

Research article

Open Access

Oestrogen receptor α gene haplotype and postmenopausal breast cancer risk: a case control study

Sara Wedrén¹, Lovisa Lovmar², Keith Humphreys¹, Cecilia Magnusson¹, Håkan Melhus², Ann-Christine Syvänen², Andreas Kindmark², Ulf Landegren³, Maria Lagerström Fermér², Fredrik Stiger², Ingemar Persson^{1,4}, John Baron⁵ and Elisabete Weiderpass^{1,6}

¹Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden

²Department of Medical Sciences, Uppsala University, Sweden

³Department of Genetics and Pathology, Uppsala University, Sweden

⁴Swedish Medical Products Agency, Uppsala, Sweden

⁵Dartmouth Medical School, Hanover, New Hampshire, USA

⁶International Agency for Research on Cancer, Lyon, France

Corresponding author: Sara Wedrén, sara.wedren@meb.ki.se

Received: 5 Mar 2004 Revisions requested: 14 Apr 2004 Revisions received: 5 May 2004 Accepted: 11 May 2004 Published: 4 Jun 2004

Breast Cancer Res 2004, **6**:R437-R449 (DOI 10.1186/bcr811)

© 2004 Wedrén *et al.*; licensee BioMed Central Ltd. This is an Open Access article: verbatim copying and redistribution of this article are permitted in all media for any purpose, provided this notice is preserved along with the article's original URL.

Abstract

Introduction Oestrogen receptor α , which mediates the effect of oestrogen in target tissues, is genetically polymorphic. Because breast cancer development is dependent on oestrogenic influence, we have investigated whether polymorphisms in the oestrogen receptor α gene (*ESR1*) are associated with breast cancer risk.

Methods We genotyped breast cancer cases and age-matched population controls for one microsatellite marker and four single-nucleotide polymorphisms (SNPs) in *ESR1*. The numbers of genotyped cases and controls for each marker were as follows: TA_n, 1514 cases and 1514 controls; c.454-397C → T, 1557 cases and 1512 controls; c.454-351A → G, 1556 cases and 1512 controls; c.729C → T, 1562 cases and 1513 controls; c.975C → G, 1562 cases and 1513 controls. Using logistic regression models, we calculated odds ratios (ORs) and 95% confidence intervals (CIs). Haplotype effects were estimated in an exploratory analysis, using expectation-maximisation algorithms for case-control study data.

Results There were no compelling associations between single polymorphic loci and breast cancer risk. In haplotype analyses, a common haplotype of the c.454-351A → G or c.454-397C → T and c.975C → G SNPs appeared to be associated with an increased risk for ductal breast cancer: one copy of the c.454-351A → G and c.975C → G haplotype entailed an OR of 1.19 (95% CI 1.06–1.33) and two copies with an OR of 1.42 (95% CI 1.15–1.77), compared with no copies, under a model of multiplicative penetrance. The association with the c.454-397C → T and c.975C → G haplotypes was similar. Our data indicated that these haplotypes were more influential in women with a high body mass index. Adjustment for multiple comparisons rendered the associations statistically non-significant.

Conclusion We found suggestions of an association between common haplotypes in *ESR1* and the risk for ductal breast cancer that is stronger in heavy women.

Keywords: breast cancer, oestrogen receptor α , gene, haplotype, polymorphism

Introduction

Oestrogenic stimulation is central in breast carcinogenesis and oestrogen receptor α is the most important mediator of the response to stimulation by oestrogens in classical target tissues, such as the breast epithelium. There are several genetic variants of the oestrogen receptor α gene

(*ESR1*), many of which have been studied in relation to breast cancer. Linkage was first described between an intron 1 variant (c.454-351A → G) and breast cancer in a family with late-onset cases [1]. However, later case-control studies have not found any convincing evidence for an association between *ESR1* variants and breast cancer risk

BMD = bone mineral density; BMI = body mass index; CI = confidence interval; EM = expectation-maximisation; *ESR1* = oestrogen receptor α gene; FPRP = false positive report probability; HWE = Hardy-Weinberg equilibrium; OR = odds ratio; PCR = polymerase chain reaction; SNP = single-nucleotide polymorphism.

Table 1

Previously published studies about oestrogen receptor α gene polymorphisms and breast cancer

Reference	Type of study	Polymorphism	Study size	Population	Main effect on risk	Other results
Zuppan [1]	Linkage analysis	c.454-351A→G	11 families	American late-onset familial cases	N/A	Nothing overall, linkage in one family
Hill [46]	Tumor and normal samples	c.454-397C→T	188 tumors, 53 reference samples	Not specified	N/A	c.454-397C→T C allele associated with ER-
Parl [47]	Cases only	c.454-397C→T, c.454-351A→G	59 cases	56 'American-white', 3 African-American, 33-87 years	N/A	c.454-397C→T TT were younger at diagnosis (more often poorly differentiated)
Yaich [13]	Cases only (case-control)	c.454-397C→T	257/140	American pre- and postmenopausal	None	c.454-397C→T TT were younger at diagnosis (slightly more often ER-)
Andersen [41]	Case-control	c.454-397C→T c.454-351A→G	360/672	Norwegian Caucasian, pre- and postmenopausal (27-94 years)	c.454-351A→G: increased risk with G allele c.454-397C→T: None	c.454-351A→G G were older at diagnosis, $P = 0.25$ c.454-397C→T T more often PR-
Roodi [11]	Cases only	c.975C→G, c.729C→T	118 ER+, 70 ER-	American	N/A	c.975C→G G associated with family history of breast cancer, $P = 0.0005$
Iwase [43]	Cases only (case-control)	c.975C→G	13 ER/PR+, 57/30	British	Increased risk with c.975C→G G, $P = 0.057$	
Southey [49]	Case-control	c.975C→G	388/294	Australian early-onset (<40 years)	None	c.975C→G GG v. CC, OR 1.59 (95% CI 0.7-3.6)
Schubert [48]	Familial cases, case-control	c.975C→G, c.729C→T	31+139 familial, 105/151	Caucasian-American and African-American	c.975C→G: No increased risk for familial disease	28% G in relatives with breast cancer and 24% in unaffected relatives, $P = 0.18$
Curran [30]	Cross-sectional association	c.975C→G	125/125	Caucasian-Australian	None	83% c.975C→G C in cases v. 77% in controls, $P = 0.14$
Vasconcelos [44]	Case-control	c.975C→G	70/69	Portuguese	Increased risk with c.975C→G G or GG, OR 2.3 (1.10-5.1)	Fewer lymph node metastases with c.975C→G G or GG, $P = 0.038$
Kang [57]	Case-control	c.975C→G	110/45	Korean	None	More often ER+, PR+ and p53 - with c.975C→G G allele
Comings [50]	Case-control	c.454-351A→G	67/145	Mixed American	Did not contribute to breast cancer variance in a multi-gene model	
Shin [42]	Case-control	c.454-397C→T, c.454-351A→G	205/205	Korean	Decreased risk with c.454-351A→G A allele 0.4 (0.3-0.6)	Association stronger among postmenopausal
Cai [45]	Case-control	c.454-397C→T, c.454-351A→G	1069/1166	Shanghai Chinese	Increased risk with c.454-397C→T, TC OR 1.3 (1.0-1.7) and CC OR 1.4 (1.1-1.8)	

Ranges in parentheses are 95% confidence intervals. ER/PR+/-, oestrogen receptor/progesterone receptor positive/negative; FH, family history; N/A, not applicable; OR, odds ratio.

(Table 1). There is nevertheless an established positive association between breast cancer risk and bone mineral density (BMD) [2], which in turn seems to be influenced by *ESR1* variation [3-5]. Hence, a role of *ESR1* polymorphism in the aetiology of breast cancer seems biologically plausible, although it might have been overlooked in earlier studies because of small sample size (Table 1). We have genotyped five *ESR1* variants in a large, population-based case-control study; a dinucleotide repeat (TA_n) polymorphism (microsatellite) in the promoter region, two single-nucleotide polymorphisms (SNPs) in intron 1 (c.454-397C → T and c.454-351A → G, also known as *PvuII* or IVS1-401 and *XbaI* or IVS1-354, respectively), and one silent SNP each in exon 3 (c.729C → T) and exon 4 (c.975C → G). We estimated the overall influence of *ESR1* genotypes and haplotypes in histopathological subgroups, and according to oestrogen-related breast cancer risk factors, on breast cancer risk.

Methods

Parent study

This nationwide population-based case-control study encompassed all incident cases of primary invasive breast cancer among women 50 to 74 years of age resident in Sweden between October 1993 and March 1995, as previously described in detail [6]. Cases of breast cancer *in situ* were not included. Breast cancer patients were identified at diagnosis through the six Swedish regional cancer registries, to which the reporting of all malignant tumors is mandatory. All Swedish residents are assigned a unique national registration number. This number is recorded in all registries, including the Total Population Register. It is possible for researchers, provided that the appropriate permissions are granted, to approach the authority in charge of the Total Population Register (currently the Tax Authority) and ask for the national registration numbers and addresses of people that fulfil certain criteria specified by the researcher. Control women were randomly selected from the general population according to the expected age frequency distribution (in 5-year age groups) of cases.

Cases were asked to participate in the study by their respective physicians. When patients consented, they received a mailed questionnaire asking for detailed information about intake of menopausal hormones and oral contraceptives, weight, height, reproductive history, medical history, and other lifestyle factors. Controls were contacted directly with the questionnaire. Eighty-four percent of eligible cases ($n = 3345$) and 82% of the controls ($n = 3454$) ultimately participated in the parent study. Among the participating controls, 455 who failed to return the mailed questionnaire were interviewed by phone. Results from the parent study are available in previous publications [6-8].

Selection of present study population

We randomly selected 1500 women with invasive breast cancer and 1500 controls (frequency-matched by age) among postmenopausal participants without any previous malignancy (except *in situ* cervix carcinoma or non-melanoma skin cancer) in the parent study. To increase statistical power in subgroup analyses, we additionally selected all remaining eligible cases and controls who had taken menopausal hormone treatment (either medium-potency oestrogen treatment only or medium-potency oestrogen in combination with progestin) for at least 4 years (191 cases and 108 controls) and all women with self-reported diabetes mellitus (110 cases and 104 controls). In total, 1801 cases and 1712 controls were selected. In addition, 345 controls from the parent study selected for a parallel endometrial cancer study who fulfilled the inclusion criteria could be added to our sample of breast cancer-free controls. The present study was approved by the respective Institutional Review Boards at Karolinska Institutet and Uppsala University and was performed in compliance with the Helsinki Declaration.

Collection of biological samples

We contacted all selected living women by mail, and those who gave informed consent received a blood sampling kit by mail. Whole blood samples were drawn at a primary health care facility close to the woman's home. Breast cancer cases who declined to donate a blood sample were asked to authorise our use of archived paraffin-embedded tissue taken at breast cancer surgery. We also attempted to retrieve archived tissue samples from all deceased breast cancer cases. We obtained blood samples from 1322 and archived tissue samples for 247 breast cancer patients (87% of all selected). Among the chosen control women, 1524 (74%) contributed blood samples. Reasons for non-participation included lack of interest in or scepticism about genetic research, old age and, in some instances, severe disease or death. We thus obtained final population-based participation rates of 73% in cases and 61% in controls.

We isolated DNA from 3 ml of whole blood with the Wizard Genomic DNA Purification Kit (Promega, Madison, WI) in accordance with the manufacturer's instructions. From non-malignant cells in paraffin-embedded tissue we extracted DNA by using a standard phenol/chloroform nmmsdzz protocol [9].

Genetic analyses

SNPs and microsatellite markers

We selected the *ESR1* polymorphisms to be analyzed from the literature [10-13]. All primers and probes for the 5' promoter TA -repeat at - 1174 base pairs upstream of exon 1 [HGVBbase STR000063453], the intron 1 SNPs c.454-397C → T (previously known as *PvuII* or IVS1-401,

[dbSNP rs2234693]) and c.454-351A → G (previously known as *Xba*I or IVS1-354, [dbSNP rs9340799]), and the exonic c.729C → T (codon 243 CGC → CGT, synonymous Arg, [dbSNP rs4986934]) and c.975C → G SNPs (codon 325 CCC → CCG, synonymous Pro, [dbSNP rs1801132]) were designed on the basis of the reference sequences [GenBank AF082876], [GenBank AF326912] and [GenBank NM_000125], respectively. For the fluorescence polarisation and minisequencing assays [14], we designed minisequencing primers complementary to the sequence immediately adjacent to the SNPs. For the Molecular Beacon assay [15], we designed two fluorescently labelled allele-specific probes for each SNP, carrying the variable position in the middle of the loop region [16]. We used a web-based DNA folding program (mfold) [17] to estimate the stability of the stem and loop structure of the Molecular Beacon probes. A full description of the laboratory protocol and also the primer and probe sequences and their modifications are given in Additional file 1.

Minisequencing assay with fluorescence polarisation detection

The region containing the intron 1 SNPs, c.454-397C → T and c.454-351A → G, was amplified in one polymerase chain reaction (PCR) fragment from 20 ng of genomic DNA. The PCR products were treated enzymatically to remove remaining primers and nucleotides, and the minisequencing reactions were performed with the two dideoxynucleotides relevant for the particular SNP fluorescently labelled and included at a 1:5 ratio relative to unlabelled dideoxynucleotides. The fluorescence signals were read on an Analyst AD™ (Molecular Devices Corporation, Sunnyvale, CA); genotypes were assigned by the software supplied with the instrument, and a custom-made Excel macro. In each run, positive controls for the three genotypes and negative controls were included. Both DNA polarities were analysed and the results were concordant in all samples. About 30% of the assays were repeated; the results were identical. In addition, 3% of the genotypes were validated by solid-phase minisequencing.

Molecular Beacon assay

To analyse the exonic SNPs c.729C → T and c.975C → G, 10 ng of genomic DNA was amplified with real-time PCR monitoring of fluorescence signals with an ABI Prism 7700 Sequence Detection System (Applied Biosystems). The increase in fluorescent signal was registered during the annealing step of the reaction, and the end-point signals were used to assign the genotypes as previously described [16]. In each run, positive controls for the three genotypes and negative controls were included. Both possible nucleotides of the SNP were interrogated in the same reaction. About 3% of the assays were repeated; the results were identical. Two percent of the assays were val-

idated by solid-phase minisequencing. The Molecular Beacon assay has previously been quantitatively validated in our laboratory [16].

Solid-phase minisequencing assay

Solid-phase minisequencing in microtitre plates [18] was used in part to genotype the c.729C → T SNP. The assay also served as a reference method for the other SNP assays.

Microsatellite assay

The TA-repeat region was amplified, with one PCR primer fluorescently labelled, using an ABI-877 Integrated Thermal cycler PCR robot with standard reagents (Applied Biosystems) [19]. The PCR products were separated by using a 96-well ABI 377 automatic sequencer and analysed with software supplied with the instrument (all Applied Biosystems). In each run, two or three negative controls were included. About 1% of the assays were repeated; they gave concordant results. After calculation of actual number of TA repeats there was no difference between repeated runs of the same sample. The assay was validated by control sequencing of three different repeat lengths.

Statistical methods

We tested for Hardy–Weinberg equilibrium (HWE) among cases and controls separately. We considered the data for TA_n in its original form, as eight categories, or dichotomised into long and short (not more than 14 repeats or more than 14 repeats), because the TA_n lengths were bimodally distributed with peaks at 11 and 18 repeats, with a dip at 14 repeats.

We also estimated all pairwise linkage disequilibrium values $|D'|$ and r^2 [20,21]. Although there are disadvantages associated with r^2 compared with $|D'|$ for linkage disequilibrium mapping [22,23], r^2 is arguably the most relevant measure for association analysis, because there is a simple inverse relationship between r^2 and the sample size required to detect association between susceptibility loci and SNPs [21].

For association analyses, using single loci genotypes, we used conditional logistic regression [24] to calculate odds ratios (ORs) and 95% confidence intervals (CIs). We conditioned on the variables used for selection, namely age in 5-year categories, long-term use of menopausal hormones, and diabetes mellitus (see the section on Selection of present study population above). We evaluated possible associations between *ESR1* polymorphisms and other exposures/covariates by scrutinising $2 \times k$ tables and calculating χ^2 statistics. Continuous variables were categorised for this purpose. Where there seemed to be plausible evidence of association, the exposure was considered either a potential confounder or a factor in the causal path-

way between *ESR1* and breast cancer, and was tested as such by introducing it into the logistic regression model.

We tested for disease–haplotype association by using likelihood-based approaches. We used the software EH plus [25] as well as routines written in the S-PLUS (Insightful) programming language. Most available software (including EH plus) for testing association between haplotypes and case/control status assume multiplicative penetrance; that is, the OR comparing two haplotype copies against none is assumed to be the square of the OR comparing one against none. Under this assumption, together with the assumption of HWE in the population and the assumption of a rare disease, it can be shown that cases, as well as controls, will be in HWE. This means that it is possible to perform a likelihood ratio test by comparing a model that infers haplotype phase for cases and controls separately (both under HWE, with different haplotype frequencies) with a model that infers haplotype phase for cases and controls jointly (that is, under HWE, identical haplotype frequencies). For the most part we have used this approach (using EH plus). We have also written our own program to estimate haplotype–case/control status association. The method in this program is in essence identical to that used by Stram and colleagues [26]. This makes use of sampling fractions that are assumed to be known (in practice estimated from population register data; our program requires as input the ratio of sampling fractions between cases and controls). With this ascertainment information we can in principle estimate any model (we are not restricted to particular penetrance assumptions). For our program we consider the likelihood $\text{pr}(y|x,s = 1)$, where y is case/control status, x represents covariate information and s is an indicator variable for whether a subject is selected to the sample. We essentially adapt the approach of Neuhaus [27], specifically to the situation in which there is missing covariate information, in this case haplotype phase. The program is able to estimate models other than multiplicative penetrance and enables adjustment for other covariates. Given that the present study is based on large age strata we wished to adjust for this by including age group as a covariate. A likelihood ratio test did not give evidence against a model of multiplicative penetrance (we compared the goodness of fit of the multiplicative penetrance model to the goodness of fit of a model that specifies an individual risk for each unique haplotype pair). Our program also represents a convenient framework within which to estimate haplotype–environment interaction. Results reported here for models involving haplotypes are based on our own program.

We fit a variety of models of association between *ESR1* variants and breast cancer risk. We chose to fit models separately for lobular and ductal cases because there is an indication in the literature that these two histotypes have

partly different aetiologies. For single-locus genotype effects we opted to include a parameter for each genotype (AA, Aa, aa) rather than test a battery of models with a specified penetrance (dominant, recessive, multiplicative penetrance/allele counting), which have one degree of freedom less. For haplotype effects we fitted multiplicative penetrance models (see above).

Because we fit several models we also make adjustments for multiplicity. We use a permutation-based approach that controls the family-wise error rate (probability of rejecting one or more true null hypotheses of no association). This is based on the permutation step-down procedure of Westfall and Young [28], and takes into account the dependence structure of the polymorphisms/hypotheses. The permutation approach to multiple testing is computationally demanding, particularly when haplotype phase has to be inferred, and for this reason we adjust only P values for the ductal cases versus control group tests. This is reasonable because the power of the individual tests for lobular cases versus controls is weak; unlike Bonferroni procedures, for the permutation approach that we use, a completely meaningless test (power ≈ 0) does not affect the other tests at all because it will never affect the distribution of the minimum P -value statistic in the lower α tail. We additionally remove the TA repeat, because of deviance from HWE (see below), and c.729C \rightarrow T, because of low rare-allele frequency, when adjusting for multiplicity. This means that we apply a multiplicity adjustment to P values from tests of seven dependent hypotheses.

Results

General

We were able to genotype the SNPs in more than 99% of the biological samples. We obtained complete information in all five markers from 1512 cases and 1511 controls.

For those participating by means of a blood sample or by means of a normal tissue sample, the mean time between breast cancer diagnosis and enrolment in the present study was 5.1 years (range 6.8–3.5 years) and 5.9 years (range 7.5–4.6 years), respectively. Breast cancer stage was more advanced in those who participated by means of normal tissue samples; 58% of tissue samples were from cases with stage 2 or more advanced stages, whereas the corresponding number was 38% in cases who donated blood ($P < 0.0001$). The mean age of those who donated a blood sample was similar in cases and controls, whereas those for whom we used tissue were on average 1.6 years older. Non-participants were on average slightly older than genotyped participants. There were no notable differences or trends in genotype frequencies over quartiles of time between breast cancer diagnosis and time of blood donation (data not shown); neither were there any differences in genotype frequencies between those who donated a blood

Table 2

Odds ratios for ductal and lobular breast cancer in relation to single locus genotype

Genotype ^a Polymorphism	00			01			11			<i>P</i> _{HWE} ^c
	Cases/ Controls	OR ^b	95% CI	Cases/ Controls	OR ^b	95% CI	Cases/ Controls	OR ^b	95% CI	
Promoter TA _n										
Ductal	369/445	1	Ref	494/593	1.01	0.84–1.21	230/312	0.87	0.70–1.09	<0.001, <0.0001 ^d
Lobular	52/445	1	Ref	81/593	1.30	0.88–1.91	33/212	0.92	0.57–1.49	
Intron 1 c.454-397C→T										
Ductal	347/384	1	Ref	542/651	0.93	0.77–1.12	229/313	0.80	0.64–1.01	0.32
Lobular	43/384	1	Ref	92/651	1.42	0.95–2.11	39/313	1.18	0.73–1.90	
Intron 1 c.454-351A→G										
Ductal	522/577	1	Ref	471/610	0.86	0.72–1.02	124/161	0.86	0.66–1.12	0.88
Lobular	66/577	1	Ref	89/610	1.36	0.96–1.92	19/161	1.05	0.60–1.84	
Exon 3 c.729C→T										
Ductal	1037/ 1253	1	Ref	84/93	1.07	0.78–1.45	1/3	0.41	0.04–4.00	0.20
Lobular	164/1253	1	Ref	10/93	0.92	0.46–1.83	0/3	-	-	
Exon 4 c.975C→G										
Ductal	722/805	1	Ref	346/477	0.81	0.68–0.96	54/67	0.93	0.64–1.35	0.87
Lobular	106/805	1	Ref	59/477	0.94	0.66–1.34	9/67	1.00	0.48–2.10	

OR, odds ratio. ^aThe genotypes are as follows: TA_n, 00 ≤ 14/≤ 14, 01 ≤ 14/>14, 11 >14/>14; c.454-397C → T, 00 T/T, 01 T/C, 11 C/C; c.454-351A → G, 00 A/A, 01 A/G, 11 G/G; c.729C → T, 00 C/C, 01 C/T, 11 T/T; c.975C → G, 00 C/C, 01 C/G, 11 G/G. ^bLong-term users of menopausal hormones and diabetics were oversampled; thus, logistic regression models are adjusted for age and sampling category, which means that only those with complete information about hormone use and diabetes mellitus were included. ^c*P* value for χ^2 test of deviation from the assumption of Hardy–Weinberg equilibrium, among controls only. ^dFor this test for deviation from Hardy–Weinberg equilibrium either all microsatellite alleles were considered or lengths were categorised into ≤10, 11, 12 or 13, 14–16, 17, 18, 19, or ≥ 20.

sample and those for whom we used normal tissue (data not shown). Further, there were no significant differences in genotype frequencies over breast cancer stages at diagnosis. As expected, mean age at first birth and mean number of births reflected known case-control differences but were largely the same between participants and non-participants. The allele and genotype frequencies in our study population were similar to what has previously been found in Caucasian populations [11,29,30].

Among controls, we did not find any convincing associations between any of the studied *ESR1* polymorphisms singly and height, body mass index (BMI), smoking, diabetes mellitus, age at menarche, age at first full-term birth, parity, age at menopause, menopausal hormone use, weight gain during adult life, alcohol intake, history of benign breast disease, first-degree family history, or use of oral contraceptives (data not shown).

Association with breast cancer

Single loci

There were no compelling overall relations between any of the studied polymorphisms singly and breast cancer risk (Table 2). If anything, there seemed to be a slight negative association between homozygosity for any of the two intron 1 SNPs and heterozygosity for the c.975C → G SNP and ductal cancer risk, but only one estimate was significant. The association patterns between single loci genotype and lobular cancer risk differed somewhat from those regarding ductal cancer risk, but the lack of power precluded a conclusive comparison. In secondary analyses, we tested different cut-points for the TA_n but found that no choice of cut-point would have resulted in a significant association between this marker and breast cancer risk.

Haplotypes

Haplotypes describe the genetic make-up more thoroughly than SNPs. Thus, our further analyses of the association between *ESR1* polymorphism and breast cancer develop-

Table 3**Pairwise linkage between polymorphisms in the oestrogen receptor α**

r^2	$ D' $	TA_n^a	c.454-397C→T	c.454-351A→G	c.729C→T	c.975C→G
	TA_n^a	-	0.784	0.727	0.122	0.161
	c.454-397C→T	0.569	-	0.998	0.895	0.329
	c.454-351A→G	0.337	0.587	-	0.872	0.242
	c.729C→T	0.001	0.028	0.016	-	0.848
	c.975C→G	0.006	0.029	0.009	0.097	-

^aFor $|D'|$ TA_n repeat lengths were divided into eight categories, ≤ 10 , 11, 12 or 13, 14–16, 17, 18, 19, or >20 . This ensured that there were few cells that contained fewer than five observations. For r^2 the TA_n repeat is dichotomised. ^bAll $|D'|$ values are statistically significant at $\alpha = 0.001$.

Table 4**Distribution of ESR1 four-locus haplotype frequencies as estimated through expectation-maximisation algorithms**

Locus				Proportion		
c.454-397C→T	c.454-351A→G	c.729C→T	c.975C→G	Controls ($n = 1511$)	Ductal cases ($n = 1148$)	Lobular cases ($n = 184$)
T	A	C	C	0.369	0.404	0.373
C	G	C	C	0.285	0.267	0.276
C	A	C	C	0.114	0.117	0.125
T	A	C	G	0.123	0.117	0.123
C	G	C	G	0.058	0.049	0.076
T	A	T	G	0.032	0.027	0.021
Proportion accounted for by the above most common haplotypes:				0.981	0.981	0.994

ment, overall and stratified according to hormonal factors, are confined to the influence of different haplotypes.

There is no evidence of extensive historical recombination between the typed SNPs (pairwise $|D'|$ values 0.998, 0.872, 0.848; Table 3). However, r^2 values indicate that the allelic association is not strong.

Table 4 shows the estimated haplotype frequencies based on the c.454-397C → T, c.454-351A → G, c.729C → T and c.975C → G SNPs. Of the controls, 65% carried either of the two most common haplotypes. Out of 16 possible haplotypes, only 6 were represented among more than 98% of the women in our study. Although the genotype frequencies of the four SNPs were in accordance with HWE, the genotype frequency of the TA_n was not, neither when all alleles were considered nor when the repeat lengths were categorised ($P < 0.0001$). We therefore excluded the TA_n marker from haplotype reconstruction because it did not fulfil the necessary assumptions. Furthermore, the c.729C → T marker was expelled from further

haplotype analyses, because the minimal prevalence of the rare allele would not allow meaningful inference.

In Table 5 we present P values, unadjusted and adjusted for multiple comparisons, for the main effects tested, beginning with single-locus associations. Next we explored the prevalence of three two-locus haplotypes among cancer cases versus controls (Table 5). We found an association between *ESR1* and ductal, but not lobular, cancer risk in a haplotype analysis based on the c.975C → G marker in combination with either c.454-397C → T or c.454-351A → G (Table 5). Likelihood-ratio tests evaluating models in which each haplotype carries its own risk (that is, a variable with four categories) were statistically significant ($P = 0.019$ and 0.022 , respectively; $df = 3$). The corresponding P values adjusted for multiple comparisons were 0.07 and 0.08, respectively.

The three-SNP haplotype composed of the c.454-397C → T, c.454-351A → G and c.975C → G SNPs was not more strongly associated with breast cancer risk than any of the

Table 5

P values from single genotype and haplotype association tests

SNP/haplotype tested	Unadjusted		Adjusted for multiple testing Control v. ductal	df ^a
	Control v. ductal	Control v. lobular		
Single loci				
c.454-397C→T	0.154	0.458	0.157	2
c.454-351A→G	0.128	0.292	0.235	2
c.729C→T	0.524	0.440	.b	2
c.975C→G	0.037	0.845	0.106	2
Two-locus haplotypes				
c.454-397C→T and c.454-351A→G	0.152	0.918	0.219	3
c.454-397C→T and c.975C→G	0.019	0.921	0.071	3
c.454-351A→G and c.975C→G	0.022	0.930	0.077	3
Three-locus haplotype				
c.454-397C→T and c.454-351A→G and c.975C→G	0.097	0.995	0.178	7

^aDegrees of freedom. Genotype and haplotype frequency distributions are compared between cases and controls without prior assumptions about high-risk or low-risk genotypes or haplotypes. ^bExcluded from adjustments for multiple testing because of low allele frequency (see the text).

two-SNP haplotypes based on the c.975C → G marker in combination with either c.454-397C → T or c.454-351A → G (Table 5).

Table 6 accounts for the age-adjusted relative risk, overall and stratified according to breast cancer risk factors, for ductal breast cancer in relation to the c.454-351A → G or c.454-397C → T and c.975C → G haplotypes (AC and TC, respectively). Under a model of multiplicative penetrance, the OR for carrying two copies of the haplotype compared with none is the square of the estimates for carrying one copy of the haplotype compared with carrying no copy. To save space we present only the latter estimates. The OR for carrying one copy of the AC haplotype was 1.19 (95% CI 1.06–1.33) compared with carrying no copy. The association was confined to overweight women (BMI > 25) and seemed more pronounced among those with a BMI of more than 28 kg/m². There was a similar pattern of risk for the TC haplotype. Further stratification revealed an even stronger relation; when considering only women with BMI > 30 with one AC compared with none, we found an OR of 1.48 (95% CI 1.08–2.03) and among those with BMI > 32 an OR of 1.60 (95% CI 1.02–2.49). The corresponding ORs for the TC haplotype were 1.59 (95% CI 1.15–2.21) for BMI > 30 and 1.71 (95% CI 1.08–2.70) for BMI > 32. The effect of the AC or TC haplotype was more pronounced in uniparous women but there was no trend over number of pregnancies. There were no indications of interaction between AC or TC haplotype and menopausal hormone use, family history, years of menstruation or age at menopause.

The AC haplotype seemed to be more common in controls with low BMI (Table 6). A test of association between BMI in two categories (less than 28 and 28 or more) and AC haplotype versus the other three possible haplotypes yielded $P = 0.09$. There was no association between BMI and TC haplotype ($P = 0.48$). In a model with BMI in two categories the P value for interaction between haplotype and BMI was 0.12 for the AC and 0.047 for the TC haplotype.

Discussion

Our results indicate that natural allelic variation in *ESR1* might be associated with postmenopausal ductal breast cancer risk; common haplotypes, composed of weakly linked markers, were in our data associated with increased breast cancer risk. The associations seemed the most pronounced in groups with high BMI.

Our study has strengths in that it is population-based, large, and set in a comparatively homogenous population with regard to ethnicity and menopausal status. In addition, we had some possibility of evaluating potential interaction with other breast cancer risk factors.

To our knowledge, no previous study has considered the influence of three-marker haplotypes in *ESR1* on breast cancer risk. Two of the loci studied are in untranslated regions, and one locus is a base-pair exchange in the third codon position that does not alter the resulting amino acid. The polymorphisms could theoretically have direct regulatory roles (see below) but can also be regarded as markers, potentially in linkage disequilibrium with a functional locus or loci.

Table 6**Ductal breast cancer risk in relation to high-risk haplotypes, stratified by breast cancer risk factors**

Factor	Number, cases/ controls	c.454-351A→G c.975C→G AC haplotype		c.454-397C→T c.975C→G TC haplotype	
		Haplotype proportion, cases/controls	OR (95% CI) ^a , 1 v. 0 copies ^b	Haplotype proportion, cases/controls	OR (95% CI) ^a , 1 v. 0 copies ^b
All ductal cancer cases	1148/1511	0.531/0.484	1.19 (1.06–1.33)	0.412/0.373	1.19 (1.06–1.34)
BMI					
<25	549/761	0.525/0.517	1.04 (0.89–1.23)	0.404/0.388	1.08 (0.91–1.27)
25–<28	299/429	0.531/0.458	1.32 (1.06–1.64)	0.391/0.348	1.21 (0.97–1.51)
≥ 28	300/321	0.541/0.456	1.41 (1.12–1.77)	0.449/0.357	1.48 (1.17–1.88)
Parity					
0	158/134	0.495/0.472	1.14 (0.82–1.60)	0.399/0.388	1.11 (0.80–1.54)
1	248/258	0.535/0.439	1.43 (1.10–1.84)	0.422/0.349	1.35 (1.04–1.74)
2+	706/1043	0.535/0.496	1.17 (1.01–1.34)	0.413/0.377	1.17 (1.02–1.35)
Menopausal hormone treatment					
<2 years	861/1169	0.533/0.482	1.23 (1.08–1.40)	0.416/0.374	1.21 (1.06–1.38)
≥ 2 years	251/266	0.515/0.490	1.08 (0.84–1.40)	0.391/0.369	1.08 (0.84–1.40)
Family history					
No	935/1246	0.528/0.486	1.18 (1.04–1.33)	0.413/0.368	1.21 (1.07–1.30)
Yes	183/128	0.548/0.505	1.17 (0.84–1.63)	0.403/0.425	0.93 (0.67–1.30)
Age at menopause					
≤ 49 years	324/506	0.555/0.496	1.25 (1.02–1.54)	0.431/0.385	1.25 (1.01–1.54)
49.5–51.5 years	445/503	0.511/0.470	1.19 (0.98–1.43)	0.397/0.356	1.19 (0.98–1.45)
≥ 52 years	372/489	0.534/0.494	1.18 (0.97–1.44)	0.411/0.372	1.17 (0.96–1.44)

BMI, body mass index. ^aAge-adjusted odds ratio and 95% confidence interval. ^bThe odds ratios (ORs) and CIs for two copies versus none are not shown because under a model of multiplicative penetrance the OR for two copies versus none is the square of the OR for one copy versus none.

If the underlying model for the disease is that combinations of SNP alleles are causally important, it is essential to define haplotypes and to use them as the unit of exposure in the analysis. A possible explanation for the observed association, apart from its being a chance finding, is that there is such a functional combination of SNP alleles; that is, that the two markers (or loci linked to the markers) in combination alter the function of the gene, such as RNA stability or the translation machinery [31]. Functional combinations of SNP alleles composed of polymorphisms in non-coding regions have been shown to exist [32].

If in contrast it seems likely that there is only one disease locus, haplotypes are useful in that they can help to make the study of a gene more efficient through reducing the number of SNPs that need to be typed in the entire study population. This can be accomplished by making use of the linkage disequilibrium between markers in a gene and

selecting for subsequent typing the SNPs that best capture the haplotype structure of a gene, the so-called haplotype tagging SNPs.

Our four SNP markers and one dinucleotide repeat were not specifically selected to define *ESR1* haplotypes. Rather, they were chosen because they were known to be polymorphic at the time when the study was planned. Clearly, the ability to capture the gene's haplotype diversity is limited by the use of only four markers, of which only three exist with a substantial prevalence. Our design is not optimal in the light of current knowledge. It would be more unbiased to address the problem by using strategies in which a comprehensive set of SNPs are identified and validated and subsequently a smaller set are chosen specifically to tag the important haplotypes in the gene. Our intention here was merely to make the best possible use of data that were already available.

A weakness in our study is that, despite its size, we had only a limited potential to investigate possible interactions between haplotype and other breast cancer risk factors. Because it is plausible that moderate genetic effects are manifested only in the presence of other exposures, it is highly desirable to be able to perform analyses of interaction.

The error of haplotype frequency estimation by the expectation-maximisation (EM) algorithm has been shown by to be low under various conditions with regard to, for example, heterozygosity, haplotype frequency distributions, and linkage disequilibrium [33]. Haplotype frequency estimation with EM algorithms entails assuming that the single marker genotype frequencies are in HWE. This assumption is easily testable. In this study we could not establish HWE for the TA_n among controls; this polymorphism was therefore not included in haplotype reconstruction. The reason for deviation of the TA_n from HWE is unclear. If the reason for the deviation, contrary to our belief, were due to genotyping error and if the error were random with regard to case-control status, association between the TA_n and breast cancer would be underestimated.

Another supposition used for our disease-haplotype association estimation is a model of multiplicative penetrance. Multiplicative penetrance, rather than dominant, is likely in the action of common genetic variants because their effect is thought to be modest. If instead a recessive model were correct, our estimates would be conservative. In our data, a model of multiplicative penetrance could not be rejected.

Selection bias is a potential problem in case-control studies, and our participation rates, calculated by using those eligible for the parent study in the denominator (73% and 61% in cases and controls, respectively), could lead to spurious findings but only if participation were related to genotype or haplotype. Survivor bias might be a concern in our study because death and severe disease were reasons for non-participation. If a genetic variant is associated with severe breast cancer but not with less severe breast cancer, and if the more severe cases of breast cancer are less likely to participate because they have died, any association with breast cancer overall would be biased towards the null. One can also view this as an issue of generalisability; in other words, our findings do not pertain to more aggressive breast cancer. However, in our data genotype frequencies did not vary according to time between diagnosis and enrolment through blood versus tissue donation (the latter mainly representing deceased cases). Neither were genotype associations appreciably different across stages at breast cancer diagnosis. We therefore conclude that survivor bias is not a major problem in our study.

Some observations make functional consequences of some of the five studied *ESR1* polymorphisms plausible. The TA_n microsatellite is located in the promoter region and, as directly exemplified by Enattah and colleagues [34] in a study about lactose intolerance, promoter variants can severely affect gene function. Herrington and colleagues [35] found that the IVS-401 C allele contains a potential binding site for Myb transcription factors and showed that this can augment transcription up to 10-fold. In general, our knowledge about the functional significance of non-coding DNA sequences is not fully formed, but there is evidence that upstream, downstream and intronic sequences have important regulatory roles [36].

There are some reports that suggest connections between *ESR1* variants and breast cancer risk factors. Ushiroyama and colleagues found that Japanese women with the c.454-397C → T CT genotype had higher plasma and serum oestradiol levels than those with the TT genotype [37]. In a Dutch population, c.454-397C → T TT was associated with an earlier onset of menopause, which would entail a decreased risk for breast cancer [38]. In contrast, Deng and colleagues [39] found c.454-397C → T TT to be associated with higher BMI and against a tendency to gain weight with age, characteristics previously shown to be associated with an increased risk for breast cancer. These associations between the c.454-397C → T variant and onset of menopause, BMI, and weight gain were not corroborated by our data. However, we did find a tentative association between BMI and a haplotype constructed by the c.454-351A → G SNP and the c.975C → G SNP. In a previous study about endometrial cancer, a disease closely related to oestrogen exposure, we found that the c.454-397C → T CC and c.454-351A → G GG genotypes were associated with a decreased risk for endometrial cancer [40].

The previous literature of *ESR1* polymorphism in relation to breast cancer risk is inconclusive (Table 1). In contrast to our findings, one study showed an increased risk with c.454-351A → G G allele [41] or decreased risk with the c.454-351A → G A allele [42], others showed an increased risk with the c.975C → G G allele [43,44]. In a recent large Chinese study [45] the c.454-397C → T C allele was associated with an increased breast cancer risk, which is contrary to the tendencies in our data for ductal cancers but in line with those for lobular cancers. Histotypes were not reported in the Chinese study but the majority were most probably ductal cancers. The remaining investigators did not find any association with breast cancer risk [11,13,30,46-50].

BMI is the most important determinant of oestrogen levels in postmenopausal women. Our finding that the association grew stronger as we considered women with increasing

BMI is biologically plausible and would, if it is real, indicate that *ESR1* variation is more influential in the presence of higher levels of oestrogen. Contradicting this theory, however, is the fact that we could not see any corresponding influence of duration of menopausal hormone treatment.

The association between BMD and *ESR1* polymorphism is by far the most closely investigated, in particular with regard to the intron 1 SNPs. Any variant in *ESR1* that seems to increase bone mass or to decrease risk for osteoporosis and fractures would be expected to increase breast cancer risk because it might indicate a greater influence of oestrogen. However, the results of studies of BMD and *ESR1* polymorphism are conflicting. Among the largest studies, some have shown, in line with the suggestions in our data, that the c.454-397C → T T allele or TT genotype confers the highest BMD, that is, the strongest oestrogenic influence [3,51], whereas others found the highest BMD in those with the C allele or CC genotype [52]. Similarly, for the c.454-351A → G locus, although some, supported by our findings, found the A variant to be associated with higher BMD [3,51,53], others found high BMD with the G variant [54].

The quest for disease-causing genes has grown markedly over the years. Nevertheless, the efforts have yielded few lucid results. It is likely that the genetic alterations we are looking for are of low penetrance and thus implicitly need co-action or interaction with other exposures to show strong associations. Another plausible scenario is that breast cancers arise from various genetically distinct causal mechanisms, in other words that genetic heterogeneity is present, which acts to blur the gene–disease associations under study. Population history might also contribute to an explanation of previous diverging results in different populations such that the causative variant might be linked to one marker allele in one population but with the other allele in another population. A contributing fact is that many studies have lacked sufficient power to capture such biological complexity either because of small sample size or, as has been argued by some, because of inappropriate study design [55].

When initiating this study we had strong belief in the possibility of *ESR1* variation's being involved in breast carcinogenesis; there is undisputed evidence of oestrogens' causal role in breast carcinogenesis, and oestrogenic action is mediated through oestrogen receptor α . Furthermore, there were existing data indicating that certain *ESR1* variants are associated with oestrogenic action. Wacholder and colleagues [56] suggest making use of prior belief in a genetic variant when trying to determine whether a statistically significant finding is noteworthy. They recommend calculating the false positive report prob-

ability (FPRP) by using the power of the study to detect the particular finding, the *P* value for the association, and the prior belief. Our prior estimate was that the chance of association between *ESR1* and breast cancer was at least 10%. The power for detecting an odds ratio of 1.19 for the AC haplotype at an α level of 0.05 in our sample was 0.79 when we compared the AC with the next most common haplotype (AC being the most common). *P* for the association was 0.02. With our own prior of 0.1 the FPRP was 0.18, which makes our finding noteworthy. However, many previous studies have failed to establish a role for *ESR1* in breast carcinogenesis, which decreases our priors. Yet, because most previous studies were small and because no other study has considered haplotypes between the SNPs in intron 1 and exon 4 the influence on the prior would not be large. Nevertheless, our design might not have captured the full haplotype diversity of *ESR1*. The belief in haplotypes based on the few markers that we have chosen is clearly lower. If, say, prior belief is then 0.01, the FPRP becomes a substantial 0.71.

Conclusions

We found suggestions of an association between *ESR1* haplotypes and the risk for postmenopausal ductal breast cancer of mild to moderate severity, although this is conceivably a false positive finding. This association seemed stronger as we considered women with higher BMI. If these haplotypes truly entail an increased breast cancer risk, owing to their high population prevalence, they have the potential for substantial role in breast cancer aetiology overall.

Competing interests

None declared.

Additional files

The following Additional files are available online:

Additional File 1

Description of methods used for *ESR1* analysis. See <http://breast-cancer-research.com/content/supplementary/bcr811-s1.doc>. See <http://www.biomedcentral.com/content/supplementary/bcr811-S1.doc>

Acknowledgements

We thank all the women who participated in this study. We also thank Anna Christensson and Boel Bissmarck for their endurance during study subject recruitment; Maria Branting, Anders Westermark, Erika Svensson, Kristina Larsson and Birgitta Sundelin for excellent technical assistance; and Hans-Olov Adami for initiating the study and guiding it through to the end. Furthermore, we thank primary health care centers and pathology departments all over Sweden for their cooperation. This

study was supported by the National Institutes of Health, grant number 5 RO1 CA 77973-03, by the United States Army, Department of Defense, award number DAMD17-97-1-7322, and by the K&A Wallenberg foundation (Wallenberg Consortium North). KH was supported by a grant from the Wallenberg Consortium North.

References

- Zuppan P, Hall JM, Lee MK, Ponglikitmongkol M, King MC: **Possible linkage of the estrogen receptor gene to breast cancer in a family with late-onset disease.** *Am J Hum Genet* 1991, **48**:1065-1068.
- Newcomb PA, Trentham-Dietz A, Egan KM, Titus-Ernstoff L, Baron JA, Storer BE, Willett WC, Stampfer MJ: **Fracture history and risk of breast and endometrial cancer.** *Am J Epidemiol* 2001, **153**:1071-1078.
- Yamada Y, Ando F, Niino N, Ohta S, Shimokata H: **Association of polymorphisms of the estrogen receptor alpha gene with bone mineral density of the femoral neck in elderly Japanese women.** *J Mol Med* 2002, **80**:452-460.
- Qin YJ, Shen H, Huang QR, Zhao LJ, Zhou Q, Li MX, He JW, Mo XY, Lu JH, Recker RR, Deng HW: **Estrogen receptor alpha gene polymorphisms and peak bone density in Chinese nuclear families.** *J Bone Miner Res* 2003, **18**:1028-1035.
- Ioannidis JP, Stavrou I, Trikalinos TA, Zois C, Brandi ML, Gennari L, Albagha O, Ralston SH, Tsatsoulis A: **Association of polymorphisms of the estrogen receptor alpha gene with bone mineral density and fracture risk in women: a meta-analysis.** *J Bone Miner Res* 2002, **17**:2048-2060.
- Magnusson C, Baron JA, Correia N, Bergstrom R, Adami HO, Persson I: **Breast-cancer risk following long-term oestrogen- and oestrogen-progestin-replacement therapy.** *Int J Cancer* 1999, **81**:339-344.
- Magnusson C, Baron J, Persson I, Wolk A, Bergstrom R, Trichopoulos D, Adami HO: **Body size in different periods of life and breast cancer risk in post-menopausal women.** *Int J Cancer* 1998, **76**:29-34.
- Magnusson CM, Persson IR, Baron JA, Ekblom A, Bergstrom R, Adami HO: **The role of reproductive factors and use of oral contraceptives in the aetiology of breast cancer in women aged 50 to 74 years.** *Int J Cancer* 1999, **80**:231-236.
- Isola J, DeVries S, Chu L, Ghazvini S, Waldman F: **Analysis of changes in DNA sequence copy number by comparative genomic hybridization in archival paraffin-embedded tumor samples.** *Am J Pathol* 1994, **145**:1301-1308.
- McGuire WL, Chamness GC, Fuqua SA: **Abnormal estrogen receptor in clinical breast cancer.** *J Steroid Biochem Mol Biol* 1992, **43**:243-247.
- Roodi N, Bailey LR, Kao WY, Verrier CS, Yee CJ, Dupont WD, Parl FF: **Estrogen receptor gene analysis in estrogen receptor-positive and receptor-negative primary breast cancer.** *J Natl Cancer Inst* 1995, **87**:446-451.
- Sano M, Inoue S, Hosoi T, Ouchi Y, Emi M, Shiraki M, Orimo H: **Association of estrogen receptor dinucleotide repeat polymorphism with osteoporosis.** *Biochem Biophys Res Commun* 1995, **217**:378-383.
- Yaich L, Dupont WD, Cavener DR, Parl FF: **Analysis of the PvuII restriction fragment-length polymorphism and exon structure of the estrogen receptor gene in breast cancer and peripheral blood.** *Cancer Res* 1992, **52**:77-83.
- Chen X, Levine L, Kwok PY: **Fluorescence polarization in homogeneous nucleic acid analysis.** *Genome Res* 1999, **9**:492-498.
- Tyagi S, Bratu DP, Kramer FR: **Multicolor molecular beacons for allele discrimination.** *Nat Biotechnol* 1998, **16**:49-53.
- Tapp I, Malmberg L, Rennel E, Wik M, Syvanen AC: **Homogeneous scoring of single-nucleotide polymorphisms: comparison of the 5'-nuclease TaqMan assay and Molecular Beacon probes.** *Biotechniques* 2000, **28**:732-738.
- Zuker M: **Mfold web server for nucleic acid folding and hybridization prediction.** *Nucleic Acids Res* 2003, **31**:3406-3415.
- Syvanen AC, Sajantila A, Lukka M: **Identification of individuals by analysis of biallelic DNA markers, using PCR and solid-phase minisequencing.** *Am J Hum Genet* 1993, **52**:46-59.
- Holgerson S, Karlsson JA, Kihlgren A, Rosen B, Savolainen P, Gyllenstein U: **Fluorescent-based typing of the two short tandem repeat loci HUMTH01 and HUMACTBP2: reproducibility of size measurements and genetic variation in the Swedish population.** *Electrophoresis* 1994, **15**:890-895.
- Lewontin RC, Feldman MW: **A general asymptotic property of two-locus selection models.** *Theor Popul Biol* 1988, **34**:177-193.
- Wall JD, Pritchard JK: **Haplotype blocks and linkage disequilibrium in the human genome.** *Nat Rev Genet* 2003, **4**:587-597.
- Devlin B, Risch N: **A comparison of linkage disequilibrium measures for fine-scale mapping.** *Genomics* 1995, **29**:311-322.
- Hedrick PW: **Gametic disequilibrium measures: proceed with caution.** *Genetics* 1987, **117**:331-341.
- Breslow NE, Day NE: **Statistical methods in cancer research. Volume I - The analysis of case-control studies.** Lyon, France: International Agency for Research on Cancer 1987.
- Zhao JH, Curtis D, Sham PC: **Model-free analysis and permutation tests for allelic associations.** *Hum Hered* 2000, **50**:133-139.
- Stram DO, Leigh Pearce C, Bretsky P, Freedman M, Hirschhorn JN, Altshuler D, Kolonel LN, Henderson BE, Thomas DC: **Modeling and E-M estimation of haplotype-specific relative risks from genotype data for a case-control study of unrelated individuals.** *Hum Hered* 2003, **55**:179-190.
- Neuhaus J: **Closure of the class of binary generalized linear models in some non-standard settings.** *J R Statist Soc B* 2000, **62**:839-846.
- Westfall P, Young S: **Resampling-based Multiple Testing: Examples and Methods for P-value Adjustment** New York: John Wiley & Sons; 1993.
- Herrington DM, Howard TD, Hawkins GA, Reboussin DM, Xu J, Zheng SL, Brosnihan KB, Meyers DA, Bleeker ER: **Estrogen-receptor polymorphisms and effects of estrogen replacement on high-density lipoprotein cholesterol in women with coronary disease.** *N Engl J Med* 2002, **346**:967-974.
- Curran JE, Lea RA, Rutherford S, Weinstein SR, Griffiths LR: **Association of estrogen receptor and glucocorticoid receptor gene polymorphisms with sporadic breast cancer.** *Int J Cancer* 2001, **95**:271-275.
- Hoehle MR: **Haplotypes and the systematic analysis of genetic variation in genes and genomes.** *Pharmacogenomics* 2003, **4**:547-570.
- Drysdale CM, McGraw DW, Stack CB, Stephens JC, Judson RS, Nandabalan K, Arnold K, Ruano G, Liggett SB: **Complex promoter and coding region beta 2-adrenergic receptor haplotypes alter receptor expression and predict in vivo responsiveness.** *Proc Natl Acad Sci USA* 2000, **97**:10483-10488.
- Fallin D, Schork NJ: **Accuracy of haplotype frequency estimation for biallelic loci, via the expectation-maximization algorithm for unphased diploid genotype data.** *Am J Hum Genet* 2000, **67**:947-959.
- Enattah NS, Sahi T, Savilahti E, Terwilliger JD, Peltonen L, Jarvela I: **Identification of a variant associated with adult-type hypoparathyroidism.** *Nat Genet* 2002, **30**:233-237.
- Herrington DM, Howard TD, Brosnihan KB, McDonnell DP, Li X, Hawkins GA, Reboussin DM, Xu J, Zheng SL, Meyers DA, Bleeker ER: **Common estrogen receptor polymorphism augments effects of hormone replacement therapy on E-selectin but not C-reactive protein.** *Circulation* 2002, **105**:1879-1882.
- Majewski J, Ott J: **Distribution and characterization of regulatory elements in the human genome.** *Genome Res* 2002, **12**:1827-1836.
- Ushiroyama T, Heishi M, Ikeda A, Ueki M: **Does the estrogen receptor gene polymorphism relate to undefined menopausal symptoms?** *Res Commun Mol Pathol Pharmacol* 2001, **109**:3-14.
- Weel AE, Uitterlinden AG, Westendorp IC, Burger H, Schuit SC, Hofman A, Helmerhorst TJ, van Leeuwen JP, Pols HA: **Estrogen receptor polymorphism predicts the onset of natural and surgical menopause.** *J Clin Endocrinol Metab* 1999, **84**:3146-3150.
- Deng HW, Li J, Li JL, Dowd R, Davies KM, Johnson M, Gong G, Deng H, Recker RR: **Association of estrogen receptor-alpha genotypes with body mass index in normal healthy postmenopausal Caucasian women.** *J Clin Endocrinol Metab* 2000, **85**:2748-2751.

40. Weiderpass E, Persson I, Melhus H, Wedren S, Kindmark A, Baron JA: **Estrogen receptor alpha gene polymorphisms and endometrial cancer risk.** *Carcinogenesis* 2000, **21**:623-627.
41. Andersen TI, Heimdal KR, Skrede M, Tveit K, Berg K, Borresen AL: **Oestrogen receptor (ESR) polymorphisms and breast cancer susceptibility.** *Hum Genet* 1994, **94**:665-670.
42. Shin A, Kang D, Nishio H, Lee MJ, Park SK, Kim SU, Noh DY, Choe KJ, Ahn SH, Hirvonen A, Kim JH, Yoo KY: **Estrogen receptor alpha gene polymorphisms and breast cancer risk.** *Breast Cancer Res Treat* 2003, **80**:127-131.
43. Iwase H, Greenman JM, Barnes DM, Hodgson S, Bobrow L, Mathew CG: **Sequence variants of the estrogen receptor (ER) gene found in breast cancer patients with ER negative and progesterone receptor positive tumors.** *Cancer Lett* 1996, **108**:179-184.
44. Vasconcelos A, Medeiros R, Veiga I, Pereira D, Carrilho S, Palmeira C, Azevedo C, Lopes CS: **Analysis of estrogen receptor polymorphism in codon 325 by PCR-SSCP in breast cancer: association with lymph node metastasis.** *Breast J* 2002, **8**:226-229.
45. Cai Q, Shu XO, Jin F, Dai Q, Wen W, Cheng JR, Gao YT, Zheng W: **Genetic polymorphisms in the estrogen receptor alpha gene and risk of breast cancer: results from the Shanghai Breast Cancer Study.** *Cancer Epidemiol Biomarkers Prev* 2003, **12**:853-859.
46. Hill SM, Fuqua SA, Chamness GC, Greene GL, McGuire WL: **Estrogen receptor expression in human breast cancer associated with an estrogen receptor gene restriction fragment length polymorphism.** *Cancer Res* 1989, **49**:145-148.
47. Parl FF, Cavener DR, Dupont WD: **Genomic DNA analysis of the estrogen receptor gene in breast cancer.** *Breast Cancer Res Treat* 1989, **14**:57-64.
48. Schubert EL, Lee MK, Newman B, King MC: **Single nucleotide polymorphisms (SNPs) in the estrogen receptor gene and breast cancer susceptibility.** *J Steroid Biochem Mol Biol* 1999, **71**:21-27.
49. Southey MC, Batten LE, McCredie MR, Giles GG, Dite G, Hopper JL, Venter DJ: **Estrogen receptor polymorphism at codon 325 and risk of breast cancer in women before age forty.** *J Natl Cancer Inst* 1998, **90**:532-536.
50. Comings DE, Gade-Andavolu R, Cone LA, Muhleman D, MacMurray JP: **A multigene test for the risk of sporadic breast carcinoma.** *Cancer* 2003, **97**:2160-2170.
51. van Meurs JB, Schuit SC, Weel AE, van der Klift M, Bergink AP, Arp PP, Colin EM, Fang Y, Hofman A, van Duijn CM, van Leeuwen JP, Pols HA, Uitterlinden AG: **Association of 5' estrogen receptor alpha gene polymorphisms with bone mineral density, vertebral bone area and fracture risk.** *Hum Mol Genet* 2003, **12**:1745-1754.
52. Giguere Y, Dodin S, Blanchet C, Morgan K, Rousseau F: **The association between heel ultrasound and hormone replacement therapy is modulated by a two-locus vitamin D and estrogen receptor genotype.** *J Bone Miner Res* 2000, **15**:1076-1084.
53. Patel MS, Cole DE, Smith JD, Hawker GA, Wong B, Trang H, Vieth R, Meltzer P, Rubin LA: **Alleles of the estrogen receptor alpha gene and an estrogen receptor cotranscriptional activator gene, amplified in breast cancer-1 (AIB1), are associated with quantitative calcaneal ultrasound.** *J Bone Miner Res* 2000, **15**:2231-2239.
54. Lau EM, Young RP, Lam V, Li M, Woo J: **Estrogen receptor gene polymorphism and bone mineral density in postmenopausal Chinese women.** *Bone* 2001, **29**:96-98.
55. Terwilliger JD, Haghghi F, Hiekkalinna TS, Goring HH: **A bias-ed assessment of the use of SNPs in human complex traits.** *Curr Opin Genet Dev* 2002, **12**:726-734.
56. Wacholder S, Chanock S, Garcia-Closas M, El Ghormli L, Rothman N: **Assessing the probability that a positive report is false: an approach for molecular epidemiology studies.** *J Natl Cancer Inst* 2004, **96**:434-442.
57. Kang HJ, Kim SW, Kim HJ, Ahn SJ, Bae JY, Park SK, Kang D, Hirvonen A, Choe KJ, Noh DY: **Polymorphisms in the estrogen receptor-alpha gene and breast cancer risk.** *Cancer Lett* 2002, **178**:175-180.