Common Genetic Variants and Modification of Penetrance of BRCA2-Associated Breast Cancer

Gaudet, Mia M.; Kirchhoff, Tomas; Green, Todd; Vijai, Joseph; Korn, Joshua M.; Guiducci, Candace; Segre, Ayellet V.; McGee, Kate; McGuffog, Lesley; Kartsonaki, Christiana; Morrison, Jonathan; Healey, Sue; Sinilnikova, Olga M.; Stoppa-Lyonnet, Dominique; Mazoyer, Sylvie; Gauthier-Villars, Marion; Sobol, Hagay; Longy, Michel; Frenay, Marc; Hogervorst, Frans B. L.; Rookus, Matti A.; Collee, J. Margriet; Hoogerbrugge, Nicoline; van Roodendaal, Kees E. P.; Piedmonte, Marion; Rubinstein, Wendy; Nerenstone, Stacy; Van Le, Linda; Blank, Stephanie V.; Caldes, Trinidad; de la Hoya, Miguel; Nevanlinna, Heli; Aittomaki, Kristiina; Lazaro, Conxi; Blanco, Ignacio; Arason, Adalgeir; Johannsson, Oskar T.; Barkardottir, Rosa B.; Devilee, Peter; Olopade, Olofunmilayo I.; Neuhausen, Susan L.; Wang, Xianshu; Fredericksen, Zachary S.; Peterlongo, Paolo; Manoukian, Siranoush; Barile, Monica; Viel, Alessandra; Radice, Paolo; Phelan, Catherine M.; Narod, Steven

Published in:
PLoS Genetics

DOI:
10.1371/journal.pgen.1001183

2010

Link to publication

Citation for published version (APA):
Common Genetic Variants and Modification of Penetrance of BRCA2-Associated Breast Cancer


1 Department of Epidemiology and Population Health and Department of Obstetrics and Gynecology and Women’s Health, Albert Einstein College of Medicine, New York, New York, United States of America, 2 Clinical Genetics Service, Department of Medicine, Sloan-Kettering Institute, Memorial Sloan-Kettering Cancer Center, New York, New York, United States of America, 3 Cancer Biology and Genetics Program, Sloan-Kettering Institute, Memorial Sloan-Kettering Cancer Center, New York, New York, United States of America, 4 Broad Institute of Harvard and Massachusetts Institute of Technology, Cambridge, Massachusetts, United States of America, 5 Program in Medical and Population Genetics, Broad Institute of Harvard and Massachusetts Institute of Technology, Cambridge, Massachusetts, United States of America, 6 Center for Human Research, Massachusetts General Hospital, Boston, Massachusetts, United States of America, 7 Department of Molecular Biology, Massachusetts General Hospital, Boston, Massachusetts, United States of America, 8 Center for Cancer Research, Cancer Inflammation Program, Human Genetics Section, National Cancer Institute – Frederick, Frederick, Maryland, United States of America, 9 Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, Cambridge, United Kingdom, 10 Unité Mixte de Génétique Constitutionnelle des Cancers Fréquents, Centre Hospitalier Universitaire de Lyon/Centre Léon Bérard, Lyon, France, 11 Equipe labellisée LIGUE 2008, UMRIS201 CNRS, Centre Léon Bérard, Université de Lyon, Lyon, France, 12 Institut Curie, Service de Génétique, INSERM U830, F-75248, Université Paris Descartes, Paris, France, 13 Service de Génétique Oncologique, Institut Curie, Paris, France, 14 Département Oncologie génétique, Prévention et Dépistage, INSERM CIC-P9502, Institut Paoli-Calmettes/Université d’Aix-Marseille II, Marseille, France, 15 Institut Bergonié, Bordeaux, France, 16 Centre Antoine Lacassagne, Nice, France, 17 GEMO Study - Cancer Genetics Network “Groupe Génétique et Cancer”, Fédération Nationale des Centres de Lutte Contre le Cancer, Paris, France, 18 Family Cancer Clinic, Netherlands Cancer Institute, Amsterdam, The Netherlands, 19 Department of Epidemiology, Netherlands Cancer Institute, Amsterdam, The Netherlands, 20 Department of Medical Oncology, Rotterdam Family Cancer Clinic, Erasmus University Medical Center, Rotterdam, The Netherlands, 21 Department of Human Genetics, Radboud University Nijmegen Medical Centre, Nijmegen, The Netherlands, 22 Department of Clinical Genetics, University Medical Center, Maastricht, The Netherlands, 23 Gynecologic Oncology Group Statistical and Data Center, Roswell Park Cancer Institute, Buffalo, New York, United States of America, 24 NorthShore University Health System, Evanston, Illinois, United States of America, 25 Central Connecticut Cancer Consortium, Hartford Hospital, Hartford, Connecticut, United States of America, 26 University of North Carolina, Chapel Hill, North Carolina, United States of America, 27 New York University School of Medicine, New York, New York, United States of America, 28 Molecular Oncology Laboratory, Hospital Clinico San Carlos, Madrid, Spain, 29 Department of Obstetrics and Gynecology, Helsinki University Central Hospital, Helsinki, Finland, 30 Department of Clinical Genetics, Helsinki University Central Hospital, Helsinki, Finland, 31 Hereditary Cancer Program, Catalan Institute of Oncology, Barcelona, Spain, 32 Department of Oncology, Landskapsläkar—LSH, Reykjavik, Iceland, 33 Department of Pathology, Landslipslät—LSH, Reykjavik, Iceland, 34 Faculty of Medicine, University of Iceland, Reykjavik, Iceland, 35 Department of Human Genetics and Department of Pathology, Leiden University Medical Center, Leiden, The Netherlands, 36 Center for Clinical Cancer Genetics and Global Health, Department of Medicine, University of Chicago Medical Center, Chicago, Illinois, United States of America, 37 Department of Population Sciences, the Beckman Research Institute of the City of Hope, Duarte, California, United States of America, 38 Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, Minnesota, United States of America, 39 Department of Health Sciences Research, Mayo Clinic, Rochester, Minnesota, United States of America, 40 Unit of Genetic Susceptibility to Cancer, Department of Experimental Oncology and Molecular Medicine, Fondazione IRCCS Istituto Nazionale Tumori (INT), Milan, Italy, 41 IFOM, Fondazione Istituto FIRC di Oncologia Molecolare, Milan, Italy, 42 Unit of Medical Genetics, Department of Preventive and Predictive Medicine, Fondazione IRCCS Istituto Nazionale dei Tumori (INT), Milan, Italy, 43 Division of Cancer
Abstract

The considerable uncertainty regarding cancer risks associated with inherited mutations of BRCA2 is due to unknown factors. To investigate whether common genetic variants modify penetrance for BRCA2 mutation carriers, we undertook a two-staged genome-wide association study in BRCA2 mutation carriers. In stage 1 using the Affymetrix 6.0 platform, 592,163 filtered SNPs genotyped were available on 899 young (~40 years) affected and 804 unaffected carriers of European ancestry. Associations were evaluated using a survival-based score test adjusted for familial correlations and stratified by country of the study and BRCA2*6174delT mutation status. The genomic inflation factor (λ) was 1.011. The stage 1 association analysis revealed multiple variants associated with breast cancer risk: 3 SNPs had p-values <10^{-5} and 39 SNPs had p-values<10^{-4}. These variants included several previously associated with sporadic breast cancer risk and two novel loci on chromosome 20 (rs311499) and chromosome 10 (rs16917302). The chromosome 10 locus was in another variant that has recently been associated with breast cancer in an independent study of unselected cases. To investigate whether common genetic variants modify penetrance for BRCA2*6174delT, the combined per allele HR of the minor allele for the novel loci rs16917302 was 0.75 (95% CI 0.66–0.86, p = 3.8 × 10^{-5}) and for rs311499 was 0.72 (95% CI 0.61–0.85, p = 6.6 × 10^{-4}). FGFR2 rs2981575 had the strongest association with breast cancer risk (per allele HR = 1.28, 95% CI 1.18–1.39, p = 1.2 × 10^{-5}). These results indicate that SNPs that modify BRCA2 penetrance identified by an agnostic approach thus far are limited to variants that also modify risk of sporadic BRCA2 wild-type breast cancer.


Editor: James M. Ford, Stanford University School of Medicine, United States of America

Received June 8, 2010; Accepted September 28, 2010; Published October 28, 2010

This is an open-access article distributed under the terms of the Creative Commons Public Domain declaration which stipulates that, once placed in the public domain, this work may be freely reproduced, distributed, transmitted, modified, built upon, or otherwise used by anyone for any lawful purpose.

Funding: This study was supported by the Starr Cancer Research Consortium and the Breast Cancer Research Foundation, as well as by NIH NCI P20CA103694-3 and the Lymphoma Foundation. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: olfftk@mskcc.org

† These authors contributed equally to this work.

‡ Membership of each consortium is provided in the Acknowledgements.


**Author Summary**

The risk of breast cancer associated with BRCA2 mutations varies widely. To determine whether common genetic variants modify the penetrance of BRCA2 mutations, we conducted the first genome-wide association study of breast cancer among women with BRCA2 mutations using a two-stage approach. The major finding of the study is that only those loci known to be associated with breast cancer risk in the general population, including FGFR2 (rs2981575), modified BRCA2-associated risk in our high-risk population. Two novel loci, on chromosomes 10 in ZNF365 (rs16917302) and chromosome 20 (rs311499), were shown to modify risk in BRCA2 mutation carriers, although not at a genome-wide level of significance. However, the ZNF365 locus has recently independently been associated with breast cancer risk in sporadic tumors, highlighting the potential significance of this zinc finger-containing gene in breast cancer pathogenesis. Our results indicate that it is unlikely that other common variants have a strong modifying effect on BRCA2 penetrance.

**Introduction**

After more than a decade of clinical testing for mutations of BRCA1 and BRCA2, there remains considerable uncertainty regarding cancer risks associated with inherited mutations of these genes. This variable penetrance is most striking for BRCA2 [1–4], and it affects medical management [5]. Women with the same BRCA2 mutation may develop breast, ovarian or other cancers at different ages or not at all [6]. In a segregation analysis of families identified through breast cancer cases diagnosed before age 55, the residual familial clustering after accounting for BRCA1 and BRCA2 mutations could be explained by a large number of low penetrance genes with multiplicative effects on breast cancer risk [7,8]. A candidate gene approach in BRCA2 mutation carriers led to the discovery of loci that modify the penetrance of BRCA2 mutations, such as RAD51 135 G>C [9] and perhaps CASP8 [10,11] and IGFBP2 [12], if replicated. To investigate whether other common single nucleotide polymorphisms (SNP), copy number variants (CNV), or copy number polymorphisms (CNP) modify penetrance for BRCA2 mutation carriers, we undertook a two-staged genome-wide association study (GWAS) in BRCA2 mutation carriers from the international Consortium for Investigators of Modifiers of BRCA1/2 (CIMBA) and other international studies. We hypothesized that an agnostic search for breast cancer loci in an enriched population of BRCA2 mutation carriers, the first among this high risk population, would provide greater power than a sporadic population of equal number, and would yield associations specific to BRCA2 carriers and/or the general population.

**Results**

**Stage 1 and Stage 2 Genotyping**

In stage 1, genotype data were available for 899 young (<40 years) affected and 804 older (>40 years) unaffected carriers of European ancestry after quality control filtering and removal of ethnic outliers (Figure S1). A total of 592,163 filtered SNPs genotyped using the Affymetrix Genome-Wide Human SNP Array 6.0 platform passed quality control assessment. In stage 1, comparison of the observed and expected distributions (quantile-quantile plot: Figure S2) showed little evidence for an inflation of the test statistics (genomic inflation factor \( \lambda = 1.01 \)), thereby excluding the possibility of significant hidden population substructure, cryptic relatedness among subjects or differential genotype calling between BRCA2 affected and BRCA2 unaffected carriers. Multiple variants were found to be associated with breast cancer risk (Figure S3); 3 SNPs had \( p < 10^{-7} \) and 39 SNPs had \( p < 10^{-5} \). The most significant association \( (p = 3.6 \times 10^{-6}) \) was observed for FGFR2 rs2981575 (Table 1), a variant previously shown to be associated with increased risk of BRCA2-related breast cancer [13]. A positive association was also observed with rs3093662 (Table 1), near TOX3, which has also been associated with sporadic breast cancer risk [13].

Using the stage 1 data, we also performed a GSEA as implemented in MAGENTA [14] to evaluate whether a functionally-related set of genes relevant to BRCA2 function (Table S1) was enriched for relative risk associations (see Statistical Methods). The 59 genes selected are related to the Fanconi anemia pathway [15] as well as other pathways reported in the literature to regulate or interact with BRCA1/2 [16]. These showed no enrichment of associations with the breast cancer risk \( (p = 0.56) \). In addition, eight of 125 known cancer susceptibility alleles identified by previous GWAS of other cancers [17] were associated with BRCA2 modification in the current study, a number not greater than expected (Kolmogorov-Smirnov \( p = 0.60 \)) by chance alone. Of the 113 most significantly associated SNPs \( (p < 10^{-5}) \) in our study, three showed significant association \( (p < 0.05) \) with BRCA1-associated breast cancer risk in a complimentary GWAS [18].

In the combined stage 1 and stage 2 results, four independent SNPs (pairwise \( r^2 < 0.80 \)) were associated with increased risk of breast cancer risk with \( p \)-values \( < 10^{-4} \) (Table 1). Previously identified breast cancer susceptibility loci [13,19,20] had the most significant associations among BRCA2 mutation carriers (FGFR2: per allele \( p - \text{value} = 1.2 \times 10^{-3} \) and TOX3: per allele \( p - \text{value} = 4.9 \times 10^{-5} \)). Novel loci, rs16917302 on chromosome 10 and rs311499 on chromosome 20, had HRs in stage 2 that were in the same direction as those observed for stage 1 (Figure 1, Table 1), but were smaller in magnitude (HR = 0.67 (95% CI:0.56-0.80) vs. 0.85 (95% CI: 0.70–1.04) for rs16917302; HR = 0.60 (95%CI:0.50-0.70) vs. 0.84 (95%CI: 0.67–1.06) for rs311499) perhaps reflecting a “winner’s curse” effect [21]. The associations for these SNPs were not statistically significant in stage 2 (Table 1). In the combined stage 1 and stage 2 dataset, the C allele of rs16917302 was associated with lower risk of breast cancer (per allele HR = 0.75, 95% CI 0.66-0.86; \( p = 2.7 \times 10^{-5} \); Table 1), and the C allele of rs311499 was associated with a reduced risk (per allele HR = 0.72, 95% CI 0.61–0.85; \( p = 7.9 \times 10^{-5} \); Table 1). A full list of stage 2 results can be found in Table S2. Using the combined stage 1 and stage 2 data, there was no evidence that the HR for SNP rs16917302 changes with age \( (p = 0.63) \), but there was some evidence that the per-allele HR for rs311499 may increase with age \( (p = 0.034) \).

**Copy Number Variant Analysis**

We also examined the association of both high-frequency CNPs and low-frequency CNVs to case-control status using the stage 1 data. After performing standard quality control measures including a minor allele frequency (MAF) threshold of 5%, we identified 191 polymorphisms with reliable genotypes. No associations were found between CNVs and the phenotype; there was no inflation or deflation of the test statistic, and the best \( p \)-value was \( 4 \times 10^{-3} \). We similarly assessed less common CNPs, and found neither the overall burden of events (or any subclass thereof, such as large deletions overlapping genes) nor any specific locus associated with breast cancer risk (Figure S4).
Discussion

In this GWAS of BRCA2 mutation carriers, the first in this high risk population, we found previously identified breast cancer susceptibility loci modified risk of BRCA2-associated breast cancer with similar magnitude of association. Although FGFR2 (rs2981575) was the only locus to reach genome-wide statistical significance, novel loci, rs16917302 and rs10509168 were each associated with breast cancer risk.

rs16917302 is located on chromosome 10, in the zinc finger protein 365 gene (ZNF365). A recent multistage GWAS of 15,992 sporadic breast cancer cases and 16,891 controls also observed an inverse association (per allele OR = 0.82, 95% CI 0.82–0.91, p = 5.1 × 10⁻¹⁵) between breast cancer risk and rs10509168, a SNP 18kb from rs16917302 (pairwise r² = 0.1) and located in intron 4 of ZNF365 [24]. Of the 3,659 cases and 4,897 controls in stage 1 of that study, imputation revealed that the locus identified in our BRCA2 study, rs16917302, was significantly associated with risk for breast cancer (p = 0.02) (Easton DF, personal communication). The second novel SNP in the current study, rs311499, is located on chromosome 20, within a region containing several possible candidate genes including GMEB2, SRMS, PTK6, STMN3, and TNFRSF6 among others. The combined stage 1 and stage 2 analyses were also stratified by stage.

The region also includes other possible genes including SRMS, PTK6, STMN3, and TNFRSF6 among others. The functional significance of these regions with breast carcinogenesis is unknown; further research is warranted.

There was some evidence that the HR associated with rs311499 may change with age. We also observed that the stage 1 HR for this SNP was larger in magnitude compared to the stage 2 HR, consistent with a winner’s curse effect [21]. Since stage 1 of our experiment included mostly BRCA2 mutation carriers diagnosed at a young age, and stage 2 mutation carriers diagnosed an older age, the “winner’s curse” and age-specific effects are confounded and may be difficult to distinguish. Fitting the age-dependent HR model for SNP rs311499 using the stage 2 data yielded no significant variation in the HR by age (p = 0.47), but the sample size for this analysis was relatively small. Future larger studies should aim to clarify this.

Mutations in known genes (BRCA1, BRCA2, TP53, CHEK2, PTEV, and ATM) explain only 20–25% of the familial clustering of breast cancer; the residual familial clustering may be explained by the existence of multiple common, low-penetrance alleles (‘polygenes’) [25]. Perhaps because the majority of BRCA2-associated breast tumors are estrogen receptor (ER)-positive, as are the majority of non-hereditary breast cancers [26], risk alleles for sporadic breast cancer are more likely to be modifiers of risk of BRCA2-associated hereditary breast cancer. Of the seven GWAS-identified breast cancer-associated SNPs examined in a BRCA2 background [13,19,20], SNPs in FGFR2 (rs2981575), TOX3 (rs3803662), MAP3K1 (rs889312), and LSP1 (rs3817198) have been shown to modify BRCA2 penetrance, in contrast with BRCA1 tumors, in which only two of these same SNPs (based on a 2 degrees of freedom model) modified risk of these largely ER-negative tumors [26]. As previously noted [13,20], the stage 1 HRs among BRCA2 mutation carriers, reported here, were nearly identical to odds ratio estimates observed in sporadic breast cancer studies, consistent with a simple multiplicative interaction between the BRCA2 mutant alleles and the common susceptibility SNPs. If
Figure 1. Association signals, genetic structure, and linkage disequilibrium of the novel modifier loci of BRCA2 penetrance in the regions surrounding rs1691730 on chromosome 10 and rs311499 on chromosome 20. The color of the dots indicates linkage disequilibrium (LD; based on r^2 values) in the CEU population (as per scale). Triangle plots below represent LD from actual data of BRCA2 carriers in the study.

doi:10.1371/journal.pgen.1001183.g001
replicated, the two additional SNPs identified here would only explain about 1.7% of the variance in breast cancer risk among BRCA2 mutation carriers. Taken together, the combined effects of all the common and putative risk modifiers in this study only account for ~4% of the variance of BRCA2 mutations, compared with 1.1% for the single RAD51 135 G>C variant, which is rare and biologically-linked to BRCA2 function, as shown by candidate gene studies [9]. Thus, the common alleles that modify risk in BRCA1 and BRCA2 backgrounds appear to have comparable associated risks in sporadic ER-positive and ER-negative tumors, respectively [18]. While individual SNPs are unlikely to be used to guide radiographic screening and risk-reducing surgical strategies, the combined effect of these SNPs may ultimately be used for the tailoring management of subsets of BRCA1 mutation carriers [5].

While we took great efforts to collect all of the possible known BRCA2 mutation carriers, there were insufficient numbers to stratify by race and BRCA2 mutations with the exception of BRCA2*6174delT mutations. Due to the small numbers of women of non-European ancestry who have participated in the individual studies represented here, the current analysis was based only on women who had genetic backgrounds consistent with HapMap CEU samples. While we expect that SNPs identified among women of European ancestry might also be applicable to women of other genetic backgrounds, additional research in these populations will be needed. Similarly, the observed associations represented across all types of mutations, and specifically a weighted average of BRCA2*6174delT and non-delT mutations. It is possible that the observed associations may only modify the penetrance of specific BRCA2 mutations due to differential effects on function or differences in genetic background. Our analysis was stratified on the basis of the most common BRCA2 mutation, BRCA2*6174delT, which is prevalent in individuals with an Ashkenazi Jewish ancestry. Large numbers of mutation carriers will be necessary to calculate mutation-specific estimates. In addition, there was a drop-out of SNPs in the two phases of this study. While we were able to achieve a representative coverage of the genome, it is also possible that additional studies using denser arrays may provide further information.

As expected, we observed associations with some of the major common genetic variants seen in genome-wide scans of breast cancer in a non-BRCA1/2 mutation background. However, we found no evidence for loci with stronger effects than FGFR2. Although we observed an association with a novel locus at 2q37.35 that appears also to be a risk factor for sporadic breast cancer, overall, our results suggest that there are no common variants with major effects (i.e., OR>2.0) that are specific in BRCA2 carriers. Similarly, in a recent report of SNPs from sporadic breast cancer GWAS genotyped in a restricted set of BRCA1/2 carriers [27], loci in LOC134997 (rs9393597: per allele HR = 1.55, 95% CI 1.25–1.92, p = 6.0 x 10^{-5}) and FBXL7 (rs12652447: HR = 1.37, 95% CI 1.16–1.62, p = 1.7 x 10^{-4}) were associated with BRCA2 breast cancer risk with p-values weaker than FGFR2 reported here (per allele p-value = 1.2 x 10^{-8}), although the magnitudes of the associations were slightly stronger than FGFR2 (HR = 1.28). Although these SNPs were not in our genotyped panel of SNPs at stage 1, imputation results indicate that SNP rs9393597 has a p-value of 0.008 and SNP rs12652447 a p-value of 0.04 for association with breast cancer risk for the BRCA2 mutation carriers in our stage1. However, there is substantial overlap between our study and the study of Wang et al. [27].

Replication in larger datasets will be necessary to precisely estimate the magnitude of the associations of suspected loci identified from our study, candidate gene analyses [10–12], and other selection approaches [27]. It is of interest, however, that when utilizing an agnostic approach in BRCA2 mutation carriers in this study, the major determinants of risk variation in mutation carriers are those that also modify risk in subsets of sporadic, BRCA1/2 wild type, breast cancer. However, it remains possible that unique variants with smaller effects, or rarer variants (not evaluated in this experiment), may be specific modifiers of breast cancer risk in BRCA2 carriers. Their detection would require study populations much larger than the current analysis, which is presently the largest such cohort assembled.

Materials and Methods

Study Subjects

Ethics statement. All carriers were recruited to studies (Table 2) at the host institutions under IRB-approved protocols.

Selection of affected individuals and controls. A total of 6,272 BRCA2 carriers from 39 studies (Table 2) and 14 countries contributed DNA samples for this project. With the exception of NICC, all studies are members of the Consortium of Investigators of Modifiers of BRCA1/2 (CIMBA) [28]. Recruitment of carriers were conducted predominantly through cancer genetics clinics, and enrolled through national or regional efforts. Other studies were recruited through population-based or community-based ascertainment. All subjects provided written informed consent. Eligible female carriers were aged 18 years or older, were self-reported ‘white’, and had mutations in BRCA2. Data were available on age at study recruitment, age at cancer diagnosis, age of bilateral prophylactic mastectomy, BRCA1/2 mutation description, and self-reported ethnicity. Only a limited number of cases had detailed information on tumor characteristics (e.g., estrogen and progesterone receptor status); therefore, subtype analyses were not performed at this stage.

Genotyping and Quality Control

Stage 1 Affymetrix genotyping. All eligible DNA samples provided by participating centers were subjected to a rigorous quality control assessment, including measures of overall DNA quality and quantity. A total of 1,156 young (≤50 years) affected women and 1,038 unaffected women with high quality DNA samples were selected (Table 2). For time efficiency, stage 1 genotyping occurred in two phases: phase 1 included 421 cases and 404 controls and phase 2 included 735 cases and 634 controls.

Prior to the genome-wide scan, we genotyped five SNPs previously genotyped by the CIMBA study centers as a pre-filter for sample identification. Thirty-one samples (Figure S1) were discordant in the two genotyping rounds and were excluded from further analysis.

The genotyping for the stage 1 GWAS was performed on 2,163 eligible carriers using the Affymetrix 6.0 GeneChip array that included 906,622 SNPs (Figure S1). To further monitor the identity of the DNA samples, a fingerprinting panel of 14 SNPs with a minor allele frequency >10% in HapMap European individuals were genotyped on all samples, using Sequenom iPLEX, before and after Affymetrix genotyping. The AMG gender assay was used for gender assessment. As an additional quality control measure, cases and controls were interleaved on each plate to eliminate technical bias. Each plate also included one HapMap CEU DNA sample.

The DNA samples and genotyping calls for both phases of stage 1 were filtered through a series of data quality control parameters using the Birsseed module of the Birdsuite software developed at Broad Institute [29]. Among the 2,163 samples genotyped in the stage 1 GWAS, 253 failed to hybridize to the chip due to poor DNA quality and were excluded (Figure S1). Fifty-five samples were dropped with call rates <95%. Three samples were
Table 2. Description of affected and unaffected carriers selected for BRCA2 GWAS Stage 1 and 2.

<table>
<thead>
<tr>
<th>Factor</th>
<th>Stage 1</th>
<th></th>
<th>Stage 2</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Affected (n = 1,156)</td>
<td>Unaffected</td>
<td>Affected (n = 1,524)</td>
<td>Unaffected</td>
</tr>
<tr>
<td></td>
<td>N</td>
<td>%</td>
<td>N</td>
<td>%</td>
</tr>
<tr>
<td>Age at Censoring</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;40</td>
<td>763</td>
<td>66.7</td>
<td>368</td>
<td>23.7</td>
</tr>
<tr>
<td>40–44</td>
<td>308</td>
<td>26.9</td>
<td>225</td>
<td>14.5</td>
</tr>
<tr>
<td>45–49</td>
<td>72</td>
<td>6.3</td>
<td>334</td>
<td>21.5</td>
</tr>
<tr>
<td>50–54</td>
<td>1</td>
<td>0.1</td>
<td>286</td>
<td>18.4</td>
</tr>
<tr>
<td>55–59</td>
<td>0</td>
<td>0.0</td>
<td>164</td>
<td>10.5</td>
</tr>
<tr>
<td>60+</td>
<td>0</td>
<td>0.0</td>
<td>178</td>
<td>11.4</td>
</tr>
<tr>
<td>Self-reported Ethnicity</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unknown</td>
<td>125</td>
<td>10.9</td>
<td>329</td>
<td>21.2</td>
</tr>
<tr>
<td>Caucasian</td>
<td>873</td>
<td>76.3</td>
<td>1037</td>
<td>66.7</td>
</tr>
<tr>
<td>Ashkenazi Jewish</td>
<td>146</td>
<td>12.8</td>
<td>189</td>
<td>12.2</td>
</tr>
<tr>
<td>DelfT Mutation</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Carrier</td>
<td>161</td>
<td>14.1</td>
<td>233</td>
<td>15.0</td>
</tr>
<tr>
<td>Non-carrier</td>
<td>983</td>
<td>85.9</td>
<td>1322</td>
<td>85.0</td>
</tr>
<tr>
<td>Country of Study</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Australia</td>
<td>109</td>
<td>9.5</td>
<td>149</td>
<td>9.6</td>
</tr>
<tr>
<td>Canada</td>
<td>98</td>
<td>8.6</td>
<td>55</td>
<td>3.5</td>
</tr>
<tr>
<td>Denmark</td>
<td>0</td>
<td>0.0</td>
<td>43</td>
<td>2.8</td>
</tr>
<tr>
<td>France</td>
<td>52</td>
<td>4.5</td>
<td>172</td>
<td>11.1</td>
</tr>
<tr>
<td>Finland</td>
<td>27</td>
<td>2.4</td>
<td>32</td>
<td>2.1</td>
</tr>
<tr>
<td>Germany</td>
<td>68</td>
<td>5.9</td>
<td>116</td>
<td>7.5</td>
</tr>
<tr>
<td>Iceland</td>
<td>25</td>
<td>2.2</td>
<td>81</td>
<td>5.2</td>
</tr>
<tr>
<td>Israel</td>
<td>49</td>
<td>4.3</td>
<td>77</td>
<td>5.0</td>
</tr>
<tr>
<td>Italy</td>
<td>110</td>
<td>9.6</td>
<td>98</td>
<td>6.3</td>
</tr>
<tr>
<td>Spain</td>
<td>107</td>
<td>9.4</td>
<td>99</td>
<td>6.4</td>
</tr>
<tr>
<td>Sweden</td>
<td>13</td>
<td>1.1</td>
<td>11</td>
<td>0.7</td>
</tr>
<tr>
<td>The Netherlands</td>
<td>15</td>
<td>1.3</td>
<td>117</td>
<td>7.5</td>
</tr>
<tr>
<td>United Kingdom</td>
<td>181</td>
<td>15.8</td>
<td>125</td>
<td>8.0</td>
</tr>
<tr>
<td>USA</td>
<td>290</td>
<td>25.4</td>
<td>380</td>
<td>24.3</td>
</tr>
<tr>
<td>Study</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BCFR-Australia</td>
<td>19</td>
<td>1.7</td>
<td>12</td>
<td>0.8</td>
</tr>
<tr>
<td>BCFR-NCCC</td>
<td>12</td>
<td>1.0</td>
<td>5</td>
<td>0.3</td>
</tr>
<tr>
<td>BCFR-Ontario</td>
<td>29</td>
<td>2.5</td>
<td>16</td>
<td>1.0</td>
</tr>
<tr>
<td>BCFR-UT</td>
<td>18</td>
<td>1.6</td>
<td>11</td>
<td>0.7</td>
</tr>
<tr>
<td>BCFR-FCCC</td>
<td>2</td>
<td>0.2</td>
<td>14</td>
<td>0.9</td>
</tr>
<tr>
<td>BCFR-NY</td>
<td>4</td>
<td>0.3</td>
<td>26</td>
<td>1.7</td>
</tr>
<tr>
<td>BIDMC</td>
<td>10</td>
<td>0.9</td>
<td>7</td>
<td>0.5</td>
</tr>
<tr>
<td>CBCS</td>
<td>0</td>
<td>0.0</td>
<td>43</td>
<td>2.8</td>
</tr>
<tr>
<td>CGB_NCI</td>
<td>7</td>
<td>0.6</td>
<td>14</td>
<td>0.9</td>
</tr>
<tr>
<td>CNIO</td>
<td>49</td>
<td>4.3</td>
<td>40</td>
<td>2.5</td>
</tr>
<tr>
<td>COH</td>
<td>30</td>
<td>2.6</td>
<td>21</td>
<td>1.4</td>
</tr>
<tr>
<td>DFCI</td>
<td>14</td>
<td>1.2</td>
<td>10</td>
<td>0.6</td>
</tr>
<tr>
<td>DKFZ</td>
<td>7</td>
<td>0.6</td>
<td>8</td>
<td>0.5</td>
</tr>
<tr>
<td>EMBRACE</td>
<td>178</td>
<td>15.6</td>
<td>123</td>
<td>7.9</td>
</tr>
<tr>
<td>FCCC</td>
<td>14</td>
<td>1.2</td>
<td>12</td>
<td>0.8</td>
</tr>
<tr>
<td>GC-HBOC</td>
<td>61</td>
<td>5.3</td>
<td>108</td>
<td>6.9</td>
</tr>
</tbody>
</table>
contaminated, 43 were identified by genotyping to be duplicates, and 4 were male; all were dropped from analyses.

SNPs were also filtered using Birdseed and were removed if monomorphic or 10% missing (n = 38,962), genotype call rates 95% (n = 50,810), minor allele frequencies 1% (n = 104,792), departures from Hardy-Weinberg Equilibrium (p 10^{-5}; n = 1,090), differential missingness with respect to phenotype (p 10^{-3}; n = 275), and differential missingness with respect to nearby SNPs (p 10^{-10}; n = 22,065). A total of 6,212 SNPs had different missingness patterns in phase 1 compared to phase 2, and were excluded. Since we found that significant missingness correlated to SNPs mapping to longer fragments of Affymetrix 6.0 digestion products, we also removed the SNPs on fragments longer than 1000bp (n = 85,990).

With the remaining 1,805 carriers and 596,426 SNPs, an iterative process proceeded to drop all individuals with low call rates (<95%), high autosomal heterozygosity rates (false discovery rate >0.1%), and high identity by descent scores (≥0.95) and to drop all SNPs with minor allele frequencies <1% and SNP call rates <95% until the final run contained individuals above the individual and SNP filter thresholds (n = 1,747 samples and 592,566 SNPs). A more stringent HWE filter (p 10^{-7}) was then applied and 403 additional SNPs were dropped. Nine individuals with missing mutation descriptions were removed.

Finally, principal components analysis was used to identify the ethnic outliers (Figure S5). A total of 1,743 BRCA2 mutation carriers and the HapMap3 data for 210 individuals of European (CEU), Han Chinese (CHB), and Yoruba (YRI) African descent were available for multidimensional scaling using the genomic kinship matrix estimated using a set of 53,641 autosomal and uncorrelated SNPs. A cut-off of >11% was used to exclude samples with non-CEU ancestry (n = 35). Genotype-phenotype association analyses were based on 1,703 (899 young affected and 804 unaffected) BRCA2 mutation carriers and 592,163 SNPs, covering 85% of the common HapMap 3 SNPs (imputed with r^2 = 0.8 (see below), including 64% of the markers that were removed in the QC process).

Stage 2 Sequenom iPLEX genotyping. The primary SNP selection strategy was based on the results of the kinship-adjusted score test of 592,163 GWAS genotyped SNPs. From stage 1, a total of 79 top independent regions (p ≤ 1.5 × 10^{-4}) with pairwise r^2 values <0.80 were selected for genotyping in stage 2 (Figure S6). For the top 10 SNPs if available, an additional correlated SNP (pairwise r^2 values >0.8; n = 5) was selected to serve as genotyping
backup. The remaining SNPs for stage 2 were selected based on two alternate strategies. First, we added the 14 (as well as FGFR2 counted in the top 10 SNPs above) confirmed breast cancer SNPs from prior independent GWAS of sporadic breast cancer. Second, we also selected the 15 top independent regions (pairwise \( r^2 < 0.50 \)) based on the ranking of the p-values from a logistic regression analysis of 1.5 million imputed SNPs. In total for stage 2 replication phase, we selected 113 SNPs and 1,524 breast cancer carriers and 1,508 control carriers (Table 2) for genotyping using the Sequenom iPLEX platform.

Samples were excluded for call rates \( \leq 95\% \) (\( n = 476 \)), duplication in stage 2 (identity by state (IBS); \( 1:0 \); \( n = 43 \)), duplication in stage 1 and 2 (IBS; \( n = 25 \)), lack of complete phenotype data (\( n = 7 \)), and insufficient country-specific numbers (\( n = 1 \); Figure S6). A total of 100 SNPs were successfully multiplexed into three pools; the remaining 13 SNPs were not genotyped. Genotyping QC filters excluded 15 SNPs due to call rates \( \leq 90\% \) (\( n = 14 \)) and MAF\(<1\% \) (\( n = 1 \)). In summary, the final association analyses in stage 2 were based on 2,486 carriers (1,264 affected and 1,222 unaffected carriers) and 85 SNPs.

**Statistical Methods**

**Defining time at risk.** Carriers were censored at the first breast or ovarian cancer or bilateral prophylactic mastectomy, whichever occurred first. Carriers who developed any cancer were censored at time of bilateral prophylactic mastectomy if it occurred more than a year prior to the cancer diagnosis (to avoid censoring at bilateral mastectomies related to diagnosis in which rounded ages were used). The remaining carriers were censored at the age of last observation. This was defined either by the age/date at interview or age at follow-up depending on the information provided by the participating center. Carriers censored at diagnosis of breast cancer were considered cases in the analysis. Mutation carriers censored at ovarian cancer diagnosis were considered unaffected. Carriers with a censoring/last follow-up age older than age 80 were censored at age 80 because there are no reliable cancer incidence rates for BRCA1/2 carriers beyond age 80.

**Genotype–phenotype associations.** Analyses, based on 1,703 BRCA2 mutation carriers and 592,163 SNPs, were performed within a survival analysis framework. Since the mutation carriers were not selected at random with respect to their disease status, standard methods of survival (e.g., Cox regression) may lead to biased estimates of relative risk [30]. Therefore, analyses were conducted by modeling the retrospective likelihood of the observed genotypes conditional on the disease phenotypes. The associations between genotype and breast cancer risk at both stages were assessed using the 1-degree of freedom score test statistic based on this retrospective likelihood, as previously described [9,18]. All models were stratified by country of study and 6174delT (c.5946delT) mutation status, and used calendar-year- and cohort-specific breast cancer incidence rates for BRCA2 [25]. The combined stage 1 and stage 2 analyses were also stratified by stage. Parameter estimates were obtained by maximizing the retrospective likelihood. To allow for the non-independence among related mutation carriers, we used a robust variance estimation approach in order to obtain standard errors for the parameters [39,40]. Related individuals were identified through a unique family identifier.

**Copy number variant association.** We also examined the association of both high-frequency and low-frequency copy number variants (CNV) to the age of diagnosis of breast cancer as a dichotomous trait using the stage 1 data [29]. We called known, common variants (copy number polymorphisms, CNPs) with Canary [29], CNP alleles lower than 1% in frequency were removed, to maximize the number of the CNPs that were bi-allelic instead of multi-allelic. CNPs were removed that had for call rate <95%, differential missingness by genotype \( (p<10^{-3}) \), or departure from Hardy-Weinberg proportions \( (p<10^{-3}) \). Post-QC, we had 191 high-quality genotyped polymorphisms. We used PLINK to assess association using logistic regression and the same ancestry covariates of no interest as with SNPs. We similarly assessed less common CNVs discovered by Birdseye [29] for association with age at diagnosis using PLINK [41]. Finally, we also looked specifically at CNVs overlapping the BRCA2 gene itself using LOD scores and Birdseye.

**Haplotype sharing analysis.** We looked for evidence of excess sharing across the genome and the BRCA2 region. Using GERMILINE [23], shared segments of greater than 5 cm were computed based on the imputed genotype dataset among both Ashkenazi \( (n = 304) \) and non-Jewish \( (n = 1,331) \) samples compared to samples from an autism study \( (n = 808) \) (Figure S3). Examining sites across the genome every 2.5 cm (excluding telomere and centromere regions), we computed the mean of the proportion, standard deviation, and the maximum values for non-Jewish and Ashkenazi women, respectively.

**Gene Set Enrichment Analysis.** We tested whether 59 genes known to regulate or interact with BRCA2 [16] (Table S1) were enriched for associations with age of onset of breast cancer in BRCA2 mutation carriers, using a new implementation of Gene Set Enrichment Analysis (GSEA) called Meta-Analysis Gene-Set Enrichment of variaNT Associations (MAGENTA) [14]. The 59 genes were compiled using a PubMed abstract mining software, Chilibot [42], and were selected if they were related to the Fanconi anemia pathway [15] as well as others reported from literature to regulate or interact with BRCA1/2 [43]. An association p-value was calculated for each gene in the genome, defined as the most-significant association p-value of all genotyped SNPs that lie within 110 kb upstream and 40 kb downstream to the gene’s most extreme transcript boundaries, followed by correction for gene score confounders (gene size, number of SNPs per gene and linkage disequilibrium related properties). SNP association p-values were taken from the stage 1 GWAS. To compute a GSEA p-value for the BRCA gene set, the fraction of genes with an association p-value more significant than the 95 percentile of all
gene p-values in the genome was compared to a null distribution, generated by randomly sampling gene-sets of identical size from the genome 10,000 times. Of the 59 BRCA4 interactors, two genes were assigned the same most significant SNP due to physical proximity in the genome. To prevent potential over-estimation of gene set enrichment due to physical clustering of genes in a gene set, we retained only one gene of each subset of genes assigned the same best SNP (the gene with the most significant gene p-value) for the analysis of both the real and permuted gene sets.

Supporting Information

Figure S1 Data filtering of stage 1 BRCA2 GWAS.
Found at: doi:10.1371/journal.pgen.1001183.s001 (0.57 MB TIF)

Figure S2 Quantile-quantile plot comparing expected distribution of chi-square values and observed chi-square values from a genome-wide scan of breast cancer cases and unaffected BRCA2 carriers.
Found at: doi:10.1371/journal.pgen.1001183.s002 (0.31 MB TIF)

Figure S3 Manhattan plot of p-values by chromosomal position from a genome-wide scan of breast cancer cases and unaffected BRCA2 carriers [Visualized using SVS7 (Goldenhelix)].
Found at: doi:10.1371/journal.pgen.1001183.s003 (0.51 MB TIF)

Figure S4 Quantile-quantile plot comparing expected distribution of p-values and observed p-values of association of common copy number polymorphisms (CNPs) from a genome-wide scan of breast cancer cases and unaffected BRCA2 carriers.
Found at: doi:10.1371/journal.pgen.1001183.s004 (0.32 MB TIF)

Figure S5 Principal components analysis, including all eligible (after filtering) BRCA2 stage 1 samples and HapMap samples.
Found at: doi:10.1371/journal.pgen.1001183.s005 (2.39 MB TIF)

Figure S6 Data filtering of stage 2 BRCA2 GWAS.
Found at: doi:10.1371/journal.pgen.1001183.s006 (0.35 MB TIF)

Table S1 List of 59 BRCA4 interactors or regulators and their gene association p-values to breast cancer age of onset in BRCA2 mutation carriers.
Found at: doi:10.1371/journal.pgen.1001183.s007 (0.13 MB DOC)

Table S2 Ranked results for the 85 SNPs successfully genotyped in stage 2, BRCA2 GWAS.
Found at: doi:10.1371/journal.pgen.1001183.s008 (0.13 MB DOC)

Acknowledgments
HEBON, kConFab, and GEMO were listed authors. The GEMO study is represented by the named authors: Olga M. Sinilnikova, Dominique Stoppa-Lyonnet, Sylvie Mazoyer, Marion Gauthier-Villars, Hayag Sobol, Michel Longy, and Marc Frenay. We wish to thank all the GEMO interactors or regulators and their gene association p-values to breast cancer age of onset in BRCA2 mutation carriers.

References


