FULL PAPER

BJC

British Journal of Cancer (2013) 108, 2610–2622 | doi: 10.1038/bjc.2013.277

Keywords: breast cancer; segregation analysis; penetrance; common alleles; parent-of-origin effects; retrospective likelihood

Estimating single nucleotide polymorphism associations using pedigree data: applications to breast cancer

D R Barnes[1], D Barrowdale[1], J Beesley[2], X Chen[2], kConFab Investigators[3], Australian Ovarian Cancer Study Group[3,4,5], P A James[6,7], JL Hopper[8], D Goldgar[9], G Chenevix-Trench[2], A C Antoniou*[,1,10] and G Mitchell[6,7,10]

[1]*Department of Public Health and Primary Care, Centre for Cancer Genetic Epidemiology, University of Cambridge, Cambridge, UK;* [2]*Department of Genetics, Institute of Medical Research, Brisbane, Queensland, Australia;* [3]*Peter MacCallum Cancer Centre, Melbourne, Victoria, Australia;* [4]*Department of Gynaecological Oncology, Westmead Hospital, Westmead Institute for Cancer Research, University of Sydney at Westmead Millennium Institute, Sydney, New South Wales, Australia;* [5]*Queensland Institute of Medical Research, Brisbane, Queensland, Australia;* [6]*Familial Cancer Centre, Peter MacCallum Cancer Centre, St Andrews Place, East Melbourne, Victoria 3002, Australia;* [7]*Sir Peter MacCallum Department of Oncology, University of Melbourne, Melbourne, Victoria 3010, Australia;* [8]*Centre for Molecular, Environmental, Genetic and Analytic Epidemiology, Melbourne School of Population Health, University of Melbourne, Melbourne, Victoria, Australia and* [9]*Department of Dermatology, University of Utah School of Medicine, Salt Lake City, UT, USA*

Background: Pedigrees with multiple genotyped family members have been underutilised in breast cancer (BC) genetic-association studies. We developed a pedigree-based analytical framework to characterise single-nucleotide polymorphism (SNP) associations with BC risk using data from 736 BC families ascertained through multiple affected individuals. On average, eight family members had been genotyped for 24 SNPs previously associated with BC.

Methods: Breast cancer incidence was modelled on the basis of SNP effects and residual polygenic effects. Relative risk (RR) estimates were obtained by maximising the retrospective likelihood (RL) of observing the family genotypes conditional on all disease phenotypes. Models were extended to assess parent-of-origin effects (POEs).

Results: Thirteen SNPs were significantly associated with BC under the pedigree RL approach. This approach yielded estimates consistent with those from large population-based studies. Logistic regression models ignoring pedigree structure generally gave larger RRs and association *P*-values. SNP rs3817198 in *LSP1*, previously shown to exhibit POE, yielded maternal and paternal RR estimates that were similar to those previously reported (paternal RR = 1.12 (95% confidence interval (CI): 0.99–1.27), $P = 0.081$, one-sided $P = 0.04$; maternal RR = 0.94 (95% CI: 0.84–1.06), $P = 0.33$). No other SNP exhibited POE.

Conclusion: Our pedigree-based methods provide a valuable and efficient tool for characterising genetic associations with BC risk or other diseases and can complement population-based studies.

Large genome-wide association studies have identified several common genetic variants associated with complex diseases. To date, more than 60 common breast cancer (BC) susceptibility alleles have been identified (Cox *et al*, 2007; Easton *et al*, 2007; Stacey *et al*, 2007, 2010; Ahmed *et al*, 2009; Thomas *et al*, 2009; Antoniou *et al*, 2010; Turnbull *et al*, 2010; Fletcher *et al*, 2011;

*Correspondence: Dr AC Antoniou; E-mail: antonis@srl.cam.ac.uk
[10]Joint senior authors.

Received 21 November 2012; revised 6 May 2013; accepted 9 May 2013; published online 11 June 2013

Milne *et al*, 2011; Ghoussaini *et al*, 2012; Hein *et al*, 2012; Michailidou *et al*, 2013). At the time of the present analysis, 24 common alleles were known to be involved in BC susceptibility. However, recent studies based on genotyping of the iCOGS custom array have since identified 47 additional common BC susceptibility alleles (Couch *et al*, 2013; Garcia-Closas *et al*, 2013; Gaudet *et al*, 2013; Michailidou *et al*, 2013).

Genome-wide association studies have usually used samples of unrelated cases and unrelated controls to evaluate evidence of associations and obtain relative risk (RR) estimates. Family-based data, where several family members are genotyped, could be an additional resource to assess such associations and for characterising the risks conferred by genetic susceptibility variants, yet they are underutilised (Galvan *et al*, 2010). This approach is appealing because common alleles conferring increased disease risk are expected to cluster in families exhibiting disease family history (FH). Furthermore, with pedigree data it is possible to estimate genetic parent-of-origin specific risks depending on whether a risk allele was inherited from the father or mother, which is not possible under a population-based study design. Standard case–control analysis methods are not optimal for estimating the risks conferred by single-nucleotide polymorphisms (SNPs) in situations where families are ascertained on the basis of multiple disease cases. Analysing pedigree data using standard analytical methods (e.g., logistic regression) could lead to biased association estimates as they do not account for correlations in genotypes between related individuals. In addition, they do not adjust for the fact that families may be ascertained on the basis of multiple affected family members and that SNPs (or other genetic factors) are expected to be correlated with FH of the disease. The retrospective likelihood (RL) approach has been shown to adjust for ascertainment bias when ascertainment of individuals or families is non-random with respect to disease phenotype (Carayol and Bonaïti-Pellié, 2004). This approach involves modelling the likelihood of the observed family genotypes conditional on family disease phenotypes. We developed pedigree RL methods for assessing associations with genetic variants and estimating the associated risks in the context of genetic susceptibility to BC. This approach takes the form of a modified segregation analysis that accounts for explicit correlations in genotypes between related individuals while adjusting for ascertainment.

At the time of analysis, 24 SNPs had been shown to be associated with BC risk, primarily through large population-based case–control studies (Supplementary Table 1) (Cox *et al*, 2007; Easton *et al*, 2007; Stacey *et al*, 2007, 2010; Ahmed *et al*, 2009; Thomas *et al*, 2009; Antoniou *et al*, 2010; Turnbull *et al*, 2010; Fletcher *et al*, 2011; Milne *et al*, 2011; Ghoussaini *et al*, 2012; Hein *et al*, 2012). We applied the pedigree RL approach to estimate SNP associations with BC risk using data from 736 families recruited on the basis of strong FH of BC and a set of unrelated unaffected controls. Our results were contrasted to those obtained from standard analytical methods such as logistic regression.

There has been criticism of the assumption in association studies that maternally and paternally inherited alleles are functionally equivalent (Guilmatre and Sharp, 2012). Three mechanisms to describe parent-of-origin effects (POEs) have been suggested: (i) the influence of the maternal intrauterine environment on fetal developments; (ii) expression of genetic variation from the maternally inherited mitochondrial genome; and (iii) epigenetic regulation of gene expression, for example, genomic imprinting (suppression of gene expression that has been passed from one parent's germline) (Falls *et al*, 1999; Haghighi and Hodge, 2002; Rampersaud *et al*, 2008). Classic examples of imprinting are Prader–Willi and Angelman syndromes, which can occur when the same region on chromosome 15 is either maternally or paternally imprinted, respectively (Falls *et al*, 1999). A previous study found that one of the BC susceptibility variants that we analysed, SNP rs3817198 in the 11p15 region (*LSP1* gene), displayed POE with BC risk (Kong *et al*, 2009). Analysing data under a POE-type analysis, the paternally inherited allele expressed a significant association (OR = 1.17, 95% CI: 1.05–1.30, $P = 0.0038$), whereas the maternally inherited allele did not (OR = 0.91, 95% CI: 0.81–1.02, $P = 0.11$). These observations are consistent with reports that the 11p15 region hosts a cluster of imprinted genes, some of which may be related to BC risk (Berteaux *et al*, 2008). The results presented by Kong *et al* (2009) indicate a paternal effect of this locus on BC risk. These findings have not yet been replicated. We extended our pedigree RL framework to examine POE by estimating RRs separately for a maternally and paternally inherited risk allele. This is not possible under a standard case–control analytical design. We evaluated these associations for all BC susceptibility alleles investigated.

We further used the available genotype data to compute a combined observed genotype risk score to investigate whether this risk score can discriminate between women with FH of BC and unaffected women.

MATERIALS AND METHODS

Study sample. The Kathleen Cuningham Foundation Consortium for Research into Familial Breast Cancer (kConFab) enrols families with multiple cases of breast and/or ovarian cancer from Australia and New Zealand (Kathleen Cuningham Foundation Consortium for research into Familial Breast Cancer (kConFab), 2012). To date, kConFab has enrolled over 1400 families. The Australian Ovarian Cancer Study (AOCS) has recruited over 1800 ovarian cancer cases and 1000 population-based controls (Australian Ovarian Cancer Study (AOCS), 2012).

Our analyses considered data from 798 kConFab families. Eligibility was restricted to families with at least one family member genotyped for the SNPs of interest. Families were systematically screened for and excluded if found to contain a mutation in *BRCA1*, *BRCA2* or *ATM*. We excluded families if at least one family member was found to have a mutation in any of the *CHEK2*, *TP53*, *PTEN*, *RAD51C*, *MLH1* or *MSH2* genes, but screening of these genes was less systematic. In total, 736 families were eligible for analysis. A total of 897 unaffected population-based controls from AOCS were also included.

Mendelian inconsistencies in genotype transmission from parents to offspring were tested using PedCheck (O'Connell and Weeks, 1998). Detected Mendelian inconsistencies were rectified by first clarifying family relationships. Where this was not possible we replaced inconsistent genotypes as missing such that as little genetic data were lost and Mendelian consistency throughout the remainder of the pedigree held.

Genotyping. SNPs were genotyped using MALDI-TOF spectrophotometric mass determination of allele-specific primer extension products with Sequenom MassARRAY platform Sequenom, Inc., San Diego, CA, USA and iPLEX Gold technology (Sequenom, Inc.,). Primer design was carried out according to Sequenom guidelines using MassARRAY Assay Design software (version 3.0). Multiplex PCR amplification of fragments containing target SNPs was performed using Qiagen HotStart Taq Polymerase (QIAGEN, Hilden, Germany) and a PerkinElmer GeneAmp 2400 thermal cycler (PerkinElmer, Waltham, MA, USA) with 10 ng genomic DNA in 384 well plates. Shrimp Alkaline Phosphatase and allele-specific primer extension reactions were carried out according to the manufacturer's instructions for iPLEX Gold chemistry. Assay data were analysed using Sequenom TYPER software (version 3.4). Cluster plots were visually inspected and standard quality-control measures were checked, including Hardy–Weinberg equilibrium

$P \geqslant 0.01$, plate call rate $\geqslant 95\%$ and duplicate concordance rate $\geqslant 98\%$ (of 5% duplicated samples).

Analytical framework. We assumed an underlying genetic model where BC susceptibility is explained by the genetic variant of interest and a residual polygenic component that represents the multiplicative effects of several loci, each of which have small contributions to disease risk. The disease incidence, $\lambda_i(t)$, was assumed to depend on the genetic effects through a model of the form:

$$\lambda_i(t)=\lambda_0(t)\exp[\beta g_i+P_i]$$

where $\lambda_0(t)$ is the baseline incidence, β is the per-allele log RR, $g_i=\{0,1,2\}$ is the SNP genotype for individual i and P_i is the polygenic component assumed to be normally distributed:

$$P_i \sim N(0,\sigma_R^2)$$

where σ_R^2 is the residual polygenic variance. Because all families were found to segregate *BRCA1* or *BRCA2* mutations, as well as some rarer mutations in other susceptibility genes were excluded, this model is plausible for the families we analysed. We constrained the sum of the variance of the measured locus of interest, σ_K^2, and the residual polygenic variance, σ_R^2, such that they agree with external estimates of the total polygenic variance σ_P^2 (Antoniou *et al*, 2002). Hence,

$$\sigma_P^2=\sigma_K^2+\sigma_R^2 \quad (1)$$

This is in line with a multiplicative assumption between the measured locus and polygenic component. A previous segregation analysis estimated $\sigma_P=1.29$ (Antoniou *et al*, 2002). Under the polygenic model, $\exp(\sigma_P^2)$ is the coefficient of variation in incidences (Risch, 1990). $\exp(\sigma_P^2)$ is also the familial RR (FRR) to the monozygotic twin of an affected individual (λ_M), such that $\lambda_M=\exp(\sigma_P^2)$. Under the assumed model, it has previously been shown that the variance of the locus of interest, $\exp(\sigma_K^2)$, will be given by $\log(\lambda_{MK})$ where λ_{MK} is the FRR to a monozygotic twin due to the locus on its own (Risch, 1990; Antoniou and Easton, 2003). Therefore, the known component of the polygenic variance was calculated as;

$$\sigma_K^2=\log\left[\frac{\sum_g \tau_g \exp[2\beta g]}{\left(\sum_g \tau_g \exp[\beta g]\right)^2}\right]=\log\left[\frac{\tau_0+\tau_1\exp[2\beta]+\tau_2\exp[4\beta]}{(\tau_0+\tau_1\exp[\beta]+\tau_2\exp[2\beta])^2}\right] \quad (2)$$

where τ_g is the frequency of genotype $g=\{0,1,2\}$ calculated under the Hardy–Weinberg equilibrium assumption (Antoniou and Easton, 2003). The polygenic component was approximated by the hypergeometric polygenic model (Fernando *et al*, 1994; Lange, 1997; Antoniou *et al*, 2001).

We assumed a censoring process such that an individual was followed from birth until the age at first BC diagnosis, age of death, age at last observation or at 80 years of age, whichever occurred first. Individuals censored at 80 years of age were censored as unaffected at this time point. We assumed men were not at risk of developing BC. In the instance of no available censoring age, we censored at 0 years.

The BC incidences were constrained over all genetic effects (Antoniou *et al*, 2001) to agree with the Australian female BC incidences for the 1993–1997 calendar period (International Agency for Research on Cancer (IARC), 2010).

Retrospective likelihood segregation models. Because families were ascertained on the basis of multiple affected family members, we modelled the RL of observing family genotypes conditional on family disease phenotypes. The likelihood was parameterised in terms of the allele frequency and per-allele log RRs (β). To obtain parameter estimates, we maximised the likelihood over the genotype frequencies and log RR. We also fitted models where no residual polygenic effect was assumed in order to investigate the effect on parameter estimates when no assumptions were made about the residual familial clustering of BC.

Parent-of-origin effects. The pedigree RL framework was extended to account for POE. Here we simultaneously model the risk associated with a maternally inherited allele and paternally inherited allele. We denote the maternal log RR as β_m, the paternal log RR as β_p, a maternally inherited risk allele indicator variable taking values 0 if no maternally inherited risk allele is present and 1 if a maternally inherited risk allele is present as g_{i_m} and similarly a paternally inherited risk allele as g_{i_p}. Under this model, the disease incidence had the form:

$$\lambda_i(t)=\lambda_0(t)\exp[\beta_m g_{i_m}+\beta_p g_{i_p}]$$

We jointly maximised the likelihood over allele frequencies and both the maternal and paternal log RRs to obtain estimates for these parameters.

We evaluated evidence for POE by testing for differences between the maternal log RR and paternal log RR using a likelihood ratio test. For this purpose, the likelihood obtained from the POE model was compared with the likelihood under a single gene model that estimated a single per-allele HR assuming the same effect for maternally and paternally inherited risk alleles.

As the primary aim of the POE analysis was to test for equality in the paternal and maternal log RRs, the polygenic component was omitted. This was in order to reduce the computational complexity.

Logistic regression analyses. Standard logistic regression analyses were performed for comparison purposes. To account for relatedness within families, we estimated robust s.e. (Huber, 1967; White, 1980, 1982). Two types of analyses were undertaken: (i) unaffected AOCS controls *vs* all affected kConFab female family members and (ii) unaffected AOCS controls *vs* one selected affected kConFab female per family (usually the female family member that led to family ascertainment).

Assessing discrimination based on SNP profiles. To evaluate the ability of SNP profiles to discriminate between unaffected women and affected women with FH of BC, we computed an observed risk score (ORS) for each individual. The score, S_i, for individual i based on the combined effects of all SNPs was given by:

$$s_i=\sum_{j=1}^{S}\hat{\beta}_j g_{ji}$$

where S is the number of SNPs, $\hat{\beta}_j$ is the published population-based estimate of the per-allele log OR (Supplementary Table 1) and $g_{ji}=\{0,1,2\}$ is the observed genotype for individual i at SNP j. The ORS was calculated for a single affected female family member who had been genotyped for all SNPs and all controls. The discriminatory ability of the ORS was evaluated using receiver operating characteristic (ROC) curves by calculating the area under the curve (AUC).

Statistical software. Logistic regression and ROC analyses were performed using Stata version 11.1 (StataCorp LP, 2009). The segregation and POE models were implemented using pedigree analysis software MENDEL (Lange *et al*, 1988).

RESULTS

Study population. After quality-control checks, 736 kConFab families with at least one genotyped individual, comprising 45 822 individuals, and 897 unrelated unaffected controls from AOCS were eligible for analyses. Sample characteristics are summarised in Table 1. In brief, 6907 individuals were genotyped for at least one

SNP. Of these, 1673 (24.2%) were male and 5234 (75.8%) were female. In total, 1590 (30.4%) affected females and 3644 (69.6%) unaffected females were genotyped. The average number of individuals genotyped in these families was eight.

Single SNP association results using logistic regression and segregation analyses. Tables 2 and 3 display logistic regression and segregation analysis results. Figure 1 shows a comparison of log RR estimates under different analytical models.

Single gene models. Fourteen SNPs were significantly associated with BC risk at the 5% significance level when data were analysed under a single gene model that does not allow for residual polygenic effects. The most significant association was *FGFR2* SNP rs2981582 (HR = 1.20, 95% CI: 1.13–1.27, $P = 6.75 \times 10^{-10}$).

Incorporating residual polygenic effects. Thirteen SNPs were significantly associated with BC risk (5% significance level) when data were analysed under the model allowing for residual familial clustering in terms of a polygenic component. All these SNPs were significantly associated when the data were analysed under the single gene model. C6orf97 SNP rs12662670 was the only SNP significantly associated under the single gene model that was not associated with risk under the model that incorporates polygenic background (single gene $P = 3.64 \times 10^{-4}$; polygenic $P = 0.086$). Overall, *P*-values of association were similar under both pedigree analysis models (Figure 2). As with the single gene model, *FGFR2* SNP rs2981582 provided the strongest association with BC risk (HR = 1.26, 95% CI: 1.17–1.36, $P = 9.04 \times 10^{-10}$). For SNPs providing evidence of association ($P < 0.05$), the effect size estimates were somewhat larger under the model allowing for polygenic background but the strength of association was generally similar. The estimated HRs under the polygenic model were closer to OR estimates obtained from population-based studies than the estimates under the model that did not allow for polygenic background (Figure 1).

SNPs that were significantly associated with risk accounted for between 0.20 and 1.62% of the total polygenic variance, but most SNPs accounted for <1%. Only two SNPs, rs2981582 in *FGFR2* and rs13387042 at 2q35, accounted for >1% of the total polygenic variance.

A comparison of estimates of association from the segregation analyses to those obtained from the naive standard case–control analyses revealed that logistic regression typically overestimated associations. For almost all SNPs, the absolute value of the estimated log OR from the logistic regression comparing AOCS controls against all female cases exceeded those obtained under the segregation models. Moreover, the estimated ORs more often lay outside the CIs of the population-based OR estimates compared with the segregation analysis models (Supplementary Figure 1).

Parent-of-origin effects. The POE segregation analyses were performed assuming no residual polygenic background. This is a reasonable assumption as the primary aim was to test for differences in paternal and maternal HRs. Moreover, the pedigree analysis becomes complex because of the implementation of the hypergeometric approximation to the polygenic model. Results for POE analyses are given in Table 4.

Two SNPs showed significant associations with the paternally inherited allele only. Five SNPs yielded significant associations with the maternally inherited allele only. The HR estimate for the paternally inherited allele of SNP rs3817198 in *LSP1* was 1.12 (95% CI: 0.99–1.27, $P = 0.081$). Under a one-sided hypothesis testing HR >1, the *P*-value was 0.04.

One SNP, rs13387042 at 2q35, showed statistically significant associations for both a paternally inherited (HR = 1.20, 95% CI: 1.04–1.37, $P = 0.0096$) and maternally inherited risk allele (HR = 1.16, 95% CI: 1.03–1.31, $P = 0.014$). No SNP exhibited significant differences between HR estimates for the maternally and paternally inherited allele (*P*-value range: 0.07–0.95).

Risk score comparisons. Two SNPs at 19p13 (rs2363956 and rs8170) were excluded when constructing risk scores as they are primarily associated with ER-negative BC risk (Antoniou *et al*, 2010). The mean (s.d.) ORS was 2.47 (0.40) in 1147 individuals (715 unaffected and 432 affected) genotyped for all 22 SNPs. There was a significant difference in the mean ORS between unaffected (mean ORS (s.d.) = 2.40 (0.39)) and affected (2.60 (0.39)) women ($P = 6.38 \times 10^{-17}$). The estimated AUC was 0.642 (95% CI: 0.610–0.675) (Figure 3).

As expected, the distribution of the ORS for unaffected women from the kConFab families, that is women with FH of BC, lies between the risk distributions of the population-based controls and affected women (Supplementary Figure 2).

Table 1. Summary of the kConFab and AOCS study populations

	Study	
	kConFab	**AOCS**
n	45 822	897
Males/female	23 415/22 407	0/897
Pedigrees	736	897
Unaffected/affected	42 709/3113	897/0
Unaffected/affected (females only)	19 294/3113	897/0
n genotyped (at least one SNP)	6010	897
Male/female	1673/4337	0/897
Unaffected/affected	4420/1590	897/0
Unaffected/affected (females only)	2747/1590	897/0
n genotyped (22 risk prediction SNPs)	574	715
Male/female	14/560	0/715
Unaffected/affected	79/495	0/715
Unaffected/affected (females only)	65/495	0/715
n genotyped (all 24 SNPs)	564	714
Male/female	14/550	0/714
Unaffected/affected	79/485	0/714
Unaffected/affected (females only)	65/485	0/714
Mean (s.d.) censoring age (unaffected)	45.00 (23.75)	57.37 (11.62)
Censored aged ⩾18 years	52.29 (18.83)	57.37 (11.62)
Females only	38.41 (27.10)	57.37 (11.62)
Females censored aged ⩾18 years	51.90 (19.23)	57.37 (11.62)
Mean (s.d.) censoring age (affected)	51.50 (12.12)	N/A
Censored aged ⩾18 years	51.50 (12.12)	N/A
Females only	51.50 (12.12)	N/A
Females censored aged ⩾18 years	51.50 (12.12)	N/A

Abbreviations: AOCS = Australian Ovarian Cancer Study; kConFab = Kathleen Cuningham Foundation Consortium for Research into Familial Breast Cancer; *n* = number of individuals in sample. Censoring age in years.

DISCUSSION

In this article, we developed an analytical framework to estimate associations between SNPs and BC risk within a pedigree setting. This approach provides an efficient method for investigating

Table 2. Logistic regression analysis results

		AOCS controls *vs* all female cases				AOCS controls *vs* one selected female case per family			
SNP	Pedigrees	Affected/ unaffected	RAF[a]	OR (95% CI)[b]	*P*-value	Affected/ unaffected	RAF[a]	OR (95% CI)[b]	*P*-value
rs2981582	1577	1460/892	0.398	1.44 (1.26–1.65)	1.46×10^{-7}	837/892	0.398	1.36 (1.18–1.58)	0.00003
rs1975930	1515	1504/813	0.106	0.75 (0.60–0.94)	0.01115	812/813	0.106	0.74 (0.58–0.95)	0.01788
rs10941679	1485	719/873	0.258	1.16 (0.99–1.36)	0.07327	601/873	0.258	1.19 (1.01–1.41)	0.03899
rs3803662	1558	1461/872	0.274	1.30 (1.13–1.50)	0.00024	845/872	0.274	1.31 (1.12–1.52)	0.00069
rs2046210	1563	1471/874	0.348	1.14 (1.00–1.31)	0.05218	846/874	0.348	1.17 (1.01–1.35)	0.03836
rs614367	1607	1571/891	0.158	1.19 (1.01–1.40)	0.04325	877/891	0.158	1.20 (1.00–1.44)	0.05003
rs10509168	1608	1563/892	0.474	0.80 (0.71–0.92)	0.00124	846/892	0.474	0.80 (0.69–0.92)	0.00205
rs1292011	1433	823/812	0.432	0.95 (0.82–1.10)	0.50408	574/812	0.432	0.95 (0.81–1.11)	0.50687
rs13387042	1551	1452/863	0.483	1.41 (1.24–1.61)	2.61×10^{-7}	790/863	0.483	1.38 (1.20–1.60)	0.00001
rs13281615	1561	1422/874	0.402	1.19 (1.04–1.35)	0.01090	838/874	0.402	1.16 (1.01–1.34)	0.03900
rs865686	1511	1426/812	0.393	0.90 (0.78–1.03)	0.13578	775/812	0.393	0.86 (0.73–1.00)	0.04650
rs11249433	1565	1474/875	0.413	1.11 (0.97–1.27)	0.11657	847/875	0.413	1.08 (0.94–1.25)	0.29256
rs2823093	1511	1489/813	0.252	0.97 (0.84–1.14)	0.74598	792/813	0.252	0.92 (0.78–1.09)	0.35286
rs3817198	1562	1463/873	0.324	1.10 (0.96–1.26)	0.18494	846/873	0.324	0.98 (0.84–1.14)	0.81915
rs889312	1560	1462/871	0.280	1.13 (0.98–1.31)	0.08430	840/871	0.280	1.14 (0.97–1.33)	0.10425
rs1011970	1608	1573/892	0.161	1.12 (0.95–1.32)	0.18912	860/892	0.161	1.08 (0.89–1.30)	0.43456
rs17468277	1564	1468/875	0.141	0.79 (0.65–0.96)	0.01540	853/875	0.141	0.78 (0.63–0.97)	0.02331
rs999737	1528	1492/831	0.746	1.16 (0.99–1.35)	0.05823	853/831	0.746	1.16 (0.99–1.37)	0.07231
rs2380205	1580	898/891	0.408	1.15 (1.00–1.32)	0.05295	636/891	0.408	1.09 (0.95–1.26)	0.22967
rs4973768	1558	1439/873	0.455	1.22 (1.07–1.39)	0.00289	814/873	0.455	1.22 (1.06–1.41)	0.00507
rs6504950	1563	1468/875	0.302	0.92 (0.80–1.06)	0.26037	858/875	0.302	0.96 (0.82–1.11)	0.56381
rs2363956	1507	1388/813	0.502	1.06 (0.92–1.21)	0.42933	775/813	0.502	1.06 (0.92–1.23)	0.43906
rs8170	1512	1496/813	0.191	1.02 (0.86–1.21)	0.78295	806/813	0.191	1.04 (0.87–1.26)	0.64778
rs12662670	1514	1500/813	0.065	1.60 (1.25–2.04)	0.00016	799/813	0.065	1.67 (1.28–2.17)	0.00015

Abbreviations: AOCS = Australian Ovarian Cancer Study; CI = confidence interval; OR = odds ratio; RAF = risk allele frequency; SNP, single-nucleotide polymorphism.
[a]RAF is the observed risk allele frequency in unaffected individuals.
[b]Per-allele OR is reported such that the effect allele is the same as those from the population-based studies (Supplementary Table 1).

associations of polymorphisms on disease risk. We extended these methods to estimate parent-of-origin associations by separately estimating HRs for maternally and paternally inherited risk alleles. This is the first time POE have been evaluated for most of the common genetic variants found to be associated with BC risk. Although we demonstrate these methods in the context of evaluating associations with BC risk, the principles are applicable to other cancers but also other complex diseases that exhibit familial aggregation.

We applied these methods to family data from kConFab, a family-based study in which families were recruited through multiple relatives diagnosed with breast and/or breast/ovarian cancer. Analysing such associations using standard analytical methods could yield biased association estimates due to non-random ascertainment of families with respect to disease phenotype and that genetic variants are likely to be correlated with FH of disease. Analysing data within a pedigree RL framework accounts for relatedness and adjusts for ascertainment bias.

Our results demonstrate that standard logistic regression analyses applied in this context generally overestimate the magnitude of disease associations when compared with estimates published by large collaborative studies. More often, those were outside the published CIs. However, estimates from the modified segregation analysis were, generally, very close and within the CIs of the reported estimates by the population-based studies (Cox *et al*, 2007; Easton *et al*, 2007; Stacey *et al*, 2007, 2010; Ahmed *et al*, 2009; Thomas *et al*, 2009; Antoniou *et al*, 2010; Turnbull *et al*, 2010; Fletcher *et al*, 2011; Milne *et al*, 2011; Ghoussaini *et al*, 2012; Hein *et al*, 2012).

In addition, the segregation models generally yielded smaller *P*-values for association than those obtained through the logistic regression analysis. This suggests that this approach has greater power to detect associations than using standard case–control analysis that ignores pedigree structure. Likely explanations include the fact that pedigree analysis methods model exact genetic correlations between relatives, and the additional information is extracted by phenotypes of family members that had not been genotyped. Additional gains in power would be expected by the use of pedigree-based methods in settings where a clear ascertainment process exists, which would involve conditioning on the phenotypes of all family members. Therefore, a family-based approach is a useful and efficient method to investigate the contribution of genetic variants to disease risk.

Our models used external data on population BC incidences and for the magnitude of the assumed polygenic variance in the polygenic model. Sensitivity analysis by misspecifying the assumed

Table 3. Single gene and polygenic segregation analysis results

		Single gene model					Polygenic model						
SNP	**Pedigrees**	**RAF (s.e.)[a]**	**HR (95% CI)[b]**	***P*-value**	**logL**	**AIC**	**RAF (s.e.)[a]**	**HR (95% CI)[b]**	***P*-value**	**VE (%)**	σ_R^2	**logL**	**AIC**
rs2981582	1626	0.400 (0.008)	1.20 (1.13–1.27)	6.75×10^{-10}	−5904.28	11812.56	0.400 (0.008)	1.26 (1.17–1.36)	9.04×10^{-10}	1.62	1.637	−5903.33	11 810.67
rs1975930	1549	0.113 (0.005)	0.85 (0.77–0.94)	0.002	−2803.64	5611.28	0.113 (0.005)	0.82 (0.72–0.93)	0.002	0.41	1.657	−2804.21	5612.41
rs10941679	1505	0.257 (0.010)	1.09 (1.00–1.19)	0.039	−1480.82	2965.63	0.255 (0.010)	1.13 (1.02–1.26)	0.023	0.38	1.658	−1480.57	2965.14
rs3803662	1606	0.268 (0.007)	1.16 (1.09–1.23)	2.41×10^{-6}	−5203.61	10411.22	0.269 (0.007)	1.20 (1.11–1.30)	3.92×10^{-6}	0.87	1.650	−5202.84	10 409.68
rs2046210	1608	0.348 (0.007)	1.08 (1.02–1.14)	0.011	−5675.98	11355.96	0.349 (0.007)	1.09 (1.01–1.18)	0.028	0.21	1.661	−5676.56	11 357.13
rs614367	1627	0.152 (0.005)	1.12 (1.04–1.20)	0.002	−3941.45	7886.89	0.152 (0.005)	1.15 (1.05–1.27)	0.003	0.35	1.658	−3940.83	7885.65
rs10509168	1628	0.466 (0.008)	0.90 (0.85–0.96)	0.001	−5958.28	11920.57	0.464 (0.008)	0.89 (0.83–0.96)	0.002	0.39	1.658	−5959.16	11 922.32
rs1292011	1455	0.434 (0.011)	0.97 (0.90–1.04)	0.349	−2177.39	4358.78	0.433 (0.011)	0.96 (0.87–1.06)	0.400	0.05	1.663	−2177.43	4358.85
rs13387042	1598	0.489 (0.008)	1.18 (1.11–1.25)	3.84×10^{-8}	−5904.92	11813.85	0.489 (0.008)	1.23 (1.14–1.32)	1.08×10^{-7}	1.25	1.643	−5905.31	11 814.62
rs13281615	1608	0.401 (0.008)	1.09 (1.03–1.16)	0.002	−5585.14	11174.29	0.402 (0.008)	1.12 (1.04–1.20)	0.004	0.36	1.658	−5585.28	11 174.56
rs865686	1547	0.386 (0.008)	0.95 (0.90–1.01)	0.112	−5397.41	10798.83	0.387 (0.008)	0.93 (0.86–1.01)	0.069	0.15	1.662	−5397.31	10 798.63
rs11249433	1609	0.414 (0.008)	1.05 (1.00–1.12)	0.072	−5949.54	11903.07	0.415 (0.008)	1.05 (0.98–1.13)	0.174	0.08	1.663	−5950.01	11 904.01
rs2823093	1549	0.261 (0.007)	0.95 (0.89–1.02)	0.164	−4882.64	9769.28	0.260 (0.007)	0.95 (0.87–1.04)	0.249	0.06	1.663	−4882.86	9769.72
rs3817198	1607	0.337 (0.007)	1.02 (0.96–1.08)	0.489	−5451.60	10907.21	0.338 (0.007)	1.02 (0.94–1.10)	0.678	0.01	1.664	−5451.68	10 907.36
rs889312	1605	0.283 (0.007)	1.07 (1.01–1.14)	0.023	−5181.52	10367.04	0.283 (0.007)	1.09 (1.01–1.18)	0.028	0.20	1.661	−5181.43	10 366.87
rs1011970	1628	0.172 (0.006)	1.02 (0.95–1.10)	0.544	−4031.16	8066.33	0.171 (0.006)	1.03 (0.94–1.14)	0.499	0.02	1.664	−4031.11	8066.21
rs17468277	1609	0.141 (0.005)	0.88 (0.80–0.96)	0.004	−3351.37	6706.75	0.141 (0.005)	0.84 (0.75–0.94)	0.002	0.40	1.657	−3351.03	6706.07
rs999737	1566	0.743 (0.007)	1.10 (1.02–1.17)	0.008	−4767.76	9539.52	0.743 (0.007)	1.12 (1.03–1.22)	0.008	0.29	1.659	−4767.38	9538.76
rs2380205	1599	0.412 (0.010)	1.05 (0.98–1.13)	0.141	−2370.80	4745.61	0.412 (0.010)	1.07 (0.98–1.17)	0.156	0.12	1.662	−2370.65	4745.29
rs4973768	1607	0.462 (0.008)	1.09 (1.03–1.15)	0.005	−5904.62	11813.24	0.462 (0.008)	1.12 (1.04–1.21)	0.003	0.38	1.658	−5904.53	11 813.06
rs6504950	1609	0.288 (0.007)	0.98 (0.92–1.05)	0.578	−5224.70	10453.40	0.289 (0.007)	0.97 (0.89–1.05)	0.432	0.03	1.664	−5224.80	10 453.61
rs2363956	1548	0.508 (0.008)	1.02 (0.96–1.08)	0.515	−5362.61	10729.23	0.508 (0.008)	1.02 (0.95–1.11)	0.533	0.02	1.664	−5362.60	10 729.20
rs8170	1549	0.182 (0.006)	1.04 (0.97–1.12)	0.241	−4063.20	8130.40	0.182 (0.006)	1.06 (0.96–1.16)	0.233	0.06	1.663	−4063.24	8130.48
rs12662670	1549	0.076 (0.004)	1.18 (1.08–1.29)	3.64×10^{-4}	−2573.28	5150.57	0.077 (0.004)	1.20 (1.06–1.36)	0.003	0.34	1.659	−2574.68	5153.35

Abbreviations: AIC = Akaike Information Criterion (Akaike, 1974); CI = confidence interval; HR = hazard ratio; logL = model maximum log-likelihood; RAF = risk allele frequency; VE = percentage of the total polygenic variance explained by the locus of interest; σ_R^2 = residual polygenic variance. The total polygenic variance estimated by a previous segregation analysis was $\sigma_P^2 = 1.6641$ (Antoniou *et al*, 2002).

[a]RAF is the maximum likelihood estimate of the risk allele frequency from the segregation analysis model.

[b]Per-allele HR is reported such that the effect allele is the same as those from the population-based studies (Supplementary Table 1).

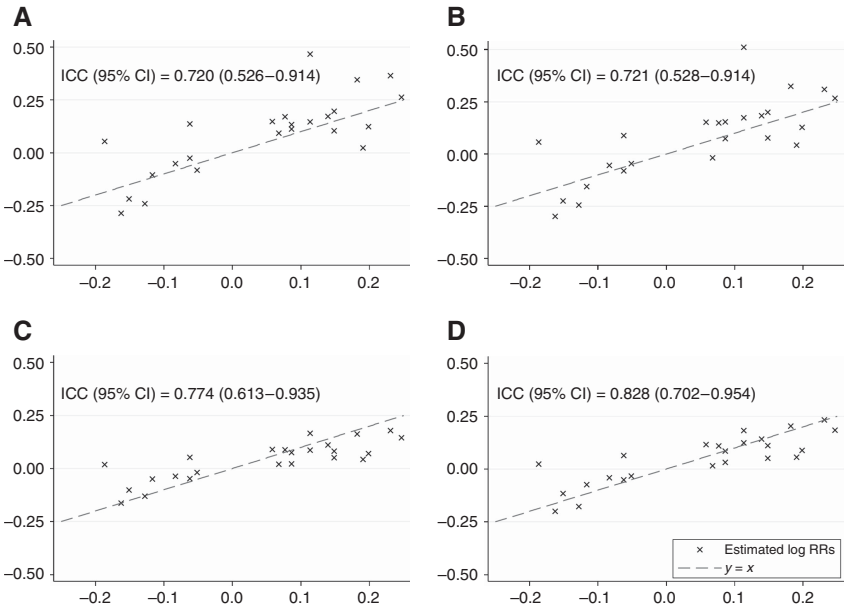

Figure 1. Scatter plots of log RR estimates from published population-based studies (Supplementary Table 1) (all *x*-axes) *vs*: (**A**) logistic regression estimates comparing AOCS controls against all familial cases (Table 2); (**B**) logistic regression estimates comparing AOCS controls against one selected female case per family (Table 2); (**C**) single gene segregation model estimates (Table 3); and (**D**) polygenic segregation model estimates (Table 3). The dashed line is $y = x$, the line of equality. ICC = intraclass correlation coefficient.

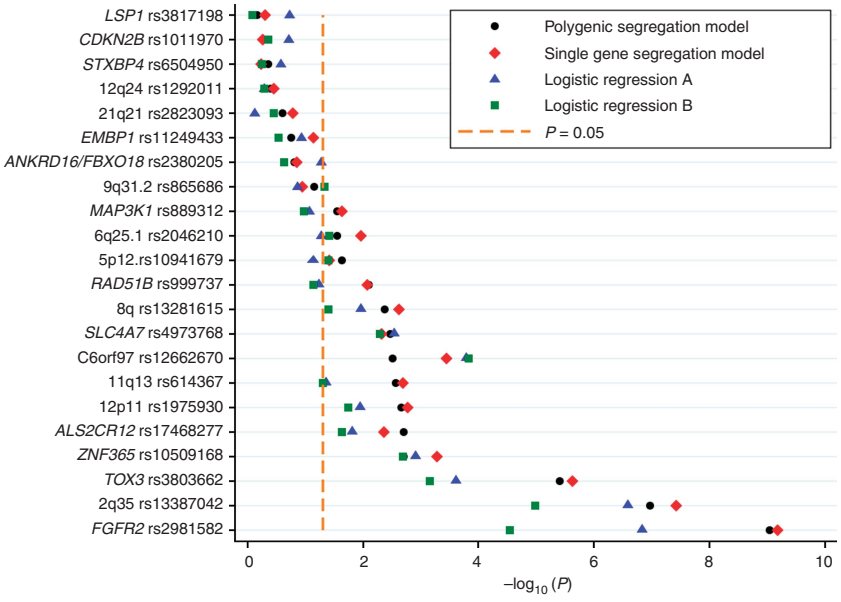

Figure 2. Scatter plot of $-\log_{10}$ *P*-values from the: (i) polygenic segregation model (Table 3); (ii) single gene segregation model (Table 3); (iii) logistic regression A: logistic regression estimates comparing AOCS controls against all familial cases (Table 2); and (iv) logistic regression B: logistic regression estimates comparing AOCS controls against one selected female case per family (Table 2). The dashed line represents a *P*-value of 0.05, the nominal significance level. SNPs are ordered by the *P*-values of the polygenic segregation analysis model. The segregation models generally yielded smaller *P*-values, indicating that these models have greater power to detect associations. 19p13 SNPs rs2363956 and rs8170 are not displayed as they are associated with ER-negative BC.

population incidences to be half or double the true population incidences revealed small deviations in the RR estimates (relative bias $<3\%$). Similarly, varying the assumed polygenic variance to be up to 80% of the assumed polygenic variance in our models had a negligible effect on the RR estimates (relative bias $<1\%$). This suggests that the estimates obtained under the methods presented are robust against misspecifications in the external model parameters.

Alternative association methods using pedigree data have been suggested. A case-only pedigree RL approach had been suggested and applied to the analysis of associations with prostate cancer risk (Schaid *et al*, 2010). However, this differs from our approach in that it does not consider genotype data from unaffected family members. Our approach allows for estimation of allele frequencies and RR parameters simultaneously, whereas Schaid *et al* used external allele frequency estimates. Unlike Schaid *et al*, our analyses incorporated all genetic information provided from all family members, therefore providing more information in the estimation process. The genetic model employed by Schaid *et al*

Table 4. Segregation analysis results allowing for parent-of-origin effects

			Maternal		Paternal		Difference				
SNP	Pedigrees	RAF (s.e.)[a]	HR (95% CI)[b]	*P*-value	HR (95% CI)[b]	*P*-value	δ (s.e.)	χ^2	*P*-value	logL	AIC
rs2981582	1626	0.401 (0.008)	1.26 (1.09–1.46)	0.002	1.13 (0.95–1.34)	0.157	−0.107 (0.150)	0.504	0.478	−5904.026	11814.05
rs1975930	1549	0.113 (0.005)	0.70 (0.55–0.90)	0.006	1.00 (0.83–1.20)	0.981	0.348 (0.194)	3.180	0.075	−2802.051	5610.10
rs10941679	1505	0.257 (0.010)	0.91 (0.62–1.32)	0.611	1.31 (0.94–1.81)	0.108	0.366 (0.347)	0.684	0.408	−1480.474	2966.95
rs3803662	1606	0.268 (0.007)	1.11 (0.99–1.25)	0.080	1.21 (1.07–1.38)	0.003	0.088 (0.109)	0.654	0.419	−5203.282	10412.56
rs2046210	1608	0.348 (0.007)	1.03 (0.91–1.16)	0.637	1.14 (1.00–1.30)	0.051	0.102 (0.112)	0.830	0.362	−5675.565	11357.13
rs614367	1627	0.152 (0.005)	1.23 (1.08–1.39)	0.001	0.98 (0.83–1.17)	0.863	−0.220 (0.134)	2.700	0.100	−3940.095	7886.19
rs10509168	1628	0.466 (0.008)	0.84 (0.75–0.95)	0.007	0.98 (0.86–1.12)	0.769	0.151 (0.118)	1.588	0.208	−5957.49	11920.98
rs1292011	1455	0.433 (0.011)	1.17 (0.92–1.49)	0.205	0.78 (0.59–1.03)	0.082	−0.400 (0.252)	1.628	0.202	−2176.575	4359.15
rs13387042	1598	0.489 (0.008)	1.16 (1.03–1.31)	0.014	1.20 (1.04–1.37)	0.010	0.030 (0.115)	0.068	0.795	−5904.889	11815.78
rs13281615	1608	0.401 (0.008)	1.09 (0.96–1.22)	0.178	1.10 (0.96–1.27)	0.155	0.017 (0.116)	0.022	0.882	−5585.132	11176.26
rs865686	1547	0.386 (0.008)	0.99 (0.87–1.14)	0.934	0.91 (0.77–1.06)	0.214	−0.093 (0.136)	0.468	0.494	−5397.179	10800.36
rs11249433	1609	0.414 (0.008)	1.06 (0.94–1.19)	0.346	1.05 (0.92–1.20)	0.486	−0.009 (0.114)	0.006	0.937	−5949.532	11905.06
rs2823093	1549	0.261 (0.007)	0.89 (0.77–1.02)	0.104	1.03 (0.89–1.20)	0.669	0.152 (0.133)	1.292	0.256	−4881.996	9769.99
rs3817198	1607	0.337 (0.007)	0.94 (0.84–1.06)	0.330	1.12 (0.99–1.27)	0.081	0.170 (0.108)	2.468	0.116	−5450.369	10906.74
rs889312	1605	0.283 (0.007)	1.01 (0.90–1.15)	0.811	1.14 (1.01–1.30)	0.042	0.119 (0.111)	1.152	0.283	−5180.945	10367.89
rs1011970	1628	0.172 (0.006)	1.05 (0.91–1.21)	0.489	0.99 (0.84–1.17)	0.912	−0.059 (0.136)	0.186	0.666	−4031.07	8068.14
rs17468277	1609	0.141 (0.005)	0.90 (0.72–1.14)	0.389	0.85 (0.65–1.10)	0.224	−0.062 (0.233)	0.072	0.788	−3351.337	6708.67
rs999737	1566	0.743 (0.007)	1.09 (0.95–1.25)	0.242	1.11 (0.94–1.30)	0.219	0.018 (0.136)	0.018	0.894	−4767.749	9541.50
rs2380205	1599	0.412 (0.010)	0.94 (0.76–1.16)	0.587	1.19 (0.95–1.49)	0.128	0.233 (0.210)	1.100	0.294	−2370.253	4746.51
rs4973768	1607	0.462 (0.008)	1.09 (0.96–1.23)	0.168	1.08 (0.94–1.24)	0.269	−0.009 (0.119)	0.006	0.937	−5904.616	11815.23
rs6504950	1609	0.288 (0.007)	1.02 (0.90–1.16)	0.724	0.94 (0.81–1.08)	0.379	−0.088 (0.122)	0.514	0.473	−5224.443	10454.89
rs2363956	1548	0.508 (0.008)	1.04 (0.92–1.19)	0.506	0.99 (0.86–1.14)	0.914	−0.051 (0.124)	0.170	0.680	−5362.528	10731.06
rs8170	1549	0.182 (0.006)	1.10 (0.95–1.27)	0.194	0.98 (0.82–1.17)	0.812	−0.116 (0.146)	0.620	0.431	−4062.888	8131.78
rs12662670	1549	0.076 (0.004)	1.26 (1.08–1.46)	0.004	1.08 (0.88–1.34)	0.460	−0.148 (0.161)	0.854	0.355	−2572.856	5151.71

Abbreviations: AIC = Akaike Information Criterion (Akaike, 1974); CI = confidence interval; χ^2 = 1 df test statistic based on likelihood ratio test between the POE model and the standard major gene segregation model; δ = difference between maternal and paternal log HRs; HR = hazard ratio; logL = model maximum log-likelihood; RAF = risk allele frequency.
[a]RAF is the maximum likelihood estimate of the risk allele frequency from the segregation analysis model.
[b]Per-allele HR is reported such that the effect allele is the same as those from the population-based studies (Supplementary Table 1).

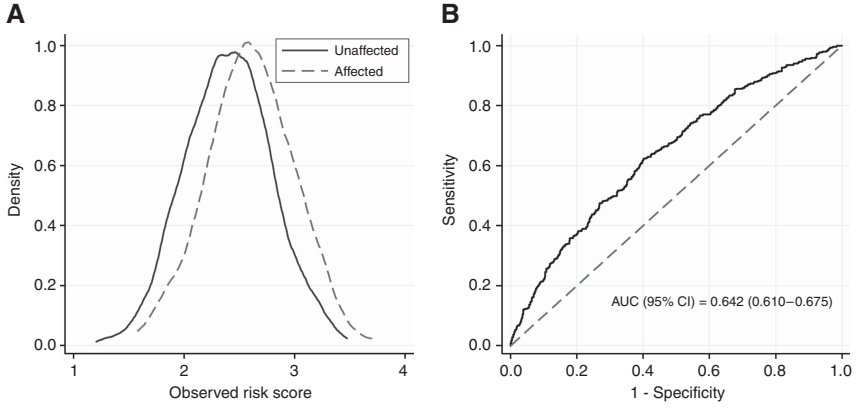

Figure 3. (**A**) Density plots of the ORS based on 22 SNPs for women with FH of BC (n = 432) and controls (n = 715). (**B**) ROC curve for the ability of the ORS based on 22 SNPs to discriminate between cases with FH and controls. The *x*-axis is 1-specificity (false-positive rate) and the *y*-axis is the sensitivity (true-positive rate). The dashed line represents an AUC of 0.50, indicating prediction no better than chance alone.

was similar to our model by allowing for residual correlations between family members using a random baseline risk parameter. Schaid *et al* found that RRs estimated under the pedigree RL were consistent with ORs estimated by large case–control studies, agreeing with our findings.

After accounting for ascertainment and the residual polygenic variance, the RR estimates for the known common BC susceptibility alleles were similar to those obtained from population-based case–control studies (Cox *et al*, 2007; Easton *et al*, 2007; Stacey *et al*, 2007, 2010; Ahmed *et al*, 2009; Thomas *et al*, 2009; Antoniou

et al, 2010; Turnbull *et al*, 2010; Fletcher *et al*, 2011; Milne *et al*, 2011; Ghoussaini *et al*, 2012; Hein *et al*, 2012). This observation suggests that the polygenic model of inheritance provides a good fit to the observed familial aggregation of BC. First, it implies that the residual genetic susceptibility to BC is unlikely to be due genes conferring large contributions to the familial risk of the disease of magnitude similar to that of *BRCA1* or *BRCA2* mutations. Instead, the residual genetic variability is likely to be due to genetic effects that have small contributions to the BC familial risk. That is, either common alleles conferring low risks or rare variants conferring moderate risks. Second, our findings suggest a general model of genetic susceptibility where the joint effects of the common alleles studied in the present study and other, as yet unidentified, BC susceptibility variants are multiplicative. Therefore, we can infer that interactions between the studied common alleles and other residual genetic effects are unlikely.

The pedigree RL was adapted to estimate parent-specific genetic effects for each common allele. This was achieved by separately estimating the risk for a maternally and paternally inherited risk allele. Although other methods have been suggested for evaluating POE, those involve direct genotyping of parents and offspring, and they may not make full use of multigenerational pedigree data or do not adjust adequately for ascertainment (Haghighi and Hodge, 2002; Belonogova *et al*, 2009; Kong *et al*, 2009; Feng *et al*, 2011; He *et al*, 2011; Li *et al*, 2011).

Our analyses suggested no significant differences between estimated HRs for maternally and paternally inherited alleles for any of the 24 SNPs. The *LSP1* SNP rs3817198 had previously been shown to display POE with BC risk where the paternally inherited allele was associated with increased BC risk (OR = 1.17, 95% CI: 1.05–1.30, $P = 0.0038$) (Kong *et al*, 2009). They also found a decreased BC risk if the risk allele was maternally inherited, but this was not significant (OR = 0.91, 95% CI: 0.81–1.02, $P = 0.11$). The magnitude and direction of our estimates for this SNP are comparable to those reported by Kong *et al* (paternal HR = 1.12, 95% CI: 0.99–1.27, $P = 0.081$; maternal HR = 0.94, 95% CI: 0.84–1.06, $P = 0.33$). Our analyses did not detect a significant difference between the maternal and paternal effect ($P = 0.11$). This is possibly because of the much greater sample size employed by Kong *et al* – 34 909 controls and 1803 BC cases, all of whom were genotyped or had imputed genotype data available. Our analyses included 5251 unaffected individuals and 1463 BC cases. It is worth noting that the paternal HR for *LSP1* SNP rs3817198 was significant under a one-sided test for the hypothesis that the paternal HR > 1 ($P = 0.04$). We meta-analysed our *LSP1* SNP RR estimates with those reported by Kong *et al* (Supplementary Table 2). The meta-analysis yielded a maternal RR = 0.93 (95% CI: 0.85–1.01, $P = 0.066$) and a paternal RR = 1.15 (95% CI: 1.06–1.24, $P = 7.8 \times 10^{-4}$). These analyses suggest no association with the maternally inherited C allele but provides stronger evidence of association with the paternally inherited C allele. Although no significant differences were observed between the estimates for the paternally and maternally inherited alleles at other loci, we observed associations for several SNPs with either the maternally or paternally inherited alleles. The current approach for evaluating POE could, potentially, be useful in the fine mapping efforts of these loci in determining causal variants.

Recent studies have estimated the ROC AUC to investigate the effect of SNPs on discriminating between affected and unaffected women. Wacholder *et al* (2010) used a modified Gail model to demonstrate an increase in AUC from 0.580 to 0.618 when the effects of the (at the time) 10 known genetic variants associated with BC risk were incorporated into the model. Sawyer *et al* (2012) have described the largest AUC (0.654, 95% CI: 0.628–0.680) based purely on genetic factors. Their analyses included 22 genetic variants in women with FH of BC in the absence of a known *BRCA1* or *BRCA2* mutation. We describe a similar AUC when considering the ORS as the sole risk predictor for individuals genotyped for all 22 SNPs. This is consistent with the fact that women with FH of BC are expected to have a higher polygenic load due to familial aggregation of the disease. This suggests that a high polygenic score in combination with a FH of the disease could jointly provide a way to identify those who may be at higher risk of developing the disease, rather than SNPs alone.

In summary, we have presented a novel analytical framework for evaluating associations between common genetic variants and disease risk that harnesses the power and efficiency of family data. Although the methods have been presented in the context of BC susceptibility, the general principles are applicable to other cancers and other complex diseases that have a heritable component. We applied these techniques to data on common susceptibility alleles, although, in principle, the methods could be applied to analyse rare variants conferring moderate cancer risks. We have further demonstrated that combined SNP profiles discriminate more effectively BC-affected status in individuals with FH of the disease compared with the general population, taking us closer to the goal of incorporating SNP profiling into clinical practice.

ACKNOWLEDGEMENTS

This research was funded by Cancer Australia's Priority-driven Collaborative Cancer Research Scheme no. 566791. DRB is funded by a Cancer Research UK studentship (C12292/A11168). ACA is a Cancer Research UK Senior Cancer Research Fellow. GC-T is an NHMRC Senior Principal Research Fellow. We thank Heather Thorne, Eveline Niedermayr, all the kConFab research nurses and staff, the heads and staff of the Family Cancer Clinics, and the Clinical Follow Up Study (funded 2001–2009 by NHMRC and currently by the National Breast Cancer Foundation and Cancer Australia no. 628333) for their contributions to this resource, and the many families who contribute to kConFab. kConFab is supported by grants from the National Breast Cancer Foundation, the National Health and Medical Research Council (NHMRC) and by the Queensland Cancer Fund, the Cancer Councils of New South Wales, Victoria, Tasmania and South Australia, and the Cancer Foundation of Western Australia. We also thank the Australian Ovarian Cancer Study Management Group (D. Bowtell, P.M. Webb, A. deFazio, D. Gertig, A. Green, P. Parsons, N. Hayward and D. Whiteman) for the AOCS data.

REFERENCES

Ahmed S, Thomas G, Ghoussaini M, Healey CS, Humphreys MK, Platte R, Morrison J, Maranian M, Pooley KA, Luben R, Eccles D, Evans DG, Fletcher O, Johnson N, dos Santos Silva I, Peto J, Strat-ton MR, Rahman N, Jacobs K, Prentice R, Anderson GL, Rajkovic A, Curb JD, Ziegler RG, Berg CD, Buys SS, McCarty CA, Feigelson HS, Calle EE, Thun MJ, Diver WR, Bojesen S, Nordestgaard BG, Flyger H, Dork T, Schurmann P, Hillemanns P, Karstens JH, Bogdanova NV, Antonenkova NN, Za-lutsky IV, Bermisheva M, Fedorova S, Khusnutdinova E, SEARCH, Kang D, Yoo K, Noh DY, Ahn S, Devilee P, van Asperen CJ, Tollenaar RAEM, Seynaeve C, Garcia-Closas M, Lissowska J, Brinton L, Peplonska B, Nevanlinna H, Heikkinen T, Aittomaki K, Blomqvist C, Hopper JL, Southey MC, Smith L, Spurdle AB, Schmidt MK, Broeks A, van Hien RR, Cornelissen S, Milne RL, Ribas G, Gonzalez-Neira A, Benitez J, Schmutzler RK, Burwinkel B, Bartram CR, Meindl A, Brauch H, Justenhoven C, Hamann U. The GENICA Consortium, Chang-Claude J, Hein R, Wang-Gohrke S, Lindblom A, Margolin S, Man-nermaa A, Kosma V, Kataja V, Olson JE, Wang X, Fredericksen Z, Giles GG, Severi G, Baglietto L, English DR, Hankinson SE, Cox DG, Kraft P, Vatten LJ, Hveem K, Kumle M, Sigurdson A, Doody M, Bhatti P, Alexander BH, Hooning MJ, van den Ouweland AMW, Oldenburg RA, Schutte M, Hall P, Czene K, Liu J, Li Y, Cox A, Elliott G, Brock I, Reed MWR, Shen C, Yu J, Hsu G, Chen S, Anton-Culver H, Ziogas A, Andrulis IL, Knight JA, kConFab, Australian

Ovarian Cancer Study Group, Beesley J, Goode EL, Couch F, Chenevix-Trench G, Hoover RN, Ponder BAJ, Hunter DJ, Pharoah PDP, Dunning AM, Chanock SJ, Easton DF (2009) Newly discovered breast cancer susceptibility loci on 3p24 and 17q23.2. *Nat Genet* **41**(5): 585–590.

Antoniou AC, Easton DF (2003) Polygenic inheritance of breast cancer: implications for design of association studies. *Genet Epidemiol* **25**(3): 190–202.

Antoniou AC, Pharoah PDP, McMullan G, Day NE, Ponder BAJ, Easton D (2001) Evidence for further breast cancer susceptibility genes in addition to BRCA1 and BRCA2 in a population-based study. *Genet Epidemiol* **21**(1): 1–18.

Antoniou AC, Pharoah PDP, McMullan G, Day NE, Stratton MR, Peto J, Ponder BJ, Easton DF (2002) A comprehensive model for familial breast cancer incorporating BRCA1, BRCA2 and other genes. *Br J Cancer* **86**(1): 76–83.

Akaike H (1974) A new look at the statistical model identification. *IEEE Trans Auto Control* **19**(6): 716–723.

Antoniou AC, Wang X, Fredericksen ZS, McGuffog L, Tarrell R, Sinilnikova M, Healey S, Morrison J, Kartsonaki C, Lesnick T, Ghoussaini M, Barrowdale D, EMBRACE, Peock S, Cook M, Oliver C, Frost D, Eccles D, Evans DG, Eeles R, Izatt L, Chu C, Douglas F, Paterson J, Stoppa-Lyonnet D, Houdayer C, Mazoyer S, Giraud S, Lasset C, Remenieras A, Caron O, Hardouin A, Berthet P, GEMO Study Collaborators, Hogervorst FBL, Rookus MA, Jager A, van den Ouweland A, Hoogerbrugge N, van der Luijt RB, Meijers-Heijboer H, Gmez Garca EB, HEBON, Devilee P, Vreeswijk MPG, Lubinski J, Jakubowska A, Gronwald J, Huzarski T, Byrski T, Grski B, Cybulski C, Spurdle AB, Holland H, kConFab, Goldgar DE, John EM, Hopper JL, Southey M, Buys SS, Daly MB, Terry MB, Schmutzler RK, Wappenschmidt B, Engel C, Meindl A, Preisler-Adams S, Arnold N, Niederacher D, Sutter C, Domchek SM, Nathanson KL, Rebbeck T, Blum JL, Piedmonte M, Rodriguez GC, Wakeley K, Boggess JF, Basil J, Blank SV, Friedman E, Kaufman B, Laitman Y, Milgrom R, Andrulis IL, Glendon G, Ozcelik H, Kirchhoff T, Vijai J, Gaudet MM, Altshuler D, Guiducci C, SWE-BRCA, Loman N, Harbst K, Rantala J, Ehrencrona H, Gerdes A, Thomassen M, Sunde L, Peterlongo P, Manoukian S, Bonanni B, Viel A, Radice P, Caldes T, de la Hoya M, Singer CF, Fink-Retter A, Greene MH, Mai PL, Loud JT, Guidugli L, Lindor NM, Hansen TVO, Nielsen FC, Blanco I, Lazaro C, Garber J, Ramus SJ, Gayther SA, Phelan C, Narod S, Szabo CI, MOD SQUAD, Benitez J, Osorio A, Nevanlinna H, Heikkinen T, Caligo MA, Beattie MS, Hamann U, Godwin AK, Montagna M, Casella C, Neuhausen SL, Karlan BY, Tung N, Toland AE, Weitzel J, Olopade O, Simard J, Soucy P, Rubinstein WS, Arason A, Rennert G, Martin NG, Montgomery GW, Chang-Claude J, Flesch-Janys D, Brauch H, GENICA, Severi G, Baglietto L, Cox A, Cross SS, Miron P, Gerty SM, Tapper W, Yannoukakos D, Fountzilas G, Fasching PA, Beckmann MW, dos Santos Silva I, Peto J, Lambrechts D, Paridaens R, Riidigger T, Forsti A, Winqvist R, Pylkas K, Diasio RB, Lee AM, Eckel-Passow J, Vachon C, Blows F, Driver K, Dunning A, Pharoah PPD, Offit K, Pankratz VS, Hakonarson H, Chenevix-Trench G, Easton DF, Couch FJ (2010) A locus on 19p13 modifies risk of breast cancer in BRCA1 mutation carriers and is associated with hormone receptor-negative breast cancer in the general population. *Nat Genet* **42**(10): 885–892.

Australian Ovarian Cancer Study (AOCS) (2012) Aocs. URL http://www.aocstudy.org.

Belonogova NM, Axenovich TI, Aulchenko YS (2009) A powerful genome-wide feasible approach to detect parent-of-origin effects in studies of quantitative traits. *Eur J Hum Genet* **18**(3): 379–384.

Berteaux N, Aptel N, Cathala G, Genton C, Coll J, Daccache A, Spruyt N, Hondermarck H, Dugi-mont T, Curgy JJ, Forne T, Adriaenssens E (2008) A novel H19 antisense RNA overexpressed in breast cancer contributes to paternal IGF2 expression. *Mol Cell Biol* **28**(22): 6731.

Carayol J, Bonaïti-Pellié C (2004) Estimating penetrance from family data using a retrospective likelihood when ascertainment depends on genotype and age of onset. *Genet Epidemiol* **27**(2): 109–117.

Couch FJ, Wang X, McGuffog L, Lee A, Olswold C, Kuchenbaecker KB, Soucy P, Fredericksen Z, Bar-rowdale D, Dennis J, Gaudet MM, Dicks E, Kosel M, Healey S, Sinilnikova OM, Lee A, Bacot F, Vincent D, Hogervorst FB, Peock S, Stoppa-Lyonnet D, Jakubowska A, kConFab Investigators, Radice P, Schmut-zler RK, SWE-BRCA, Domchek SM, Piedmonte M, Singer CF, Friedman E, Thomassen M, Ontario Cancer Genetics Network, Hansen TV, Neuhausen SL, Szabo CI, Blanco I, Greene MH, Karlan BY, Garber J, Phelan CM, Weitzel JN, Montagna M, Olah E, Andrulis IL, Godwin AK, Yannoukakos D, Goldgar DE, Caldes T, Nevanlinna H, Osorio A, Terry MB, Daly MB, van Rensburg EJ, Hamann U, Ramus SJ, Toland AE, Caligo MA, Olopade OI, Tung N, Claes K, Beattie MS, Southey MC, Imyanitov EN, Tis-chkowitz M, Janavicius R, John EM, Kwong A, Diez O, Balmana J, Barkardottir RB, Arun BK, Rennert G, Teo SH, Ganz PA, Campbell I, van der Hout AH, van Deurzen CH, Seynaeve C, Gomez Garcia EB, van Leeuwen FE, Meijers-Heijboer HE, Gille JJ, Ausems MG, Blok MJ, Ligtenberg MJ, Rookus MA, Devilee P, Verhoef S, van Os TA, Wijnen JT, HEBON, EMBRACE, Frost D, Ellis S, Fineberg E, Platte R, Evans DG, Izatt L, Eeles RA, Adlard J, Eccles DM, Cook J, Brewer C, Douglas F, Hodgson S, Morrison PJ, Side LE, Donaldson A, Houghton C, Rogers MT, Dorkins H, Eason J, Gregory H, McCann E, Murray A, Calender A, Hardouin A, Berthet P, Delnatte C, Nogues C, Lasset C, Houdayer C, Leroux D, Rouleau E, Prieur F, Damiola F, Sobol H, Coupier I, Venat-Bouvet L, Castera L, Gauthier-Villars M, Leone M, Pujol P, Ma-zoyer S, Bignon YJ, GEMO Study Collaborators, Zlowocka-Perlowska E, Gronwald J, Lubinski J, Durda K, Jaworska K, Huzarski T, Spurdle AB, Viel A, Peissel B, Bonanni B, Melloni G, Ottini L, Papi L, Varesco L, Tibiletti MG, Peterlongo P, Volorio S, Manoukian S, Pensotti V, Arnold N, Engel C, Deissler H, Gadz-icki D, Gehrig A, Kast K, Rhiem K, Meindl A, Niederacher D, Ditsch N, Plendl H, Preisler-Adams S, Engert S, Sutter C, Varon-Mateeva R, Wappenschmidt B, Weber BH, Arver B, Stenmark-Askmalm M, Loman N, Rosenquist R, Einbeigi Z, Nathanson KL, Rebbeck TR, Blank SV, Cohn DE, Rodriguez GC, Small L, Friedlander M, Bae-Jump VL, Fink-Retter A, Rappaport C, Gschwantler-Kaulich D, Pfeiler G, Tea MK, Lindor NM, Kaufman B, Shimon Paluch S, Laitman Y, Skytte AB, Gerdes AM, Pedersen IS, Moeller ST, Kruse TA, Jensen UB, Vijai J, Sarrel K, Robson M, Kauff N, Mulligan AM, Glendon G, Ozcelik H, Ejlertsen B, Nielsen FC, Jonson L, Andersen MK, Ding YC, Steele L, Foretova L, Teule A, Lazaro C, Brunet J, Pujana MA, Mai PL, Loud JT, Walsh C, Lester J, Orsulic S, Narod SA, Herzog J, Sand SR, Tognazzo S, Agata S, Vaszko T, Weaver J, Stavropoulou AV, Buys SS, Romero A, de la Hoya M, Aittomaki K, Muranen TA, Duran M, Chung WK, Lasa A, Dorfling CM, Miron A, BCFR, Benitez J, Senter L, Huo D, Chan SB, Sokolenko AP, Chiquette J, Tihomirova L, Friebel TM, Agnarsson BA, Lu KH, Lejbkowicz F, James PA, Hall P, Dunning AM, Tessier D, Cunningham J, Slager SL, Wang C, Hart S, Stevens K, Simard J, Pastinen T, Pankratz VS, Offit K, Easton DF, Chenevix-Trench G, Antoniou AC, CIMBA (2013) Genome-wide association study in brca1 mutation carriers identifies novel loci associated with breast and ovarian cancer risk. *PLoS Genet* **9**(3): e1003212.

Cox A, Dunning AM, Garcia-Closas M, Balasubramanian S, Reed MWR, Pooley KA, Scollen S, Baynes C, Ponder BAJ, Chanock S, Lissowska J, Brinton L, Peplonska B, Southey MC, Hopper JL, McCredie MRE, Giles GG, Fletcher O, Johnson N, dos Santos Silva I, Gibson L, Bojesen SE, Nordestgaard BG, Axelsson CK, Torres D, Hamann U, Justenhoven C, Brauch H, Chang-Claude J, Kropp S, Risch A, Wang-Gohrke S, Schurmann P, Bogdanova N, Dork T, Fagerholm R, Aaltonen K, Blomqvist C, Nevanlinna H, Seal S, Renwick A, Stratton MR, Rahman N, Sangrajrang S, Hughes D, Odefrey F, Brennan P, Spurdle AB, Chenevix-Trench G, The Kathleen Cunningham Foundation Con-sortium for Research into Familial Breast Cancer, Beesley J, Mannermaa A, Hartikainen J, Kataja V, Kosma V, Couch FJ, Olson JE, Goode EL, Broeks A, Schmidt MK, Hogervorst FBL, Vant Veer LJ, Kang D, Yoo K, Noh D, Ahn S, Wedren S, Hall P, Low Y, Liu J, Milne RL, Ribas G, Gonzalez-Neira A, Benitez J, Sigurdson AJ, Stredrick DL, Alexander BH, Struewing JP, Pharoah PDP, Easton DF, on behalf of the Breast Cancer Association Consortium (2007) A common coding variant in CASP8 is associated with breast cancer risk. *Nat Genet* **39**(3): 352–358.

Easton DF, Pooley KA, Dunning AM, Pharoah PDP, Thompson D, Ballinger DG, Struewing JP, Morrison J, Field H, Luben R, Wareham N, Ahmed S, Healey CS, Bowman R, the SEARCH collaborators, Meyer KB, Haiman CA, Kolonel LK, Henderson BE, Le Marchand L, Brennan P, Sangrajrang S, Gaborieau V, Odefrey F, Shen C, Wu P, Wang H, Eccles D, Evans DG, Peto J, Fletcher O, Johnson N, Seal S, Stratton MR, Rahman N, Chenevix-Trench G, Bojesen SE, Nordestgaard BG, Axelsson CK, Garcia-Closas M, Brinton L, Chanock S, Lissowska J, Peplonska B, Nevanlinna H, Fagerholm R, Eerola H, Kang D, Yoo K, Noh D, Ahn S, Hunter DJ, Han-kinson SE, Cox DG, Hall P, Wedren S, Liu J, Low Y, Bogdanova N, Schurmann P, Dork T, Tollenaar RAEM, Jacobi CE, Devilee P, Klijn JGM, Sigurdson AJ, Doody MM, Alexander BH, Zhang J, Cox A, Brock IW, MacPherson G, Reed MWR, Couch FJ, Goode EL, Olson JE, Meijers-Heijboer H, van den Ouweland A, Uitterlinden A, Rivadeneira F, Milne

RL, Ribas G, Gonzalez-Neira A, Benitez J, Hopper JL, McCredie M, Southey M, Giles GG, Schroen C, Justen-hoven C, Brauch H, Hamann U, Ko Y, Spurdle AB, Beesley J, Chen X, kConFab, AOCS Management Group, Mannermaa A, Kosma V, Kataja V, Hartikainen J, Day NE, Cox DR, Ponder BAJ (2007) Genome-wide association study identifies novel breast cancer susceptibility loci. *Nature* **447**(7148): 1087–1093.

Falls JG, Pulford DJ, Wylie AA, Jirtle RL (1999) Genomic imprinting: implications for human disease. *Am J Pathol* **154**(3): 635.

Feng R, Wu Y, Jang GH, Ordovas JM, Arnett D (2011) A powerful test of parent-of-origin effects for quantitative traits using haplotypes. *PloS One* **6**(12): e28909.

Fernando RL, Stricker C, Elston RC (1994) The finite polygenic mixed model: an alternative formulation for the mixed model of inheritance. *Theor Appl Genet* **88**(5): 573–580.

Fletcher O, Johnson N, Orr N, Hosking FJ, Gibson LJ, Walker K, Zelenika D, Gut I, Heath S, Palles C, Coupland B, Broderick P, Schoemaker M, Jones M, Williamson J, Chilcott-Burns S, Tom-czyk K, Simpson G, Jacobs KB, Chanock SJ, Hunter DJ, Tomlinson IP, Swerdlow A, Ashworth A, Ross G, dos Santos Silva I, Lathrop M, Houlston RS, Peto J (2011) Novel breast cancer susceptibility locus at 9q31.2: results of a genome-wide association study. *J Natl Cancer Inst* **103**(5): 425–435.

Galvan A, Ioannidis J, Dragani TA (2010) Beyond genome-wide association studies: genetic heterogeneity and individual predisposition to cancer. *Trends Genet* **26**(3): 132–141.

Garcia-Closas M, Couch FJ, Lindstrom S, Michailidou K, Schmidt MK, Brook MN, Orr N, Rhie SK, Riboli E, Feigelson HS, Le Marchand L, Buring JE, Eccles D, Miron P, Fasching PA, Brauch H, Chang-Claude J, Carpenter J, Godwin AK, Nevanlinna H, Giles GG, Cox A, Hopper JL, Bolla MK, Wang Q, Dennis J, Dicks E, Howat WJ, Schoof N, Bojesen SE, Lambrechts D, Broeks A, Andrulis IL, Guenel P, Burwinkel B, Sawyer EJ, Hollestelle A, Fletcher O, Winqvist R, Brenner H, Mannermaa A, Hamann U, Meindl A, Lindblom A, Zheng W, Devillee P, Goldberg MS, Lubinski J, Kristensen V, Swerdlow A, Anton-Culver H, Dork T, Muir K, Matsuo K, Wu AH, Radice P, Teo SH, Shu XO, Blot W, Kang D, Hartman M, Sangrajrang S, Shen CY, Southey MC, Park DJ, Hammet F, Stone J, Veer LJ, Rutgers EJ, Lophatananon A, Stewart-Brown S, Siriwanarangsan P, Peto J, Schrauder MG, Ekici AB, Beckmann MW, Dos Santos Silva I, Johnson N, Warren H, Tomlinson I, Kerin MJ, Miller N, Marme F, Schneeweiss A, Sohn C, Truong T, Laurent-Puig P, Kerbrat P, Nordestgaard BG, Nielsen SF, Flyger H, Milne RL, Perez JI, Menendez P, Muller H, Arndt V, Stegmaier C, Lichtner P, Lochmann M, Justenhoven C, Ko YD, Gene ENvironmental Interaction and breast CAncer (GENICA) Network, Muranen TA, Aittomaki K, Blomqvist C, Greco D, Heikkinen T, Ito H, Iwata H, Yatabe Y, Antonenkova NN, Margolin S, Kataja V, Kosma VM, Hartikainen JM, Balleine R, kConFab Investigators, Tseng CC, Berg DV, Stram DO, Neven P, Dieudonne AS, Leunen K, Rudolph A, Nickels S, Flesch-Janys D, Peterlongo P, Peissel B, Bernard L, Olson JE, Wang X, Stevens K, Severi G, Baglietto L, McLean C, Coetzee GA, Feng Y, Henderson BE, Schumacher F, Bogdanova NV, Labreche F, Dumont M, Yip CH, Taib NA, Cheng CY, Shrubsole M, Long J, Pylkas K, Jukkola-Vuorinen A, Kauppila S, Knight JA, Glendon G, Mulligan AM, Tollenaar RA, Seynaeve CM, Kriege M, Hooning MJ, van den Ouweland AM, van Deurzen CH, Lu W, Gao YT, Cai H, Balasubramanian SP, Cross SS, Reed MW, Signorello L, Cai Q, Shah M, Miao H, Chan CW, Chia KS, Jakubowska A, Jaworska K, Durda K, Hsiung CN, Wu PE, Yu JC, Ashworth A, Jones M, Tessier DC, Gonzalez-Neira A, Pita G, Alonso MR, Vincent D, Bacot F, Ambrosone CB, Bandera EV, John EM, Chen GK, Hu JJ, Rodriguez-Gil JL, Bernstein L, Press MF, Ziegler RG, Millikan RM, Deming-Halverson SL, Nyante S, Ingles SA, Waisfisz Q, Tsimiklis H, Makalic E, Schmidt D, Bui M, Gibson L, Muller-Myhsok B, Schmutzler RK, Hein R, Dahmen N, Beckmann L, Aaltonen K, Czene K, Irwanto A, Liu J, Turnbull C, Familial Breast Cancer Study (FBCS), Rahman N, Meijers-Heijboer H, Uitterlinden AG, Rivadeneira F, Australian Breast Cancer Tissue Bank (ABCTB) Investigators, Olswold C, Slager S, Pilarski R, Ademuyiwa F, Konstantopoulou I, Martin NG, Montgomery GW, Slamon DJ, Rauh C, Lux MP, Jud SM, Bruning T, Weaver J, Sharma P, Pathak H, Tapper W, Gerty S, Durcan L, Trichopoulos D, Tumino R, Peeters PH, Kaaks R, Campa D, Canzian F, Weider-pass E, Johansson M, Khaw KT, Travis R, Clavel-Chapelon F, Kolonel LN, Chen C, Beck A, Hankinson SE, Berg CD, Hoover RN, Lissowska J, Figueroa JD, Chasman DI, Gaudet MM, Diver WR, Willett WC, Hunter DJ, Simard J, Benitez J, Dunning AM, Sherman ME, Chenevix-Trench G, Chanock SJ, Hall P, Pharoah PD, Vachon C, Easton DF, Haiman CA, Kraft P (2013) Genome-wide association studies identify four er negative-specific breast cancer risk loci. *Nat Genet* **45**(4): 392–398.

Gaudet MM, Kuchenbaecker KB, Vijai J, Klein RJ, Kirchhoff T, McGuffog L, Barrowdale D, Dunning AM, Lee A, Dennis J, Healey S, Dicks E, Soucy P, Sinilnikova OM, Pankratz VS, Wang X, Eldridge RC, Tessier DC, Vincent D, Bacot F, Hogervorst FB, Peock S, Stoppa-Lyonnet D, kConFab Investigators, Peterlongo P, Schmutzler RK, Nathanson KL, Piedmonte M, Singer CF, Thomassen M, Ontario Cancer Genetics Network, Hansen TV, Neuhausen SL, Blanco I, Greene MH, Garber J, Weitzel JN, Andrulis IL, Goldgar DE, D'Andrea E, Caldes T, Nevanlinna H, Osorio A, van Rensburg EJ, Arason A, Rennert G, van den Ouweland AM, van der Hout AH, Kets CM, Aalfs CM, Wijnen JT, Ausems MG, HEBON, EMBRACE, Frost D, Ellis S, Fineberg E, Platte R, Evans DG, Jacobs C, Adlard J, Tischkowitz M, Por-teous ME, Damiola F, GEMO Study Collaborators, Golmard L, Barjhoux L, Longy M, Belotti M, Ferrer SF, Mazoyer S, Spurdle AB, Manoukian S, Barile M, Genuardi M, Arnold N, Meindl A, Sutter C, Wappenschmidt B, Domchek SM, Pfeiler G, Friedman E, Jensen UB, Robson M, Shah S, Lazaro C, Mai PL, Benitez J, Southey MC, Schmidt MK, Fasching PA, Peto J, Humphreys MK, Wang Q, Michailidou K, Sawyer EJ, Burwinkel B, Guenel P, Bojesen SE, Milne RL, Brenner H, Lochmann M, GENICA Network, Aittomaki K, Dork T, Margolin S, Mannermaa A, Lambrechts D, Chang-Claude J, Radice P, Giles GG, Haiman CA, Winqvist R, Devillee P, Garcia-Closas M, Schoof N, Hooning MJ, Cox A, Pharoah PD, Jakubowska A, Orr N, Gonzalez-Neira A, Pita G, Alonso MR, Hall P, Couch FJ, Simard J, Altshuler D, Easton DF, Chenevix-Trench G, Antoniou AC, Offit K (2013) Identification of a brca2-specific modifier locus at 6p24 related to breast cancer risk. *PLoS Genet* **9**(3): e1003173.

Ghoussaini M, Fletcher O, Michailidou K, Turnbull C, Schmidt MK, Dicks E, Dennis J, Wang Q, Humphreys MK, Luccarini C, Baynes C, Conroy D, Maranian M, Ahmed S, Driver K, Johnson N, Orr N, dos Santos Silva I, Waisfisz Q, Meijers-Heijboer H, Uitterlinden AG, Rivadeneira F, Netherlands Collaborative Group on Hereditary Breast, Ovarian Cancer (HEBON), Hall P, Czene K, Irwanto A, Liu J, Nevanlinna H, Aittomaki K, Blomqvist C, Meindl A, Schmutzler RK, Miiller-Myhsok B, Lichtner P, Chang-Claude J, Hein R, Nickels S, Flesch-Janys D, Tsimiklis H, Makalic E, Schmidt D, Bui M, Hopper JL, Apicella C, Park DJ, Southey M, Hunter DJ, Chanock SJ, Broeks A, Verhoef S, Hogervorst FBL, Fasching PA, Lux MP, Beckmann MW, Ekici AB, Sawyer E, Tomlinson I, Kerin M, Marme F, Schneeweiss A, Sohn C, Burwinkel B, Guénel P, Truong T, Cordina-Duverger E, Menegaux F, Bojesen SE, Nordestgaard BG, Nielsen SF, Flyger H, Milne RL, Alonso MR, Gonzalez-Neira A, Benitez J, Anton-Culver H, Ziogas A, Bernstein L, Dur CC, Brenner H, Miller H, Arndt V, Stegmaier C, Familial Breast Cancer Study (FBCS), Justenhoven C, Brauch H, Bruining T, The Gene Environment Interaction of Breast Cancer in Germany (GENICA) Network, Wang-Gohrke S, Eilber U, Drk T, Schuirmann P, Bremer M, Hillemanns P, Bogdanova NV, Antonenkova NN, Rogov YI, Karstens JH, Bermisheva M, Prokofieva D, Khusnutdinova E, Lindblom A, Margolin S, Mannermaa A, Kataja V, Kosma V, Hartikainen JM, Lambrechts D, Yesilyurt BT, Floris G, Leunen K, Manoukian S, Bonanni B, Fortuzzi S, Peterlongo P, Couch FJ, Wang X, Stevens K, Lee A, Giles GG, Baglietto L, Severi G, McLean C, Alnas GG, Kristensen V, Borrensen-Dale A, John EM, Miron A, Winqvist R, Pylkias K, Jukkola-Vuorinen A, Kauppila S, Andrulis IL, Glendon G, Mulligan AM, Devilee P, van Asperen CJ, Tollenaar RAEM, Seynaeve C, Figueroa JD, Garcia-Closas M, Brinton L, Lissowska J, Hooning MJ, Hollestelle A, Oldenburg RA, van den Ouweland AMW, Cox A, Reed MWR, Shah M, Jakubowska A, Lubinski J, Jaworska K, Durda K, Jones M, Schoemaker M, Ashworth A, Swerdlow A, Beesley J, Chen X, kConFab Investi-gators, Australian Ovarian Cancer Study Group, Muir KR, Lophatananon A, Rattanamongkongul S, Chaiwerawattana A, Kang D, Yoo K, Noh D, Shen C, Yu J, Wu P, Hsiung C, Perkins A, Swann R, Velentzis L, Eccles DM, Tapper WJ, Gerty SM, Graham NJ, Ponder BAJ, Chenevix-Trench G, Pharoah PDP, Lathrop M, Dunning AM, Rahman N, Peto J, Easton DF (2012) Genome-wide association analysis identifies three new breast cancer susceptibility loci. *Nat Genet* **44**(3): 312–318.

Guilmatre A, Sharp AJ (2012) Parent of origin effects. *Clin Genet* **81**(3): 201–209.

Haghighi F, Hodge SE (2002) Likelihood formulation of parent-of-origin effects on segregation analysis, including ascertainment. *Am J Hum Genet* **70**(1): 142–156.

He F, Zhou JY, Hu YQ, Sun F, Yang J, Lin S, Fung WK (2011) Detection of parent-of-origin effects for quantitative traits in complete and

incomplete nuclear families with multiple children. *Am J Epidemiol* **174**(2): 226–233.

Hein R, Maranian M, Hopper JL, Kapuscinski MK, Southey MC, Park DJ, Schmidt MK, Broeks A, Hogervorst FB, Bueno-de Mesquit HB, Muir KR, Lophatananon A, Rattanamongkongul S, Puttawibul P, Fasching PA, Hein A, Ekici AB, Beckmann MW, Fletcher O, Johnson N, dos Santos Silva I, Peto J, Sawyer E, Tomlinson I, Kerin M, Miller N, Marmee F, Schneeweiss A, Sohn C, Burwinkel B, Guénel P, Cordina-Duverger E, Menegaux F, Truong T, Bojesen SE, Nordestgaard BG, Flyger H, Milne RL, Perez JI, Zamora MP, Benitez J, Anton-Culver H, Ziogas A, Bernstein L, Clarke CA, Brenner H, Muiller H, Arndt V, Stegmaier C, Rahman N, Seal S, Turnbull C, Renwick A, Meindl A, Schott S, Bartram CR, Schmut-zler RK, Brauch H, Hamann U, Ko YD, The GENICA Network, Wang-Gohrke S, Diork T, Schuirmann P, Karstens JH, Hillemanns P, Nevanlinna H, Heikkinen T, Aittomiaki K, Blomqvist C, Bogdanova NV, Zalutsky IV, Antonenkova NN, Bermisheva M, Prokovieva D, Farahtdinova A, Khusnutdinova E, Lind-blom A, Margolin S, Mannermaa A, Kataja V, Kosma VM, Hartikainen J, Chen X, Beesley J, KConFab Investigators, AOCS Group, Lambrechts D, Zhao H, Neven P, Wildiers H, Nickels S, Flesch-Janys D, Radice P, Peterlongo P, Manoukian S, Barile M, Couch FJ, Olson JE, Wang X, Fredericksen Z, Giles GG, Baglietto L, McLean CA, Severi G, Offit K, Robson M, Gaudet MM, Vijai J, Alnaes GG, Kristensen V, Borresen-Dale AL, John EM, Miron A, Winqvist R, Pylkas K, Jukkola-Vuorinen A, Grip M, Andrulis IL, Knight JA, Glendon G, Mulligan AM, Figueroa JD, Garcia-Closas M, Lissowska J, Sherman ME, Hooning M, Martens JW, Seynaeve C, Collee M, Hall P, Humpreys K, Czene K, Liu J, Cox A, Brock IW, Cross SS, Reed MW, Ahmed S, Ghoussaini M, Pharoah PD, Kang D, Yoo KY, Noh DY, Jakubowska A, Jaworska K, Durda K, Zlowocka E, Sangrajrang S, Gaborieau V, Brennan P, McKay J, Shen CY, Yu JC, Hsu HM, Hou MF, Orr N, Schoemaker M, Ashworth A, Swerdlow A, Trentham-Dietz A, Newcomb PA, Titus L, Egan KM, Chenevix-Trench G, Antoniou AC, Humphreys MK, Morrison J, Chang-Claude J, Easton DF, Dunning AM (2012) Comparison of 6q25 breast cancer hits from asian and european genome wide association studies in the breast cancer association consortium (bcac). *PloS One* **7**(8): e42380.

Huber PJ (1967) The behavior of maximum likelihood estimates under nonstandard conditions. In Lucien M le Cam, Jerzy Neyman (eds) *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*. University of California Press, Berkeley, **Vol. 1**: pp 221–233.

International Agency for Research on Cancer (IARC) (2010) Cancer in five continents. URL http://ci5.iarc.fr/CI5i-ix/ci5i-ix.htm.

Kathleen Cuningham Foundation Consortium for research into Familial Breast Cancer (kConFab) (2012) kconfab. URL www.kconfab.org/Index.shtml.

Kong A, Steinthorsdottir V, Masson G, Thorleifsson G, Sulem P, Besenbacher S, Jonasdottir A, Sig-urdsson A, Kristinsson KT, Jonasdottir A, Frigge ML, Gylfason A, Olason PI, Gudjonsson SA, Sverrisson S, Stacey SN, Sigurgeirsson B, Benediktsdottir KR, Sigurdsson H, Jonsson T, Benediktsson R, Olafsson JH, Johannsson OT, Hreidarsson AB, Sigurdsson G, the DIAGRAM Consortium, Ferguson-Smith AC, Gudbjartsson DF, Thorsteinsdottir U, Stefansson K (2009) Parental origin of sequence variants associated with complex diseases. *Nature* **462**(7275): 868–874.

Lange K (1997) An approximate model of polygenic inheritance. *Genetics* **147**(3): 1423–1430.

Lange K, Weeks D, Boehnke M, MacCluer J (1988) Programs for pedigree analysis: MENDEL, FISHER, and dGENE. *Genet Epidemiol* **5**(6): 471–472, ISSN 1098-2272.

Li Y, Guo Y, Wang J, Hou W, Chang MN, Liao D, Wu R (2011) A statistical design for testing transgenerational genomic imprinting in natural human populations. *PloS One* **6**(2): e16858.

Michailidou K, Hall P, Gonzalez-Neira A, Ghoussaini M, Dennis J, Milne RL, Schmidt MK, Chang-Claude J, Bojesen SE, Bolla MK, Wang Q, Dicks E, Lee A, Turnbull C, Rahman N, Breast and Ovarian Cancer Susceptibility Collaboration, Fletcher O, Peto J, Gibson L, Dos Santos Silva I, Nevanlinna H, Muranen TA, Aittomaki K, Blomqvist C, Czene K, Irwanto A, Liu J, Waisfisz Q, Meijers-Heijboer H, Adank M, Hereditary Breast and Ovarian Cancer Research Group Netherlands (HEBON), van der Luijt RB, Hein R, Dahmen N, Beckman L, Meindl A, Schmutzler RK, Muller-Myhsok B, Lichtner P, Hopper JL, Southey MC, Makalic E, Schmidt DF, Uitterlinden AG, Hofman A, Hunter DJ, Chanock SJ, Vincent D, Bacot F, Tessier DC, Canisius S, Wessels LF, Haiman CA, Shah M, Luben R, Brown J, Luccarini C, Schoof N, Humphreys K, Li J, Nordestgaard BG, Nielsen SF, Flyger H, Couch FJ, Wang X, Vachon C, Stevens KN, Lambrechts D, Moisse M, Paridaens R, Christiaens MR, Rudolph A, Nickels S, Flesch-Janys D, Johnson N, Aitken Z, Aaltonen K, Heikkinen T, Broeks A, Veer LJ, van der Schoot CE, Guenel P, Truong T, Laurent-Puig P, Menegaux F, Marme F, Schneeweiss A, Sohn C, Burwinkel B, Zamora MP, Perez JI, Pita G, Alonso MR, Cox A, Brock IW, Cross SS, Reed MW, Sawyer EJ, Tomlinson I, Kerin MJ, Miller N, Henderson BE, Schumacher F, Le Marchand L, Andrulis IL, Knight JA, Glendon G, Mulligan AM, kConFab Investigators, Australian Ovarian Cancer Study Group, Lindblom A, Margolin S, Hooning MJ, Hollestelle A, van den Ouweland AM, Jager A, Bui QM, Stone J, Dite GS, Api-cella C, Tsimiklis H, Giles GG, Severi G, Baglietto L, Fasching PA, Haeberle L, Ekici AB, Beckmann MW, Brenner H, Muller H, Arndt V, Stegmaier C, Swerdlow A, Ashworth A, Orr N, Jones M, Figueroa J, Lis-sowska J, Brinton L, Goldberg MS, Labreche F, Dumont M, Winqvist R, Pylkas K, Jukkola-Vuorinen A, Grip M, Brauch H, Hamann U, Bruning T, GENICA (Gene Environment Interaction and Breast Cancer in Germany) Network, Radice P, Peterlongo P, Manoukian S, Bonanni B, Devilee P, Tollenaar RA, Seynaeve C, van Asperen CJ, Jakubowska A, Lubinski J, Jaworska K, Durda K, Mannermaa A, Kataja V, Kosma VM, Hartikainen JM, Bogdanova NV, Antonenkova NN, Dork T, Kristensen VN, Anton-Culver H, Slager S, Toland AE, Edge S, Fostira F, Kang D, Yoo KY, Noh DY, Matsuo K, Ito H, Iwata H, Sueta A, Wu AH, Tseng CC, Van Den Berg D, Stram DO, Shu XO, Lu W, Gao YT, Cai H, Teo SH, Yip CH, Phuah SY, Cornes BK, Hartman M, Miao H, Lim WY, Sng JH, Muir K, Lophatananon A, Stewart-Brown S, Siriwanarangsan P, Shen CY, Hsiung CN, Wu PE, Ding SL, Sangrajrang S, Gaborieau V, Bren-nan P, McKay J, Blot WJ, Signorello LB, Cai Q, Zheng W, Deming-Halverson S, Shrubsole M, Long J, Simard J, Garcia-Closas M, Pharoah PD, Chenevix-Trench G, Dunning AM, Benitez J, Easton DF (2013) Large-scale genotyping identifies 41 new loci associated with breast cancer risk. *Nat Genet* **45**(4): 353–361.

Milne RL, Goode EL, Garcia-Closas M, Couch FJ, Severi G, Hein R, Fredericksen Z, Malats N, Zamora MP, Perez JIA, Benitez J, Dork T, Schurmann P, Karstens JH, Hillemanns P, Cox A, Brock IW, Elliot G, Cross SS, Seal S, Turnbull C, Renwick A, Rahman N, Shen C, Yu J, Huang C, Hou M, Nordestgaard BG, Bojesen SE, Lanng C, Alnaes GG, Kristensen V, B0rrensen-Dale A, Hopper JL, Dite GS, Apicella C, Southey MC, Lambrechts D, Yesilyurt BT, Floris G, Leunen K, Sangrajrang S, Gaborieau V, Brennan P, McKay J, Chang-Claude J, Wang-Gohrke S, Radice P, Peterlongo P, Manoukian S, Barile M, Giles GG, Baglietto L, John EM, Miron A, Chanock SJ, Lissowska J, Sherman ME, Figueroa JD, Bogdanova NV, Antonenkova NN, Zalutsky IV, Rogov YI, Fasching PA, Bayer CM, Ekici AB, Beckmann MW, Brenner H, Muller H, Arndt V, Stegmaier C, Andrulis IL, Knight JA, Glendon G, Mulligan AM, Mannermaa A, Kataja V, Kosma V, Hartikainen JM, Meindl A, Heil J, Bartram CR, Schmutzler RK, Thomas GD, Hoover RN, Fletcher O, Gibson LJ, dos Santos Silva I, Peto J, Nickels S, Flesch-Janys D, Anton-Culver H, Ziogas A, Sawyer E, Tomlinson I, Kerin M, Miller N, Schmidt MK, Broeks A, Vant Veer LJ, Tollenaar RAEM, Pharoah PDP, Dunning AM, Pooley KA, Marme F, Schneeweiss A, Sohn C, Burwinkel B, Jakubowska A, Lubinski J, Jaworska K, Durda K, Kang D, Yoo K, Noh D, Ahn S, Hunter DJ, Hankinson SE, Kraft P, Lindstrom S, Chen X, Beesley J, Hamann U, Harth V, Justenhoven C, for the GENICA Network, Winqvist R, Pylkas K, Jukkola-Vuorinen A, Grip M, Hooning M, Hollestelle A, Oldenburg RA, Tilanus-Linthorst M, Khusnutdinova E, Bermisheva M, Prokofieva D, Farahtdinova A, Olson JE, Wang X, Humphreys MK, Wang Q, for the kConFab Investigatorsm, Chenevix-Trench G, the AOCS Group, Easton DF (2011) Confirmation of 5p12 as a susceptibility locus for progesterone-receptor-positive, lower grade breast cancer. *Cancer Epidemiol Biomarkers Prev* **20**(10): 2222–2231.

O'Connell JR, Weeks DE (1998) Pedcheck: a program for identification of genotype incompatibilities in linkage analysis. *Am J Hum Genet* **63**(1): 259–266.

Rampersaud E, Mitchell BD, Naj AC, Pollin TI (2008) Investigating parent of origin effects in studies of type 2 diabetes and obesity. *Curr Diab Rev* **4**(4): 329.

Risch Neil (1990) Linkage strategies for genetically complex traits. i. multilocus models. *A J Hum Genet* **46**(2): 222.

Sawyer S, Mitchell G, McKinley J, Chenevix-Trench G, Beesley J, Chen X, Bowtell D, Trainer AH, Harris M, Lindeman GJ, James PA. A role for common genomic variants in the assessment of familial breast cancer (2012) *J Clin Oncol* **30**(35): 4330–4336.

Schaid DJ, McDonnell SK, Riska SM, Carlson EE, Thibodeau SN. Estimation of genotype relative risks from pedigree data by retrospective likelihoods. *Genet Epidemiol* **34**(4): 287–298 (2010) .

Stacey SN, Manolescu A, Sulem P, Rafnar T, Gudmundsson J, Gudjonsson SA, Masson G, Jakob-sdottir M, Thorlacius S, Helgason A, Aben KK, Strobbe LJ, Albers-Akkers MT, Swinkels DW, Henderson BE, Kolonel LN, Le Marchand L, Millastre E, Andres R, Godino J, Garcia-Prats MD, Polo E, Tres A, Mouy M, Saemundsdottir J, Backman VM, Gudmundsson L, Kristjansson K, Bergthorsson JT, Kostic J, Frigge ML, Geller F, Gudbjartsson D, Sigurdsson H, Jonsdottir T, Hrafnkelsson J, Johannsson J, Sveinsson T, Myrdal G, Grimsson HN, Jonsson T, von Holst S, Werelius B, Margolin S, Lindblom A, Mayordomo JI, Haiman CA, Kiemeney LA, Johanns-son OT, Gulcher JR, Thorsteinsdottir U, Kong A, Stefansson K. Common variants on chromosomes 2q35 and 16q12 confer susceptibility to estrogen receptor-positive breast cancer. *Nat Genet* **39**(7): 865–869 (2007) .

Stacey SN, Sulem P, Zanon C, Gudjonsson SA, Thorleifsson G, Helgason A, Jonasdottir A, Be-senbacher S, Kostic JP, Fackenthal JD, Huo D, Adebamowo C, Ogundiran T, Olson JE, Fredericksen ZS, Wang X, Look MP, Sieuwerts AM, Martens JWM, Pajares I, Garcia-Prats MD, Ramon-Cajal JM, de Juan A, Panadero A, Ortega E, Aben KKH, Vermeulen SH, Asadzadeh F, van Engelenburg KCA, Margolin S, Shen C, Wu P, Forsti A, Lenner P, Henriksson R, Johansson R, Enquist K, Hallmans G, Jonsson T, Sigurdsson H, Alexiusdottir K, Gudmundsson J, Sigurdsson A, Frigge ML, Gudmundsson L, Kristjansson K, Halldorsson BV, Styrkarsdottir U, Gulcher JR, Hemminki K, Lindblom A, Kiemeney LA, Mayordomo JI, Foekens JA, Couch FJ, Olopade OI, Gudbjartsson DF, Thorsteinsdottir U, Rafnar T, Johannsson OT, Stefansson K (2010) Ancestry-shift refinement mapping of the C6orf97-ESR1 breast cancer susceptibility locus. *PLoS Genet* **6**(7): e1001029.

StataCorp LP (2009) *Stata statistical software: Release 11*. Stata Corporation: College Station, TX.

Thomas G, Jacobs KB, Kraft P, Yeager M, Wacholder S, Cox DG, Hankinson SE, Hutchinson A, Wang Z, Yu K, Chatterjee N, Garcia-Closas M, Gonzalez-Bosquet J, Prokunina-Olsson L, Orr N, Willett WC, Colditz GA, Ziegler RG, Berg CD, Buys SS, McCarty CA, Spencer Feigelson H, Calle EE, Thun MJ, Diver R, Prentice R, Jackson R, Kooperberg C, Chlebowski R, Lissowska J, Peplonska B, Brinton LA, Sigurdson A, Doody M, Bhatti P, Alexander BH, Buring J, Lee I, Vatten LJ, Hveem K, Kumle M, Hayes RB, Tucker M, Gerhard DS, Fraumeni Jr JF, Hoover RN, Chanock SJ, Hunter DJ (2009) A multistage genome-wide association study in breast cancer identifies two new risk alleles at 1p11.2 and 14q24.1 (RAD51L1). *Nat Genet* **41**(5): 579–584.

Turnbull C, Ahmed S, Morrison J, Pernet D, Renwick A, Maranian M, Seal S, Ghoussaini M, Hines S, Healey CS, Hughes D, Warren-Perry M, Tapper W, Eccles D, Evans DG, The Breast Cancer Susceptibility Collaboration (UK), Hooning M, Schutte M, van den Ouweland A, Houlston R, Ross G, Langford C, Pharoah PDP, Stratton MR, Dunning AM, Rahman N, Easton DF (2010) Genome-wide association study identifies five new breast cancer susceptibility loci. *Nat Genet* **42**(6): 504–507.

Wacholder S, Hartge P, Prentice R, Garcia-Closas M, Spencer Feigelson H, Diver WR, Thun MJ, Cox DG, Hankinson SE, Kraft P, Rosner B, Berg CD, Brinton LA, Lissowska J, Sherman ME, Chlebowski R, Kooperberg C, Jackson RD, Buckman DW, Hui P, Pfeiffer R, Jacobs KB, Thomas GD, Hoover RN, Gail MH, Chanock SJ, Hunter DJ (2010) Performance of common genetic variants in breast-cancer risk models. *N Engl J Med* **362**(11): 986–993.

White H (1980) A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica* **48**(4): 817–838.

White H (1982) Maximum likelihood estimation of misspecified models. *Econometrica* **50**(1): 1–25.

Supplementary Information accompanies this paper on British Journal of Cancer website (http://www.nature.com/bjc)