BMC Cancer

Research article

SNP-SNP interactions in breast cancer susceptibility

Venüs Ümmiye Onay[1,3], Laurent Briollais[1,2,5], Julia A Knight[1,2,5], Ellen Shi[4], Yuanyuan Wang[1,2], Sean Wells[1,3], Hong Li[1,3], Isaac Rajendram[1,3], Irene L Andrulis[1,3,4,6,7] and Hilmi Ozcelik*[1,3,7]

Address: [1]Fred A. Litwin Centre for Cancer Genetics, Samuel Lunenfeld Research Institute, Mount Sinai Hospital, Toronto, Ontario, Canada, [2]Prosserman Centre for Health Research, Samuel Lunenfeld Research Institute, Mount Sinai Hospital, Toronto, Ontario, Canada, [3]Department of Pathology and Laboratory Medicine, Mount Sinai Hospital, Toronto, Ontario, Canada, [4]Ontario Cancer Genetics Network, Cancer Care Ontario, Toronto, Ontario, Canada, [5]Department of Public Health Sciences, University of Toronto, Toronto, Ontario, Canada, [6]Department of Molecular and Medical Genetics, University of Toronto, Toronto, Ontario, Canada and [7]Department of Laboratory Medicine and Pathobiology, University of Toronto, Toronto, Ontario, Canada

Email: Venüs Ümmiye Onay - onay@mshri.on.ca; Laurent Briollais - laurent@mshri.on.ca; Julia A Knight - knight@mshri.on.ca; Ellen Shi - juqingshi@hotmail.com; Yuanyuan Wang - annywan@hotmail.com; Sean Wells - sean@mshri.on.ca; Hong Li - li@mshri.on.ca; Isaac Rajendram - idr1978@hotmail.com; Irene L Andrulis - andrulis@mshri.on.ca; Hilmi Ozcelik* - ozcelik@mshri.on.ca

* Corresponding author

Abstract

Background: Breast cancer predisposition genes identified to date (e.g., BRCA1 and BRCA2) are responsible for less than 5% of all breast cancer cases. Many studies have shown that the cancer risks associated with individual commonly occurring single nucleotide polymorphisms (SNPs) are incremental. However, polygenic models suggest that multiple commonly occurring low to modestly penetrant SNPs of cancer related genes might have a greater effect on a disease when considered in combination.

Methods: In an attempt to identify the breast cancer risk conferred by SNP interactions, we have studied 19 SNPs from genes involved in major cancer related pathways. All SNPs were genotyped by TaqMan 5'nuclease assay. The association between the case-control status and each individual SNP, measured by the odds ratio and its corresponding 95% confidence interval, was estimated using unconditional logistic regression models. At the second stage, two-way interactions were investigated using multivariate logistic models. The robustness of the interactions, which were observed among SNPs with stronger functional evidence, was assessed using a bootstrap approach, and correction for multiple testing based on the false discovery rate (FDR) principle.

Results: None of these SNPs contributed to breast cancer risk individually. However, we have demonstrated evidence for gene-gene (SNP-SNP) interaction among these SNPs, which were associated with increased breast cancer risk. Our study suggests cross talk between the SNPs of the DNA repair and immune system (XPD-[Lys751Gln] and IL10-[G(-1082)A]), cell cycle and estrogen metabolism (CCND1-[Pro241Pro] and COMT-[Met108/158Val]), cell cycle and DNA repair (BARD1-[Pro24Ser] and XPD-[Lys751Gln]), and within carcinogen metabolism (GSTP1-[Ile105Val] and COMT-[Met108/158Val]) pathways.

Conclusion: The importance of these pathways and their communication in breast cancer predisposition has been emphasized previously, but their biological interactions through SNPs have not been described. The strategy used here has the potential to identify complex biological links among breast cancer genes and processes. This will provide novel biological information, which will ultimately improve breast cancer risk management.

Background

The existence of dominant predisposition alleles/mutations, conferring a high breast cancer risk, has been confirmed with the discovery of BRCA1 and BRCA2 [1,2]. The functionally defective mutations of BRCA1 and BRCA2 are strongly associated with dramatically increased breast cancer risk, however, such mutations are found to be rare (<5%) in unselected breast cancer cases [1-3]. Besides these rare mutations, commonly occurring single nucleotide polymorphisms (SNPs) have also been shown to incrementally contribute to breast cancer risk, however, their individual contributions are relatively small [4-6].

SNPs have been historically classified as commonly occurring (>1%) genetic variation in the general population, whereas the rare variants with obvious functional consequences on the protein have been classified as mutations. Compared to mutations, SNPs have been perceived as functionally insignificant, however, current evidence emphasizes that a considerable fraction affects the intrinsic properties and the function of the proteins to a variable degree [7-9]. Although the effect of an individual SNP is generally small, the genetic effect of combinations of functionally relevant SNPs may additively or synergistically contribute to increased breast cancer risk. Epistasis or gene-gene interaction is likely to be a ubiquitous component of the genetic architecture of common diseases, such as breast cancer. The effects of epistasis could dictate functional outcomes over the independent effects of any one susceptibility gene [10]. Polygenic models have also been proposed to explain the joint effect of many susceptibility alleles on breast cancer, but without considering specifically their possible interactions [11-14].

To estimate breast cancer risk conferred by individual SNPs, as well SNP-SNP interactions, we have studied 19 SNPs from 18 key cancer genes, involved in DNA repair (XPD, PTEN, GADD45), cell cycle (CCND1, p27, BARD1), carcinogen/estrogen metabolism (ESR1, CYP17, COMT, GSTP1, GSTM3, MTHFR), immune system (IL1a, IL10, IL13, TNFa, G-CSF) and others (MMP1) (Table 1). SNPs were initially selected from the best evidence from published studies in the beginning of the project, in year 2000, and subsequently classified under three categories (high-, medium- and low- rank), representing SNPs with a wide range of functional evidence. High-rank SNPs were supported by studies, which demonstrated the effect of the SNP on the regulation of expression or protein function. The medium-rank category is more likely to include functionally relevant SNPs, as the substitutions are predicted to significantly affect function, although this was not confirmed experimentally. This category also includes SNPs, which were associated with breast cancer risk factors. The low ranking category, on the other hand, contained SNPs with no functional information. Among the

SNPs studied, XPD-[Lys751Gln], MTHFR-[Ala222Val], COMT-[Met108/158Val], GSTP1-[Ile105Val] and CCND1-[Pro241Pro], have been shown to alter the function or post-translational modification of their encoded protein [15-27]. MMP1-[1G(-1607)2G] and IL10-[G(-1082)A] have been shown to alter the transcription and expression of these genes [28-32]. IL13-[Arg130Gln] has been suggested to have functional consequences, while GSTM3-[4595 (3bp ins/del)] was predicted to create a YY1 transcription factor binding site [33,34]. The TNFA-[G(-308)A] forms a haplotype with some nearby SNPs and some studies observed increased haplotype dependent transcriptional activity change while some others do not [35-38]. CYP17-[C518T], and IL13-[Arg130Gln] were found to be associated with other cancer related variables, such as serum estrogen and IgE levels, respectively [39-41]. BARD1-[Pro24Ser] changes a structurally important non-polar proline residue to a positively charged serine. There were no functional speculations for ESR1-[Ser10Ser], ESR1-[Pro325Pro], PTEN-[(IVS4+109)ins/delACTAA], IL1A-[Ala114Ser], G-CSF-[Leu185Leu]and GADD45-[C(IVS3+168)T]. Thus, the 19 SNPs studied represent SNPs with a wide range of functional knowledge and evidence. The range of the minor allele frequencies of 19 SNPs studied varied between 15-48% in the general population. The SNPs studied were selected to represent more commonly occurring variants, in order to gain statistical power to detect SNP-SNP interactions.

Methods

Subject population

A case control study was conducted using biospecimens and data from the Ontario Familial Breast Cancer Registry (OFBCR) a participating site in the NIH-funded Breast Cancer Family Registry [42]. Written informed consent was obtained from all subjects, and the study protocol was approved by Mount Sinai Hospital Research Ethics Board.

Cases of invasive breast cancer, pathologically confirmed and diagnosed between 1996 and 1998 in the province of Ontario were identified from the population-based Ontario Cancer Registry. All female cases under 55, a random sample (35%) of female cases aged 55 to 69, and all male cases under age 80 were identified. Physician permission to contact patients was granted for 91% of cases (7668 of 8453). Patients were then mailed a cancer family history questionnaire and 65% (4957) completed it. All respondents who met a defined set of genetic risk criteria (i.e., Ashkenazi Jewish; diagnosed before age 36 years; previous ovarian or breast diagnosis; one or more first- or two or more second-degree relatives with breast or ovarian cancer; one or more second- or third-degree relatives with either breast cancer diagnosed before age 36 years, ovarian cancer diagnosed before age 61 years, multiple breast or breast and ovarian primaries, or male breast can-

Table 1: SNPs selected for the study and their functional consequences.

SNPs	Minor Allele (%)	Interacting SNPs*	Alteration	Function	Breast Cancer Association§	Selected Biological Processes¶
Higher Rank (Direct Functional Evidence)						
XPD-[Lys751Gln]	C: 32.3	Yes	Missense	Gln (C) allele has decreased DNA repair capacity [15-18]	Yes [83]	transcription-coupled nucleotide-excision repair, ATP-dependent DNA helicase activity
COMT-[Met108/158Val]	G: 47.3	Yes	Missense	Met (A) allele has lower enzymatic activity [20,21]	Yes [84]	neurotransmitter catabolism, catecholamine metabolism
GSTP1-[Ile105Val]	G: 31.2	Yes	Missense	Val (G) allele is associated with reduced enzymatic activity [22,23]	Yes [85]	Metabolism, glutathione transferase activity
MTHFR-[Ala222Val]	T: 35.6	No	Missense	Val (T) allele is associated with reduced enzyme activity [19]	Yes† [86]	Folate metabolism, methylene-tetrahydrofolate reductase (NADPH) activity
CCND1-[Pro241Pro]	A: 46.1	Yes	Splice Variant	(A) allele affects protein stability [24-27]	Yes [87]	G1/S transition of mitotic cell cycle
MMP1-[1G(-1607)2G]	Ins: 49.7	No	Regulatory	(Ins) allele shows increased transcription [28,29]	No	collagen catabolism, interstitial collagenase activity
IL10-[G(-1082)A]	G: 47.6	Yes	Regulatory	G allele is associated with increased expression [30-32]	No	cell-cell signaling, B-cell differentiation and proliferation, anti-apoptosis.
Medium Rank (Indirect Functional Evidence)						
BARD1-[Pro24Ser]	T: 37.5	Yes	Missense	Proline to serine change is a significant alteration	No	response to DNA damage, protein ubiquitination, regulation of apoptosis
IL13-[Arg130Gln]	A : 18.3	No	Missense	Gln allele associated with increased IgE levels [41]	No	immune response, inflammatory response, signal transduction,
p27-[Val109Gly]	G: 19.4	No	Missense	Possible function [33]	No	regulation of cyclin dependent protein kinase activity
GSTM3-[4595 (3bp ins/del)]	Del: 16.8	No	UTR	creates recognition site for transcription factor YY1 [34]	No	Metabolism, glutathione transferase activity
TNFA-[G(-308)A]	A: 17.6	No	UTR	increased transcriptional activity; also no functional change [35-38]	No	inflammatory response, signal transduction, regulation of transcription, apoptosis
CYP17-[C(518)T]	C: 34.3	No	Regulatory	Associated with increased serum estradiol [39,40]	Yes [4]	C21-Steroid hormone metabolism,

Table 1: SNPs selected for the study and their functional consequences. *(Continued)*

				Lower Rank (No Functional Evidence)			
IL1A-[Ala114Ser]	T : 27.4	No	Missense	no published functional evidence	No	MAPK signaling pathway, regulation of progression through cell cycle
GADD45-[C(IVS3+168)T]	C: 31.2	No	Intronic	no published functional evidence	No	regulation of cyclin dependent protein kinase activity, DNA repair, apoptosis
PTEN-[(IVS4+109)ins/del5	Ins: 30.1	No	Intronic	no published functional evidence	No	negative regulation of cell cycle, protein tyrosine/serine/threonine phosphatase activity
ESR1-[Ser10Ser]	C: 48.5	No	Silent	no published functional evidence	No	steroid hormone receptor activity, signal transduction, regulation of transcription
G-CSF-[Leu185Leu]	G: 38	No	Silent	no published functional evidence	No	immune response, cell-cell signaling, positive regulation of cell proliferation
ESR1-[Pro325Pro]	G: 24.1	No	Silent	no published functional evidence	No	steroid hormone receptor activity, signal transduction, regulation of transcription

*No and Yes indicates whether the SNP has been shown to be "interacting" or "not interacting" with other SNPs in this study; § The studies that showed statistically significant overall SNP-disease risk associations were considered (N_{cases} > 250 and $N_{controls}$ > 250).

†Cases are women diagnosed with breast cancer before age 40; ¶ GeneCards [88]

cer; three or more first-degree relatives with any combination of breast, ovarian, colon, prostate, or pancreatic cancer or sarcoma, with at least one diagnosis before age 51 years) were included in the study [43] and a random sample of 25% of those not meeting criteria were selected to continue to participate in the OFBCR (n= 2580). This participation included providing a blood sample (provided by 62% of all eligible, n = 1601). For the current study, we restricted the sample to women who identified themselves as Caucasian and were less than 55 years old. As we had randomly sampled 25% of those who did not meet genetic risk criteria, we also randomly sampled 25% of those who did meet genetic risk criteria in order to create a more representative sample of cases. Therefore, the cases should better represent all cases without enrichment for genetic risk criteria such as family history. In Table 2, 21.6% of cases in the present study had a first-degree family history of breast cancer, which is consistent with the 17 to 22% frequency reported in cases in a number of large case-control studies [44-46]. Of 459 Caucasian breast cancer cases with blood available, 398 were successfully genotyped and included in the study.

Controls were identified by calling randomly selected residential telephone numbers from across the province of Ontario and were frequency-matched to all female OFBCR cases by 5-year age group. The number of telephone numbers was 14,653, but 1101 (8%) were invalid and no contact could be made for 841 (6%). Of the 12,711 households contacted, 7829 (62%) did not have an eligible individual. No information on eligibility was provided for 2194 (17%) households. Of the 2688 eligible individuals identified on the telephone, 1726 (64%) completed the mailed risk factor questionnaire and 75% of these agreed to be contacted about providing a blood sample. The 676 women under age 55 who had agreed to be approached about blood sampling were asked to provide a blood sample and 419 (62%) did so. Individuals who were not Caucasian were excluded from the analysis, as were those with insufficient DNA or those subsequently found to be ineligible because of age. The remaining 372 population controls were successfully genotyped in this study.

Molecular genotyping
All SNPs were analyzed by TaqMan 5'nuclease assay [47] using the ABI PRISM 7900 HT Sequence Detection System (version 2.0). Oligonucleotide primers and the dual labeled allele specific probes were designed using PrimerExpress version 2.0 (PE Biosystems). Positions of primers for and probes in their appropriate accession numbers are given in Additional file 1.

A panel of DNA samples were sequenced for each SNP region initially, in order to identify control genotypes to

be used in each experiment. PCRs were performed in 96 well plates (AXYGEN) with each plate containing four control samples for each possible genotype. Genomic DNA (10 ng) was amplified in a total volume of 10 ul in the presence of 100 uM of each of the dNTPs, 3 pmoles of each of the appropriate primers, 2 pmoles of each of the corresponding dual labelled probes, and 0.025 units of Platinum Taq DNA Polymerase (InVitrogen). PCR cycling conditions consisted of 40 cycles of 94°C for 15 sec, 55–60°C for 15 sec and 72°C for 15 sec. The optimal $MgCl_2$ concentrations and annealing temperatures for each SNP are given in Additional file 2. The reliability of the results was determined by re-genotyping a randomly selected 10% portion of the total study population.

Statistical analyses
We sought evidence of association between each of the 19 SNPs and breast cancer risk in a multi-step process. At the first stage, we calculated crude allele and genotype frequencies for each individual polymorphism and evaluated Hardy-Weinberg equilibrium using a one-degree of freedom goodness-of-fit test among controls [48]. The association between the case-control status and each individual SNP, measured by the odds ratio (OR) and its corresponding 95% confidence interval, was estimated using unconditional logistic regression after adjustment for age. Several epidemiological risk factors were also assessed for association with breast cancer including age, BMI, education status, smoking status, family history, menopausal status, age at menarche, age at menopause, parity and age at first live birth (Table 2). Some of our analyses were also carried out adjusting the SNP main effect for the statistically significant epidemiological risk factors.

All analyses were performed assuming a dominant, recessive and co-dominant effect for each polymorphism. In the dominant model, both the heterozygous variant and the rare homozygous variant were combined. In the recessive model, the variant was defined as only the rare homozygous genotype and in the co-dominant model both rare homozygous and heterozygous variant effects were estimated using two dummy variables. In all analyses, the common homozygote genotype in the control population was defined as the reference category. Age was considered as a continuous variable. The likelihood ratio test was used to test the effect of each SNP at the nominal 5% significance level. Akaïke's information criterion [49] was also used to select the best genetic effect for each SNP.

At the second stage, two-way interactions were investigated using multivariate logistic models. More specifically, we tested all SNP-SNP interactions. We assumed a multiplicative interaction effect on the logit scale. Statistically significant interactions were selected using a forward stepwise selection procedure to evaluate evidence that

Table 2: Baseline characteristics of breast cancer cases and controls.

	Cases		Controls		P value#
	n	%	*n*	%	
Overall	398		372		
Age *					
≤ 40	79	19.8	69	18.6	0.08
41–45	106	26.6	98	26.4	
46–50	144	36.2	113	30.4	
51–55	69	17.3	91	24.5	
Missing	0	-	1	-	
Mean (SD)	44.8	(6.2)	45.2	(6.6)	0.37
BMI					
<18.5	7	2.2	4	1.1	0.05
18.5 – 25	178	55.1	166	46.0	
25 – 30	81	25.1	116	32.1	
> 30	57	17.6	75	20.8	
Missing	75	-	11	-	
Mean (SD)	25.7	(5.4)	26.3	(5.5)	0.14
Education					
Up to high school	122	35.4	101	27.2	0.06
Technical school	27	7.8	34	9.2	
Some college/ University	196	56.8	236	63.4	
Missing	53	-	1	-	
Smoking status					
Ever smoked	184	53.2	209	56.3	0.40
Never smoked	162	46.8	162	43.7	
Missing	52	-	1	-	
Family History					
Yes‡	86	21.6	34	9.2	<10^-5
No	312	78.4	337	90.8	
Missing	0	-	1	-	
Menopausal status					
Pre-menopausal	296	84.0	261	82.1	0.42
Post-menopausal	56	16.0	57	17.9	
Missing†	46	-	54	-	
Age at menarche					
Age ≤ 12	154	45.2	177	48.1	0.43
Age > 12	187	54.8	191	51.9	
Missing	57	-	4	-	
Age period stopped					
Pre-menopausal	296	84.6	261	82.1	0.68
Age ≤ 49	39	11.1	42	13.2	
Age > 49	15	4.3	15	4.7	
Missing	48	-	54	-	
Parity (# of live births)					
Nulliparous	68	19.6	63	17.0	0.85
1 child	56	16.1	57	15.4	
2 children	145	41.8	157	42.3	

Table 2: Baseline characteristics of breast cancer cases and controls. *(Continued)*

≥ 3 children	78	22.4	94	25.3	
Missing	51	-	1	-	
Age at 1st live birth					
Nulliparous	68	19.7	63	17.0	0.63
≤ 24 years	126	36.4	136	36.7	
> 24 years	152	43.9	172	46.4	
Missing	52	-	1	-	

*Age of first breast cancer diagnosis for cases and age at interview for controls
†Uncertain or missing data
‡Women with first-degree relatives with breast cancer
#χ^2 test for categorical variables or Student's t test (equal variance) for continuous variables

specific interactions were independently associated with breast cancer. The initial model included all SNPs and age as main effects, and then searched for the most significant candidate interactions to enter into the model based on the score statistics at the 5% level. Backward elimination of variables was then performed using the likelihood ratio test (LRT) also at the level of 5%. Forward stepwise selection procedure has proven to be efficient in assessing interaction effects as compared to backward elimination when testing multiple interactions. First, it is more time efficient and second, when using backward elimination, a relatively large number of predictor variables may increase the risk of complete separation of the two outcome groups, which would result in numerical problems in estimating the model parameters [50]. Since the genetic risk model is uncertain for most of the SNPs considered, we performed these tests on the co-dominant models only. Therefore, tests for SNP-SNP interactions have four degrees of freedom. All these analyses were also performed adjusting the interaction effects for the risk factors found to be associated with breast cancer risk at a significance level of 5% (BMI and family history).

We have also estimated the amount of linkage disequilibrium (LD) between the two ESR1 SNPs separated by about 140 kb on chromosome 6 and investigated their haplotype effect on breast cancer using the software "Unphased" from Dudbridge [51].

The large number of interactions (n = 171) analyzed could lead to false positive results, therefore, we adopted two different strategies to avoid this problem: The first approach included the assessment of the selection procedure using bootstrapping and the second one included an adjustment for multiple testing using detection rate (FDR). The bootstrap approach selects random samples of size n (n_1 cases + n_2 controls) with replacement from the original data [52]. Repeating the sampling procedure a large number of times provides information on the variability and validity of the parameter estimate and model selection. We repeated the selection procedure on 1,000

random samples (each random sample comprising 398 cases and 372 controls), generated from the original sample and the number of times a particular interaction was selected was reported. The achieved significance level (ASL) from the bootstrap test of hypotheses was also computed. Following Efron and Tibshirani [52], the ASL was obtained by comparing the observed LRT statistic for a specific interaction to its null distribution, evaluated by randomly assigning the case-control status in 1,000 bootstrap samples. The second approach tries to correct formally for the multiple testing problem using the FDR principle [53]. This procedure does not control the experiment-wise error rate like the Bonferroni-type correction (which is known to be conservative) but estimates the proportion of errors among the rejected null hypotheses. FDR was applied to both main effect models and interaction models using bootstrap *P*-values. For these latter models, bootstrap *P*-values and FDR-adjusted *P*-values correspond to interaction effects in multivariate logistic models that include all main effects and only the interaction of interest. This is equivalent to the test performed at the first step of the forward stepwise regression. We also computed the probability of no true association between an interaction and the disease status given a statistically significant result (i.e. the false positive report probability, FPRP) proposed by Wacholder et al. [54]. This statistic depends on the observed *P*-value but also on both the prior probability that the association between the SNP-SNP interaction and the disease is real and the statistical power of the test. The power of the test was determined by computing the expected value of the likelihood ratio test statistic, assuming our data were analyzed by the unconditional age-adjusted logistic model and with the specific coding of the SNP-SNP interactions using four dummy variables (see above). This computation was implemented into an R program, following the method described by Gauderman [55,56]. We used informative prior probabilities using the functional studies presented in Table 1 to classify the importance of each SNP. Following Wacholder's recommendations [54], the probability assigned to each SNP was 0.10, 0.01 and 0.001 for the

high-, medium-, and low-ranked SNPs respectively. The prior joint probability for each pair of SNPs was just the product of the individual SNP probability. Finally, we used the bootstrap *P*-values in the FPRP computation.

Results

Table 1 gives the minor allele frequencies of 19 SNPs estimated in our control population. None of the SNP distributions showed deviation from Hardy-Weinberg equilibrium in this sample. The distribution of selected epidemiologic risk factors in cases and controls is shown in Table 2. Cases and controls were similar with respect to the distribution of smoking status, menopausal status, age at menarche, age at menopause, parity, and age at first birth. Controls tended to have a higher BMI ($p = 0.05$) and level of education ($p = 0.06$) than cases. Cases were also more likely to have a positive family history of breast cancer than controls, and this difference was highly significant ($p = < 10^{-5}$). Logistic regression analysis was performed for all SNPs in the context of recessive, dominant and co-dominant models after adjustment for age. The estimated ORs and 95% CIs for all SNPs under co-dominant models are shown in Table 3.

Among the 19 SNPs studied, XPD-[Lys751Gln] was the only one showing a significant main effect in our sample based on the crude *P*-value. However, after correction for multiple testing using FDR, the effect was not significant. Our results remained unchanged when the models were also adjusted for BMI and family history. Results of two-way interaction analyses are shown in Table 4. Since the genetic risk models are uncertain for most of the SNPs considered, we performed these tests on the co-dominant models only. A total of nine SNP-SNP interactions were consistently selected in at least 30% of the random samples by the stepwise procedure. Interactions were observed more frequently for XPD-[Lys751Gln] and IL10-[G(-1082) (68%), and COMT-[Met108/158Val] and CCND1-[Pro241Pro] (61%). Interactions between GSTP1-[Ile105Val] and COMT-[Met108/158Val], CYP17-[C(518)T] and GADD45-[C(IVS3+168)T], and BARD1-[Pro24Ser] and ESR1-[Pro325Pro] selected in 54%, 53% and 51% of the random samples, respectively. All interactions, except BARD1-[Pro24Ser] and ESR1-[Pro325Pro], and, BARD1-[Pro24Ser] and p27-[Val109Gly], were statistically significant (P < 0.05) based on the bootstrap *P*-values.

After correction for multiple testing using FDR principle, four interactions remained significant at the 5% level; XPD-[Lys751Gln] and IL10-[G(-1082)A] (p = 0.007), GSTP1-[Ile105Val] and COMT-[Met108/158Val] (p = 0.007), COMT-[Met108/158Val] and CCND1-[Pro241Pro] (p = 0.014), and BARD1-[Pro24Ser] and XPD-[Lys751Gln] (p = 0.014). Based on the False Positive

Report Probability (FPRP) approach, computed using the functional importance of each SNP, we found that three interactions were noteworthy at the 0.2 FPRP level (Table 4); XPD-[Lys751Gln] and IL10-[G(-1082)A] (FPRP = 0.092), GSTP1-[Ile105Val] and COMT-[Met108/158Val] (FPRP = 0.169), and COMT-[Met108/158Val] and CCND1-[Pro241Pro] (FPRP = 0.093). The effect size of each SNP-SNP genotype combination for the four significant interactions is given in Table 5.

None of these interactions were significant after the more conservative Bonferroni adjustment. The Bonferroni adjusted *P*-values were 0.19 for both XPD-[Lys751Gln] and IL10-[G(-1082)A], and GSTP1-[Ile105Val] and COMT-[Met108/158Val] interactions. The *P*-values for the COMT-[Met108/158Val] and CCND1-[Pro241Pro], and BARD1-[Pro24Ser] and XPD-[Lys751Gln] interactions were 0.38.

The results of our multivariate analyses adjusted for age, BMI and family history confirmed the role of the most important interactions. The bootstrap *P*-values associated with XPD-[Lys751Gln] and IL10-[G(-1082)A], COMT-[Met108/158Val] and CCND1-[Pro241Pro], GSTP1-[Ile105Val] and COMT-[Met108/158Val], and BARD1-[Pro24Ser] and XPD-[Lys751Gln] were all significant (respectively, $P = 0.014$, $P = 0.020$, $P = 0.022$ and $P = 0.020$), however the significance of the tests decreased due to the high proportion of individuals missing BMI information. Therefore, only the analyses adjusted for age are presented in Table 4.

The amount of LD between the two ESR1 SNPs was relatively small, with a D' [57] of 0.07 in cases and 0.15 in controls, and none of the four haplotypes was significantly associated with breast cancer. Therefore, only the interaction effect between the two SNPs is presented.

Discussion

In this study, we have analysed the contribution of 19 SNPs from 18 cancer-related genes, to breast cancer risk in a case-control study of 398 breast cancer cases and 372 population controls, sampled from the population-based OFBCR. All cases and controls were Caucasian women under age 55. We found that among 19 SNPs, XPD-[Lys751Gln] substitution was the only one showing a significant association with breast cancer risk. However, after correction for multiple testing, the effect became insignificant, suggesting that this finding might be due to chance. Overall we found little evidence of breast cancer risk conferred by individual commonly occurring SNPs in this dataset.

Our main focus was to understand the contribution to breast cancer risk of functionally relevant SNP-SNP inter-

Table 3: Analysis of individual SNP effects on breast cancer. Estimated odds-ratios (OR) and 95% confidence intervals (CI) associated with each of the 19 polymorphisms selected under co-dominant age-adjusted main effect models (the significant associations are in bold).

SNP	Genotype	Cases n (%)	Controls n (%)	OR*	95% CI
COMT Met108/158Val	AA	94 (23.6)	96 (25.8)	1.0.	-
	AG	202 (50.8)	196 (52.7)	1.3	(0.9–1.9)
	GG	102 (25.6)	80 (21.5)	1.3	(0.9–2.0)
CCND1 Pro241Pro	GG	104 (26.1)	114 (30.6)	1.0.	-
	AG	203 (51.0)	178 (47.8)	1.3	(0.9–1.9)
	AA	91 (22.9)	80 (21.5)	1.3	(0.9–1.8)
IL13 Arg130Gln	GG	240 (60.3)	252 (67.8)	1.0.	-
	AG	144 (36.2)	105 (28.2)	0.8	(0.4–1.8)
	AA	14 (3.5)	15 (4.0)	1.2	(0.9–1.6)
IL1A Ala114Ser	GG	204 (51.3)	179 (48.1)	1.0.	-
	GT	155 (38.9)	164 (44.1)	1.0	(0.7–1.3)
	TT	39 (9.8)	29 (7.8)	1.7	(1.0–2.8)
G-CSF Leu185Leu	AA	146 (36.7)	138 (37.1)	1.0.	-
	AG	193 (48.5)	182 (48.9)	1.0	(0.7–1.3)
	GG	59 (14.8)	52 (14.0)	1.1	(0.7–1.8)
ESR1 Ser10Ser	TT	107 (26.9)	97 (26.1)	1.0.	-
	CT	200 (50.3)	187 (50.3)	1.1	(0.7–1.6)
	CC	91 (22.9)	88 (23.7)	1.1	(0.7–1.4)
TNF-A G(-308)A	GG	274 (68.8)	237 (63.7)	1.0.	-
	AG	113 (28.4)	120 (32.2)	0.9	(0.4–2.1)
	AA	11 (2.8)	15 (4.0)	0.9	(0.7–1.2)
CYP17 C(518)T	TT	170 (42.7)	172 (46.2)	1.0.	-
	CC	60 (15.1)	49 (13.2)	1.5	(1.0–2.4)
	CT	168 (42.2)	151 (40.6)	1.3	(0.9–1.7)
BARD1 Pro24Ser	CC	174 (43.7)	142 (38.2)	1.0.	-
	CT	188 (47.2)	184 (49.5)	0.8	(0.6–1.1)
	TT	36 (9.0)	46 (12.4)	0.8	(0.5–1.2)
XPD Lys751Gln	AA	146 (36.7)	165 (44.3)	1.0.	-
	AC	194 (48.7)	167 (44.9)	1.3	(0.9–1.7)
	CC	**58 (14.6)**	**40 (10.8)**	**1.6**	**(1.0–2.6)**
IL10 G(-1082)A	AA	90 (22.6)	107 (28.7)	1.0.	-
	AG	205 (51.5)	194 (52.2)	1.1	(0.7–1.5)
	GG	103 (25.9)	71 (19.1)	1.1	(0.7–1.6)
ESR1 Pro325Pro	CC	243 (61.1)	213 (57.3)	1.0.	-
	CG	126 (31.7)	138 (37.1)	1.0	(0.7–1.3)
	GG	29 (7.3)	21 (5.6)	1.0	(0.5–1.8)
MTHFR Ala222Val	CC	162 (40.7)	160 (43.0)	1.0.	-
	CT	170 (42.7)	161 (43.3)	1.1	(0.8–1.5)
	TT	66 (16.6)	51 (13.7)	1.3	(0.8–2.0)
GSTP1 Ile105Val	AA	208 (52.3)	175 (47.0)	1.0.	-
	AG	149 (37.4)	161 (43.3)	1.1	(0.8–1.5)
	GG	41 (10.3)	36 (9.7)	1.3	(0.8–2.0)

Table 3: Analysis of individual SNP effects on breast cancer. Estimated odds-ratios (OR) and 95% confidence intervals (CI) associated with each of the 19 polymorphisms selected under co-dominant age-adjusted main effect models (the significant associations are in bold). *(Continued)*

GADD45	*TT*	189 (47.5)	177 (47.6)	1.0.	-
C(IVS3+168)T	*CC*	32 (8.0)	43 (11.6)	0.8	(0.5–1.3)
	CT	177 (44.5)	152 (40.9)	0.9	(0.7–1.2)
P27	*TT*	256 (64.3)	235 (63.2)	1.0.	-
Val109Gly	*GG*	17 (4.3)	15 (4.0)	1.0	(0.5–2.0)
	GT	125 (31.4)	122 (32.8)	1.0	(0.7–1.3)
MMP1	*Del/del*	110 (27.6)	104 (28.0)	1.0.	-
1G(-1607)2G	*Ins/ins*	92 (23.1)	94 (22.6)	0.8	(0.6–1.3)
	Ins/del	196 (49.2)	184 (49.5)	1.0	(0.7–1.4)
PTEN	*Del/del*	171 (43.0)	196 (52.7)	1.0.	-
(IVS4+109) ins/ delACTAA	*Ins/ins*	47 (11.8)	38 (10.2)	1.3	(0.8–2.1)
	Ins/del	180 (45.2)	138 (37.1)	1.2	(0.9–1.7)
GSTM3	*Ins/ins*	271 (68.1)	272 (73.1)	1.0.	-
4595 (3bp ins/del)	*Del/del*	7 (1.8)	9 (2.4)	0.9	(0.4–2.1)
	Ins/del	120 (30.1)	91 (24.5)	0.9	(0.7–1.3)

*Age-adjusted odds ratios from unconditional logistic regression analyses.

actions within and between different cancer pathways. Recently, there has been increasing evidence regarding the joint effect of commonly occurring SNPs on cancer risk, supported by polygenic breast cancer models [11-14]. Although this model was originally based on the additive effects of multiple risk alleles [14,58] (each with a low to moderate risk) it can be generalized by considering interaction effects between the susceptibility alleles.

In this study, we have shown several statistically significant interactions between SNP pairs. Despite the low false discovery rates observed for certain interactions and small-unadjusted *P*-values, caution should be used when interpreting these results. First, the statistical modeling of interactions through a product term might not correspond to or reflect any biological interaction such as synergy or antagonism [59,60]. Second, as in any association study, epidemiologic limitations such as selection bias or confounding have the potential to lead to false-positive results. Cases were selected from a population-based cancer registry and although selection may have occurred, there was no evidence of selection related to family history of breast cancer [43,61]. This analysis was also restricted to Caucasians who had the highest response rates and also minimizes the potential for population stratification.

To assess the robustness of interaction models, we used two different strategies: an internal validation procedure based on bootstrap re-sampling methods [52] and a correction for multiple testing using the FDR principle [53].

The first approach allowed us to prioritize nine "candidate" SNP-SNP interactions that were consistently selected by the stepwise procedure across the bootstrap samples (i.e. in more than 30% of the 1,000 random samples) and that were significant based on the age adjusted bootstrap *P*-values. This method can be used to assess the variability of our model selection but does not control formally for the multiple testing problem. Although correction for multiple testing can be performed under the bootstrap framework [62] or using the familiar Bonferroni correction, these procedures can be very conservative. We therefore used the FDR principle that does not control the experiment-wise error rate but estimates the proportion of errors among the rejected null hypotheses. Using this correction, none of the SNP main effects were significant, but four 2-way interactions had adjusted *P*-values lower than 5%. The stepwise selection procedure applied to our original data set without validation or correction would have detected 14 significant SNP-SNP interactions at the 5% level. Based on our validation procedures, ten of these are likely to be false positive results. This shows the importance of model validation in studies of gene interactions [63]. Although this validation was internal (i.e. using the same data set), we intend to replicate our results using an external data set.

To interpret our positive results, we should also remember that our SNPs were selected from genes involved in cancer, and enriched by SNPs that are likely to affect the function of the encoded protein. Using a Bayesian approach, Wacholder [54] recently showed how the probability of

Table 4: Analysis of two-way SNP interaction effects on breast cancer.

Two-way Interactions between polymorphisms	Crude P-values‡	Bootstrap frequency of stepwise variable selection*	Bootstrap P-values†	FDR adjusted P-values§	FPRP¶
XPD-[Lys751Gln] and IL10-[G(-1082)A]	0.035	68%	0.001	**0.007**	**0.092**
COMT-[Met108/158Val] and CCND1-[Pro241Pro]	0.010	61%	0.002	**0.014**	**0.169**
GSTP1-[Ile105Val] and COMT-[Met108/158Val]	0.036	54%	0.001	**0.007**	**0.093**
CYP17-[C(518)T] and GADD45-[C(IVS3+168)T]	0.024	53%	0.018	0.062	0.999
BARD1-[Pro24Ser] and ESR1-[Pro325Pro]	0.039	51%	ns‖	ns	-
TNFA-[G(-308)A] and p27-[Val109Gly]	0.016	49%	0.025	**0.079**	0.996
BARD1-[Pro24Ser] and p27-[Val109Gly]	0.021	44%	ns	ns	-
BARD1-[Pro24Ser] and XPD-[Lys751Gln]	0.024	36%	0.002	**0.014**	0.671
ESR1-[Ser10Ser] and ESR1-[Pro325Pro]	0.028	30%	0.097	ns	0.999

‡ Probability to enter the model in the stepwise multivariable model based on the score statistic in the original data set.
*Proportion of times the interaction is selected by the stepwise multivariable model in the 1,000 bootstrap samples.
†P-values estimated by bootstrap analysis in the 1,000 random samples.
§False Discovery Rate (FDR) adjusted P-values.
‖ns: p > 0.10.
¶False Positive Report Probability

no true association between a genetic variant and disease given a statistically significant result (i.e. the false positive report probability (FPRP) depends on the prior probability that the association is real and also on the statistical power of the test. Although the determination of a prior probability is quite challenging [64], selecting SNPs based on their functions clearly reduce the FPRP. Based on this approach, we found that three out of the four significant interactions were noteworthy at the 0.2 FPRP level: XPD-[Lys751Gln] and IL10-[G(-1082)A], GSTP1-[Ile105Val] and COMT-[Met108/158Val], and COMT-[Met108/158Val] and CCND1-[Pro241Pro]. More complex approaches to account for the prior knowledge of the functional importance of each SNP have also been proposed [65-68]. We intend to use some of these methods to confirm our results in future analyses. Another problem to consider is the chance for false negative results. SNPs that were considered as negative after our conservative multiple comparison adjustment might still be worthy of investigation in other data sets. Therefore, some of our results must be viewed as hypothesis generating. Validation of our results (positive and negative) in an independent data set will provide further insight into the role of these SNPs in breast cancer etiology.

The novelty of our study is the demonstration of statistically significant interactions between SNPs that did not have an effect on breast cancer risk individually. Most studies have investigated the main effects of commonly occurring SNPs and categorized them as "not associated", and thus not important in breast cancer risk. However, our study suggests that SNPs without main effects or with main effects, too small to detect, may interact with others and confer an increased risk for breast cancer. Larger studies will allow a better application of our model, in which

more complex interactions could be investigated. SNP-SNP interactions in breast cancer development have been also reported in other studies, which targeted the SNPs of the carcinogen metabolism genes, including GSTM1, GSTT1, GSTP1, GSTM3 and CYPs [69-73]. These reports support our findings regarding SNP-SNP interactions on breast cancer risk, although they were only limited to the interactions of SNPs within a single cancer pathway.

In the context of breast cancer predisposition, our study suggests there is cross talk between the alleles of proteins of different cancer pathways including DNA repair and the immune system (XPD-[Lys751Gln] and IL10-[G(-1082)A]), cell cycle and estrogen metabolism (COMT-[Met108/158Val] and CCND1-[Pro241Pro]), cell cycle and DNA repair (BARD1-[Pro24Ser] and XPD-[Lys751Gln]) as well as within a single pathway such as estrogen metabolism (GSTP1-[Ile105Val] and COMT-[Met108/158Val]). Among the four interactions identified, COMT and XPD seem to play a central role since they interacted with different proteins in our set. XPD interaction was identified with both IL-10 and BARD1; whereas COMT interaction was identified with CCND1 and GSTP1. Estimation of odds ratios for particular genotype combinations can show considerable increase in the risk associated with breast cancer. For example the combination of rare genotypes of BARD1-[Ser24] (TT genotype) and XPD-[Gln751] (CC genotype) resulted in a relative risk of 7.4 (95% CI: 1.3-12.4). This is an excellent example where rare genotype combinations are associated with increased cancer risk. Such findings would have been missed in the absence of interaction analysis. These results emphasize the need for larger studies where the risk associated with such rare genotype combinations can be validated.

Table 5: Estimated odds-ratios (ORs) and 95% confidence intervals (CIs) associated with the genotype combinations of the four significant two-way SNP-SNP interactions.

SNP-SNP interaction	SNP Genotypes		Cases n	Controls n	OR†	95% CI†
	XPD	**IL10**				
XPD-[Lys751Gln] and IL10-[G(-1082)A]	AA	AA	40	57	1.0	-
	AC	AA	45	39	1.7	[0.9–2.7]
	CC	AA	21	8	3.7	[1.6–9.7]
	AA	AG	74	89	1.2	[0.8–1.9]
	AC	AG	99	71	1.9	[1.3–3.1]
	CC	AG	22	22	1.4	[0.8–3.1]
	AA	GG	41	26	2.2	[1.3–3.9]
	AC	GG	42	50	1.2	[0.7–2.4]
	CC	GG	14	10	2.2	[0.8–5.5]
	BARD1	**XPD**				
BARD1-[Pro24Ser] and XPD-[Lys751Gln]	CC	AA	66	60	1.0	-
	CT	AA	68	90	0.7	[0.4–1.1]
	TT	AA	21	22	0.9	[0.4–1.8]
	CC	AC	82	67	1.1	[0.7–1.8]
	CT	AC	89	69	1.2	[0.7–1.9]
	TT	AC	15	24	0.6	[0.3–1.2]
	CC	CC	23	13	1.6	[0.7–3.4]
	CT	CC	27	26	1.0	[0.5–1.9]
	TT	CC	7	1	7.4	[1.3–12.4]
	COMT	**CCND1**				
COMT-[Met108/158Val] and CCND1-[Pro241Pro]	AA	GG	21	40	1.0	-
	AG	GG	57	53	1.9	[1.3–4.6]
	GG	GG	25	21	2.4	[1.0–4.9]
	AA	AG	48	36	2.6	[1.3–4.5]
	AG	AG	110	92	2.3	[1.3–4.5]
	GG	AG	38	45	1.6	[1.0–3.2]
	AA	AA	22	29	1.4	[0.7–3.3]
	AG	AA	43	37	2.2	[1.1–4.3]
	GG	AA	34	19	3.3	[1.9–7.2]
	GSTP1	**COMT**				
GSTP1-[Ile105Val] and COMT-[Met108/158Val]	AA	AA	30	49	1.0	-
	AG	AA	46	48	1.5	[0.9–2.8]
	GG	AA	15	8	3.4	[1.1–11.4]
	AA	AG	95	92	1.7	[1.0–2.6]
	AG	AG	91	69	2.2	[1.4–3.6]
	GG	AG	24	21	1.8	[0.9–4.8]
	AA	GG	54	36	2.6	[1.4–4.2]
	AG	GG	36	41	1.5	[0.9–2.7]
	GG	GG	7	8	1.6	[0.5–4.2]

†OR and CIs estimated by bootstrap analysis in the 1,000 random samples.

We have further investigated the biological relationships between XPD and COMT interactions with other SNPs using protein-protein interaction databases and a literature search. As seen in Figure 1, statistical interaction between XPD-[Lys751Gln] and BARD1-[Pro24Ser] can be accounted for by protein-protein interactions among ERCC2 (XPD), TP53, BRCA1 and BARD1 [74,75]. Similarly the statistical interaction between XPD-[Lys751Gln] and IL10-[G(-1082)A] can be accounted for by protein-protein interactions among XPD(ERCC2), BRCA1, TP53, STAT5A, JAK1, IL-10RA and IL-10 [74-77]. It is important to point out that interaction of XPD (ERCC2) with IL-10

and BARD1 revolves around the protein-protein interaction of BRCA1 and TP53. As shown previously, mutations in these two proteins lead to predisposition to hereditary breast cancer [1,78]. This supports the observation that SNPs of XPD (ERCC2), BARD1 and IL-10 may be good candidates for breast cancer predisposition, which may also modify the effect of BRCA1 in carriers. On the other hand, statistical interaction of COMT-[Met108/158Val] with GSTP1-[Ile105Val] and CCND1-[Pro241Pro] revolves around estrogen metabolism and cell proliferation (Figure 2). Estrogen is an important risk factor for breast cancer. Estrogen is broken down into reactive species by phase I enzymes, which are then inactivated by phase II enzymes such as the methylation of catechol estrogens by COMT or the conjugation of oxidized estrogen-quinones by GSTP1. The relation between COMT and GSTP presumably depends on reduced inactivation of the aforementioned reactive estrogen intermediates, because of decreased activities of both enzymes [4,22,23]. The regulated level of estrogen might in turn influence the cell proliferation through CCND1 transcription [4,24-27,79].

The biological pathways investigated in this study have been previously implicated in breast cancer development; however, their genetic interactions, detected through variant alleles (SNPs), have not been previously described. These data and the statistical approaches applied to them have the potential to assist in the identification of complex biological relationships among cancer processes during the development of breast cancer. When moving into the era of "genetic dissection of complex traits [80]", we will need to abandon the concept of single genetic deter-

Figure 2
A simplified drawing showing the roles of COMT and GSTP1 enzymes in estrogen metabolism. The figure is modified from [4] with permission.

minants to favor the idea of a "web of causation [81]" involving multiple and complex pathways, which in turn could implicate many genes and environmental factors. This study provides a possible framework for a functional SNP-SNP interaction-based model for breast cancer risk.

Conclusion

Our focus in this study has been to uncover SNP-SNP interactions, which additively or synergistically contribute to breast cancer risk. From our small pool of SNPs, we have shown significant statistical interactions suggesting biological cross talk among genes/SNPs from DNA repair, cell cycle, immune system and carcinogen metabolism pathways. Our immediate task is to apply this strategy to a larger sample, with the aim of replicating our findings and investigating more complex interactions (involving three or more SNPs). This line of research has the potential to identify important cross talk between members of the cancer pathways in the disease state. This study not only provides insight into the analysis of the multi-genic nature of breast cancer, but also provides important information regarding how cell function relates to breast cancer development. We believe that these and other interactions in breast cancer will one day be identified and used in clinics to identify individuals at increased risk of breast cancer and develop preventive strategies.

Abbreviations

SNP – single nucleotide polymorphisms

FDR – false discovery rate

FPRP – false positive report probability

OFBCR – Ontario familial breast cancer registry

OR – odds ratio

LRT – likelihood ratio test

Figure 1
Evaluation of statistical interactions of XPD with IL-10 and BARD1 on the bases of protein-protein interactions of their intermediate proteins. Solid lines and dashed lines are used wherever there is a protein-protein or a statistical interaction, respectively. The protein-protein interaction map has been obtained using a protein-protein interaction database [82].
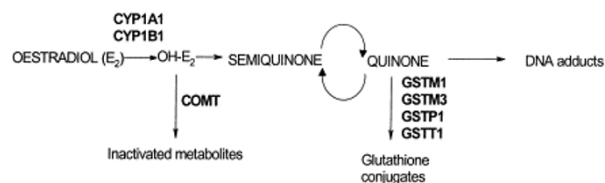
ASL – achieved significance level

LD – linkage disequilibrium

Competing interests

The author(s) declare that they have no competing interests.

Authors' contributions

VUO participated in design of the study, in data production and management, and manuscript preparation.

LB carried out the statistical analyses and wrote of the statistical sections of the manuscript.

JAK participated in formation of epidemiological study design and manuscript preparation.

ES participated in statistical analysis.

YW participated in statistical analysis.

SW participated in production of genotype data.

HL participated in production of genotype data.

IR participated in statistical analysis.

ILA provided information about study samples and OFBCR, and participated in manuscript preparation.

HO participated in the design and coordination of the study, and manuscript preparation.

All authors read and approved the final manuscript.

Additional material

Additional file 1

Primer_and_probe_information_for_19_SNPs. Positions of primers and probes used in the study, in their appropriate accession numbers, are given in this table.
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2407-6-114-S1.xls]

Additional file 2

PCR_Conditions_for_19_SNPs. The optimal $MgCl_2$ concentrations and annealing temperatures for each SNP PCR are given in this table.
Click here for file
[http://www.biomedcentral.com/content/supplementary/1471-2407-6-114-S2.xls]

References

1. Miki Y, Swensen J, Shattuck-Eidens D, Futreal PA, Harshman K, Tavtigian S, Liu Q, Cochran C, Bennett LM, Ding W: **A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1.** *Science* 1994, **266:**66-71.
2. Wooster R, Bignell G, Lancaster J, Swift S, Seal S, Mangion J, Collins N, Gregory S, Gumbs C, Micklem G: **Identification of the breast cancer susceptibility gene BRCA2.** *Nature* 1995, **378:**789-792.
3. Nelson HD, Huffman LH, Fu R, Harris EL: **Genetic risk assessment and BRCA mutation testing for breast and ovarian cancer susceptibility: systematic evidence review for the U.S. Preventive Services Task Force.** *Ann Intern Med* 2005, **143:**362-379.
4. Mitrunen K, Hirvonen A: **Molecular epidemiology of sporadic breast cancer. The role of polymorphic genes involved in oestrogen biosynthesis and metabolism.** *Mutat Res* 2003, **544:**9-41.
5. Goode EL, Ulrich CM, Potter JD: **Polymorphisms in DNA repair genes and associations with cancer risk.** *Cancer Epidemiol Biomarkers Prev* 2002, **11:**1513-1530.
6. Kang D: **Genetic polymorphisms and cancer susceptibility of breast cancer in Korean women.** *J Biochem Mol Biol* 2003, **36:**28-34.
7. Collins FS, Guyer MS, Charkravarti A: **Variations on a theme: cataloging human DNA sequence variation.** *Science* 1997, **278:**1580-1581.
8. Chakravarti A: **It's raining SNPs, hallelujah?** *Nat Genet* 1998, **19:**216-217.
9. Mehrian-Shai R, Reichardt JK: **A renaissance of "biochemical genetics"? SNPs, haplotypes, function, and complex diseases.** *Mol Genet Metab* 2004, **83:**47-50.
10. Moore JH: **The ubiquitous nature of epistasis in determining susceptibility to common human diseases.** *Hum Hered* 2003, **56:**73-82.
11. Ponder BA: **Cancer genetics.** *Nature* 2001, **411:**336-341.
12. Antoniou AC, Pharoah PD, McMullan G, Day NE, Ponder BA, Easton D: **Evidence for further breast cancer susceptibility genes in addition to BRCA1 and BRCA2 in a population-based study.** *Genet Epidemiol* 2001, **21:**1-18.
13. Antoniou AC, Pharoah PD, McMullan G, Day NE, Stratton MR, Peto J, Ponder BJ, Easton DF: **A comprehensive model for familial breast cancer incorporating BRCA1, BRCA2 and other genes.** *Br J Cancer* 2002, **86:**76-83.
14. Pharoah PD, Antoniou A, Bobrow M, Zimmern RL, Easton DF, Ponder BA: **Polygenic susceptibility to breast cancer and implications for prevention.** *Nat Genet* 2002, **31:**33-36.
15. Au WW, Salama SA, Sierra-Torres CH: **Functional characterization of polymorphisms in DNA repair genes using cytogenetic challenge assays.** *Environ Health Perspect* 2003, **111:**1843-1850.
16. Hemminki K, Xu G, Angelini S, Snellman E, Jansen CT, Lambert B, Hou SM: **XPD exon 10 and 23 polymorphisms and DNA repair in human skin in situ.** *Carcinogenesis* 2001, **22:**1185-1188.
17. Spitz MR, Wu X, Wang Y, Wang LE, Shete S, Amos CI, Guo Z, Lei L, Mohrenweiser H, Wei Q: **Modulation of nucleotide excision repair capacity by XPD polymorphisms in lung cancer patients.** *Cancer Res* 2001, **61:**1354-1357.
18. Qiao Y, Spitz MR, Shen H, Guo Z, Shete S, Hedayati M, Grossman L, Mohrenweiser H, Wei Q: **Modulation of repair of ultraviolet damage in the host-cell reactivation assay by polymorphic**

XPC and XPD/ERCC2 genotypes. *Carcinogenesis* 2002, **23**:295-299.

19. Yamada K, Chen Z, Rozen R, Matthews RG: **Effects of common polymorphisms on the properties of recombinant human methylenetetrahydrofolate reductase.** *Proc Natl Acad Sci USA* 2001, **98**:14853-14858.

20. Lotta T, Vidgren J, Tilgmann C, Ulmanen I, Melen K, Julkunen I, Taskinen J: **Kinetics of human soluble and membrane-bound catechol O-methyltransferase: a revised mechanism and description of the thermolabile variant of the enzyme.** *Biochemistry* 1995, **34**:4202-4210.

21. Lachman HM, Papolos DF, Saito T, Yu YM, Szumlanski CL, Weinshilboum RM: **Human catechol-O-methyltransferase pharmacogenetics: description of a functional polymorphism and its potential application to neuropsychiatric disorders.** *Pharmacogenetics* 1996, **6**:243-250.

22. Sundberg K, Johansson AS, Stenberg G, Widersten M, Seidel A, Mannervik B, Jernstrom B: **Differences in the catalytic efficiencies of allelic variants of glutathione transferase P1-1 towards carcinogenic diol epoxides of polycyclic aromatic hydrocarbons.** *Carcinogenesis* 1998, **19**:433-436.

23. Zimniak P, Nanduri B, Pikula S, Bandorowicz-Pikula J, Singhal SS, Srivastava SK, Awasthi S, Awasthi YC: **Naturally occurring human glutathione S-transferase GSTP1-1 isoforms with isoleucine and valine in position 104 differ in enzymic properties.** *Eur J Biochem* 1994, **224**:893-899.

24. Betticher DC, Thatcher N, Altermatt HJ, Hoban P, Ryder WD, Heighway J: **Alternate splicing produces a novel cyclin D1 transcript.** *Oncogene* 1995, **11**:1005-1011.

25. Diehl JA, Sherr CJ: **A dominant-negative cyclin D1 mutant prevents nuclear import of cyclin-dependent kinase 4 (CDK4) and its phosphorylation by CDK-activating kinase.** *Mol Cell Biol* 1997, **17**:7362-7374.

26. Alt JR, Cleveland JL, Hannink M, Diehl JA: **Phosphorylation-dependent regulation of cyclin D1 nuclear export and cyclin D1-dependent cellular transformation.** *Genes Dev* 2000, **14**:3102-3114.

27. Lu F, Gladden AB, Diehl JA: **An alternatively spliced cyclin D1 isoform, cyclin D1b, is a nuclear oncogene.** *Cancer Res* 2003, **63**:7056-7061.

28. Rutter JL, Mitchell TI, Buttice G, Meyers J, Gusella JF, Ozelius LJ, Brinckerhoff CE: **A single nucleotide polymorphism in the matrix metalloproteinase-1 promoter creates an Ets binding site and augments transcription.** *Cancer Res* 1998, **58**:5321-5325.

29. Tower GB, Coon CI, Brinckerhoff CE: **The 2G single nucleotide polymorphism (SNP) in the MMP-1 promoter contributes to high levels of MMP-1 transcription in MCF-7/ADR breast cancer cells.** *Breast Cancer Res Treat* 2003, **82**:75-82.

30. Turner DM, Williams DM, Sankaran D, Lazarus M, Sinnott PJ, Hutchinson IV: **An investigation of polymorphism in the interleukin-10 gene promoter.** *Eur J Immunogenet* 1997, **24**:1-8.

31. Suarez A, Castro P, Alonso R, Mozo L, Gutierrez C: **Interindividual variations in constitutive interleukin-10 messenger RNA and protein levels and their association with genetic polymorphisms.** *Transplantation* 2003, **75**:711-717.

32. Reuss E, Fimmers R, Kruger A, Becker C, Rittner C, Hohler T: **Differential regulation of interleukin-10 production by genetic and environmental factors – a twin study.** *Genes Immun* 2002, **3**:407-413.

33. Kibel AS, Suarez BK, Belani J, Oh J, Webster R, Brophy-Ebbers M, Guo C, Catalona WJ, Picus J, Goodfellow PJ: **CDKN1A and CDKN1B polymorphisms and risk of advanced prostate carcinoma.** *Cancer Res* 2003, **63**:2033-2036.

34. Yengi L, Inskip A, Gilford J, Alldersea J, Bailey L, Smith A, Lear JT, Heagerty AH, Bowers B, Hand P, Hayes JD, Jones PW, Strange RC, Fryer AA: **Polymorphism at the glutathione S-transferase locus GSTM3: interactions with cytochrome P450 and glutathione S-transferase genotypes as risk factors for multiple cutaneous basal cell carcinoma.** *Cancer Res* 1996, **56**:1974-1977.

35. Uglialoro AM, Turbay D, Pesavento PA, Delgado JC, McKenzie FE, Gribben JG, Hartl D, Yunis EJ, Goldfeld AE: **Identification of three new single nucleotide polymorphisms in the human tumor necrosis factor-alpha gene promoter.** *Tissue Antigens* 1998, **52**:359-367.

36. Bayley JP, de Rooij H, van den Elsen PJ, Huizinga TW, Verweij CL: **Functional analysis of linker-scan mutants spanning the -376, -308, -244, and -238 polymorphic sites of the TNF-alpha promoter.** *Cytokine* 2001, **14**:316-323.

37. Wilson AG, Symons JA, McDowell TL, McDevitt HO, Duff GW: **Effects of a polymorphism in the human tumor necrosis factor alpha promoter on transcriptional activation.** *Proc Natl Acad Sci USA* 1997, **94**:3195-3199.

38. Baseggio L, Bartholin L, Chantome A, Charlot C, Rimokh R, Salles G: **Allele-specific binding to the -308 single nucleotide polymorphism site in the tumour necrosis factor-alpha promoter.** *Eur J Immunogenet* 2004, **31**:15-19.

39. Nedelcheva KV, Haraldsen EK, Anderson KB, Lonning PE, Erikstein B, Karesen R, Gabrielsen OS, Borresen-Dale AL: **CYP17 and breast cancer risk: the polymorphism in the 5' flanking area of the gene does not influence binding to Sp-1.** *Cancer Res* 1999, **59**:2825-2828.

40. Feigelson HS, Shames LS, Pike MC, Coetzee GA, Stanczyk FZ, Henderson BE: **Cytochrome P450c17alpha gene (CYP17) polymorphism is associated with serum estrogen and progesterone concentrations.** *Cancer Res* 1998, **58**:585-587.

41. Graves PE, Kabesch M, Halonen M, Holberg CJ, Baldini M, Fritzsch C, Weiland SK, Erickson RP, von Mutius E, Martinez FD: **A cluster of seven tightly linked polymorphisms in the IL-13 gene is associated with total serum IgE levels in three populations of white children.** *J Allergy Clin Immunol* 2000, **105**:506-513.

42. John EM, Hopper JL, Beck JC, Knight JA, Neuhausen SL, Senie RT, Ziogas A, Andrulis IL, Anton-Culver H, Boyd N, Buys SS, Daly MB, O'Malley FP, Santella RM, Southey MC, Venne VL, Venter DJ, West DW, Whittemore AS, Seminara D: **The Breast Cancer Family Registry: an infrastructure for cooperative multinational, interdisciplinary and translational studies of the genetic epidemiology of breast cancer.** *Breast Cancer Res* 2004, **6**:R375-R389.

43. Knight JA, Sutherland HJ, Glendon G, Boyd NF, Andrulis IL: **Characteristics associated with participation at various stages at the Ontario site of the cooperative family registry for breast cancer studies.** *Ann Epidemiol* 2002, **12**:27-33.

44. Brinton LA, Hoover R, Fraumeni JF Jr: **Interaction of familial and hormonal risk factors for breast cancer.** *J Natl Cancer Inst* 1982, **69**:817-822.

45. Egan KM, Stampfer MJ, Rosner BA, Trichopoulos D, Newcomb PA, Trentham-Dietz A, Longnecker MP, Mittendorf R, Greenberg ER, Willett WC: **Risk factors for breast cancer in women with a breast cancer family history.** *Cancer Epidemiol Biomarkers Prev* 1998, **7**:359-364.

46. Marchbanks PA, McDonald JA, Wilson HG, Folger SG, Mandel MG, Daling JR, Bernstein L, Malone KE, Ursin G, Strom BL, Norman SA, Wingo PA, Burkman RT, Berlin JA, Simon MS, Spirtas R, Weiss LK: **Oral contraceptives and the risk of breast cancer.** *N Engl J Med* 2002, **346**:2025-2032.

47. Livak KJ: **Allelic discrimination using fluorogenic probes and the 5' nuclease assay.** *Genet Anal* 1999, **14**:143-149.

48. Weir BS: *Genetic Data Analysis II: Methods for discrete population genetic data* Sunderland, MA: Sinauer Associates; 1996.

49. Akaike H: **Information theory as an extension of the maximum likelihood principle.** In *Second International Symposium on Information Theory* Edited by: Petrov BN, Csaki F. Budapest: Akademiai Kiado; 1973:267-281.

50. Hosmer DW, Lemeshow S: *Applied Logistic Regression* New York: John Wiley & Sons; 1989.

51. Dudbridge F: **Pedigree disequilibrium tests for multilocus haplotypes.** *Genet Epidemiol* 2003, **25**:115-121.

52. Efron B, Tibshirani RJ: *An introduction to the bootstrap* London: Chapman & Hall; 1993.

53. Benjamini Y, Hochberg Y: **Controlling the false discovery rate: a practical approach and powerful approach for multiple testing.** *J R Stat Soc B* 1995, **57**:289-300.

54. 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.

55. Gauderman WJ: **Sample size requirements for matched case-control studies of gene-environment interaction.** *Stat Med* 2002, **21**:35-50.

56. Gauderman WJ: **Sample size requirements for association studies of gene-gene interaction.** *Am J Epidemiol* 2002, **155**:478-484.
57. Devlin B, Risch N: **A comparison of linkage disequilibrium measures for fine-scale mapping.** *Genomics* 1995, **29**:311-322.
58. Morton NE, MacLean CJ: **Analysis of family resemblance. 3. Complex segregation of quantitative traits.** *Am J Hum Genet* 1974, **26**:489-503.
59. Greenland S: **Modeling and variable selection in epidemiologic analysis.** *Am J Public Health* 1989, **79**:340-349.
60. Siemiatycki J, Thomas DC: **Biological models and statistical interactions: an example from multistage carcinogenesis.** *Int J Epidemiol* 1981, **10**:383-387.
61. Mancuso C, Glendon G, Anson-Cartwright L, Shi EJ, Andrulis I, Knight J: **Ethnicity, but not cancer family history, is related to response to a population-based mailed questionnaire.** *Ann Epidemiol* 2004, **14**:36-43.
62. Westfall P, Young SS: **P value adjustment for multiple tests in multivariate binomial models.** *JASA* 1989, **84**:780-785.
63. Coffey CS, Hebert PR, Krumholz HM, Morgan TM, Williams SM, Moore JH: **Reporting of model validation procedures in human studies of genetic interactions.** *Nutrition* 2004, **20**:69-73.
64. Thomas DC, Clayton DG: **Betting odds and genetic associations.** *J Natl Cancer Inst* 2004, **96**:421-423.
65. Conti DV, Cortessis V, Molitor J, Thomas DC: **Bayesian modeling of complex metabolic pathways.** *Hum Hered* 2003, **56**:83-93.
66. Cortessis V, Thomas DC: *Toxicokinetic genetics: An approach to gene-environment and gene-gene interactions in complex metabolic pathways. Mechanistic considerations in the molecular epidemiology of cancer* Edited by: Bird P, Boffetta P, Buffler P, Rice J. Lyon, France, IARC Scientific Publications; 2003.
67. Hung RJ, Brennan P, Malaveille C, Porru S, Donato F, Boffetta P, Witte JS: **Using hierarchical modeling in genetic association studies with multiple markers: application to a case-control study of bladder cancer.** *Cancer Epidemiol Biomarkers Prev* 2004, **13**:1013-1021.
68. Ulrich CM, Robien K, McLeod HL: **Cancer pharmacogenetics: polymorphisms, pathways and beyond.** *Nat Rev Cancer* 2003, **3**:912-920.
69. Levy-Lahad E, Lahad A, Eisenberg S, Dagan E, Paperna T, Kasinetz L, Catane R, Kaufman B, Beller U, Renbaum P, Gershoni-Baruch R: **A single nucleotide polymorphism in the RAD51 gene modifies cancer risk in BRCA2 but not BRCA1 carriers.** *Proc Natl Acad Sci USA* 2001, **98**:3232-3236.
70. Ritchie MD, Hahn LW, Roodi N, Bailey LR, Dupont WD, Parl FF, Moore JH: **Multifactor-dimensionality reduction reveals high-order interactions among estrogen-metabolism genes in sporadic breast cancer.** *Am J Hum Genet* 2001, **69**:138-147.
71. Ambrosone CB, Sweeney C, Coles BF, Thompson PA, McClure GY, Korourian S, Fares MY, Stone A, Kadlubar FF, Hutchins LF: **Polymorphisms in glutathione S-transferases (GSTM1 and GSTT1) and survival after treatment for breast cancer.** *Cancer Res* 2001, **61**:7130-7135.
72. Gudmundsdottir K, Tryggvadottir L, Eyfjord JE: **GSTM1, GSTT1, and GSTP1 genotypes in relation to breast cancer risk and frequency of mutations in the p53 gene.** *Cancer Epidemiol Biomarkers Prev* 2001, **10**:1169-1173.
73. Park SK, Yoo KY, Lee SJ, Kim SU, Ahn SH, Noh DY, Choe KJ, Strickland PT, Hirvonen A, Kang D: **Alcohol consumption, glutathione S-transferase M1 and T1 genetic polymorphisms and breast cancer risk.** *Pharmacogenetics* 2000, **10**:301-309.
74. Wang XW, Vermeulen W, Coursen JD, Gibson M, Lupold SE, Forrester K, Xu G, Elmore L, Yeh H, Hoeijmakers JH, Harris CC: **The XPB and XPD DNA helicases are components of the p53-mediated apoptosis pathway.** *Genes Dev* 1996, **10**:1219-1232.
75. Fabbro M, Savage K, Hobson K, Deans AJ, Powell SN, McArthur GA, Khanna KK: **BRCA1-BARD1 complexes are required for p53Ser-15 phosphorylation and a G1/S arrest following ionizing radiation-induced DNA damage.** *J Biol Chem* 2004, **279**:31251-31258.
76. Vidarsson H, Mikaelsdottir EK, Rafnar T, Bertwistle D, Ashworth A, Eyfjord JE, Valgeirsdottir S: **BRCA1 and BRCA2 bind Stat5a and suppress its transcriptional activity.** *FEBS Lett* 2002, **532**:247-252.
77. Ho AS, Wei SH, Mui AL, Miyajima A, Moore KW: **Functional regions of the mouse interleukin-10 receptor cytoplasmic domain.** *Mol Cell Biol* 1995, **15**:5043-5053.
78. Phillips KA, Nichol K, Ozcelik H, Knight J, Done SJ, Goodwin PJ, Andrulis IL: **Frequency of p53 mutations in breast carcinomas from Ashkenazi Jewish carriers of BRCA1 mutations.** *J Natl Cancer Inst* 1999, **91**:469-473.
79. Sabbah M, Courilleau D, Mester J, Redeuilh G: **Estrogen induction of the cyclin D1 promoter: involvement of a cAMP response-like element.** *Proc Natl Acad Sci USA* 1999, **96**:11217-11222.
80. Lander ES, Schork NJ: **Genetic dissection of complex traits.** *Science* 1994, **265**:2037-2048.
81. Rothman KJ: **Causes. 1976.** *Am J Epidemiol* 1995, **141**:90-95.
82. **NETPRO:Protein-protein Interaction Database** [http://www.biobase-international.com/]
83. Terry MB, Gammon MD, Zhang FF, Eng SM, Sagiv SK, Paykin AB, Wang Q, Hayes S, Teitelbaum SL, Neugut AI, Santella RM: **Polymorphism in the DNA repair gene XPD, polycyclic aromatic hydrocarbon-DNA adducts, cigarette smoking, and breast cancer risk.** *Cancer Epidemiol Biomarkers Prev* 2004, **13**:2053-2058.
84. Cheng TC, Chen ST, Huang CS, Fu YP, Yu JC, Cheng CW, Wu PE, Shen CY: **Breast cancer risk associated with genotype polymorphism of the catechol estrogen-metabolizing genes: a multigenic study on cancer susceptibility.** *Int J Cancer* 2005, **113**:345-353.
85. Egan KM, Cai Q, Shu XO, Jin F, Zhu TL, Dai Q, Gao YT, Zheng W: **Genetic polymorphisms in GSTM1, GSTP1, and GSTT1 and the risk for breast cancer: results from the Shanghai Breast Cancer Study and meta-analysis.** *Cancer Epidemiol Biomarkers Prev* 2004, **13**:197-204.
86. Campbell IG, Baxter SW, Eccles DM, Choong DY: **Methylenetetrahydrofolate reductase polymorphism and susceptibility to breast cancer.** *Breast Cancer Res* 2002, **4**:R14.
87. Ceschi M, Sun CL, Van Den BD, Koh WP, Yu MC, Probst-Hensch N: **The effect of cyclin D1 (CCND1) G870A-polymorphism on breast cancer risk is modified by oxidative stress among Chinese women in Singapore.** *Carcinogenesis* 2005, **26**:1457-1464.
88. **GeneCards:An Academic Website of the Weizmann Institute of Science** [http://www.genecards.org]

Pre-publication history

The pre-publication history for this paper can be accessed here:

http://www.biomedcentral.com/1471-2407/6/114/prepub