Genetic variation in the genome-wide predicted estrogen response element-related

sequences is associated with breast cancer development

Jyh-Cherng Yu[1], Chia-Ni Hsiung[2], Huan-Ming Hsu[1], Bo-Ying Bao[3], Shou-Tung Chen[4],

Giu-Cheng Hsu[5], Wen-Cheng Chou[2], Ling-Yueh Hu[2], Shian-ling Ding[6], Chun-Wen Cheng[7],

Pei-Ei, Wu[2], Chen-Yang Shen[2,8*]

[1]Departments of Surgery, Tri-Service General Hospital, Ming-Cheng East Road, Taipei, 114,

Taiwan; [2]Institute of Biomedical Sciences, Academia Sinica, Academy Street, Taipei, 115,

Taiwan; [3]Department of Pharmacy, China Medical University, Hsueh-Shih Road, Taichung,

404, Taiwan; [4]Department of Surgery, Changhua Christian Hospital, Nanxiao Street,

Changhua, 500, Taiwan; [5]Department of Radiology, Tri-Service General Hospital,

Ming-Cheng East Road, Taipei, 114, Taiwan; [6]Department of Nursing, Kang-Ning Junior

College of Medical Care and Management, Kangning Road, Taipei, 114, Taiwan, [7]Institute of

Biochemistry and Biotechnology, Chung Shan Medical University, Jianguo North Road,

Taichung, 402, Taiwan; [8]Graduate Institute of Environmental Sciences, China Medical

University, Hsueh-Shih Road, Taichung, 404, Taiwan

*Correspondence should be addressed:

Chen-Yang Shen (Institute of Biomedical Sciences, Academia Sinica, Academy Road,

Taipei, Taiwan; e-mail: bmcys@ibms.sinica.edu.tw; Phone:+886-2-27899036;

FAX:+886-2-27823047)

Abstract.

Introduction: Estrogen forms a complex with the estrogen receptor (ER) that binds to estrogen response elements (EREs) in the promoter region of estrogen-responsive genes, regulates their transcription, and consequently, mediates physiological or tumorigenic effects. Thus, sequence variants in EREs have the potential to affect the estrogen-ER-ERE interaction. In this study, we examined the hypothesis that genetic variations of EREs are associated with breast cancer development.

Methods. This case-control study involved 815 patients of Asian descent with incident breast cancer and 821 healthy female controls. A total of 13,737 ERE sites in the whole genome predicted by a genome-wide computational algorithm were blasted with single-nucleotide polymorphisms (SNPs) sequences. Twenty-one SNPs located within 2000 base-pairs upstream or within introns 1/2 of putative genes and with a minor allele frequency greater than 5% were identified and genotyped. Frequencies of SNPs were compared between cases and controls to identify SNPs associated with cancer susceptibility.

Results. A significant combined effect of rs12539530, an ERE SNP in intron 2 of *NRCAM* which codes for a cell adhesion molecule, and SNPs of *ESR1*, the gene coding for estrogen receptor, on breast cancer risk was found. Interestingly, this combined effect was more significant in women who had experienced a longer period of life-time estrogen exposure, supporting a hormonal etiology of this SNP in breast tumorigenesis.

Conclusions: Our findings provide support for a role of genetic variation in ERE-*ESR1* in determining susceptibility of breast cancer development.

Introduction.

The roles of estrogen receptor α (ERα) in breast cancer in initiating tumor development, regulating progression, and determining therapeutic protocols and efficacy are well documented [1-3]. Although ERα can be activated in an estrogen-independent manner, the classical activation mechanism involves the ERα binding to estrogen and other coactivator proteins to form the estrogen-bound ER complex, which functions as a transcriptional regulator [4, 5]. The DNA-binding domain of ERα binds to estrogen response elements (EREs) in the promoter region of estrogen-responsive genes, activating or repressing their transcription and, consequently, mediating physiological or tumorigenic effects. Given that sequence variants, such as single-nucleotide polymorphisms (SNPs), located in the promoters of genes have the potential to affect the protein (transcription factor)-DNA (promoter) interaction, resulting in altered expression of target genes [6, 7], it was meaningful to examine the hypothesis that genetic variations of EREs are associated with breast cancer development.

Early work on the *Xenopus* vitellogenin gene identified a minimal ERE core sequence of 5'-GGTCANNNTGACC-3' [8]. Since then, several computational approaches have been used to map EREs on a genome-wide level, based on the presence of EREs within promoter proximal regions [9, 10]. By specifically focusing on promoter regions, 12,515 EREs have been identified in the human genome [10-12]. In order to distinguish between real binding sites and noise, several attempts have been made to improve the specificity of prediction. For instance, by eliminating EREs that are not conserved between the human and mouse genomes, the number of gene proximal EREs has been reduced to 660 [10, 11]. In this study, we used PReMod [13], a new database of

3

genome-wide *cis*-regulatory modules, to predict all possible EREs in the genome. The

prediction algorithm of PReMod takes into account the fact that, in higher eukaryotes,

cis-regulatory regions often contain several phylogenetically conserved binding sites for

different transcription factors [13, 14], and thus has proven to be more reliable than other

methods. Using the SNPs of the human genome available in databases, we searched for

SNPs within these genome-wide predicted EREs and explored their association with

breast cancer.

Materials and Methods.

Study subjects

This case-control study is part of an ongoing cooperative study aimed at understanding the causes of breast cancer in Taiwan, which is characterized by low incidence, early tumor onset, hormone dependency, and novel genomic alterations [15-17]. We studied 815 female breast cancer patients with pathologically confirmed incident primary breast cancer seen at the Tri-Service General Hospital or the Changhua Christian Hospital between March 2002 and August 2007. The 821 healthy female controls were selected from women attending the health examination clinics of the same hospitals during the same period. The characteristics of these study participants have already been described in detail [18-21], and some (551 cases and 727 controls) have recently been genotyped for polymorphism of *ESR1* [21], the gene coding for the ER. This study was approved by the Ethics Committee of the Institutional Review Board of the Academia Sinica, Taiwan, and informed consent was obtained from all study participants before collection of epidemiologic data by personal interview. Considerations regarding methodological issues in the present study (such as study design, sampling scheme, and potential bias) have been described in detail [18-21].

Questionnaire

Experienced research nurses were assigned to administer a structured questionnaire to both cases and controls. The information collected has been described, and the validity of the questionnaire addressed and confirmed, in our previous studies [18-21].

Specimen collection and SNP selection and genotyping.

At the end of the interview, blood was taken for DNA isolation and genotyping. All samples were examined blind by laboratory personnel. DNA was extracted from the peripheral blood samples of patients and controls using DNA purification kits (Promega, USA).

The PReMod database is a genome-wide/transcription factor-wide collection of more than 100,000 computational predicted transcriptional regulatory modules within the human genome [13, 14]. These modules are specific sequences potentially regulated by 229 transcription factor families, and the PReMod algorithm predicts that a total of 13,737 sites within the human genome are bound/regulated by the ER [13, 14]. We compared these sites with conventional SNP databases [22-24], and identified ER-binding sites potentially harboring SNPs. The following three criteria were then used to determine the SNPs to be genotyped: (a) due to statistical considerations (consideration of study power), the minor allele frequency of the selected SNPs had to be higher than 0.05; (b) for biological considerations, the selected SNPs had to be located within 2000 base-pairs upstream or located within introns 1/2 of putative genes; and (c) due to technical considerations, SNPs having the potential to yield a false signal using the iPLEX high-throughput genotyping platform were excluded. As a result, a total of 21 SNPs were chosen for genotyping.

SNPs were genotyped in all samples using Sequenom iPLEX (Sequenom, Hamburg, Germany) technology. Positive, negative, and duplicate controls were included on all plates, with genotypes being autocalled by specialized software (MassARRAY Typer v3.4) and subsequently confirmed by visual assessment of the data. All assays were performed by individuals blinded to the case-control status of the samples. As a quality control, we

repeated the genotyping on 10% of the samples, and all genotype scoring was performed

and checked separately by one reviewer unaware of the case-control status. The

concordance rate for replicate samples was 100%.

Statistical analysis.

Univariate and multivariate analyses were used to determine the risk factors for

breast cancer in this series of study subjects, and the odds ratio (ORs) and corresponding

95% confidence intervals (CIs) were estimated. For individual ERE SNPs, genotype

frequencies were assessed for departure from Hardy-Weinberg equilibrium using either a

X^2 goodness-of-fit test or an exact test. The X^2 test for 2x3 contingency table was used to

compare genotype frequency between cases and controls. To take account of multiple

comparisons, these associations were also assessed using the permutation test provided in

Haploview, run using 10,000 permutations. The association of susceptibility genotypes

and breast cancer risk was further evaluated with simultaneous consideration of

established risk factors for breast cancer or other significant risk factors, in a multivariate

logistic regression models. Biologic plausibility was the most important criterion for

inclusion of variables in the model; therefore, we included all established risk factors of

breast cancer in the statistical models: age, family history of breast cancer, age at

menarche, parity, and age at first full-term pregnancy (FFTP). Adjusted ORs (aORs) and

95% CIs for genotypes were then estimated.

We made use of the information on the *ESR1* polymorphism of our subjects that we

published recently [21] and explored the effect of a possible *ESR1*-ERE interaction or

estrogen-*ESR1*-ERE interaction in determining breast cancer development. A

combination of the joint method and stratified analysis [18-21] was applied to determine

whether this interaction between *ESR1* and ERE was associated with breast cancer formation. A joint effect of *ESR1*-ERE on increased breast cancer risk was explored using conventional logistic regression, a test evaluating whether a statistically significant increase in risk was observed with specific combinations of putative high-risk genotypes in these SNPs (measured by the ß estimates from this regression model). In addition, we stratified our subjects on the basis of their *ESR1* genotype and examined whether breast cancer risk associated with ERE SNPs was particularly significant in specific *ESR1* genotype subsets of women. Because we were especially interested in the relationship between the joint effect of *ESR1*-ERE and breast cancer risk within categories of risk factors representing different levels of estrogen exposure, we performed stratified analysis to test this hypothesis. Therefore, if the identified joint effect of *ESR1*-ERE SNPs initiated breast cancer by the formation of the estrogen-ER complex, then the relationship between breast cancer risk and the joint effect would not be the same in women who had experienced different lengths of estrogen exposure; this was evaluated by calculating the risk (adjusted OR, aOR) of breast cancer associated with the joint effect of *ESR1*-ERE SNPs in women with a longer or a shorter period of total estrogen exposure. For menopausal women, total estrogen exposure was calculated using the formula [age at menopause – age at menarche - years of full-term pregnancy], and, for premenopausal women, age at menopause in this formula was replaced by age at recruitment into this study.

Results.

The risk profile of breast cancer in our study participants was similar to that found in our previous studies [18-21] and in other breast cancer studies. The development of breast cancer was found to be highly associated with reproductive risk factors, including early menarche, nulliparity, lower number of full-term pregnancies (FTPs), and older age at FFTP (Table 1). Compared to controls, cases were younger at menarche ($\leqq$14 years *vs.* >14 years, aOR, 1.52; 95%CI, 1.18-1.96) and older at FFTP (>23 years *vs.* $\leqq$23 years, aOR, 1.30; 95%CI, 1.00-1.69). Significant protection was conferred by a history of FTP (parious women vs. nulliparious women, aOR, 0.67; 95%CI, 0.46-0.97) and a greater number of FTPs (history of >2 FTPs *vs.* history of <=2 FTPs, aOR, 0.46, 95%CI, 0.34-0.61). No association was found between cancer risk and a history of oral contraceptive use or between cancer risk and body mass index. Significant risk factors were included in the multivariate logistic regression models when we examined the association between SNPs and cancer risk. More importantly, these significant associations between reproductive risk factors and breast cancer reveal the importance of the estrogen-related etiology of breast cancer in our participants, providing the opportunity to examine the contribution of EREs during breast tumorigenesis.

A total of 13,737 sites (274,740 bp of DNA) in the whole genome were predicted to be ERE-related sequences using the PReMod algorithm. After blasting these data with online information available from SNP datasets (UCSC, NCBI and HapMap), 322 ER-binding sites were identified as potentially harboring SNPs. One hundred and ten of these 322 SNPs were found to contain no variant allele in the Chinese population, resulting in 212 SNPs, 21 of which met our criteria and were genotyped in cases and

controls (Figure 1).

To determine the breast tumorigenic contribution of ERE SNPs, we examined
whether the genotypic distribution of individual SNPs differed between the cases and
controls (Table S1 in Additional file 1). The frequencies of all SNPs in the controls
agreed with those expected under Hardy-Weinberg equilibrium, suggesting that
genotyping error was relatively unlikely. The result for the genotypic analysis of two
SNPs (i.e. rs12539530 and rs9527676) was important, as it showed that women carrying
the homozygous variant genotype had an significantly increased OR ($P<0.05$) compared
to women carrying the homozygous wild-type genotype, and that the carrying of one
additional risk allele was associated with a significant increase in risk (Table S1 in
Additional file 1). The possibility of false positives due to multiple testing is less likely,
because the permutation test based on 10,000 random permutations showed that these two
associations were borderline significant or significant ($P=0.08$ and 0.03) (Table S1 in
Additional file 1). Both SNPs are located in regulatory regions of genes coding for cell
adhesion molecules, as rs12539530 is located in intron 2 of *NRCAM*, a gene coding for a
neuron-related cell adhesion molecule, and rs9527676 is located in intron 1 of *PCDH17*,
coding for protocadherin-17. To gain initial clues for further analysis, we examined the
expression of these two genes in breast cancer cell lines. To this end, we checked the
expression of *NRCAM* and *PCDH17* in breast cancer cell lines expressing *ESR1* (MCF7
cells) or not expressing *ESR1* (MDA-MB-231 cells) and examined whether the
expression of these putative ER-regulated genes was ER-dependent. The results for
NRCAM were more promising, as they showed that this gene could be expressed in an
ER-positive breast cancer cell (Figure 2), which is consistent with previous findings using

microarrary technology [25, 26]. In addition, cell-cell adhesion, in which *NRCAM* is known to play a part, has been well-documented as involved in cancer formation [27, 28]. All these lines of evidence support the biological plausibility of our findings, suggesting that rs12539530, found in an ERE-related sequence and possibly regulating *NRCAM* expression, is associated with breast cancer susceptibility.

The well-known mechanism in which ER binds to EREs to mediate the expression of ER-regulated genes [2-4] prompted us to speculate whether the SNPs of EREs and *ESR1* are jointly associated with breast cancer. We made use of the information on the *ESR1* genotype of our subjects that we published recently [21], and examined this possibility by both stratified analysis and the joint method. Three SNPs (rs3778609, rs12665044, and rs827421), located in one cluster in intron 1 within the sequence coding for the AF1 domain of the ER, and one SNP (rs7739506) located in intron 4 within the sequence coding for the AF2/ligand binding domain, have been previously found to show significant/borderline significant associations with breast cancer susceptibility in our population [21]. If this ERE SNP were linked to breast cancer susceptibility via the suspected ER-related mechanism, the association between rs12539530 and breast cancer should differ between women harboring different *ESR1* genotypes. Our findings are consistent with this speculation, and the association between high-risk genotypes of rs12539530 and an increased breast cancer risk was only significant in one subset of women carrying specific genotypes of *ESR1* but not in the other subset (Table S2 in Additional file 1). Furthermore, based on a very stringent multiplicative model, the borderline significance p values for the interaction between rs12539530 and rs827421 (*P*= 0.07) and between rs12539530 and rs7739506 (*P*= 0.09) are in line with the

suggestion that an interaction between rs12539530 and *ESR1* polymorphism is linked to breast cancer risk. To confirm this interaction, we used the joint method and calculated the risk of breast cancer associated with both rs12539530 and *ESR1* SNPs using a set of dummy variables representing different combinations of genotypes of *ESR1* and rs12539530. In contrast to the increased, but not significant, risk associated with either a high-risk genotype of *ESR1* or rs12539530 alone, the greatest risk was found in those harboring high-risk genotypes of both rs12539530 and *ESR1* SNPs in all women combined (Figure 3). More interestingly, estrogen promotes breast tumorigenesis by forming a complex with the ER which then binds to EREs [1, 2], and, thus the relationship between breast cancer risk and the joint effect of rs12539530 and *ESR1* polymorphism might be modified by estrogen exposure, and this is consistent with what we observed (Figure 3). The risk of breast cancer associated with the joint effect of polymorphisms of rs12539530 and *ESR1* was evaluated in different groups of women stratified by years of total estrogen exposure and a significant and increased joint effect was only seen in the subgroup of women with more than 30 years of estrogen exposure (Figure 3). In contrast, in the subset of women with less than 30 years of estrogen exposure, the same joint effect was associated with a non-significant aOR (Figure 3).

Discussion.

Of the predicted ERE-related sequences found throughout the whole genome, the present study identified genetic variation of rs12539530, a SNP located in the putative ERE site in intron 2 of *NRCAM*, as an important factor determining susceptibility to breast cancer development. We attempted to address the possibility of false positives and the effects of multiple testing by demonstrating a borderline significant p value in the permutation test. Even so, given the lack of strongly significant results, the power of the present study should be an issue of particular concern, and our suggestion that rs12539530 is involved in determining breast cancer susceptibility certainly needs to be confirmed in other studies with larger sample sizes.

In considering whether our findings represent a true association between this SNP in ERE and breast cancer, the most important issue is the interpretation of the identified association between SNPs and the trait. Because the SNP identified is in an intron, it does not affect amino acid coding and therefore probably does not directly affect protein function; in addition, the observed association between breast cancer and this SNP could also be due to the presence of linkage disequilibrium (LD) between this SNP and other SNPs in exons (resulting in functional change) or in regulatory regions (affecting the expression of these genes). However, the first of these two possibilities is less likely, as we checked the LD block in which rs12539530 is located in the Chinese population in HapMap [24] , which spans 22 kb in chromosome region 7q31, and found that the whole block is within intron 2 of *NRCAM*. In other words, no well-defined genes are in the same haplotype block as this SNP. To address the second possiblity, we examined the sequence of this LD block, and found that it contains more than 200 predicted transcription factor

binding sites, including those for Sp1 and AP1. After blasting these binding sites with the sequences of 30 SNPs known to reside in this block as reported by HapMap, some SNPs were found to be located within these binding sites, so the second possibility cannot be totally excluded. However, the findings (i) that rs12539530 and *ESR1* SNPs jointly increased breast cancer susceptibility, and (ii) that this joint effect was more significant in women with a longer period of estrogen exposure, prompt us to suggest the breast tumorigenic contribution of rs12539530. In addition, this suggestion is mechanistically plausible and is similar to the finding that rs10736303, the SNP generating a putative ERE in intron 2 of *FGFR2*, has been identified as the most significant gene determining breast cancer susceptibility in recent GWAS [29-31]. In addition, expression of *NRCAM*, the gene putatively regulated by rs12539530, has been suggested to be upregulated in ER-positive, but not in ER-negative, breast cancer cell lines [25, 26] and this was confirmed by the present study (Figure 2).

Recent GWAS, based on technological advances in high-throughput genotyping, information regarding the LD of neighboring polymorphisms that makes it possible to use a few hundred thousand SNPs as tags for all other variants, and the biobanking of tens of thousands of specimens allowing their immediate use in whole genome research, have led to the mapping of novel susceptibility loci for many common traits, including breast cancer [for review ref. 30, 31]. Interestingly, none of these loci identified as significant in determining breast cancer risk in GWAS have been identified as significant in previous association studies based on a candidate-gene approach, and very few of the genes involved in the most plausible mechanisms of breast tumorigenesis, including those involved in DNA repair and sex hormone synthesis and metabolism pathways, have been

14

reported as important in GWAS. Region 7q31, in which *NRCAM* is located, has never been reported to be important in breast cancer risk determination in GWAS. This might be partially explained by the relatively low-penetrance effect of SNPs of estrogen-regulating genes. Distinct from those alleles identified by current GWAS, these polymorphic alleles of estrogen-regulating genes would only predispose carriers to a moderately increased risk of developing cancer. Thus, the significance of such SNPs depends not only on their own effect, but also on the interaction between the target genes and other functionally-related genes (e.g. *ESR1*) or the promoting effect of reproductive risk factors reflecting estrogen exposure. Our findings are in line with this suggestion.

This study is a hybrid of candidate-gene and genome-wide approaches, in which we took advantages of both designs. The well-defined roles of the ER during breast tumorigenesis makes it reasonable to assume that polymorphic genetic variants of EREs, a central node in the ER pathway, might underline the variation seen between individuals in their susceptibility to breast cancer. This candidate mechanism lends critical support to the biological plausibility and tumorigenic relevance of our findings. In addition, our genotyping of SNPs, based on genome-wide predicted EREs, provided a unique opportunity to comprehensively examine putative ERE sites without depending on a prior hypothesis. The successful identification of rs12539530 by these combined methods suggests that this is a promising approach to identifying breast tumorigenic contribution of EREs on a genome-wide scale, based on information generated by recent technological advances. For example, in our ongoing study, we are exploring the breast tumorigenic role of putative EREs, identified by chromatin immunoprecipitation-based methods [25, 32, 33]. The simultaneous consideration of these ER-associated EREs and ER-regulating

genes that drive breast tumorigenesis might be critical in the development of new

anticancer drug targets and new therapeutic and diagnostic approaches.

Conclusions.

Though the promise of personalized medicine, in which the risk and the course of diseases and the efficacy of treatment protocols would be predicted on the basis of a person's genotypes, must been tempered with caution, validated molecular tests assessing the patient's germline DNA already drive therapeutic decision-making. Given the well-documented role of ER in breast cancer development and progression, this study explored whether genetic variations in EREs, the sequences bound by ER to activate transcriptional regulation of target genes, are associated with susceptibility of breast cancer. Notably, these ERE sites genotyped were based on genome-wide prediction, providing a unique opportunity to comprehensively examine putative ERE sites without depending on a prior hypothesis. A significant combined effect of rs12539530, an ERE SNP in intron 2 of *NRCAM*, which codes for a cell adhesion molecule, and SNPs of *ESR1*, the gene coding for ER, on breast cancer risk was found. Our findings provide support for a role of *ESR1*-ERE polymorphism in determining susceptibility of breast cancer development. This knowledge will be helpful for directing the focus of future experimental studies.

Abbreviations:

CI: confidence interval; ER: estrogen receptor; ERE: estrogen response element; ERα: estrogen receptor α; FFTP: first full-term pregnancy; FTP: full-term pregnancy; GWAS: genome-wide association studies; LD: linkage disequilibrium; OR: odds ratio; SNP: single nucleotide polymorphism.

Author details.

[1]Departments of Surgery, Tri-Service General Hospital, Ming-Cheng East Road, Taipei, 114, Taiwan; [2]Institute of Biomedical Sciences, Academia Sinica, Academy Street, Taipei, 115, Taiwan; [3]Department of Pharmacy, China Medical University, Hsueh-Shih Road, Taichung, 404, Taiwan; [4]Department of Surgery, Changhua Christian Hospital, Nanxiao Street, Changhua, 500, Taiwan; [5]Department of Radiology, Tri-Service General Hospital, Ming-Cheng East Road, Taipei, 114, Taiwan; [6]Department of Nursing, Kang-Ning Junior College of Medical Care and Management, Kangning Road, Taipei, 114, Taiwan, [7]Institute of Biochemistry and Biotechnology, Chung Shan Medical University, Jianguo North Road, Taichung, 402, Taiwan; [8]Graduate Institute of Environmental Sciences, China Medical University, Hsueh-Shih Road, Taichung, 404, Taiwan

Author's contribution.

JCY and CYS participated in the study design and coordination, carried out the recruitment and drafted the manuscript. CNH and PEW carried out the genotyping and data analysis. BYB contributed to search ERE sites using the PReMod. HMH, STC and GCH carried out the recruitment. WCC and LYH contributed to the validation of *NRCAM* expression in breast cancer cell lines. SlD and CWC performed *ESR1* genotyping.

Competing interests.

The authors declare that they have no competing interest.

References

1. Ali S, Coombes RC: **Endocrine-responsive breast cancer and strategies for combating resistance.** *Nat Rev Cancer* 2002,2:101–112.

2. Yager JD, Davidson NE: **Estrogen carcinogenesis in breast cancer.** *N Engl J Med* 2006,354:270–282.

3. Osborne CK, Schiff R: **Estrogen-receptor biology: continuing progress and therapeutic implications.** *J Clin Oncol* 2005,23:1616–622.

4. Green KA, Carroll JS: **Oestrogen-receptor-mediated transcription and the influence of co-factors and chromatin state.** *Nat Rev Cancer* 2007,7:713-722.

5. Shang Y: **Molecular mechanisms of oestrogen and SERMs in endometrial carcinogenesis.** *Nat Rev Cancer* 2006,6:360-368.

6. Bond GL, Hu W, Bond EE, Robins H, Lutzker SG, Arva NC, Bargonetti J, Bartel F, Taubert H, Wuerl P, Onel K, Yip L, Hwang SJ, Strong LC, Lozano G, Levine AJ: **A single nucleotide polymorphism in the MDM2 promoter attenuates the p53 tumor suppressor pathway and accelerates tumor formation in humans.** *Cell* 2004,119:591-602.

7. Li LC, Chui RM, Sasaki M, Nakajima K, Perinchery G, Au HC, Nojima D, Carroll P, Dahiya R: **A single nucleotide polymorphism in the E-cadherin gene promoter alters transcriptional activities.** *Cancer Res* 2000,60:873-876.

8. Klein-Hitpass L, Schorpp M, Wagner U, Ryffel GU: **An estrogen-responsive element derived from the 5' flanking region of the Xenopus vitellogenin A2 gene functions in transfected human cells.** *Cell* 1986,46:1053-1061.

9. Bajic VB, Tan SL, Chong A, Tang S, Ström A, Gustafsson JA, Lin CY, Liu ET: **Dragon ERE Finder version 2: a tool for accurate detection and analysis of estrogen response elements in vertebrate genomes.** *Nucleic Acids Res* 2003,31:3605–3607.

10. Bourdeau V, Deschenes J, Metivier R, Nagai Y, Nguyen D, Bretschneider N, Gannon F, White JH, Mader S: **Genome-wide identification of high-affinity estrogen response elements in human and mouse.** *Mol Endocrinol* 2004, 18:1411–1427.

11. Welboren WJ, Stunnenberg HG, Sweep FC, Span PN: **Identifying estrogen receptor target genes.** *Mol Oncol* 2007,1:138-143.

12. Carroll JS, Brown M: **Estrogen receptor target gene: an evolving concept**. *Mol Endocrinol* 2006,20:1707-1714.

13. Ferretti V, Poitras C, Bergeron D, Coulombe B, Robert F, Blanchette M: **PReMod: a database of genome-wide mammalian cis-regulatory module predictions.** *Nucleic Acids Res* 2007,35:D122-126.

14. Blanchette M, Bataille AR, Chen X, Poitras C, Laganière J, Lefèbvre C, Deblois G, Giguère V, Ferretti V, Bergeron D, Coulombe B, Robert F : **Genome-wide computational prediction of transcriptional regulatory modules reveals new insights into human gene expression.** *Genome Res* 2006,16:656–668.

15. Lo YL, Yu JC, Huang CS, Tseng SL, Chang TM, Chang KJ, Wu CW, Shen CY: **Allelic loss of the BRCA1 and BRCA2 genes and other regions on 17q and 13q in breast cancer among women from Taiwan (area of low incidence but early onset).** *Int J Cancer* 1998,79:580-587.

16. Shen CY, Yu JC, Lo YL, Kuo CH, Yue CT, Jou YS, Huang CS, Lung JC, Wu CW: **Genome-wide search for loss of heterozygosity using laser capture microdissected tissue of breast carcinoma: an implication for mutator phenotype and breast cancer pathogenesis.** *Cancer Res* 2000,60:3884–3892.

17. Cheng TC, Chen ST, Huang CS, Fu YP, Yu JC, Cheng CW, Wu PE, Shen CY: **Breast cancer risk associated with genotype polymorphism of the catechol estrogen-metabolizing genes: a multigenic study on cancer susceptibility.** *Int J Cancer* 2005,113:345-353.

18. Ding SL, Yu JC, Chen ST, Hsu GC, Kuo SJ, Lin YH, Wu PE, Shen CY: **Genetic variants of BLM interact with RAD51 to increase breast cancer susceptibility.** *Carcinogenesis* 2009,30:43–49.

19. Ding SL, Yu JC, Chen ST, Hsu GC, Shen CY: **Genetic variation in the premature aging gene WRN: a case-control study on breast cancer susceptibility.** *Cancer Epidemiol Biomarkers Prev* 2007,16:263–269.

20. Yu JC, Ding SL, Chang CH, Kuo SH, Chen ST, Hsu GC, Hsu HM, Hou MF, Jung LY, Cheng CW, Wu PE, Shen CY: **Genetic susceptibility to the development and progression of breast cancer associated with polymorphism of cell-cycle and ubiquitin ligase genes.** *Carcinogenesis* 2009,30:1562–1570.

21. Ding SL, Yu JC, Chen ST, Hsu HM, Ho JY, Lin YH, Chang CC, Fann CS, Cheng CW, Wu PE, Shen CY: **Diverse associations between ESR1 polymorphism and breast cancer development and progression.** *Clin Cancer Res* 2010,16:3473-384.

22. UCSC Genome Bioinformatics. [http://genome.ucsc.edu/]

23. NCBI PubMed. [http://www.ncbi.nlm.nih.gov/sites/entrez]

24. HapMap Project. [http://hapmap.ncbi.nlm.nih.gov/index.html.en/]

25. Lin CY, Vega VB, Thomsen JS, Zhang T, Kong SL, Xie M, Chiu KP, Lipovich L, Barnett DH, Stossi F, Yeo A, George J, Kuznetsov VA, Lee YK, Charn TH, Palanisamy N, Miller LD, Cheung E, Katzenellenbogen BS, Ruan Y, Bourque G, Wei CL, Liu ET: **Whole-genome cartography of estrogen receptor alpha binding sites.** *PLoS Genet* 2007,3:e87.

26. Rae JM, Johnson MD, Scheys JO, Cordero KE, Larios JM, Lippman ME: **GREB 1 is a critical regulator of hormone dependent breast cancer growth.** *Breast Cancer Res Treat* 2005,92:141-149.

27. Micalizzi DS, Farabaugh SM, Ford HL: **Epithelial-mesenchymal transition in cancer: parallels between normal development and tumor progression.** *J Mammary Gland Biol Neoplasia* 2010,15:117-134.

28. Thiery JP, Acloque H, Huang RY, Nieto MA: **Epithelial-mesenchymal transitions in development and disease.** *Cell* 2009,139:871-890.

29. Easton DF, Pooley KA, Dunning AM, Pharoah PD, Thompson D, Ballinger DG, Struewing JP, Morrison J, Field H, Luben R, Wareham N, Ahmed S, Healey CS, Bowman R, Meyer KB, Haiman CA, Kolonel LK, Henderson BE, Le Marchand L, Brennan P, Sangrajrang S, Gaborieau V, Odefrey F, Shen CY, Wu PE, Wang HC, Eccles D, Evans DG, Peto J, Fletcher O, *et al.*: **Genome-wide association study identifies novel breast cancer susceptibility loci.** *Nature* 2007, 447:1087-1093.

30. Varghese SJ, Easton DF: **Genome-wide association studies in common cancers-what have we learnt?** *Genet Develop* 2010, 20:201-209.

31. Mavaddat N, Antoniou AC, Easton DF, Garcia-Closas M: **Genetic susceptibility to breast cancer.** *Mol Oncol* 2010, 4:174-191.

32. Welboren WJ, van Driel MA, Janssen-Megens EM, van Heeringen SJ, Sweep FC, Span PN, Stunnenberg HG: **ChIP-Seq of ERalpha and RNA polymerase II defines genes differentially responding to ligands.** *EMBO J* 2009,28:1418-1428.

33. Welboren WJ, Sweep FC, Span PN, Stunnenberg HG: **Genomic actions of estrogen receptor alpha: what are the targets and how are they regulated?** *Endocr Relat Cancer* 2009,16:1073-1089.

Figure legends.

Figure 1. Selection from the genome-wide-predicted estrogen response element-related sequences of single-nucleotide polymorphisms (SNPs) for genotyping.

Figure 2. Transcription (mRNA) of *NRCAM*, detected by reverse-transcriptase-PCR (RT-PCR), is seen in the estrogen-receptor (ER)-positive breast cancer cell line, MCF-7, but not the ER-negative breast cancer cell line, MDA-MB-231. *GAPDH*, the positive (mRNA) control of RT-PCR. For RT-PCR, total RNA was extracted from cell lines using the PicoPur RNA isolation kit (Arcturus, USA), and, to convert RNA into cDNA, reverse transcription was performed for 70 min at 42°C in a reaction volume of 20 μl containing 1 μg of RNA, 10 mM random oligo-dT primer (Promega, Madison, WI), and 5 units of Superscript II reverse transcriptase (Gibco-BRL, Gaithersburg, MD, USA). The presence of a cDNA band of the appropriate molecular weight was determined on 1% agarose gel after electrophoresis.

Figure 3. Association of the joint effect of genotypes of rs12539530, an SNP associated with estrogen response elements (EREs), and of SNPs of *ESR1*, the estrogen receptor gene, with breast cancer risk in all women combined and in women stratified by total years of estrogen exposure (indicated as "Estrogen" in the figure).

Table 1. Frequencies of risk factors in breast cancer cases and controls and the adjusted odds ratio (aOR) in relation to breast cancer risk

Risk factor	%,Cases(N=814)	%,Controls(N=821)	aOR(95%CI)[*]
Age at menarche			
>14 years	34.1	37.6	1.00(ref.)
≤14 years	65.9	62.4	1.52(1.18-1.96)
Menopause			
No	49.6	59.5	1.00(ref.)
Yes	50.4	40.5	0.60(0.42-0.84)
Family history of breast cancer in the first degree relatives			
No	88.7	90.7	1.00(ref.)
Yes	11.3	9.3	1.20(0.83-1.73)
Nulliparity			
No	11.6	10.2	1.00(ref.)
Yes	88.4	89.8	0.67(0.46-0.97)
No. of full-term pregnancy			
≤2	51.0	45.6	1.00(ref.)
>2	49.0	54.4	0.46(0.34-0.61)
Age of first full-term pregnancy			
≤23 years	62.6	64.0	1.00(ref.)
>23 years	37.4	36.0	1.30(1.00-1.69)
Body mass index			
≤23	56.9	59.1	1.00(ref.)
>23	43.1	40.9	0.90(0.71-1.15)
Use of oral contraceptive			
No	81.4	82.1	1.00(ref.)
Ever	18.6	17.9	1.05(0.78-1.40)

[*]The adjusted odds ratio (aOR) and 95% confidence interval (95%CI) were estimated in a logistic regression model, in which the age of study participants was included.

Additional files

Additional file 1: Table S1 to Table S2. Table S1 presents genotype frequencies of sequence variants of estrogen-response- element (ERE)-related sequences in breast cancer cases and controls and the adjusted odds ratio (aOR) in relation to breast cancer risk. Table S2 presents breast cancer risk associated with genotypic polymorphism of rs12539530, an SNP in the estrogen response element-related sequence, stratified by genotypes of *ESR1*.