melanoma susceptibility locus. *Nat Genet* **8**, 23-26 (1994).

22. Bishop,D.T. *et al.* Genome-wide association study identifies three loci associated with melanoma risk. *Nat Genet* **41**, 920-925 (2009).

23. Falchi,M. *et al.* Genome-wide association study identifies variants at 9p21 and 22q13 associated with development of cutaneous nevi. *Nat Genet* **41**, 915-919 (2009).

24. Stacey,S.N. *et al.* New common variants affecting susceptibility to basal cell carcinoma. *Nat Genet* **41**, 909-914 (2009).

25. Shete,S. *et al.* Genome-wide association study identifies five susceptibility loci for glioma. *Nat Genet* **41**, 899-904 (2009).

26. Wrensch,M. *et al.* Variants in the CDKN2B and RTEL1 regions are associated with high-grade glioma susceptibility. *Nat Genet* **41**, 905-908 (2009).

27. Zeggini E *et al.* Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes. *Science* **316**, 1336-1341 (2007).

28. The Wellcome Trust Case Control Consortium Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* **447**, 661-678 (2007).

29. Liu,Y. *et al.* INK4/ARF transcript expression is associated with chromosome 9p21 variants linked to atherosclerosis. *PLoS One* **4**, e5027 (2009).

30. Karlseder,J. *et al.* Patterns of DNA amplification at band q13 of chromosome 11 in human breast cancer. *Genes. Chromosomes. Cancer* **9**, 42-48 (1994).

31. Ornitz,D.M. *et al.* Receptor specificity of the fibroblast growth factor family. *J Biol Chem.* **271**, 15292-15297 (1996).

32. Gianfrancesco,F. *et al.* Identification of a novel gene and a common variant associated with uric acid nephrolithiasis in a Sardinian genetic isolate. *Am J Hum Genet* **72**, 1479-1491 (2003).

33. Barrett,J.C. *et al.* Genome-wide association defines more than 30 distinct susceptibility loci for Crohn's disease. *Nat Genet* **40**, 955-962 (2008).

34. Ng S.B. *et al.* Exome sequencing identifies the cause of a mendelian disorder *Nat.Genet.* **42**, 30-35 (2010)

Appendices

None

Supporting Data

None

Table 1. Associations in current study at previously known breast cancer loci

Locus	Strongest association in current study				Published association				Association for published SNP in current study			
	Best SNP	Per-allele OR (95%CI)[a]	Alleles (freq.)	*P*	Published SNP	Alleles (freq.)	$r^{2\,b}$	Published OR	Best tag in GWAS (r^2)[c]	Alleles (freq.)	Per-allele OR (95%CI)	*P*
FGFR2	rs2981579	1.43 (1.35-1.53)	G/A (0.42)	$3.6x10^{-31}$	rs2981582[d]	G/A (0.38)	1.0	1.26 (1.22-1.29)[1]	rs2981579 (r^2=1.0)	G/A (0.42)	1.43 (1.35-1.53)	$3.6x10^{-31}$
TOX3	rs3803662	1.30 (1.22-1.39)	G/A 0.26	$3.2x10^{-15}$	rs3803662	G/A (0.25)	1.0	1.19 (1.15-1.23)[1]	rs3803662 G/A	(0.26)	1.30 (1.22-1.39)	$3.2x10^{-15}$
MAP3K1	rs889312	1.22 (1.14-1.30)	A/C 0.28	$4.6x10^{-9}$	rs889312	A/C (0.38)	1.0	1.12 (1.08-1.16)[1]	rs889312 A/C	(0.28)	1.22 (1.14-1.30)	$4.6x10^{-9}$
8q24 rs1562430		1.17 (1.10-1.25)	C/T (0.58)	$5.8x10^{-7}$	rs13281615	A/G (0.40)	0.42	1.08 (1.05-1.12)[1]	rs13281615 A/G	(0.41)	1.14 (1.07-1.21)	$2.2x10^{-5}$
2q35 rs13387042		1.21 (1.14-1.29)	G/A (0.49)	$2.0x10^{-10}$	rs13387042	G/A (0.49)	1.0	1.12 (1.09-1.15)[10]	rs13387042 G/A	(0.49)	1.21 (1.14-1.29)	$2.0x10^{-10}$
LSP1	rs909116	1.17 (1.10-1.24)	C/T (0.53)	$7.3x10^{-7}$	rs3817198	T/C (0.30)	0.23	1.07 (1.04-1.11)[1]	rs3817198 T/C	(0.33)	1.12 (1.05-1.19)	.0006
5p12 rs9790879		1.10 (1.03-1.17)	T/C (0.40)	.0032	rs10941679	(A/G) 0.25	0.48	1.19 (1.11-1.28)[4]	rs7716600 (0.75)	C/A (0.22)	1.11 (1.04-1.19)	.0034
6q25.1	rs3757318	1.30 (1.17-1.46)	G/A (0.07)	$2.9x10^{-6}$	rs2046210	G/A (0.34)	0.088	1.15[e] (1.03-1.28)[7]	rs6900157 (0.96)	T/C (0.35)	1.15 (1.08-1.22)	$1.8x10^{-5}$

Locus	Strongest association in current study				Published association				Association for published SNP in current study			
	Best SNP	Per-allele OR (95%CI)[a]	Alleles (freq.)	*P*	Published SNP	Alleles (freq.)	r^2 [b]	Published OR	Best tag in GWAS (r^2)[c]	Alleles (freq.)	Per-allele OR (95%CI)	*P*
SLC4A7	rs4973768	1.16 (1.10-1.24)	C/T (0.47)	$5.8x10^{-7}$	rs4973768	C/T (0.46)	1.0	1.11 (1.08-1.13)[5]	rs4973768	C/T (0.47)	1.16 (1.10-1.24)	$5.8x10^{-7}$
COX11	rs1156287	0.91 (0.85-0.97)	A/G (0.29)	.0058	rs6504950	G/A (0.27)	0.91	0.95 (0.92-0.97)[5]	rs7222197 (1.0)	G/A (0.28)	0.92 (0.86-0.99)	.021
RAD51L1	rs8009944	0.88 (0.82-0.95)	C/A (0.75)	.0004	rs999737	C/T (0.24)	0.13	0.94 (0.88-0.99)[6]	rs999737	C/T (0.25)	0.89 (0.83-0.95)	.0009
1p11.2	rs11249433	1.08 (1.02-1.15)	A/G (0.42)	.010	rs11249433	A/G (0.39)	1.0	1.16 (1.09-1.24)[6]	rs11249433	A/G (0.42)	1.08 (1.02-1.15)	.010
CASP8	rs10931936	0.88 0.82-0.94))	T/C (0.74)	.00015	rs1045485	G/C (0.13)	0.083	0.88 (0.84-0.92)[8]	rs17468277 (1.0)	C/T (0.13)	0.93 (0.85-1.02)	.14

[a] Per-allele OR for the allele correlated with the published allele (+ strand).

[b] r^2 between the published SNP and most significant SNP in this study, based on Hapmap CEU.

[c] r^2 between the published SNP and the best tagSNP in this study, based on Hapmap CEU.

[d] Note that fine-mapping and functional analysis suggest that the strongest association for breast cancer is with rs2981578[31]. It is correlated with rs2981579 and rs2981582 at r^2=0.85. No better tag for rs2981578 was typed in the GWAS.

[e] Estimated OR in Europeans. Estimated OR in Chinese was 1.36.

Table 2. Associations between genotype and breast cancer risk for six SNPs

Marker Chromosome Position	Stage[1]	Cases/Controls	MAF	Per-allele OR (95%CI)	Heterozygous OR (95%CI)	Homozygous OR (95%CI)	*P*-value[2]	
					Stage			Combined
rs1011970 9 G/T 22052134	Stage 1	3,730/4,894	.16	1.20 (1.11-1.30)	1.19 (1.08-1.31)	1.45 (1.13-1.86)	$2.6x10^{-5}$	
	Stage 2	12,253/12,000	.17	1.09 (1.04-1.14)	1.07 (1.01-1.13)	1.29 (1.12-1.50)	.00026	$2.5x10^{-8}$
rs2380205 10 C/T 5926740	Stage 1	3,730/4,895	.44	0.86 (0.81-0.92)	0.86 (0.78-0.95)	0.75 (0.66-0.85)	$7.9x10^{-5}$	
	Stage 2	12,235/11,961	.43	0.94 (0.91-0.98)	0.95 (0.90-1.01)	0.89 (0.82-0.95)	.0017	$4.6x10^{-7}$
rs10995190 10 G/A 63948688	Stage 1	3,731/4,891	.14	0.76 (0.70-0.84)	0.77 (0.69-0.86)	0.55 (0.40-0.77)	$6.1x10^{-8}$	
	Stage 2	12,261/12,000	.15	0.86 (0.82-0.91)	0.84 (0.79-0.89)	0.83 (0.69-1.00)	$1.4x10^{-8}$	$5.1x10^{-15}$
rs704010 10 G/A 80511154	Stage 1	3,726/4,893	.39	1.15 (1.09-1.23)	1.05 (0.95-1.15)	1.38 (1.22-1.57)	$3.5x10^{-6}$	
	Stage 2	12,222/11,992	.39	1.07 (1.03-1.11)	1.11 (1.05-1.17)	1.13 (1.04-1.21)	.00026	$3.7x10^{-9}$
rs614367 11 C/T 69037945	Stage 1	3,723/4,882	.15	1.30 (1.20-1.41)	1.24 (1.13-1.37)	2.02 (1.56-2.64)	$3.9x10^{-8}$	
	Stage 2	12,114/11,967	.15	1.15 (1.10-1.20)	1.16 (1.10-1.23)	1.27 (1.10-1.47)	$1.3x10^{-8}$	$3.2x10^{-15}$

[1]Stage 2 includes genotype data in SEARCH, RBCS and FBCS together with publicly available data from CGEMS.

[2]adjusted 1df *P*-trend