Original citation:

Permanent WRAP url:
http://wrap.warwick.ac.uk/49246

Copyright and reuse:
The Warwick Research Archive Portal (WRAP) makes the work of researchers of the University of Warwick available open access under the following conditions.

© 2012 Hein et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

A note on versions:
The version presented in WRAP is the published version, or, version of record, and may be cited as it appears here.

For more information, please contact the WRAP Team at: wrap@warwick.ac.uk

http://go.warwick.ac.uk/lib-publications


1 Unit of Genetic Epidemiology, Division of Cancer Epidemiology, German Cancer Research Center (DKFZ), Heidelberg, Germany, 2 PMV (Prima¨rmedizinische Versorgung) Research Group at the Department of Child and Adolescent Psychiatry and Psychotherapy, University of Cologne, Cologne, Germany, 3 Department of Oncology, University of Cambridge, Cambridge, United Kingdom, 4 Centre for Molecular, Environmental, Genetic and Analytic Epidemiology, The University of Melbourne, Melbourne, Australia, 5 Department of Pathology, The University of Melbourne, Melbourne, Australia, 6 Genetic Epidemiology Laboratory, Department of Pathology, The University of Melbourne, Melbourne, Australia, 7 Netherlands Cancer Institute, Antoni van Leeuwenhoek hospital, Amsterdam, The Netherlands, 8 Family Cancer Clinic, Netherlands Cancer Institute, Antoni van Leeuwenhoek Hospital, Amsterdam, The Netherlands, 9 National Institute for Public Health and the Environment, Bilthoven, The Netherlands, 10 Health Sciences Research Institute, Warwick Medical School, Warwick University, Coventry, Warwick, United Kingdom, 11 Department of Preventive Medicine, Srinakhrainwirot University, Ongkharak, Nakhon Nayok, Thailand, 12 Department of Surgery, Medical School, Prince Songkla University, Songkla, Thailand.

PLoS ONE | www.plosone.org 1 August 2012 | Volume 7 | Issue 8 | e42380
The 6q25.1 locus was first identified via a genome-wide association study (GWAS) in Chinese women and marked by single nucleotide polymorphism (SNP) rs2046210, approximately 180 Kb upstream of \(\text{ESR1} \). There have been conflicting reports about the association of this locus with breast cancer in Europeans, and a GWAS in Europeans identified a different SNP, tagged here by rs12662670. We examined the associations of both SNPs in up to 61,689 cases and 58,822 controls from forty-four studies collaborating in the Breast Cancer Association Consortium, of which four studies were of Asian and 39 of European descent. Logistic regression was used to estimate odds ratios (OR) and 95% confidence intervals (CI). Case-only analyses were used to compare SNP effects in Estrogen Receptor positive (ER+) versus negative (ER−) tumours. Models including both SNPs were fitted to investigate whether the SNP effects were independent. Both SNPs are significantly associated with breast cancer risk in both ethnic groups. Per-allele ORs are higher in Asian than in European studies (rs2046210: OR (A/G) = 1.36 (95% CI 1.26–1.46), \(p = 7.6 \times 10^{-15} \) in Asians and 1.09 (95% CI 1.07–1.11), \(p = 6.8 \times 10^{-18} \) in Europeans. rs12662670: OR (G/T) = 1.29 (95% CI 1.19–1.41), \(p = 1.2 \times 10^{-15} \) in Asians and 1.12 (95% CI 1.08–1.17), \(p = 3.8 \times 10^{-19} \) in Europeans). SNP rs2046210 is associated with a significantly greater risk of ER+ compared with ER− tumours in Europeans (OR (ER+) = 1.20 (95% CI 1.15–1.25), \(p = 1.8 \times 10^{-10} \) versus OR (ER−) = 1.02 (95% CI 1.04–1.11), \(p = 3.0 \times 10^{-09} \)) in these Asian studies, by contrast, there is no clear evidence of a differential association by tumour receptor status. Each SNP is associated with risk after adjustment for the other SNP. These results suggest the presence of two variants at 6q25.1 each independently associated with breast cancer risk in Asians and in Europeans. Of the two, these one tagged by rs2046210 is associated with a greater risk of ER− tumours.


**Editor:** Kelvin Yuen Kwong Chan, Tsan Yuk Hospital, Hospital Authority, China

**Received** March 26, 2012; **Accepted** July 4, 2012; **Published** August 7, 2012

**Copyright:** © 2012 Hein et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.
Introduction

A genome-wide association study (GWAS) in Chinese women by Zheng et al. [1] identified a novel breast cancer susceptibility locus at 6q25.1. The most strongly associated single nucleotide polymorphism (SNP) was rs2046210, with an estimated Odds ratio (OR) per-allele A/G = 1.29 (95% confidence interval (CI) 1.21–1.37, p = 10^{-14}). SNP rs2046210 did not show a clear association in GWAS carried out in women of European ancestry, and replication studies indicated its effect, if any, was weaker in Europeans [OR (per allele A/G) = 1.04 (95% CI 0.99–1.08), p = 0.09 in a combined analysis of European studies [2]].

More recent studies in European women suggested stronger associations with other SNPs in the region: Turnbull et al. [3] found the most significantly associated SNP to be rs3757318, which is weakly correlated with rs2046210 in Europeans (r^2 = 0.09 from in HapMap2 CEU), while Stacey et al. [2] suggested that SNPs closer to ENR1 may be more strongly associated. It is as yet unclear whether this difference in breast cancer associated SNPs between Asians and Europeans indicates the presence of a single or multiple causative variant(s) at this locus. If there is only one, it is unlikely to be highly correlated with the best tags identified from either the Asian or European GWAS and could potentially be a common variant with a small effect or a rarer one with a larger effect on breast cancer risk.

In this, by far the largest study to date, we investigate associations with SNP rs2046210, as well as with SNP rs12662670 in forty-four case-control studies within the Breast Cancer Association Consortium (BCAC). These two SNPs have been genotyped in a total of 120,511 female subjects, of which 110,265 subjects are of European ancestry and 8,559 are Asian. SNP rs2046210 is the best tag from the original Asian GWAS [1] and SNP rs12662670 is an easier to genotype surrogate for SNP rs3757318 - the best tag SNP at the 6q25.1 locus from an Italian GWAS [3]. Our aims were to compare the effects of these tags in well-powered studies of both Asian and European ancestry and to test if these known SNP associations are shared by the different ethnic groups. We have been successful in achieving these aims and our analyses provide additional insights into the nature of this locus.

Materials and Methods

Ethics Statement

Approval of the studies was obtained from the ethics committees listed in Table S1. All studies conform to the Declaration of Helsinki and all study participants gave written informed consent.

Study Populations

Data from forty-five BCAC case-control studies from Australia, Europe, North America, and South-East Asia were available for...
inclusion in this analysis (see Table S1 for a description of the individual studies). To be eligible for BCAC, studies needed to include at least 500 cases of invasive breast cancer and 500 controls, with DNA samples available for genotyping. The controls needed to be broadly from the same population as the cases (http://www.srl.cam.ac.uk/consortia/bcac/about/about.html). Some studies selected cases preferentially on the basis of age and/or family history.

All studies provided information on disease status (58,822 controls/62,861 invasive cases/2,769 in-situ cases/1,435 cases of unknown invasiveness), age at diagnosis or interview and ethnicity (Asian-European/other). Forty studies also provided information on estrogen receptor (ER) status for a total of 40,308 cases (9,978/30,630 Estrogen Receptor positive [ER+] /30,630 Estrogen Receptor negative [ER–]).

### Laboratory Methods

In most studies SNPs were assayed by Taqman™ (Applied Biosystems, Foster City, USA). Primers, probes and master mix were ordered in a single batch and aliquots shipped to each study. Reactions were performed according to manufacturer’s instructions, using the following thermal cycling profile 95°C for 15 secs, 60°C for 1 min for 40–60 cycles.

SNP rs12662670 was chosen as the most easily assayable surrogate for the best European GWAS hit, rs3757318, for which no working Taqman™ assay could be designed. These two SNPs are correlated at \( r^2 = 0.89 \) in the European samples used in the analysis (see Table S1 for information on the respective phases for SNP rs2046210 and rs12662670). All studies followed standard quality control guidelines (for details see http://www.srl.cam.ac.uk/consortia/bcac/about/about.html). Data were excluded for any sample that failed genotyping for >20% of the SNPs typed in a given phase of genotyping. All study data were excluded for any SNP with overall call rate <95% or duplicate concordance <94% (based on at least 2% of samples in each study being genotyped in duplicate) or departure of genotype distribution from Hardy-Weinberg equilibrium.

### Table 1. Association of rs2046210 and rs12662670 with breast cancer.

<table>
<thead>
<tr>
<th>Ethnicity</th>
<th>Number of cases/controls</th>
<th>Odds ratio (95% confidence interval)</th>
<th>P-valuea</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>per alleleb</td>
<td>Heterozygote</td>
</tr>
<tr>
<td>rs2046210</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Analyses adjusted for study</td>
<td>Overall</td>
<td>54,647/49,559</td>
<td>1.10 (1.08–1.13)</td>
</tr>
<tr>
<td></td>
<td>Europeans</td>
<td>49,634/46,679</td>
<td>1.09 (1.07–1.11)</td>
</tr>
<tr>
<td></td>
<td>Asians</td>
<td>2983/2332</td>
<td>1.36 (1.26–1.48)</td>
</tr>
<tr>
<td>Analyses adjusted for study and rs12662570</td>
<td>Overall</td>
<td>40,384/33,750</td>
<td>1.08 (1.05–1.11)</td>
</tr>
<tr>
<td></td>
<td>Europeans</td>
<td>36,396/31,105</td>
<td>1.08 (1.05–1.11)</td>
</tr>
<tr>
<td></td>
<td>Asians</td>
<td>3416/2420</td>
<td>1.17 (1.02–1.36)</td>
</tr>
<tr>
<td>rs12662670</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Analyses adjusted for study</td>
<td>Overall</td>
<td>42,654/40,166</td>
<td>1.15 (1.12–1.19)</td>
</tr>
<tr>
<td></td>
<td>Europeans</td>
<td>38,723/37,400</td>
<td>1.12 (1.08–1.17)</td>
</tr>
<tr>
<td></td>
<td>Asians</td>
<td>3273/2451</td>
<td>1.29 (1.19–1.41)</td>
</tr>
<tr>
<td>Analyses adjusted for study and rs2046210</td>
<td>Overall</td>
<td>40,384/33,750</td>
<td>1.10 (1.06–1.15)</td>
</tr>
<tr>
<td></td>
<td>Europeans</td>
<td>36,396/31,105</td>
<td>1.07 (1.02–1.12)</td>
</tr>
<tr>
<td></td>
<td>Asians</td>
<td>3416/2420</td>
<td>1.21 (1.04–1.40)</td>
</tr>
</tbody>
</table>

Results are presented overall and separately for Europeans and Asians. Pooled analyses adjusted for study only as well as adjusted for rs12662670 or rs2046210, respectively, in addition to study were performed.

aP-value derived from the log-additive model.

bOdds ratio per minor allele (A allele for rs2046210, G allele for rs12662670).

cOdds ratio relative to the major allele homozygous (GT) genotype.

doi:10.1371/journal.pone.0042380.t001
Weinberg equilibrium in controls (p<0.005). In addition, all genotyping centres assayed an identical plate of 80 control DNA samples (referred to as the Coriell plate; which also included 14 internal duplicates) and had to achieve call rates and duplicate concordance >98% in order for their data to be included. Data for both SNPs from one study (NBCS) were excluded from further analyses after quality control rules were applied. Quality control data for the individual studies are shown in Tables S2a and S2b. Thus, for SNP rs2046210 forty out of forty-one assayed studies (56,607 cases/49,559 controls), and for SNP rs12662670 thirty-three out of thirty-four assayed studies (47,251 cases/40,161 controls) were included in the statistical analysis.

Statistical Analyses

ORs were estimated using logistic regression. In order to provide reliable estimates of effect sizes, study-specific effect estimates of ORs were derived only for those studies that provided at least 100 cases and controls for the respective (sub-) group of interest. The primary analysis estimated ORs for the main effect of the SNP, adjusted for the studies that provided data for the respective analysis (i.e. S-1 indicator variables were entered the logistic regression model, where S was number of studies that provided data for the respective analysis). ORs adjusted for both study and age were essentially identical and we did not therefore present the age-adjusted analyses. Per allele ORs were estimated under the assumption of a log-additive mode of inheritance, i.e. the SNP was coded according to the number of minor alleles 0, 1 or 2. Additionally, ORs by genotype were calculated, i.e. two indicator variables indicating the presence of the heterozygous genotype and the genotype homozygous for the minor allele, respectively, were entered the model. The primary p-values were derived by means of a Wald-Test assuming a log-additive mode of inheritance (one degree of freedom). Following Laird and Mosteller, heterogeneity of per allele ORs between studies was assessed by the p-value derived from the Q statistic [4] and using I². Tests were two-sided.

Genetic main effects by ER status were estimated using case-control logistic regression and restricting the case sample to ER+ or ER− cases, respectively. To test for significant differences between main effects of rs2046210 or rs12662670 in ER+ versus ER− cases, logistic regression analyses were conducted in cases only. In these case-only analyses, the binary ER status was the outcome/dependent variable and the respective SNP and the indicator variables representing the studies were the independent variables.

Variation in OR by age was evaluated by testing for an interaction between age-group (<40, 40 to 49, 50 to 59, ≥60) and SNP, separately for each subgroup defined by ethnicity and ER
Table 2. Association of haplotypes composed of rs2046210 and rs12662670 with breast cancer.

<table>
<thead>
<tr>
<th>Ethnicity</th>
<th>Number of cases/controls</th>
<th>rs2046210</th>
<th>rs12662670</th>
<th>Haplotype frequency</th>
<th>Odds ratio (95% confidence interval)</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall</td>
<td>40,384/33,750</td>
<td>A&lt;sup&gt;c&lt;/sup&gt;</td>
<td>T</td>
<td>0.26</td>
<td>1.08 (1.05–1.11)</td>
<td>1.55 × 10&lt;sup&gt;-3&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td></td>
<td>G</td>
<td>G&lt;sup&gt;c&lt;/sup&gt;</td>
<td>0.01</td>
<td>1.10 (0.95–1.27)</td>
<td>0.189</td>
</tr>
<tr>
<td></td>
<td></td>
<td>A&lt;sup&gt;c&lt;/sup&gt;</td>
<td>G&lt;sup&gt;c&lt;/sup&gt;</td>
<td>0.09</td>
<td>1.23 (1.18–1.28)</td>
<td>2.20 × 10&lt;sup&gt;-4&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td></td>
<td>G&lt;sup&gt;d&lt;/sup&gt;</td>
<td>T&lt;sup&gt;d&lt;/sup&gt;</td>
<td>0.65&lt;sup&gt;d&lt;/sup&gt;</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Europeans</td>
<td>36,396/31,105</td>
<td>A&lt;sup&gt;c&lt;/sup&gt;</td>
<td>T</td>
<td>0.27</td>
<td>1.08 (1.05–1.11)</td>
<td>1.59 × 10&lt;sup&gt;-8&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td></td>
<td>G</td>
<td>G&lt;sup&gt;c&lt;/sup&gt;</td>
<td>0.01</td>
<td>1.02 (0.86–1.21)</td>
<td>0.832</td>
</tr>
<tr>
<td></td>
<td></td>
<td>A&lt;sup&gt;c&lt;/sup&gt;</td>
<td>G&lt;sup&gt;c&lt;/sup&gt;</td>
<td>0.07</td>
<td>1.16 (1.11–1.21)</td>
<td>4.35 × 10&lt;sup&gt;-11&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td></td>
<td>G&lt;sup&gt;d&lt;/sup&gt;</td>
<td>T&lt;sup&gt;d&lt;/sup&gt;</td>
<td>0.65&lt;sup&gt;d&lt;/sup&gt;</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>Asians</td>
<td>3416/2420</td>
<td>A&lt;sup&gt;c&lt;/sup&gt;</td>
<td>T</td>
<td>0.09</td>
<td>1.23 (1.05–1.45)</td>
<td>1.15 × 10&lt;sup&gt;-2&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td></td>
<td>G</td>
<td>G&lt;sup&gt;c&lt;/sup&gt;</td>
<td>0.04</td>
<td>1.23 (0.94–1.61)</td>
<td>0.134</td>
</tr>
<tr>
<td></td>
<td></td>
<td>A&lt;sup&gt;c&lt;/sup&gt;</td>
<td>G&lt;sup&gt;c&lt;/sup&gt;</td>
<td>0.27</td>
<td>1.42 (1.30–1.56)</td>
<td>4.44 × 10&lt;sup&gt;-14&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td></td>
<td>G&lt;sup&gt;d&lt;/sup&gt;</td>
<td>T&lt;sup&gt;d&lt;/sup&gt;</td>
<td>0.60&lt;sup&gt;d&lt;/sup&gt;</td>
<td>–</td>
<td>–</td>
</tr>
</tbody>
</table>

Results are presented overall and separately for Europeans and Asians. Pooled analyses adjusted for study were performed.

<sup>a</sup>Odds ratio per haplotype compared to the reference haplotype (i.e., the most frequent haplotype).
<sup>b</sup>P-value derived from the log-additive model.
<sup>c</sup>Minor allele.
<sup>d</sup>Reference haplotype.

doi:10.1371/journal.pone.0042380.t002

Figure 2. Association of rs12662670 with breast cancer in Europeans versus Asians.
doi:10.1371/journal.pone.0042380.g002
status. Thus, the multiplicative SNP by age-group interaction term entered the model in addition to the main effect terms for SNP, age-group and study.

To investigate whether the association with breast cancer risk could be explained by one SNP or whether both SNPs had independent effects on disease risk, we fitted logistic regression models which included both SNPs, in addition to indicator variables for study. Haplotypes were estimated using the expectation-maximization algorithm. All analyses were performed using R version 2.11.0 [5] and the R packages meta, rmeta and haplo.stats.

Results

Key characteristics for each participating study are shown in Table S1. In addition to the originally discovered SNP rs2046210, SNP s12662670 was genotyped as a surrogate for the best tag from Turnbull et al. [rs3757318] [3], for which no working Taqman™ assay could be designed. The genotype distributions by ethnicity and study were analyzed using logistic regression. The results are presented overall as well as separately for Europeans and Asians.

### Table 3. Association of rs2046210 and rs12662670 with risk of ER−/ER+ breast cancer.

<table>
<thead>
<tr>
<th>Ethnicity</th>
<th>Estrogen receptor status</th>
<th>Number of cases/controls</th>
<th>Odds ratio (95% confidence interval)</th>
<th>P-value</th>
<th>P-heterogeneity</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs2046210</td>
<td>Analyses adjusted for study</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overall</td>
<td>ER−*</td>
<td>7126/35,833</td>
<td>1.23 (1.18–1.28)</td>
<td>6.90 × 10^{-25}</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ER+**</td>
<td>5914/32,977</td>
<td>1.09 (1.06–1.12)</td>
<td>1.66 × 10^{-11}</td>
<td>5.93 × 10^{-7}</td>
</tr>
<tr>
<td>Europeans</td>
<td>ER−*</td>
<td>807/2334</td>
<td>1.20 (1.15–1.25)</td>
<td>1.77 × 10^{-17}</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ER+**</td>
<td>24,596/42,664</td>
<td>1.07 (1.04–1.10)</td>
<td>1.26 × 10^{-7}</td>
<td>5.10 × 10^{-6}</td>
</tr>
<tr>
<td>Asians</td>
<td>ER−*</td>
<td>22,336/39,856</td>
<td>1.45 (1.29–1.64)</td>
<td>1.70 × 10^{-9}</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ER+**</td>
<td>1307/2334</td>
<td>1.35 (1.22–1.49)</td>
<td>1.09 × 10^{-8}</td>
<td>0.513</td>
</tr>
<tr>
<td>Analyses adjusted for study and rs12662670</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overall</td>
<td>ER−*</td>
<td>5526/25,627</td>
<td>1.19 (1.13–1.26)</td>
<td>1.26 × 10^{-10}</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ER+**</td>
<td>19,502/32,010</td>
<td>1.07 (1.03–1.10)</td>
<td>7.26 × 10^{-5}</td>
<td>1.01 × 10^{-3}</td>
</tr>
<tr>
<td>Europeans</td>
<td>ER−*</td>
<td>4497/23,090</td>
<td>1.20 (1.13–1.26)</td>
<td>4.54 × 10^{-10}</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ER+**</td>
<td>17,653/29,365</td>
<td>1.07 (1.03–1.10)</td>
<td>2.16 × 10^{-4}</td>
<td>4.63 × 10^{-4}</td>
</tr>
<tr>
<td>Asians</td>
<td>ER−*</td>
<td>975/2420</td>
<td>1.16 (0.94–1.44)</td>
<td>0.169</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ER+**</td>
<td>1649/2420</td>
<td>1.13 (0.94–1.35)</td>
<td>0.194</td>
<td>0.777</td>
</tr>
<tr>
<td>rs12662670</td>
<td>Analyses adjusted for study</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overall</td>
<td>ER−*</td>
<td>5422/28,201</td>
<td>1.23 (1.15–1.32)</td>
<td>6.16 × 10^{-9}</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ER+**</td>
<td>4563/26,198</td>
<td>1.14 (1.09–1.19)</td>
<td>3.05 × 10^{-8}</td>
<td>0.074</td>
</tr>
<tr>
<td>Europeans</td>
<td>ER−*</td>
<td>808/1890</td>
<td>1.17 (1.08–1.27)</td>
<td>1.90 × 10^{-4}</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ER+**</td>
<td>20,095/34,460</td>
<td>1.09 (1.04–1.15)</td>
<td>6.08 × 10^{-4}</td>
<td>0.070</td>
</tr>
<tr>
<td>Asians</td>
<td>ER−*</td>
<td>18,519/32,349</td>
<td>1.37 (1.20–1.56)</td>
<td>3.30 × 10^{-6}</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ER+**</td>
<td>1394/1890</td>
<td>1.35 (1.21–1.51)</td>
<td>1.58 × 10^{-7}</td>
<td>0.691</td>
</tr>
<tr>
<td>Analyses adjusted for study and rs2046210</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overall</td>
<td>ER−*</td>
<td>5526/25,627</td>
<td>1.10 (1.01–1.20)</td>
<td>0.028</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ER+**</td>
<td>19,502/32,010</td>
<td>1.09 (1.03–1.15)</td>
<td>2.14 × 10^{-3}</td>
<td>0.87</td>
</tr>
<tr>
<td>Europeans</td>
<td>ER−*</td>
<td>4497/23,090</td>
<td>1.03 (0.94–1.14)</td>
<td>0.516</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ER+**</td>
<td>17,653/29,365</td>
<td>1.05 (0.99–1.11)</td>
<td>0.110</td>
<td>0.92</td>
</tr>
<tr>
<td>Asians</td>
<td>ER−*</td>
<td>975/2420</td>
<td>1.33 (1.07–1.65)</td>
<td>0.011</td>
<td></td>
</tr>
<tr>
<td></td>
<td>ER+**</td>
<td>1649/2420</td>
<td>1.26 (1.04–1.52)</td>
<td>0.017</td>
<td>0.30</td>
</tr>
</tbody>
</table>

Results are presented overall as well as separately for Europeans and Asians. Pooled analyses adjusted for the studies were performed. A log-additive genetic model was assumed.

*Estrogen receptor negative.
**Estrogen receptor positive.
*Odds ratio per minor allele (A allele for rs2046210, G allele for rs12662670) derived from case-control logistic regression restricted to ER+ or ER− cases, respectively, and the whole control sample.
**P-value derived from the log-additive model derived from case-control logistic regression restricted to ER+ or ER− cases, respectively, and the whole control sample.
**P-value for heterogeneity between estimates of genetic main effects derived from case-only analysis.

doi:10.1371/journal.pone.0042380.t003
The associations of each SNP are presented in Table 1 and as Forest plots in Figure 1 and 2, separately for Europeans and Asians. Both SNPs are significantly associated with breast cancer risk in both ethnic groups. However, the per-allele OR associated with the minor A allele of SNP rs2046210 is higher in Asian populations (OR (A/G) = 1.36 (95% CI 1.26–1.48), p = 7.6 \times 10^{-14}) than in Europeans (OR (A/G) = 1.09 (95% CI 1.07–1.11), p = 6.8 \times 10^{-10}) and this difference is statistically significant (I^2 = 0.002). SNP rs12662670 shows a similar pattern, with a higher OR associated with the minor G allele in Asian studies (OR (G/T) = 1.29 (95% CI 1.19–1.41), p = 1.2 \times 10^{-14}) than in Europeans (OR (G/T) = 1.12 (95% CI 1.06–1.17), p = 3.8 \times 10^{-9}) and again this difference is statistically significant (I^2 = 0.002). In each case there is no evidence for departure from a log-additive model (a co-dominant mode of inheritance).

Logistic regression models, which include both SNPs, indicate that the two SNPs are independently associated in both Europeans (p = 1.8 \times 10^{-18} for rs2046210, p = 2.9 \times 10^{-13} for rs12662670; Table 1) and Asians (p = 0.028 for rs2046210, p = 0.012 for rs12662670). In each ethnicity the estimated ORs for each SNP, after adjustment for the other SNP, are of similar magnitudes: For rs2046210 in Europeans OR (A/G) = 1.08 (95% CI 1.05–1.11) and for rs12662670 in Europeans OR (G/T) = 1.07 (95% CI 1.02–1.12). For rs2046210 in Asians OR (A/G) = 1.17 (95% CI 1.02–1.36) and for rs12662670 in Asians OR (G/T) = 1.21 (95% CI 1.04–1.40). Similar effect estimates are also obtained for haplotypes carrying one minor allele though estimates do not reach statistical significance for the very rare haplotype carrying the major (G) allele of rs2046210 along with the minor (G) allele of rs12662670 (Table 2). Of note, from the four observed haplotypes, effects are strongest and highly statistically significant for the haplotype carrying both minor alleles: In Europeans OR (AG) = 1.16 (95% CI 1.11–1.21). In Asians OR (AG) = 1.42 (95% CI 1.30–1.56).

The OR estimates for in-situ cancer are similar to those for invasive cancer for both SNPs in Europeans, although, due to small numbers, the effect of rs12662670 on in-situ tumours does not reach statistical significance (Table S4). For each Asian study, the number of in-situ cases is less than 100 and so effect estimates are inaccurate but do not differ from those for invasive cancer (data not shown).
The associations of these two SNPs with tumour sub-types defined by ER status (ER+ and ER−) were also investigated and are presented in Table 3 and Figures 3, 4, 5, and 6. In Europeans, SNP rs2046210 is associated with a greater OR for ER− than ER+ tumours: OR (ER−) = 1.20 (95% CI 1.15–1.25), \( p = 1.8 \times 10^{-17} \) vs. OR (ER+) = 1.07 (95% CI 1.04–1.1), \( p = 1.3 \times 10^{-7} \), \( \text{P}_{\text{heterogeneity}} = 5.1 \times 10^{-6} \). This difference remains significant after adjustment for rs12662670. A similar, although non-significant, difference is observed in European women for SNP rs12662670 (Table 3). In the Asian studies, however, there is no clear evidence of a differential association by tumour receptor status for either SNP (Table 3).

We further investigated whether the magnitudes of these SNP associations on tumour sub-types differed by age at diagnosis/interview (see Table S5). In Asian studies the data are too sparse to give meaningful results. In the combined ethnicities and the European studies alone, the magnitudes of the observed associations are greater in younger women.

Fourteen of the European studies had been designed to oversample cases with a family history of breast cancer (see Table S1), which could have led to an overestimation of the ORs relative to those expected in a population-based case-control study. However, exclusion of these studies does not materially affect the estimated ORs for either SNP (see Table S6).

**Discussion**

In this large collaborative study of up to 61,689 cases and 58,822 controls, we demonstrate a highly statistically significant association between the A allele of rs2046210 and increased breast cancer risk in women of both Asian and European ancestry, thus extending the association previously observed in Asian populations. Consistent with previous reports [1–3], the effect sizes are significantly greater in Asians than in Europeans. Our study also reveals that the G allele of SNP rs12662670 is significantly associated with increased breast cancer risk in both ethnicities. SNP rs12662670 is used here as surrogate for SNP rs3757318 - the most strongly associated SNP at this locus in the European GWAS described by Turnbull et al. [3]. In addition, and also in contrast to Stacey et al. [2], we find that the OR for rs12662670 is greater in Asians than in Europeans (Table 1, Figure 1 and 2). In contrast to previous reports, our study indicates that both SNPs (rs2046210 and rs12662670) may be independently associated with breast cancer risk – in models including both SNPs, both

---

**Figure 4. Association of rs2046210 with breast cancer in Asian ER− versus ER+ cases and controls.**

*Estrogen receptor negative; **Estrogen receptor positive.

doi:10.1371/journal.pone.0042380.g004
maintain significant ORs after adjustment for the other. Haplo-
type analyses result in effect estimates for the AT and GG
haplotypes, which carry only one minor allele, very similar to those
of the single SNP analyses for the respective minor alleles.
Furthermore, haplotype analyses show a clearly stronger effect of
the AG haplotype, carrying both minor alleles, compared to the
effects of the AT and GG haplotypes, further supporting the
hypothesis that there may be two different causative variants, one
on each haplotype carrying only one minor allele and both on the
haplotype carrying both minor alleles (i.e., the stronger effect of
the AG haplotype compared to the AT and GG haplotypes may
be explained by the joint effect of the two minor alleles on the AG
haplotype). However, the alternative conclusion that a single
causative variant may exist that is intermediate between the two
SNPs phylogenetically, i.e. on the AG haplotype and on some of
the AT haplotypes, cannot yet be excluded, since this
could also be an explanation for the stronger effect of the AG
haplotype compared to the AT and GG haplotypes.

We also find evidence that SNP rs2046210 is more strongly
associated with ER− than ER+ disease in both European and
Asian women. In the present study this differential association with
receptor status is statistically significant in European studies (and
remains after adjustment for rs12662670) but is not quite
significant in Asians which may be due to a lack of power attributable
to the comparatively small number of Asian individuals involved in
our study (Table 3). However this same SNP had previously been
reported to be more strongly associated with ER− tumours in the
original Chinese cases [1] as well as in a recent replication study in
Chinese women [6]. In line with these reports, a meta-analysis
(14,231 cases, 10,244 controls) on this SNP-disease association by ER
status in Asians, incorporating published results as well as those
presented here, reveals a significant difference in OR associated with
ER− versus ER+ tumour risk [OR (A/G - ER−) = 1.37 (95% CI
1.30–1.44), p = 3.7×10$$^{-35}$$ vs. OR (A/G- ER+) = 1.27 (95% CI
1.22–1.34), p = 2.2×10$$^{-32}$$; p heterogeneity = 0.04]. A stronger associ-
ation of SNP rs2046210 with ER− tumours is also consistent with the
report from the Consortium of Modifiers of BRCA1/2 (CIMBA) [7]
that the same allele is associated with an increased Hazard Ratio of
breast cancer in BRCA1 mutation carriers (who predominantly
develop ER− tumours). The CIMBA study also observed that this
allele conferred increased Hazard Ratios among younger mutation
carriers while we observed similar trends for greater SNP ORs at
younger age groups (Table S5). By contrast, the CIMBA consortium
reported that SNP rs9397435 (the tag they used for rs12662670;
r$$^2$$ = 0.61, r$$^2$$ = 0.50 and r$$^2$$ = 0.85 in HapMap2 CEU, JPT and CHB
samples, respectively) shows evidence of modification of risk in both

---

Figure 5. Association of rs12662670 with breast cancer in European ER− vs. ER+ cases and controls. *Estrogen receptor negative; **Estrogen receptor positive.
doi:10.1371/journal.pone.0042380.g005
BRCA1 and BRCA2 mutation carriers (who mainly develop ER- and ER+ tumours respectively) [7] whilst similarly, we find that SNP rs12662670 is associated with increased risks of both ER- and ER+ tumours.

Previous fine-scale mapping publications on this locus [2,8] have sought a single variant to explain the associations seen with all SNPs in the region: Stacey et al. [2] proposed SNP rs9397435 as a possible single causative variant since it was more strongly associated than rs2046210 in women of European, African and Asian ancestry. We are not able to comment on this variant, as it has not been genotyped in BCAC. However, our findings suggest there could be two independent associations at this locus: one, better tagged by SNP rs2046210, predisposing to ER- tumours and the second, better tagged by rs12662670, conferring similar risks of both tumour types. Although physically close, SNPs rs2046210 and rs12662670 are not highly correlated with each other, particularly in Europeans (in BCAC $r^2 = 0.12$ in Europeans and $r^2 = 0.56$ in Asians) and all four possible combinations (haplotypes) of these two SNPs clearly exist.

Examination of linkage disequilibrium plots of the regions surrounding these two SNPs in Europeans (Figure 7) reveals little, if any, physical overlap between SNPs highly correlated ($r^2 > 0.9$) with rs2046210 and those with rs12662670. If there were a single causal variant, directly responsible for the associations seen with both SNPs, it would need to be correlated with both SNPs. Such a variant has not been yet identified (e.g. by the 1000 Genomes Project). It would presumably be relatively rare. An alternative, and we think, more plausible, explanation for the pattern of associations may be the existence of two independent causative variants, one correlated with rs2046210 and another correlated with rs12662670. If this is the case, the former variant may be more strongly associated with ER- breast cancer than the latter.

The reason why both SNPs confer higher relative risks in Asians than in Europeans is unclear. Within the BCAC studies, ER- tumours are relatively more prevalent among Asian (36%) compared to European cases (23%), but this is not sufficient to explain the higher ORs in Asians, since the effects persist after stratification by ER status. It remains possible that the higher relative risks are due to differential patterns of linkage disequilibrium if the, as yet, unidentified causal variants are not strongly correlated with the SNPs identified to date. These questions may be resolved by comprehensive re-sequencing of this locus and fine scale mapping to identify the causal variant (or variants) responsible for the observed breast cancer risks. One aim of the
iCOGS Project [9], which is currently underway, is to address these questions. However it is possible that these observed differences between Asians and Europeans may reflect interactions with lifestyle risk factors or other unlinked genetic loci. Another possible explanation is that the estimated SNP effects in Asians are inflated given the phenomenon known as the “winner’s curse”, i.e. the suboptimal power of the pool of Asian studies (due to the small number of Asian individuals) together with the commonly used requirement for a published association to pass a certain predefined p-value threshold may have resulted in biased SNP effect estimates [10,11].

Although there are eleven genes within 1 Mb of this locus, attention has focused on the ESR1 gene, whose transcription start site is located approximately 180 Kb downstream of SNP rs2046210. ESR1 encodes ERα and has long been implicated in breast carcinogenesis. However, it is possible that the proximity of this SNP to ESR1 may be providing a false lead – both SNPs (rs2046210 and rs12662670) lie in the flanking region of C6orf97 and there are numerous other genes in close physical proximity (see Figure 7). It is notable however, that SNPs mapping to this region have also been identified in GWAS for bone mineral density – another phenotype in which estradiol metabolism is clearly implicated [12,13]. Furthermore, a recent paper [14] demonstrates that a number of genes, including ESR1 and C6orf97 are co-regulated at this locus although the functions of most of these co-regulated genes have not yet been elucidated. The SNP associations, presented here, may provide a basis to explore the biological role of this locus in estrogen signalling and cancer development in more detail.

Taken together our findings suggest the possibility of the presence of two different causative variants at the 6q25.1 locus and indicate that fine-scale mapping efforts aimed at finding a single variant accounting for associations with both marker SNPs, may not be successful.

**Supporting Information**

Table S1 Characteristics of 45 case-control studies within the Breast Cancer Association Consortium (BCAC). (DOC)
Acknowledgments

We thank all the individuals who took part in these studies and all the researchers, clinicians, technicians and administrative staff who have enabled this work to be carried out. In particular we thank Jonathan Morrison (BCAC); Maggie Angelakos, Judi Maskiell and Gillian Dite (BCAC); Richard van Hien, Sten Cornelissen and the NKI-AVL Family Cancer Clinic (ABCS); all the participants in the Thai Breast Cancer Study within the Asia Cancer Programme (ACP) initiative; special thanks also go to the data collection process; finally, we would like to thank Ms Sunee Erkkila (HEBCS); Eija Myo¨ha¨nen and Helena Kemila¨inen (KBCP); Kirsimari Aaltonen, Irja H. Karstens (HABCS, HMBCS, HUBCS); all the study co-ordinator (ACP); furthermore, we thank Eileen Williams, Elaine Ryder-Mills and Kara Sargas (BBCS); Niall Morrison, Andrew Rowan and Angela Jones (BIGGS); Charo Alonso, Tais Moreno, Guillermo Pita, Primitiva Menendez and Anna Gonzalez-Neira (CNIO-BCS); Hartwig Ziegler, Sonja Wolf and Volker Hermann (ESTHER); all the researchers who took part in these studies and all the researchers, clinicians, technicians and administrative staff who have enabled this work to be carried out; in particular, we thank The Welcome Trust Case Control Consortium (see the WTCCC website for a full list of contributing investigators) (FBCS); furthermore, we thank Bernl Fred, Rita K. Schmutzler and Claus R. Bartram (GC-HBCOC); Christina Justenhoven, Beate Pesch, Thomas Van Brussel and Kathleen Corthouts (LMBC); Tracy Slanger, Gilian Peuteman, Dominiek Smeets, B. Kaspereit, N. Knese and K. Smit (MARIE); Marco Pierotti, Bernard Prissel and Daniela Zaffaroni of the Fondazione IRCCS Istituto Nazionale Tumori, Bernardo Bonanni of the Istituto Europeo di Oncologia, Loris Bernard and the personnel of the Cancer Genetics Testing laboratory at the IFOM-IEO campus (MCBCS); all the people contributing to the MCCS, including the original investigators and the diligent team who recruited the participants and who continue working on follow up; finally, we would like to express our gratitude to the many thousands of Melbourne residents who continue to participate in the study in particular, we thank the study participants of the MSKCC study (MSKCC); Merri Otsukia and Kari Mononen for their contribution to OBCC (OBCC); the Ontario Family Tumor Breast Cancer Registry; Teresa Sander and Nayana Weerasooriya from Cancer Care for their contributions to OFBCR (OFBCR); Michael Stagner and Pei Chao from Information Management Services (Sliver Spring, MD, USA), for data management support (PBGS); Petra Bos, Jannet Blom, Ellen Crepin, Elisabeth Huisjens, Annette Heemskerk and the Erasmus MC Family Cancer Clinic (RBCS); Helen Cramp, Dan Comoley, Sabapathy Balasubramaniam and Sue Higham for their contribution to patient recruitment and data collection (SBBCs); the SEARCH and EPIC teams (SEARCH); Irene Masunaka (UCIBCS); John Hampton from University of Wisconsin and Michael Stagner and Pei Chao from Information Management Services (Sliver Spring, MD, USA), for data management support (US3SS).

Author Contributions

Anaalyzed the data: RH MM ACA JCC DFE AMD. Wrote the paper: RH ACA JCC DFE AMD. Participated in Design/Recruitment/Phenotype-Collection: KR AL PP FF AH AB MJB NJ ISSP JP ES IT MK NM SEI BGN HF JP MPZ JB HAC AZ LB CCDD HB HM FM AS CS BB VA CS SS CRB HB UH YDK JCC VK VMK GCT DFJ FJC JEO JW XZF GGGB LB CAM KO MR MMG J VJ VK ALBD EMR RJ KW KP AM SS CRB RK5 AJV MG JIFD JL MES AC MMRR MG PDPP DK KYY DYN CYS CYC HAC AZ MKH DFE AMD. Participated in Genotyping/ Data Quality Control: MM KRK AB OF EB SEB SWG AM HJ JB ZF LB CAM GGA VK AM MCCG PH KH KC JL AC SA MG HMH MKH MHK. Managed the database: MH JM. Revised the manuscript critically and approved the version of the manuscript to be published: All authors.

References


Table S2 Characteristics of the study populations genotyped for rs2046210 (a) and rs12662670 (b).

Table S3 Genotype frequencies of SNP rs2046210 (a) and rs12662670 (b) in the different studies.

Table S4 Association of rs2046210 and rs12662670 with in-situ/invasive breast cancer.

Table S5 Association of rs2046210 and rs12662670 with risk of ER−/ERα**breast cancer.

Table S6 Association of rs2046210 and rs12662670 with breast cancer.