Breast Cancer Res Treat (2008) 108:137–149
DOI 10.1007/s10549-007-9586-8

EPIDEMIOLOGY

Breast cancer: a candidate gene approach across the estrogen metabolic pathway

Christina Justenhoven · Ute Hamann · Falk Schubert · Marc Zapatka · Christiane B. Pierl · Sylvia Rabstein · Silvia Selinski · Tina Mueller · Katja Ickstadt · Michael Gilbert · Yon-Dschun Ko · Christian Baisch · Beate Pesch · Volker Harth · Hermann M. Bolt · Caren Vollmert · Thomas Illig · Roland Eils · Jürgen Dippon · Hiltrud Brauch

Received: 23 March 2007 / Accepted: 26 March 2007 / Published online: 23 June 2007

Abstract Polymorphisms within the estrogen metabolic pathway are prime candidates for a possible association with breast cancer risk. We investigated 11 genes encoding key proteins of this pathway for their potential contribution to breast cancer risk. Of these CYP17A1, CYP19A1, EPHX1, HSD17B1, SRD5A2, and PPARG2 participate in biosynthesis, CYP1A1, CYP1B1, COMT, GSTP1, and SOD2 in catabolism and detoxification. We performed a population-based case-control study with 688 incident breast cancer cases and 724 controls from Germany and genotyped 18 polymorphisms by matrix assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS), PCR based RFLP (restriction fragment length polymorphism), and TaqMan® allelic discrimination. Genotype frequencies were compared between cases and controls and odds ratios were calculated by conditional logistic regression. Further statistical analyses were based on cluster analysis, multifactor dimensionality reduction, logic regression, and global testing. Single factor analyses pointed to *CYP1B1*_1294_GG as a possible breast cancer risk modulator (OR = 2.57; 95% CI: 1.34–4.93) and two way stratification suggested associations between BMI $\geq$ 30 kg/m^2 and *COMT*_472_GG ($P = 0.0076$ and $P = 0.0026$), BMI < 20 kg/m^2 and *HSD17B1*_937_GG

Electronic supplementary material The online version of this article (doi:10.1007/s10549-007-9586-8) contains supplementary material, which is available to authorized users.

C. Justenhoven · H. Brauch (✉)
Molecular Mechanisms of Origin and Treatment of Breast Cancer, Dr. Margarete Fischer-Bosch-Institute of Clinical Pharmacology, Auerbachstrasse 112, 70376 Stuttgart, Germany
e-mail: hiltrud.brauch@ikp-stuttgart.de

U. Hamann · M. Gilbert
Molecular Genetics of Breast Cancer, Deutsches Krebsforschungszentrum (DKFZ), Heidelberg, Germany

F. Schubert · M. Zapatka · R. Eils
Department of Theoretical Bioinformatics, Deutsches Krebsforschungszentrum (DKFZ), Heidelberg, Germany

C. B. Pierl · S. Rabstein · B. Pesch
Berufsgenossenschaftliches Forschungsinstitut für Arbeitsmedizin (BGFA), Ruhr University Bochum, Bochum, Germany

S. Selinski · T. Mueller · K. Ickstadt
Department of Statistics, Universität Dortmund, Dortmund, Germany

S. Selinski · H. M. Bolt
Institut für Arbeitsphysiologie an der Universität Dortmund, Dortmund, Germany

Y.-D. Ko · C. Baisch · V. Harth
Department of Internal Medicine, Evangelische Kliniken Bonn gGmbH, Johanniter Krankenhaus, Bonn, Germany

C. Vollmert · T. Illig
Institute of Epidemiology, GSF-National Research Center for Environment and Health, Neuherberg, Germany

J. Dippon
Department of Mathematics, Universität Stuttgart, Stuttgart, Germany

C. Justenhoven · H. Brauch
University Tuebingen, Tuebingen, Germany

($P = 0.0082$) as well as *CYP17A1*_-34_CC and HRT use ≥10 years ($P = 0.0063$). Following correction for multiple testing none of these associations remained significant. No significant association between breast cancer risk and genetic polymorphisms was observed in multifactor analyses. The tested polymorphisms of the estrogen metabolic pathway may not play a direct role in breast cancer risk. Therefore, future association studies should be extended to other polymorphisms and other regulatory pathways.

Keywords Breast cancer · Susceptibility · Estrogen metabolism · Polymorphisms · Association analyses · Multivariate analyses

Introduction

The search for breast cancer susceptibility genes is going at an enormous speed owing to the general appreciation of breast cancer being a multifactor as well as polygenic disease. This continuous search is encouraged by the worldwide annual toll of more than one million new breast cancer cases and more than 400,000 breast cancer related deaths [1, 2]. Twin studies in Scandinavia suggested that heredity may account for about 27% of breast cancer. As of today less than 5% of familial breast cancer have been attributed to high penetrance breast cancer genes *BRCA1*, *BRCA2*, *PTEN*, and *TP53* [3–5] and rare genetic variants at *ATM*, *CHEK2*, *BRIP1* or *PALB2* that confer an approximately 2-fold increased risk [6–8]. Strong evidence for a common breast cancer susceptibility allele has been provided for *CASP8* [9].

From an epidemiological view point it is generally accepted that cumulative, excessive exposure to endogenous estrogen across a woman's life span contributes and may be causal to breast cancer [10]. High serum estrogen level has been proposed as a major risk factor for breast cancer [11]. *In vitro* and *in vivo* animal as well as patient-based studies suggested that estrogens, their metabolic compounds and the entire biochemical metabolic machinery may play a role in breast carcinogenesis [12]. Moreover, recent observational studies showed a possible influence of exogenous hormones such as hormone replacement therapy (HRT) [13–15] and oral contraceptives (OC) [16, 17].

In keeping with the polygene hypothesis of breast cancer [18], the steroid hormone metabolic pathway appears to be a prime candidate for the investigative search of breast cancer susceptibility genes. This notion draws substance from frequent polymorphisms at enzymes of biosynthesis and catabolism. A systematic search across the pathway holds the potential to comprehensively scrutinize a group of cooperating candidate genes and variants through gene-gene and gene-environment interactions. This approach has been suggested recently [19] and may be superior to common single factor analyses due to the feasibility of interaction analyses.

The GENICA interdisciplinary network of clinicians, epidemiologists, human geneticists, statisticians, and bioinformatic experts applies current state-of-the-art analytical and computational methods towards the investigation of breast cancer susceptibility genes and gene-environment interactions within a population-based case-control study from Germany [14, 20]. Analyses reported herein include six epidemiological variables and 18 polymorphisms of 11 genes encoding enzymes related to estrogen biosynthesis (CYP17A1, CYP19A1, EPHX1, HSD17B1, SRD5A2, PPARG2), catabolism and detoxification (CYP1A1, CYP1B1, COMT, GSTP1, and SOD2). We present single factor and multifactor analyses of high-order gene–gene and gene–environment interactions, and discuss the findings within the biological and currently available methodological context.

Materials and methods

Study population

The GENICA study participants of the population-based breast cancer case-control study from the Greater Bonn Region, Germany, were recruited between 08/2000 and 10/2002 as described previously [14, 20]. In brief, there are 688 incident breast cancer cases and 724 population controls matched in 5-year classes. Cases and controls were eligible if they were of Caucasian ethnicity, current residents of the study region, and below 80 years of age. Information on known and supposed risk factors was collected via in person interviews. The response rate for cases was 88% and for controls 67%. Characteristics of the study population regarding potential breast cancer risk factors include age at diagnosis (<50, ≥50 years), menopausal status (premenopausal, postmenopausal), breast cancer in mother and sisters (yes, no), OC use (never, >0–<5, 5–<10, ≥10 years), HRT use (never, >0–<10, ≥10 years), body mass index (BMI; <20, 20–<25, 25–<30, ≥30 kg/m^2), and smoking status (never, former, current) (Table 1). The GENICA study was approved by the Ethic's Committee of the University of Bonn. All study participants gave written informed consent.

DNA isolation and genotyping

Genomic DNA was extracted from heparinized blood samples (Puregene$^{\mathrm{TM}}$, Gentra Systems, Inc., Mineapolis, USA) as previously described [20]. DNA samples were available for 610 cases (89%) and 651 controls (90%).

Table 1 Selected epidemiological variables and their risk association in the GENICA study population

Variables		Cases		Controls		OR (95% CI)[a]
		n	(%)	*n*	(%)	
Age	<50 y	140	(23.0)	147	(23.5)	1.00
	≥50 y	468	(77.0)	479	(76.5)	1.03 (0.79–1.34)
Menopausal status	Pre	146	(24.3)	149	(24.1)	1.00
	Post	455	(75.7)	470	(75.9)	0.95 (0.62–1.45)
Breast cancer in mother or sisters	No	537	(88.3)	580	(92.7)	1.00
	Yes	71	(11.7)	46	(7.3)	1.67 (1.13–2.47)
OC use	Never	230	(37.9)	245	(39.3)	1.00
	>0–<5 y	101	(16.6)	97	(15.5)	1.13 (0.79–1.62)
	5–<10 y	81	(13.3)	73	(11.7)	1.21 (0.81–1.80)
	≥10 y	195	(32.2)	209	(33.5)	1.02 (0.75–1.39)
HRT use	Never	293	(48.8)	314	(50.6)	1.00
	>0–<10 y	153	(25.5)	180	(29.0)	0.91 (0.68–1.21)
	≥10 y	154	(25.7)	127	(20.4)	1.30 (0.94–1.80)
BMI	<20 kg/m^2	57	(9.4)	47	(7.5)	1.25 (0.82–1.91)
	20–<25 kg/m^2	293	(48.2)	305	(48.8)	1.00
	25–<30 kg/m^2	174	(28.6)	193	(30.9)	0.94 (0.72–1.22)
	≥30 kg/m^2	84	(13.8)	80	(12.8)	1.10 (0.77–1.55)
Smoking	Never	351	(57.8)	346	(55.3)	1.00
	Former	122	(20.1)	127	(20.3)	0.95 (0.71–1.27)
	Current	134	(22.1)	153	(24.4)	0.86 (0.64–1.15)

[a] Odds ratio conditional on age in 5-year groups

Abbreviations: BMI = body mass index, CI = confidence interval, HRT = hormone replacement therapy, OC = oral contraceptives, OR = odds ratio, y = years

Eighteen polymorphisms of 11 genes of the estrogen metabolic pathway were subjected to genotyping in DNA of 1,261 patients and controls. Genes have been selected according to their role with respect to the estrogen metabolic pathway and polymorphisms have been selected on the basis of a known or putative functional consequence as well as an allele frequency of at least 4% in a population of European descent. A detailed description of genes and polymorphisms is given in Supplementary Table S1.

MALDI-TOF MS genotyping

Genotyping of 16 single nucleotide polymorphisms (SNPs) at *CYP19A1*_790_C>T (rs700519), *EPHX1*_337_T>C (rs1051740), *EPHX1*_416_A>G (rs2234922), *HSD17B1*_937_A>G (rs6050598), *SRD5A2*_265_G>C (rs523349), *PPARG2*_-34_C>G (rs1801282), *CYP1A1*_2452_C>A (rs1799814), *CYP1A1*_2454_A>G (rs1048943), *CYP1A1*_3801_T>C (rs4646903), *CYP1B1*_1294_C>G (rs1056836), *CYP1B1*_1358_A>G (rs1800440), *COMT*_472_G>A (rs4680), *GSTP1*_313_A>G (rs947894), *GSTP1*_341_C>T (rs1799811), and *SOD2*_47_T>C (rs1799725) were performed by matrix assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) as described previously [20, 21]. In brief, 5 ng of genomic DNA was amplified by PCR using 0.1 unit HotStarTaq DNA Polymerase (Qiagen, Hilden, Germany). PCR conditions were 95°C for 15 min, followed by 44 cycles of 95°C for 30 s, 56°C for 30 s, 72°C for 1 min, finally followed by 72°C for 10 min. PCR products were treated with shrimp alkaline phosphatase (SAP, Amersham, Freiburg, Germany) for 20 min at 37°C followed by 10 min at 85°C. Base extension [homogenous MassEXTEND (hME^{TM}), Sequenom, San Diego, CA] reaction in a final volume of 10 µl contained extension primers at a final concentration of 0.54 µM and 0.6 units ThermoSequenase (Amersham, Freiburg, Germany). Reaction conditions were 94°C for 2 min, followed by 40 cycles of 94°C for 5 s, 52°C for 5 s, and 72°C for 5 s. Final base extension products were treated with SpectroCLEAN resin (Sequenom, San Diego, CA, USA). A 10 nl aliquot of reaction solution was dispensed onto a 384 format SpectroCHIP microarray (Sequenom, San Diego, CA, USA) pre-spotted with a matrix of 3-hydroxypicolinic acid. A Bruker Autoflex MALDI-TOF MS was used for data acquisitions from the SpectroCHIP. Genotyping calls were made with MASSARRAY RT software v 3.0.0.4 (Sequenom, San Diego, CA, USA). For quality control repeated analyses was performed for 10% randomly selected samples. Primers were synthesized by Metabion International AG, Martinsried, Germany.

TaqMan® assisted allelic discrimination

In case of *EPHX1*_337_T>C (rs1051740) and *EPHX1*_416_A>G (rs2234922) genotyping of 69% and 56% of the samples, respectively, was additionally performed by TaqMan® allelic discrimination as described previously [22]. In brief, PCRs were performed in a reaction volume of 25 µl containing 50 ng DNA, 1 × TaqMan® Universal PCR Master Mix (Applied Biosystems, Foster City, CA, USA) Primer and MGB probes were synthesized by MWG-Biotech AG, Ebersberg, Germany and Applied Biosystems, Foster City, CA, USA, respectively. PCR and genotype analysis was performed with ABI Prism 7700 Sequence Detection System and ABI Prism 7700 SDS software (Applied Biosystems, Foster City, CA, USA).

PCR-based fragment length polymorphism genotyping

Four polymorphisms including *COMT*_472_G>A (rs4680), *CYP17A1*_-34_T>C (rs743572), *CYP19A1*_630_delCTT, and *CYP19A1*_681_(TTTA)$_n$ were genotyped by fragment analysis. For quality controls 10% of randomly selected samples were repeated.

The *COMT*_472_G>A polymorphism creates a *Hsp*92II restriction site: PCR was performed with primers from Thompson et al. [23] and carried out in 10 µl reaction containing 50 ng genomic DNA, 1 × PCR buffer (Qiagen, Hilden, Germany), 1.5 mM $MgCl_2$, 0.1 µM of each primer, 250 µM of each dNTP (Promega, Mannheim, Germany), 0.4 U HotStarTaq DNA polymerase (Qiagen, Hilden, Germany). After an initial 15 min at 95°C, DNA was amplified by 35 cycles of 1 min at 94°C, 1 min at 58°C, 1 min at 72°C, and with a final extension step of 10 min at 72°C. Amplified DNA fragments were digested with 5 U *Hsp*92II (Promega, Mannheim, Germany) in a total volume of 20 µl, separated on a 2.5% agarose gel containing ethidium bromide (Sigma-Aldrich, Steinheim, Germany) and scored by UV visualization. Fragment sizes were 114, 29, and 26 bp for the G allele and 96, 18, 29, and 26 bp for the A allele.

The polymorphism *CYP17A1*_-34_T>C creates a *Msp*A1I restriction site: PCR was performed with primers including a D4-labeled forward primer to generate a fluorescent product, PCR reactions, conditions and restriction enzyme digestion followed the method described by Bergmann-Jungestrom et al. [24]. The sizes of the labelled fragments were 209 bp for the T allele and 123 bp for the C allele.

*CYP19A1*_630_delCTT and *CYP19A1*_681_(TTTA)$_n$ polymorphisms: The CTT insertion/deletion (ins/del) polymorphism is located upstream of the tetranucleotide repeat (TTTA)$_n$ [25]. The deletion is linked to (TTTA)$_7$ generating the two alleles delCTT_(TTTA)$_7$ and insCTT_(TTTA)$_7$. CTT ins/del was amplified using a D3-labelled forward primer [25]. PCR conditions were the same as for *COMT*_472_G>A analysis. Fragment sizes were 171 bp for the insCTT allele and 168 bp for the delCTT allele. The (TTTA)$_n$ polymorphism without the ins/del site was analyzed according to Kristensen et al. [26] using a D4-labelled forward primer. PCR reactions and conditions were as for *COMT*_472_G>A analysis, with the exception of using 2 mM $MgCl_2$ and an annealing temperature of 51°C. The sizes of the labelled fragments ranged from 302 bp for the (TTTA)$_7$ allele to 326 bp for the (TTTA)$_{13}$ allele.

For fragment analyses fluorescence labelled fragments were separated by capillary gel electrophoresis on a CEQ™ 8000 fully automated Genetic DNA Analysis System (Beckman Coulter, Krefeld, Germany). Separation of the CEQ DNA Size Standard-400 in each well allowed for CEQ automated sizing of D3- and D4-labelled PCR products. The analysis was performed in a 96-well format.

Due to their functional roles *CYP19A1*_630_delCTT and *CYP19A1*_681_(TTTA)$_n$ polymorphisms were combined for multivariate analysis. The *CYP19A1*_630_delCTT is linked to *CYP19A1*_681_(TTTA)$_{7,}$ moreover carriers of the *CYP19A1*_630_insCTT genotype can be carrier of all numbers of repeats. To increase the power of statistical analyses we combined these two polymorphisms and created three variables: (1) Carriership of *CYP19A1*_630_delCTT and *CYP19A1*_681_(TTTA)$_7$, (2) carriership of *CYP19A1*_630_insCTT and *CYP19A1*_681_(TTTA)$_{7,}$ (3) carriership of *CYP19A1*_630_insCTT and *CYP19A1*_681_(TTTA)$_{>10}$. The three groups represent hypothesized increasing levels of sex hormone levels. Accordingly, the delCTT is associated with lower estrone, estradiol, free estradiol levels and higher sex hormone binding globulin concentrations [27]. Likewise, more than seven repeats are associated with higher concentrations of estrone, estradiol, free estradiol, androstenedione as well as testosterone in the serum [27–29].

Statistical analyses

A number of statistical strategies have been applied to analyze genetic and epidemiologic variables of the GENICA study collection for the identification of main effects and interactions on the risk to develop breast cancer. Altogether we applied six methods including conditional logistic regression, haplotype analysis, cluster analysis, multifactor dimensionality reduction, logic regression, and global testing. For the analyses two datasets were used. A standard dataset included the five epidemiological data and

SNP data encoded as 1 = homozygous for major allele, 2 = heterozygous, and 3 = homozygous for minor allele. A dichotomized dataset was used to increase the power of the study with SNP data encoded as 1 = homozygous for major allele and 2 = heterozygous and homozygous for minor allele. Using the standard dataset the study had an 80% power to detect minimum ORs of 1.39–1.96 ($\alpha = 0.05$, two sided test). Using the dichotomized dataset the study had an 80% power to detect minimum ORs of 1.09–1.59 ($\alpha = 0.05$, two sided test). The range reflects the variation of minor allele frequencies of the 18 polymorphisms which were at 4–48%.

Basic single and multifactor analyses

Odds ratios (OR) and 95% confidence intervals (CI) were calculated by conditional logistic regression and *P*-values by chi square tests. All tests were two sided. To identify subgroups of patients at risk we performed stratified analyses in two directions. *P*-values were corrected for multiple testing by Holm's method. If indicated, study subjects with missing values were excluded from the analyses.

We stratified for six epidemiologic variables (menopausal status, breast cancer in mother or sisters, OC use, HRT use, BMI, and smoking) and for genotypes of all 18 polymorphisms. Risk estimation was performed by logistic regression conditional on age in 5-year groups, according to the matching scheme of controls to cases and adjusted for the five epidemiologic variables using SAS [30]. We considered $P < 0.01$ noteworthy provided that the minimal size of the group was ≥5. Conditional logistic regression was also used as a multivariate procedure and to investigate possible interaction effects by comparing models with and without interaction term and considering deviance statistic. Models were compared considering their deviance statistics.

Haplotypes were estimated using PHASE version 2.1 [31, 32]. Comparisons between observed and expected haplotype frequencies as well as comparison of frequencies between cases and controls were performed using chi square test.

Cluster analysis, higher order interactions and global associations

All these analyses were carried out using the open source statistical package R (www.r-project.org).

Cluster analysis To investigate associations of genetic and epidemiological variables a hierarchical cluster analysis of all variables was performed with Pearson's corrected coefficient of contingency as similarity measure [33, 34]. We conducted separate analyses for cases and controls using average linkage as clustering algorithm.

Multifactor dimensionality reduction The multifactor dimensionality reduction (MDR) algorithm [35, 36] performs well in detecting high-order interactions, even in the absence of any statistically significant main effects [37]. The classification accuracy is estimated using a 10-fold cross-validation. However, this estimation tends to be over optimistic because the best model of all cross-validation steps is selected. Therefore, we estimated an additional classification accuracy using an independent test and training set generated by splitting the original dataset (¼–¾). Furthermore, the dichotomized dataset was used to reduce the size of cross tables used for MDR algorithm.

Logic regression Logic regression has been developed to analyze genetic data [38]. It is an adaptive regression methodology for predicting the outcome in classification and regression problems that are based on binary (true/false) variables. Logic regression searches for the logic tree that best explains the cases. Instead of just building a single tree, multiple trees may be built and combined by a generalized linear model with logit link function. Since we focus on variable/feature selection rather than on classification, we combine the logic regression approach with a method for quantifying the importance of identified genetic or epidemiological variables and their interactions of order two and higher. The underlying measure of importance is the logicFS measure introduced in Schwender and Ickstadt [39] constructed for both the single and multiple tree logic regression approach. We used the dichotomized dataset and the analysis was carried out using the R package logicFS which is part of Bioconductor [40], a project of the analysis of genomic data.

Global testing Instead of checking the impact of each polymorphism individually we asked whether the whole set of polymorphisms cooperating within the estrogen metabolic pathway may affect case and control status. This issue can be addressed by global testing as it has been recently proposed within the context of gene microarray analysis [41]. The global test is particularly applicable for the following situations: (1) Each of several genes (in our case polymorphisms) may have a small impact on the outcome, (2) few genes may have an impact on the outcome or (3) one gene may have a major impact on the outcome. Epidemiological variables were considered as confounders and we based our analyses on the dichotomized dataset.

Furthermore, we applied non-linear and non-parametric methods such as neural networks, support vector machines, and random forests.

Results

We genotyped 1261 study participants at 18 loci for the investigation of a possible association with breast cancer risk, either alone or in combination. Call rates were >97.5%. Concordance of MALDI-TOF MS genotyping data was 99.9%, for PCR-RFLP based genotyping 99.5%, and for TaqMan® allelic discrimination 100%. For *CYP17A1*_-34_T>C and *COMT*_472_G>A duplicate analyses were performed with MALDI-TOF MS and PCR-RFLP based genotyping with a concordance rate of 99.7%. *EPHX1*_337_T>C and *EPHX1*_416_A>G were genotyped by MALDI-TOF MS and repeated by TaqMan® allelic discrimination in randomly selected samples of 69% and 56%, respectively. Concordance rate was 99.9%. Genotype frequencies of 17 polymorphisms of cases and controls are given in Table 2 and allele frequencies of *CYP19A1*_681_(TTTA)$_n$ are given in Table 3. All genotype frequencies were in HWE with the exception of *CYP1B1*_1358_A>G in cases (*P* = 0.007).

Association analyses of single genetic variants

When we calculated ORs from genotype frequencies at 17 loci (Table 2) and allele frequencies at one locus (Table 3) we identified a significantly increased breast cancer risk for carriers of the *CYP1B1*_1358_GG genotype (OR = 2.57; 95% CI: 1.34–4.93). In addition we observed a borderline decreased risk for carriers of the *HSD17B1*_937_GG genotype (OR = 0.73; 95% CI: 0.52–1.01). Following correction for multiple testing the significance of these effects vanished. No significant differences between cases and controls were observed at the remaining 16 loci (Tables 2 and 3).

Stratified analyses

We stratified in two directions and observed four effects which we considered to be noteworthy. When we stratified genetic data for epidemiological variables we observed two associations involving BMI. Women with a BMI ≥30 kg/

Table 2 Frequencies and risk estimates of polymorphic loci at genes related to the estrogen metabolic pathway

Polymorphism	Genotypes	Cases		Controls		OR (95% CI)[a]
		n	(%)	*n*	(%)	
Estrogen biosynthesis						
*CYP17A1*_-34_T>C	TT	202	(33.3)	214	(34.2)	1.00
	TC	298	(49.2)	305	(48.8)	1.01 (0.79–1.31)
	CC	106	(17.5)	106	(17.0)	1.00 (0.72–1.40)
*CYP19A1*_790_C>T	CC	549	(91.6)	561	(90.2)	1.00
	CT	49	(8.2)	60	(9.6)	0.83 (0.56–1.24)
	TT	1	(0.2)	1	(0.2)	1.25 (0.08–20.4)
*CYP19A1*_630_delCTT	ins/ins	250	(41.6)	261	(41.8)	1.00
	ins/del	281	(46.8)	287	(45.9)	1.01 (0.79–1.30)
	del/del	70	(11.6)	77	(12.3)	0.95 (0.66–1.38)
*EPHX1*_337_T>C	TT	296	(48.9)	295	(48.4)	1.00
	TC	246	(40.7)	269	(44.2)	0.93 (0.73–1.18)
	CC	63	(10.4)	45	(7.4)	1.42 (0.94–2.17)
*EPHX1*_416_A>G	AA	391	(65.1)	388	(62.2)	1.00
	AG	182	(30.3)	213	(34.1)	0.88 (0.68–1.12)
	GG	28	(4.7)	23	(3.7)	1.20 (0.68–2.14)
*HSD17B1*_937_A>G	AA	180	(30.6)	159	(25.7)	1.00
	AG	296	(50.2)	319	(51.4)	0.83 (0.64–1.09)
	GG	113	(19.2)	142	(22.9)	0.73 (0.52–1.01)[c]
*SRD5A2*_265_G>C	GG	296	(49.0)	286	(45.8)	1.00
	GC	248	(41.1)	267	(42.7)	0.93 (0.73–1.18)
	CC	60	(9.9)	72	(11.5)	0.82 (0.56–1.21)

Table 2 continued

Polymorphism	Genotypes	Cases		Controls		OR (95% CI)[a]
		n	(%)	*n*	(%)	
*PPARG2*_-34_C>G	CC	452	(76.2)	462	(74.3)	1.00
	CG	135	(22.8)	145	(23.3)	0.96 (0.74–1.27)
	GG	6	(1.0)	15	(2.4)	0.41 (0.16–1.08)
Estrogen catabolism (phase I)						
CYP1A1*_2452_C>A*	CC	542	(91.7)	548	(89.9)	1.00
	CA	48	(8.1)	60	(9.8)	0.84 (0.56–1.26)
	AA	1	(0.2)	2	(0.3)	0.47 (0.04–5.31)
*CYP1A1*_2454_A>G	AA	563	(93.2)	565	(91.3)	1.00
	AG	41	(6.8)	51	(8.2)	0.78 (0.51–1.21)
	GG	0	(0)	3	(0.5)	–
*CYP1A1*_3801_T>C	TT	492	(82.0)	516	(83.5)	1.00
	TC	105	(17.5)	98	(15.9)	1.13 (0.83–1.53)
	CC	3	(0.5)	4	(0.6)	0.83 (0.18–3.76)
*CYP1B1*_1294_C>G	CC	185	(31.1)	186	(29.9)	1.00
	CG	296	(49.8)	306	(49.2)	0.97 (0.74–1.26)
	GG	114	(19.1)	130	(20.9)	0.84 (0.60–1.17)
*CYP1B1*_1358_A>G	AA	405[b]	(67.3)	427	(69.1)	1.00
	AG	165[b]	(27.4)	177	(28.6)	1.00 (0.77–1.29)
	GG	32[b]	(5.3)	14	(2.3)	2.57 (1.34–4.93)[d]
Estrogen catabolism (phase II) and detoxification						
*COMT*_472_G>A	GG	163	(26.9)	170	(27.3)	1.00
	GA	298	(49.2)	305	(49.0)	0.99 (0.76–1.30)
	AA	145	(23.9)	147	(23.6)	1.04 (0.76–1.44)
*GSTP1*_313_A>G	AA	259	(43.4)	276	(45.2)	1.00
	AG	271	(45.4)	268	(43.9)	1.09 (0.85–1.39)
	GG	67	(11.2)	67	(11.0)	1.03 (0.70–1.52)
*GSTP1*_341_C>T	CC	507	(85.1)	522	(85.7)	1.00
	CT	82	(13.7)	82	(13.5)	0.98 (0.70–1.37)
	TT	7	(1.2)	5	(0.8)	1.35 (0.42–4.38)
*SOD2*_47_T>C	TT	159	(26.3)	163	(26.3)	1.00
	TC	312	(51.7)	313	(50.4)	1.00 (0.76–1.32)
	CC	133	(22.0)	145	(23.3)	0.92 (0.66–1.27)

[a] Odds ratios conditional on age in 5-year classes, adjusted for breast cancer in mother or sisters, OC use, HRT use, BMI, and smoking

[b] Genotype frequencies failed Hardy-Weinberg equilibrium (HWE)

[c] *HSD17B1*_937_GG: $P = 0.060$

[d] *CYP1B1*_1358_GG: $P = 0.004$

m^2 carrying *COMT*_472_GG had an increased breast cancer risk (OR = 3.84; 95% CI: 1.43–10.3; Table 4). Women with a BMI <20 kg/m^2 carrying *HSD17B1*_937_GG showed a decreased breast cancer risk (OR = 0.17; 95% CI: 0.04–0.63; Table 4). When we stratified epidemiological data for genetic variables we observed two associations involving BMI and HRT use. Carriers of *COMT*_472_GG had an increased breast cancer risk when they had a BMI ≥30 kg/m^2 (OR = 3.87; 95% CI: 1.61–9.34; Table 4). Carriers of *CYP17A1*_-34_CC had an increased breast cancer risk when they used HRT ≥10 years (OR = 3.17, 95% CI: 1.39–7.25; Table 4). Following correction for multiple testing these results were not significant.

Haplotype analyses

Haplotypes were estimated for all polymorphisms located within the same gene, i.e. *CYP1A1*, *CYP1B1*, *CYP19A1*, *EPHX1*, and *GSTP1*. For *CYP1B1* and *CYP19A1* we observed haplotype frequencies that differed significantly

Table 3 Allele frequencies of *CYP19A1*_681_(TTTA)$_n$ repeat polymorphism and combined alleles with *CYP19A1*_630_delCTT polymorphism in breast cancer cases and controls

Polymorphism	Allele	Cases		Controls		*P*
		n	(%)	*n*	(%)	
*CYP19A1*_681_(TTTA)$_n$	7	593	(49.3)	620	(49.6)	
	8	126	(10.5)	135	(11.2)	
	9	3	(0.2)	2	(0.2)	
	10	16	(1.3)	22	(1.8)	
	11	428	(35.7)	433	(34.6)	
	12	29	(2.4)	29	(2.3)	
	13	7	(0.6)	9	(0.7)	0.966
*CYP19A1*_681_(TTTA)$_n$ + *CYP19A1*_630_delCTT	(TTTA)$_7$ + delCTT	250	(27.6)	261	(27.6)	
	(TTTA)$_7$ + insCTT	436	(48.1)	462	(48.9)	
	(TTTA)$_{\geq 10}$ + insCTT	220	(24.3)	222	(23.5)	0.915

Table 4 Breast cancer risk estimation for genotypes stratified for epidemiologic variables and for epidemiologic variables stratified for genotypes (selection criteria: $P < 0.01$, minimum number of cases in strata 5)

	Genotypes	Cases		Controls		OR (95% CI)[a]
		n	(%)	*n*	(%)	
Stratification for epidemiological variables						
BMI ≥30 kg/m^2						
*COMT*_472_G>A	AA	28	(35)	31	(39)	1.00
	AG	28	(35)	40	(51)	0.80 (0.39–1.63)
	GG	25	(30)	8	(10)	3.84 (1.43–10.3)*
						*P = 0.0076
BMI <20 kg/m^2						
*HSD17B1*_937_A>G	AA	19	(35)	6	(13)	1.00
	AG	27	(49)	26	(55)	0.36 (0.12–1.06)
	GG	9	(16)	15	(32)	0.17 (0.04–0.63)**
						**P = 0.0082
Stratification for genotypes						
Carriers of *COMT*_472_GG genotype						
BMI	<20 kg/m^2	9	(6)	9	(6)	1.28 (0.47–3.52)
	≥20-<25 kg/m^2	69	(49)	82	(57)	1.00
	≥25-<30 kg/m^2	39	(27)	46	(32)	1.09 (0.63–1.87)
	≥30 kg/m^2	25	(18)	8	(5)	3.87 (1.61–9.34)***
						***P = 0.0026
Carriers of *CYP17A1*_-34_CC genotype						
HRT use	never	43	(42)	60	(57)	1.00
	>0–<10 y	30	(29)	29	(27)	1.48 (0.72–3.06)
	≥10 y	30	(29)	17	(16)	3.17 (1.39–7.25)****
						****P = 0.0063

[a] Odds ratios conditional on age in 5-year classes, adjusted for breast cancer in mother or sisters, OC use, HRT use, BMI, and smoking

from those expected from an independent assortment: *CYP1B1*_1294_C/*CYP1B1*_1358_G and *CYP1B1*_1294_G/*CYP1B1*_1358_A as well as *CYP19A1*_630_insCTT/*CYP19A1*_681_(TTTA)$_{11}$/*CYP19A1*_790_C and *CYP19A1*_630_delCTT/*CYP19A1*_681_(TTTA)$_7$/*CYP19A1*_790_C were more frequent (P = 0.004, P = 0.002, respectively; Table 5). Frequencies of *CYP1A1*, *EPHX1*, and *GSTP1* haplotypes were similar to those expected. No differences

Table 5 Haplotype frequencies of *CYP1B1* and *CYP19A1* in breast cancer cases and controls compared to those expected from an independent assortment

Polymorphic loci			Expected (%)	Cases (%)	Controls (%)	*P*
CYP1B1						
1294_C>G	1358_A>G					
C	A		46	37	38	
C	G		8	19	17	
G	A		38	44	45	
G	G		9	<1	<1	
						0.004
CYP19A1						
629_delCTT	$681\_(TTTA)]_n$	790_C>T				
ins	7	C	30	11	10	
ins	7	T	2	4	5	
ins	8	C	7	11	11	
ins	8	T	<1	0	<1	
ins	9	C	<1	<1	<1	
ins	10	C	1	1	2	
ins	11	C	21	35	35	
ins	12	C	1	3	2	
ins	13	C	<1	<1	1	
del	7	C	17	35	36	
del	7	T	1	<1	<1	
del	8	C	4	<1	<1	
del	8	T	<1	0	<1	
del	11	C	12	<1	<1	
del	11	T	<1	0	<1	
						0.002

were observed between breast cancer cases and controls. Due to their possible role in slowing down or increasing the production of estrogen metabolites, we were particularly interested in frequencies of combined haplotypes of $\mathit{CYP1A1}_{\mathrm{461Asn\text{-}462Ile}}\mathit{CYP1B1}_{\mathrm{432Val\text{-}453Asn}}\mathit{COMT}_{\mathrm{158Met}}$ and $\mathit{CYP1A1}_{\mathrm{461Thr\text{-}462Val}}\mathit{CYP1B1}_{\mathrm{432Val\text{-}453Ser}}\mathit{COMT}_{\mathrm{158Val}}$. No significant differences were observed between cases and controls.

Multifactor model for breast cancer risk

Logistic regression

In a first multivariate logistic regression model we investigated a possible influence of the five epidemiological factors under investigation. In a second model we included the polymorphisms *CYP1B1*_1358_A>G and *HSD17B1*_937_A>G (Table 6), for which a significant and a borderline significant association with breast cancer risk has been observed in the preceeding univariate analyses, respectively. The comparison of both models showed a significant improvement of the risk association in the SNP containing model ($P = 0.0044$). In detail, the epidemiological variable breast cancer in mother or sisters gave a significant *P*-value ($P = 0.01$), but other parameters showed no significance. The genetic variable *CYP1B1*_1358_A>G showed an increased risk to develop breast cancer for carriers of the rare homozygous *CYP1B1*_1358_GG genotype (OR: 2.95; 95% CI = 1.51–5.75; Table 6), and the genetic variable *HSD17B1*_937_A>G showed a borderline significant decreased risk to develop breast cancer for carriers of the rare *HSD17B1*_937_GG genotype (OR = 0.72; 95% CI: 0.51–1.00; Table 6).

Interaction analyses using logistic regression

To find possible pairwise interactions of an epidemiologic and a genetic variable or of two genetic variables we considered all possible pairs in an appropriate logistic

Table 6 Multivariate breast cancer risk model including five epidemiological and two genetic variables; overall $P = 0.0069$ ($n = 1181$)

Variable	OR (95% CI)[a]	*P*
Positive family history of breast cancer	1.68 (1.12–2.51)	0.01
OC use >0–<5 y	1.11 (0.76–1.63)	0.58
OC use 5–<10 y	1.27 (0.84–1.93)	0.26
OC use ≥10 y	1.03 (0.74–1.43)	0.87
HRT use >0–<10 y	0.90 (0.66–1.23)	0.51
HRT use ≥10 y	1.43 (1.01–2.01)	0.04
BMI <20 kg/m^2	1.29 (0.83–1.99)	0.26
BMI 25–<30 kg/m^2	0.95 (0.72–1.24)	0.69
BMI ≥30 kg/m^2	1.11 (0.77–1.62)	0.58
Current smoker	0.83 (0.61–1.12)	0.22
Former smoker	0.92 (0.67–1.25)	0.60
*HSD17B1*_937_AG	0.81 (0.62–1.06)	0.13
*HSD17B1*_937_GG	0.72 (0.51–1.00)	0.05
*CYP1B1*_1358_AG	1.03 (0.80–1.33)	0.83
*CYP1B1*_1358_GG	2.95 (1.51–5.75)	0.002

[a] Logistic regression conditional on age in 5-year groups

regression model with breast cancer risk as an outcome. However, after accounting for multiple testing no significant interaction was found.

Cluster analysis

Performing the cluster analysis with Pearson's corrected coefficient of contingency for epidemiological and genetic variables we observed no differences between the clusters in cases and controls except for *EPHX1*_416_A>G and *CYP1A1*_2452_C>A, which clustered in one group in controls whereas in the case group they appeared to be independent.

Multifactor dimensionality reduction

Multifactor dimensionality reduction (MDR) was applied to the original and dichotomized dataset to identify higher order effects. Using the original dataset a significant association of HRT use ≥10 years with breast cancer ($P = 0.0107$, sign test) was observed. For the dichotomized dataset three significant rules were found (HRT use ≥10 years, $P = 0.0107$; HRT use ≥10 years and *HSD17B1*_937_AA, $P = 0.0010$; HRT use ≥10 years and *CYP1B1*_1358_AG/GG and *HSD17B1*_937_AA, $P = 0.0010$). For an unbiased estimation of the classification performance independent test set training sets (¾ training, ¼ test) were evaluated however, none of these effects remained significant. Furthermore, none of the rule sets generated with the original dataset were significant ($P < 0.05$, chi squared statistic) for the test data in cross validation. Finally, no statistically significant interactions were found using MDR.

Logic regression

Application of the single and multiple tree version of logic regression demonstrated that no genetic or epidemiologic variable and no gene-gene or gene-epidemiologic interaction of order 2 or higher influenced the case-control status significantly.

Global testing

We investigated whether the whole set of polymorphisms associated with the estrogen pathway might be significantly related to the breast cancer risk rather than single polymorphisms. Again no significant associations were observed.

Other non-parametric and non-linear classification methods such as support vector machines did not detect any significant association between the set of considered polymorphisms and breast cancer either.

Discussion

We performed association studies at multiple polymorphic loci across the estrogen metabolic pathway to identify genetic variants that may contribute to breast cancer risk. We focused on potential risk classifiers including hereditary variables as well as endocrine and exocrine hormonal history being particularly interested in major effects and possible interactions of gene-gene and gene-environment effects. Our investigation included 18 polymorphisms of 11 genes which have been selected on the basis of a minimum minor allele frequency of 4% in Caucasians and a known or predictable functional effect. To our knowledge this is the first report in which these 18 variants have been investigated together within the same study population for joint analyses. A special feature of our study is the comprehensive statistical analysis by means of a multitude of computational approaches to identify or refute effects on breast cancer risk. We observed modest effects in single factor and subgroup analyses as well as multivariate analyses. Although, we employed additional interaction and cluster analysis, MDR, logic regression, and global testing no statistical significant association with breast cancer risk was observed.

In single factor analyses *CYP1B1* was seemingly associated with breast cancer risk in that *CYP1B1*_1358_GG carriers had more frequently breast cancer than controls. The result however was not stable upon correction for multiple testing which is in line with a recent meta-analysis [42].

For the identification of subgroups of women at breast cancer risk we performed two-way stratification analyses. Given a required threshold *P*-value of 0.002 after adjusting for all 18 polymorphisms we consider *P*-values of <0.01 and a minimum number of five subjects in the strata noteworthy. For example, obese women (BMI <30 kg/m^2) showed a fourfold increased breast cancer risk when they were homozygous carriers of COMT$_{158Met}$. Likewise, carriers of COMT$_{158Met}$ showed a fourfold increased breast cancer risk when they were obese. Although, this observation is in line with epidemiologic findings of an increased breast cancer risk among obese women [43, 44] our observed breast cancer risk association with the more efficient COMT detoxification allele remains illusive. In contrast, slim women (BMI < 20 kg/m^2) seemed to be protected when they were homozygous carriers of HSD17B1$_{313Gly}$ of which functional implications are unknown. Moreover, carriers of the promoter *CYP17A1*_ -34_C variant showed an increased breast cancer risk when they had used HRT for more than 10 years. The latter observation may spur future investigations towards the elucidation of a molecular basis of an HRT-related breast cancer risk association.

Led by our findings from single polymorphisms we further hypothesized that any potential breast cancer risk effect might be observed and become even more evident in multifactor and interaction analyses. When we employed additional statistical methods neither method revealed a significant breast cancer risk model of gene–gene or gene–environment interaction even among high order interaction. Yet, our cluster analysis and logic regression approach suggest that some variables may influence the case-control status more than others. This observation stresses the value of association analyses and points to the likely role of yet unknown risk classifiers from this or other biological pathways which may corroborate towards a measurable breast cancer risk.

Additional information on a possible association between the genome and breast cancer risk may come from haplotype analyses. A recently reported mathematical model of the mammary estrogen metabolism provided a kinetic analysis for the estrogen metabolic pathway based on the conversion of 17beta-estradiol by the enzymes CYP1A1, CYP1B1, COMT, and GSTP1 into eight metabolites [45]. The model allows the prediction of concentrations of each metabolite including the transient quinones. Moreover, it was used to simulate the kinetic effect of enzyme polymorphisms on the pathway and identified haplotypes generating the largest amounts of catechols and quinones. 17beta-estradiol-3,4-quinone has been described as the most potent carcinogenic metabolite and combined haplotypes conferring variable metabolic activities were established across *CYP1A1*, *CYP1B1*, and *COMT*. Highest metabolite production has been assigned to the combined haplotype $\mathit{CYP1A1}_{461\text{Asn-}462\text{Ile}}\mathit{CYP1B1}_{48\text{Arg-}119\text{Ser-}432\text{Val-}453\text{Asn}}\mathit{COMT}_{158\text{Met}}$, for which an increased breast cancer risk has been shown in a case-control study [45]. Led by these compelling findings derived from *in silico* and genetic data we tested for a similar risk association in our case-control study. Yet, we were unable to detect an association with the combined haplotype $\mathit{CYP1A1}_{461\text{Asn-}462\text{Ile}}\mathit{CYP1B1}_{432\text{Val-}453\text{Asn}}\mathit{COMT}_{158\text{Met}}$ or any other haplotype. In comparison to Crooke et al. who included four *CYP1B1* polymorphisms our combined haplotype was limited to two of these *CYP1B1* polymorphisms which may explain the discrepant results. However, there may be a chance that these divergent results may be due to the low frequency of the combined haplotype in both studies.

Our finding of a lack of associations between breast cancer and polymorphisms participating in the estrogen metabolic pathway is relevant since we investigated a large number of potential classifiers. These have been at least in part and/or may be still under scrutiny in various international studies. Of note is the recent recommendation by the National Cancer Institute Breast and Prostate Cancer Cohort Consortium to conduct a search for low-penetrance breast cancer genes within the steroid-hormone metabolism pathway in pooled analyses of multiple large cohort studies [19]. Considering the large body of literature with respect to reported gene-breast cancer associations and the inherent error rate recently debated by Ioannidis [46], our study with more than 1,200 cases and controls is substantial. Yet, despite its power to detect potential major effects, a single effort like ours requires replication and confirmation preferably from large collaborative efforts. This has been recently demonstrated by the Breast Cancer Association Consortium that reported null results for common genetic variants from pooled analyses in as many as 12,000–30,000 subjects [47]. Our findings are important because we may infer that genetic polymorphisms scrutinized by us do not significantly contribute to the breast cancer risk in our German and possibly other populations. Moreover, the data obtained from multifactor and interaction analyses shed new light on the strategies applied towards the identification of breast cancer risk associations. We neither found nor confirmed any suggested risk associations which supports the critical discussion reviewed in Folkerd et al. [48]. Thus, single factor analysis may be interpreted with caution because any putative classifier may drop from the list of candidates when scrutinized within the context of multifactorial analysis. We propose, that the tested polymorphisms of the estrogen metabolic pathway may not play a direct role in breast cancer risk and that future investigations should be extended to other DNA variants of these genes and to other regulatory pathways.

Acknowledgements We are indebted to all women participating in the GENICA study. We gratefully acknowledge support by interviewers as well as physicians and pathologists of the study region. High-throughput genotyping analyses were supported by Sandra Brod. This work was supported by the Federal Ministry of Education and Research (BMBF) Germany grants 01KW9975/5, 01KW9976/8, 01KW9977/0, and 01KW0114, the Robert Bosch Foundation of Medical Research, Stuttgart, Deutsches Krebsforschungszentrum, Heidelberg, Berufsgenossenschaftliches Forschungsinstitut für Arbeitsmedizin Bochum, and Medizinische Universitäts- und Poliklinik, Bonn, Germany.

References

1. Parkin DM, Bray F, Ferlay J et al (2005) Global cancer statistics, 2002. CA Cancer J Clin 55:74–108
2. Stewart BW, Kleihues P (2003) World cancer report. IARC Press, Lyon
3. Malone KE, Daling JR, Doody DR et al (2006) Prevalence and predictors of BRCA1 and BRCA2 mutations in a population-based study of breast cancer in white and black American women ages 35 to 64 years. Cancer Res 66:8297–8308
4. Walsh T, Casadei S, Coats KH et al (2006) Spectrum of mutations in BRCA1, BRCA2, CHEK2, and TP53 in families at high risk of breast cancer. Jama 295:1379–1388
5. Wooster R, Weber BL (2003) Breast and ovarian cancer. N Engl J Med 348:2339–2347
6. Meijers-Heijboer H, van den Ouweland A, Klijn J et al (2002) Low-penetrance susceptibility to breast cancer due to CHEK2(*)1100delC in noncarriers of BRCA1 or BRCA2 mutations. Nat Genet 31:55–59
7. Rahman N, Seal S, Thompson D et al (2007) PALB2, which encodes a BRCA2-interacting protein, is a breast cancer susceptibility gene. Nat Genet 39:165–167
8. The CHEK2 Breast Cancer Case-Control Consortium (2004) CHEK2*1100delC and susceptibility to breast cancer: a collaborative analysis involving 10,860 breast cancer cases and 9,065 controls from 10 studies. Am J Hum Genet 74:1175–1182
9. Cox A, Dunning AM, Garcia-Closas M et al (2007) A common coding variant in CASP8 is associated with breast cancer risk. Nat Genet 39:352–358
10. Yager JD, Davidson NE (2006) Estrogen carcinogenesis in breast cancer. N Engl J Med 354:270–282
11. Key TJ, Allen NE, Spencer EA et al (2002) The effect of diet on risk of cancer. Lancet 360:861–868
12. Zumoff B (1998) Does postmenopausal estrogen administration increase the risk of breast cancer? Contributions of animal, biochemical, and clinical investigative studies to a resolution of the controversy. Proc Soc Exp Biol Med 217:30–37
13. Beral V (2003) Breast cancer and hormone-replacement therapy in the million women study. Lancet 362:419–427
14. Pesch B, Ko Y, Brauch H et al (2005) Factors modifying the association between hormone-replacement therapy and breast cancer risk. Eur J Epidemiol 20:699–711
15. Rossouw JE, Anderson GL, Prentice RL et al (2002) Risks and benefits of estrogen plus progestin in healthy postmenopausal women: principal results from the women's health initiative randomized controlled trial. Jama 288:321–333
16. Collaborative Group on Hormonal Factors in Breast Cancer (1996) Breast cancer and hormonal contraceptives: collaborative reanalysis of individual data on 53 297 women with breast cancer and 100 239 women without breast cancer from 54 epidemiological studies. Collaborative group on hormonal factors in breast cancer. Lancet 347:1713–1727
17. Kahlenborn C, Modugno F, Potter DM et al (2006) Oral contraceptive use as a risk factor for premenopausal breast cancer: a meta-analysis. Mayo Clin Proc 81:1290–1302
18. Pharoah PD, Dunning AM, Ponder BA et al (2004) Association studies for finding cancer-susceptibility genetic variants. Nat Rev Cancer 4:850–860
19. Hunter DJ, Riboli E, Haiman CA et al (2005) A candidate gene approach to searching for low-penetrance breast and prostate cancer genes. Nat Rev Cancer 5:977–985
20. Justenhoven C, Hamann U, Pesch B et al (2004) ERCC2 genotypes and a corresponding haplotype are linked with breast cancer risk in a German population. Cancer Epidemiol Biomarkers Prev 13:2059–2064
21. Vollmert C, Windl O, Xiang W et al (2006) Significant association of a M129V independent polymorphism in the 5′ UTR of the PRNP gene with sporadic Creutzfeldt-Jakob disease in a large German case-control study. J Med Genet 43:e53
22. Jaremko M, Justenhoven C, Abraham BK et al (2005) MALDI-TOF MS and TaqMan assisted SNP genotyping of DNA isolated from formalin-fixed and paraffin-embedded tissues (FFPET). Hum Mutat 25:232–238
23. Thompson PA, Shields PG, Freudenheim JL et al (1998) Genetic polymorphisms in catechol-*O*-methyltransferase, menopausal status, and breast cancer risk. Cancer Res 58:2107–2110
24. Bergman-Jungestrom M, Gentile M, Lundin AC et al (1999) Association between CYP17 gene polymorphism and risk of breast cancer in young women. Int J Cancer 84:350–353
25. Kurosaki K, Saitoh H, Oota H et al (1997) Combined polymorphism associated with a 3-bp deletion in the 5′- flanking region of a tetrameric short tandem repeat at the CYP19 locus. Nippon Hoigaku Zasshi 51:191–195
26. Kristensen VN, Andersen TI, Lindblom A et al (1998) A rare CYP19 (aromatase) variant may increase the risk of breast cancer. Pharmacogenetics 8:43–48
27. Tworoger SS, Chubak J, Aiello EJ et al (2004) Association of CYP17, CYP19, CYP1B1, and COMT polymorphisms with serum and urinary sex hormone concentrations in postmenopausal women. Cancer Epidemiol Biomarkers Prev 13:94–101
28. Berstein LM, Imyanitov EN, Suspitsin EN et al (2001) CYP19 gene polymorphism in endometrial cancer patients. J Cancer Res Clin Oncol 127:135–138
29. Haiman CA, Hankinson SE, Spiegelman D et al (2000) A tetranucleotide repeat polymorphism in CYP19 and breast cancer risk. Int J Cancer 87:204–210
30. Changes and Enhancement Through Release 8.0 Cary, NC: SAS Institute. SAS/STAT Software 2000
31. Stephens M, Smith NJ, Donnelly P (2001) A new statistical method for haplotype reconstruction from population data. Am J Hum Genet 68:978–989
32. Stephens M, Donnelly P (2003) A comparison of bayesian methods for haplotype reconstruction from population genotype data. Am J Hum Genet 73:1162–1169
33. Ickstadt K, Muller T, Schwender H (in press) Analysing SNPs–are there needles in the haysack? Chance
34. Selinski S (2006) Similarity measures for clustering SNP and epideiological data. Technical report, SFB 475, Department of Statistics. University of Dortmund, Germany
35. Hahn LW, Ritchie MD, Moore JH (2002) Multifactor dimensionality reduction of detecting gene–gene interactions. Bioinformatics 19:376–382
36. Ritchie MD, Hahn LW, Moore JH (2003) Power of multifactor dimensionality reduction for detecting gene–gene interactions in the presence of genotyping error, missing data, phenocopy, and genetic heterogeneity. Genet Epidemiol 24:150–157
37. Ritchie MD, Hahn LW, Roodi N et al (2001) Multifactor-dimensionality reduction reveals high-order interactions among

estrogen-metabolism genes in sporadic breast cancer. Am J Hum Genet 69:138–147

38. Ruczinski I, Kooperberg C, LeBlanc M (2003) Logic regression. J Comput Graph Stat 12:475–511
39. Schwender H, Ickstadt K (2006) Identification of SNP interactions using logic regression. Technical report, SFB 475, Department of Statistics. University of Dortmund, Germany
40. Gentleman RC, Carey VJ, Bates DM et al (2004) Bioconductor: open software development for computational biology and bioinformatics. Genome Biol 5:R80
41. Goeman JJ, van de Geer SA, de KF et al (2004) A global test for groups of genes: testing association with a clinical outcome. Bioinformatics 20:93–99
42. Wen W, Cai Q, Shu XO et al (2005) Cytochrome P450 1B1 and catechol-*O*-methyltransferase genetic polymorphisms and breast cancer risk in Chinese women: results from the shanghai breast cancer study and a meta-analysis. Cancer Epidemiol Biomarkers Prev 14:329–335
43. Abu-Abid S, Szold A, Klausner J (2002) Obesity and cancer. J Med 33:73–86
44. Lorincz AM, Sukumar S (2006) Molecular links between obesity and breast cancer. Endocr Relat Cancer 13:279–292
45. Crooke PS, Ritchie MD, Hachey DL et al (2006) Estrogens, enzyme variants, and breast cancer: a risk model. Cancer Epidemiol Biomarkers Prev 15:1620–1629
46. Ioannidis JP (2006) Common genetic variants for breast cancer: 32 largely refuted candidates and larger prospects. J Natl Cancer Inst 98:1350–1353
47. Breast Cancer Association Consortium (2006) Commonly studied single-nucleotide polymorphisms and breast cancer: results from the breast cancer association consortium. J Natl Cancer Inst 98:1382–1396
48. Folkerd EJ, Martin LA, Kendall A et al (2006) The relationship between factors affecting endogenous oestradiol levels in postmenopausal women and breast cancer. J Steroid Biochem Mol Biol 102:250–255