Common breast cancer susceptibility alleles and the risk of breast cancer for BRCA1 and BRCA2 mutation carriers: implications for risk prediction

Supplementary material

Appendix 1

Genotyping Quality Control

Genotyping was performed using either the iPLEX or Taqman platforms. To ensure genotyping consistency, included at least 2% of the samples in duplicate, no template controls in every plate, and a random mixture of affected and unaffected carriers. Samples that failed for two or more of the SNPs genotyped (or ≥20% of the SNPs typed if more than three SNPs were analysed in that genotyping round) were excluded from the analysis. The genotype data for a given SNP and a given study were included in the analysis only if the call rate was >95% after samples that failed at multiple SNPs had been excluded. The concordance between duplicates had to be at least 98%. To assess the accuracy of genotyping across genotyping centers, all centers genotyped 95 DNA samples from a standard test plate (Coriell Institute) for all SNPs. If the genotyping was inconsistent for more than one sample in the test plate, the study was excluded from the analysis of that SNP. As an additional genotyping quality-control check, we also evaluated the deviation from Hardy-Weinberg equilibrium (HWE) for unrelated subjects separately for each SNP and study. Studies with a HWE p-value of less than 0.005 were excluded from the analysis. If HWE p-values were in the range 0.005-0.05 we examined the genotyping cluster plots; none revealed any unusual patterns and these studies were therefore included in all the analyses. For the three novel polymorphisms investigated, one study failed the quality control criteria and was excluded from the analysis of the rs4973768 (*SLC4A7/NEK10*) polymorphism and five studies were excluded from the

analysis of the rs10941679 (5p12) SNP. No studies were excluded from the analysis of the rs6504950 (*STXBP4*/*COX11*) SNP.

A number of samples (3344 - 5150 mutation carriers depending on SNP) were re-genotyped by iPLEX for the six previously published SNPs in a multiplex along with the new SNPs, as part of the effort to genotype additional samples from mutation carriers who were recruited subsequent to the original reports. Samples with inconsistent genotypes between the two genotyping rounds were excluded from the analysis of each polymorphism (discordance rate: 0.002-0.006 depending on the SNP).

Appendix 2

To evaluate the combined effects of the significant SNPs on breast cancer risk, for each pair of SNPs we fitted models of the form: $\lambda(t) = \lambda_0(t)\exp(\beta_1 x_1 + \beta_2 x_2 + \gamma x_1 x_2)$ where β_1 is the per-allele log-hazard ratio for SNP1, β_2 is the per-allele log-hazard ratio for SNP2, γ is a term for the interaction between SNP1 and SNP2 and x_1, x_2 represent the number of minor alleles at locus 1 and 2 respectively (e.g. 0,1,2). To test for interactions between a pair of SNPs, we tested for the significance of the interaction parameter (i.e. $\gamma \neq 0$).

Implications for risk predictions

We estimated the absolute risk of developing breast cancer based on the joint distribution of all SNPs that were significantly associated with risk for either *BRCA1* or *BRCA2* mutation carriers. To do this, we assumed that the average, age-specific breast cancer incidences for *BRCA1* and *BRCA2* mutation carriers, over all modifying loci, agreed with published penetrance estimates for *BRCA1* and *BRCA2 (1)*. Under this model, and assuming independence between the SNPs, the breast cancer incidence $\lambda(t,X)$ at age *t* for a mutation carrier with genotype vector $\mathbf{X}=(x_1, x_2, ..., x_n)$ at loci 1 to n, depends on the underlying genotypes according to : $\lambda(t,X) = \lambda_0(t)\exp(\beta^t X)$ where $\lambda_0(t)$ is the baseline incidence and $\boldsymbol{\beta} = (\beta_1, \beta_2, ..., \beta_n)$ is the vector of logHRs estimates for SNPs 1 to n. $\lambda_0(t)$ is unknown. $\exp(\beta^t X)$ represents the combined HR across all loci. If $\mu(t)$ is the average incidence at age *t* obtained from published studies, then:

$$\mu(t) = \frac{\sum_i p(X_i)\lambda(t,X_i)S(t-1,X_i)}{\sum_i p(X_i)S(t-1,X_i)} \quad (1) .$$

$p(X_i)$ is the probability of multilocus genotype i ($i=1,\ldots,3^n$) and under the assumption of linkage equilibrium between the SNPs $p(X_i) = \prod_{j=1}^{n} q(x_{ij})$ with $q(x_{ij})$ being the genotype frequency at locus j of the ith multilocus genotype . Under HWE for someone with 2,1 or 0 minor alleles at locus j, $q(x_{ij}) = \alpha_j^2, 2\alpha_j(1-\alpha_j)$ or $(1-\alpha_j)^2$ respectively, where α_j is the minor allele frequency at locus j . $S(t,X_i)$ is the survival function to age t for an individual with a multilocus genotype **X**, which can be expressed in terms of the baseline incidence and the logHRs according to standard survival analysis theory(2). Assuming $S(0,X_i)=1$ for all genotypes, it is possible to solve recursively for $\lambda_0(t)$ and hence obtain the age-specific breast cancer incidence for each multilocus genotype $\lambda(t,X)$. In these calculations, we used minor allele frequency estimates based on control data from population based studies (3-6) and the logHR estimates obtained from the present analysis. As an alternative, we used the estimated log(per-allele odds ratios) from the largest published population-based studies. To further evaluate the departure from the multiplicative model for the combined effects of all SNPs, we derived a risk score under the assumption of independence based on the HR estimated from the analysis of each SNP (that is, the risk score was of the form $\exp(\beta^t X)$). We then estimated the HR associated with the quantiles of the distribution of the risk score and compared those against the HR predicted under the multiplicative model.

Supplementary Figure 1 legend: A. Cumulative distribution function of the combined HR for breast cancer risk for *BRCA2* mutation carriers based on the Odds Ratio estimates from population based studies for all nine SNPs. **B.** Predicted cumulative risk of developing breast cancer by age 80 for *BRCA2* mutation carriers by the combined HR at the above SNPs.

Supplementary Table 1: *BRCA1* and *BRCA2* mutation carriers genotyped for at least one SNP by study

Study	Country	*BRCA1*	*BRCA2*
Breast Cancer Family Registry (BCFR)	USA, Canada, Australia	512	366
Baltic Familial Breast and Ovarian Cancer Consortium (BFBOCC)	Latvia, Lithuania	102	0
Copenhagen Breast Cancer Study (CBCS)	Denmark	175	88
Spanish National Cancer Centre (CNIO)	Spain, Greece	254	334
CONsorzio Studi ITaliani sui Tumori Ereditari Alla Mammella (CONSIT TEAM)	Italy	525	325
Deutsches Krebsforschungszentrum (DKFZ)	Germany	69	30
HEreditary Breast and Ovarian study Netherlands (HEBON)	The Netherlands	796	304
EMBRACE	UK and Eire	999	840
Fox Chase Cancer Centre (FCCC)	USA	82	54
German Consortium of Hereditary Breast and Ovarian Cancer (GC-HBOC)	Germany	985	467
Genetic Modifiers of cancer risk in *BRCA1*/2 mutation carriers (GEMO)	France, USA	1203	599
Georgetown University	USA	31	17
Gynecologic Oncology Group (GOG)	USA	390	277
Hospital Clinico San Carlos (HCSC)	Spain	132	119
Helsinki Breast Cancer Study (HEBCS)	Finland	103	105
Institut Català d'Oncologia (ICO)	Spain	112	119
International Hereditary Cancer Centre (IHCC)	Poland	350	0
Iceland Landspitali - University Hospital (ILUH)	Iceland	0	135
Interdisciplinary Health Research International Team Breast Cancer Susceptibility (INHERIT BRCAs)	Canada	73	82
Istituto Oncologico Veneto Hereditary Breast and Ovarian Cancer Study (IOVHBOCS)	Italy	118	108
kConFab	Australia	595	488
MAGIC	USA	554	313
Mayo Clinic (MAYO)	USA	243	132
Modifier Study of Quantitative Effects on Disease (ModSQuaD)	Czech Republic, Belgium	324	147
Memorial Sloane Kettering Cancer Center (MSKCC)	USA	289	211
Medical University Vienna (MUV)	Austria	278	120
National Cancer Institute (NCI)	USA	172	77
National Israeli Cancer Control Center (NICCC)	Israel	340	220
N.N. Petrov Institute of Oncology (NNPIO)	Russia	93	0
Ontario Cancer Genetics Network (OCGN)	Canada	227	172
Ohio State University Clinical Cancer Center (OSU CCG)	USA	87	49
Odense University Hospital (OUH)	Denmark	268	195
Pisa Breast Cancer Study (PBCS)	Italy	88	57
Sheba Medical Centre (SMC)- Tel Hashomer	Israel	614	301
Swedish Breast Cancer Study (SWE-BRCA)	Sweden	537	177
University of California Irvine (UCI)	USA	172	125
UK and Gilda Radner Familial Ovarian Cancer Registries (UKGRFOCR)	USA, UK	161	35
University of Pennsylvania (UPENN)	USA	313	152
Women's Cancer Research Institute (WCRI)	USA	159	69
Total		**12525**	**7409**

Supplementary Table 2: Hazard Ratio estimates after excluding prevalent cancer cases

Mutation/SNP	Unaffected/Affected	HR*	95% CI	p-value*
SLC4A7 rs4973768				
BRCA1	4427/2528	1.04	0.97-1.13	0.25
BRCA2	2639/1700	1.10	1.01-1.19	0.032
STXBP4 rs6504950				
BRCA1	4466/2574	1.04	0.97-1.11	0.32
BRCA2	2666/1720	1.02	0.92-1.12	0.74
5p12 rs10941679				
BRCA1	4040/2430	0.99	0.92-1.07	0.82
BRCA2	2458/1649	1.08	0.98-1.20	0.13
BRCA2 dominant	2458/1649	1.13	1.00-1.28	0.052

*Multiplicative model unless specified

Locus1	Locus2	Unaffected	Breast Cancer cases	Interaction		
				HR	95%CI	p-interaction
BRCA1						
TNRC9	*2q35*	3367	4117	1.09	0.95-1.24	0.22
BRCA2						
FGFR2	*TNRC9*	2046	2558	1.09	0.98-1.20	0.10
FGFR2	*MAP3K1*	2154	2706	0.94	0.85-1.03	0.17
FGFR2	*LSP1*	1989	2525	1.00	0.90-1.09	0.85
FGFR2	*2q35*	1975	2523	1.02	0.87-1.18	0.82
FGFR2	*SLC4A7*	1874	2408	0.98	0.89-1.07	0.64
FGFR2	*5p12*	1847	2406	1.04	0.92-1.19	0.53
TNRC9	*MAP3K1*	2127	2669	0.93	0.84-1.03	0.14
TNRC9	*LSP1*	1904	2408	1.04	0.66-1.63	0.86
TNRC9	*2q35*	1927	2465	1.11	0.94-1.30	0.22
TNRC9	*SLC4A7*	1823	2331	0.96	0.86-1.06	0.40
TNRC9	*5p12*	1806	2335	1.09	0.64-1.84	0.75
MAP3K1	*LSP1*	2043	2565	0.96	0.86-1.07	0.45
MAP3K1	*2q35*	2067	2621	1.06	0.90-1.23	0.48
MAP3K1	*SLC4A7*	1949	2483	1.02	0.93-1.13	0.64
MAP3K1	*5p12*	1925	2478	0.95	0.83-1.10	0.52
LSP1	*2q35*	2531	3149	1.03	0.90-1.18	0.68
LSP1	*SLC4A7*	2285	2888	0.92	0.84-1.01	0.07
LSP1	*5p12*	2213	2827	0.97	0.85-1.10	0.61
2q35	*SLC4A7*	2329	2961	1.10	0.96-1.26	0.17
2q35	*5p12*	2257	2899	0.87	0.72-1.06	0.17
SLC4A7	*5p12*	2562	3231	1.03	0.92-1.15	0.60

Supplementary Table 3: Pairwise interaction Hazard Ratio estimates and significance tests

Reference List

(1) Antoniou AC, Cunningham AP, Peto J, et al. The BOADICEA model of genetic susceptibility to breast and ovarian cancers: updates and extensions. Br J Cancer 2008;98:1457-66.

(2) Cox DR, Oakes D. Analysis of survival data. London: Chapman and Hall; 1984.

(3) Ahmed S, Thomas G, Ghoussaini M, et al. Newly discovered breast cancer susceptibility loci on 3p24 and 17q23.2. Nat Genet 2009;41:585-90.

(4) Easton DF, Pooley KA, Dunning AM, et al. Genome-wide association study identifies novel breast cancer susceptibility loci. Nature 2007;447:1087-93.

(5) Milne RL, Osorio A, Cajal TR, et al. The Average Cumulative Risks of Breast and Ovarian Cancer for Carriers of Mutations in BRCA1 and BRCA2 Attending Genetic Counseling Units in Spain. Clin Cancer Res 2008;14:2861-9.

(6) Stacey SN, Manolescu A, Sulem P, et al. Common variants on chromosome 5p12 confer susceptibility to estrogen receptor-positive breast cancer. Nat Genet 2008;40:703-6.

A.

B.

Supplementary Figure 1