Six low-penetrance SNPs for the estimation of breast cancer heritability: A family-based study in Caucasian Italian patients

SIMONA DE SUMMA1*, FRANCESCA GRAZIANO2*, BRUNELLA PILATO1, ROSAMARIA PINTO1, KATIA DANZA1, ROSANNA LACALAMITA1, SIMONA SERRATÌ1, DOMENICO SAMBIASI3, MARIO GRASSI2 and STEFANIA TOMMASI4

1Molecular Genetics Laboratory, IRCCS, Istituto Tumori ‘Giovanni Paolo II’, I-70124 Bari;
2Department of Brain and Behavioral Sciences, Section of Biostatistics, NeuroPhysiology and Psychiatry, Unit of Medical Statistics and Genomics, University of Pavia, I-27100 Pavia;
3Familial Hereditary Cancer Study Unit, IRCCS, Istituto Tumori ‘Giovanni Paolo II’, I-70124 Bari, Italy

Received December 28, 2016; Accepted March 3, 2017

DOI: 10.3892/ol.2017.6725

Abstract. Breast cancer is a malignancy with a strong heritable component. Genetic counseling has been principally focused on families carrying high-penetrance breast cancer 1/2, early onset genes. Current modeling suggests that the majority of the unexplained fraction of familial risk is likely to be explained by a polygenic model. The aim of the present study was to estimate the heritability (h^2) of breast cancer susceptibility through the analysis of 6 single nucleotide polymorphisms (SNPs), which were previously identified through GWAS studies, were considered to evaluate the additive and common environmental components that contribute to the development of breast cancer in nuclear (pedigrees including only first degree relationships) and in extended families (with at most third degree relationships). A total of 22 extended pedigrees, subsequently split into 52 nuclear pedigrees were analyzed. An example of splitting process from extended to nuclear pedigree is shown in Fig. 1. Firstly, an underline latent continuous trait (Y*) using breast cancer status and information of 6 breast cancer-associated SNPs was calculated. This novel trait summarized the susceptibility of breast cancer in each individual. Secondly, the h^2 of Y* was estimated using an additive polygenic-common environment-unique error model. h^2 was evaluated in extended and immediate pedigrees, obtaining comparable results. h^2 accounts for ~40% of the total phenotypic variance, indicating a fairly strong additive genetic effect of breast cancer susceptibility. The present study indicated the importance of the evaluation and consideration of these six SNPs, which can be used as instrumental variables in order to obtain improved genetic models that are useful for h^2 analysis.

Introduction

Breast cancer (BC) is one of the most commonly diagnosed types of malignancy following non-melanoma skin cancer worldwide, as reported in GLOBOCAN 2012 (1). BC is a disease with a strong heritable component. Risk increases with the number of affected relatives, age at diagnosis and the number of affected male relatives (2-5). Genetic counseling has principally focused on families with a history of breast cancer that have a risk of carrying alterations in high-penetrance breast cancer 1, early onset (BRCA1) and breast cancer 2, early onset (BRCA2) genes, that are associated with the highest lifetime risk. However, only 25% of the familial aggregation of BC is explained by known high- and moderate-risk genes (6). For this reason, genetic testing leads to uninformative results in a number of patients. Numerous efforts have been made by consortia, including the Consortium of Investigators of Modifiers of BRCA1 and BRCA2 (7) and the Breast Cancer Association Consortium (8), to explain the remaining heritable risk through genome-wide association studies (GWAS).

However, GWAS results only collectively explain a small part of the estimated heritability. It is of note that GWAS were conducted in the general population and that patients affected by familial BC constitute a different group of risk. Sawyer et al (9) analyzed 22 single nucleotide polymorphisms (SNPs), which were previously identified through GWAS to be involved in increasing the risk of developing BC in a cohort of 1,143 index patients with BRCA1/2 alterations and

Correspondence to: Dr Stefania Tommasi, Molecular Genetics Laboratory, IRCCS, Istituto Tumori ‘Giovanni Paolo II’, Viale Orazio Flacco 65, I-70124 Bari, Italy
E-mail: s.tommasi@oncologico.bari.it

*Contributed equally

Abbreviations: BC, breast cancer; IDC, invasive ductal carcinoma; DCIS, ductal carcinoma in situ

Key words: heritability, breast cancer risk, familial breast cancer, single nucleotide polymorphisms, familial study
BRCAX indices (patients with family history of BC but with no mutation in BRCA1/2 genes). It was demonstrated that the magnitude of the effect of the SNPs was greater compared with those described through population-based studies and, in addition, the authors identified for the first time an application of GWAS results (9). However, the study conducted by Sawyer et al (9) only considered index patients, and not their relatives and the shared environmental factors.

Family-based studies provide several opportunities for the investigation and interpretation of the numerous, but at present unidentified, genetic variations underlying complex diseases. This type of study allows heritability ($h^2$) to be estimated, which is formally defined as a ratio of variances, or more specifically as the proportion of total variance in a population for a particular measurement, obtained at a particular time or age that is attributable to variation in additive genetic values (10).

In order to investigate BC $h^2$, a family-based study analyzing 6 SNPs whose association with BC risk was previously identified by GWAS and confirmed by Sawyer et al (9) was performed. The reason for studying the $h^2$ of BC is that at present, 20 risk alleles have been identified (11-14) to be associated with BC risk, however these results have not yet been integrated into clinical practice. Common low-penetrance BC susceptibility alleles have been implemented in different BC Risk Assessment Tools (15-19), however no improvement in model performance has been exhibited (20,21). Thus, the primary aim of the present study of BC in Caucasian families was to verify the opportunity to implement a procedure for $h^2$ estimation based on GWAS-identified BC alleles.

Materials and methods

Study population. A sample of 22 patients at high risk of carrying BRCA mutations were enrolled from the Counseling Program of IRCCS Istituto Tumori ‘Giovanni Paolo II’ of Bari between March 2008 and September 2011, following the specific criteria summarized in Table I. Genealogical information and medical history were registered to each participant. Genetic information and the tumor-node-metas-tasis staging (22) of breast cancer are also available for the analysis (Table II). All patients enrolled were characterized according to pathological features and family history, and were classified as having an increased risk to carry BRCA mutations (10-20%) by BRCApro software (19). On the basis of genealogical information, 22 extended and 56 nuclear pedigrees were developed for each patient (Table II). An extended pedigree is represented by each index enrolled in genetic counseling and all relatives of all degrees, each including only first-degree relatives.

Ethics approval and patient consent. Written informed consent was obtained from all participants. The present study was approved by the Ethics Committee of the IRCCS-Istituto Tumori ‘Giovanni Paolo II’ (Prot. N56/CE; 06/05/2011) to perform molecular analyses and use data for research purposes.

BRCA genes sequencing and SNP genotyping. Genomic DNA was extracted using QI Amp DNA blood midi kit (Qiagen, Inc., Valencia, CA, USA) according to the manufacturer's protocol. DNA was quantified using a ND-8000 Spectrophotometer (NanoDrop Technologies; Thermo Fisher Scientific, Inc., Wilmington, DE, USA) and mutational screening was performed (23). If a variant was identified in the index patient, the consent to inform other family members was requested. The blood of all enrolled members, healthy or affected, were analyzed for the identified variant in the index patient. The variants identified in the sequence were characterized and compared with those present in the online databases Breast Cancer Information Core (http://www.research.nhgri.nih.gov/bic/), ENTREZ SNP (http://www.ncbi.nlm.nih.gov), Ensemble Database (http://www.ensembl.org) and the Human Genome Variation database (http://hgvdbase.cgb.ki.se/).

SNP genotyping (rs3018301, rs614367, rs6504950, rs2981582, rs3803662 and rs4973768) was performed through TaqMan assays (Thermo Fisher Scientific, Inc., Waltham, MA, USA), according to the manufacturer's protocol. SNPs information are summarized in Table III.

Statistical analysis. A generalized linear model was applied to each of the 6 SNPs. Data are presented as the mean ± standard error of the mean for continuous variable and for categorical data. The association between genotyped SNPs and BC was tested using a generalized mixed linear model using R software (version 3.1.3) (24). As the traditional variance components model for dichotomous phenotypes demands a high number of cases, a two-step approach was adopted.

First step: From binary to continuous score trait. A BC susceptibility score was obtained using a liability threshold model. This model assumed that under dichotomous disease information there was a hypothetical continuous liability composed of latent genetic and environmental factors (25). In particular, for each subject the susceptibility score $Y^*$ was calculated, which summarizes disease status (affected or unaffected) and information on six disease-associated SNPs. In order to estimate $h^2$, it is mandatory to have a continuous and normal distribution. Generally, such variables depend on the presence of a number of binary response variables from which a latent variable may be calculated (25). In the present study, a binary variable (disease status) and information regarding disease-associated SNPs were available. Thus, these were used as instrumental variables to obtain a novel variable, $Y^*$, and a Gaussian continuous variable $N (0,1)$ underlying to disease status. Specifically, the disease status $Y$ coded (0=affected, 1=not affected) is linked to $Y^*$ by a threshold $\tau$: $Y = \begin{cases} 0 & \text{if } Y^* < \tau \\ 1 & \text{if } Y^* \geq \tau \end{cases}$

$Y^*$ is linked to the six SNPs additively coded ($X=0,1,2$) by $Y^* = \sum_{i=1}^{6} \beta_i X_i + U$ In other terms, $Y^*$ includes the liability of BC predicted by the SNPs of BC and unobserved terms $U$ representing the liability of BC variation unexplained by SNPs. Computations were carried out by Mplus software (version 6.0; Muthen and Muthen, Los Angeles, CA, USA) (26), and the Mplus output
scores for each subject for the latent response $Y^*$ is the outcome variable for hereditability analysis used in the second step.

Second step: $h^2$ and $c^2$ estimations. In the second step, individuals were not considered independently, but associated with each other by drawing pedigrees. A total of 22 extended families (52 nuclear families) were used to estimate the genetic component of the susceptibility of BC. $h^2$ analysis was carried out by the variance component analysis algorithm implemented in the genetic analysis software SOLAR (version 6.0; Southwest Foundation for Biomedical Research, San Antonio, TX, USA) (27), assuming an Additive polygenic-common environment-unique error (ACE) model and adjusting for sex and age (28). SOLAR was performed with extended pedigrees and nuclear pedigrees. Therefore, a variance component ($c^2$) was adopted to denote the variance among family members due to a common cluster environment, using extended pedigrees, and the variance between siblings due to a common sibling environment, using nuclear pedigrees.

Results

The demographic characteristics and BRCA1/2 mutational status of the enrolled 144 subjects (52 patients with breast cancer) are described in Table IV. Regarding the observed odds ratio of the BRCA1/2 and the six SNPs no significance was observed.

$h^2$. $h^2$ effects were estimated in extended and nuclear families (22 and 52 pedigrees, respectively) through SOLAR using the liability scores $Y^*$, that were normally distributed, as expected. The scores were interpreted only as the liability variation of BC unexplained by SNPs, as no significant SNPs results were observed (Table IV). The results from the analysis of the extended and nuclear pedigrees are summarized in Table V.

Extended pedigrees. $h^2$ estimation in extended pedigrees is demonstrated in Table V. $h^2$ without covariates was significant and equal to 40.7%. $h^2$ was increased compared with the model without covariates and also significant when sex and age are
considered \( h^2 = 46.4\% \). Finally when BRCA1 and BRCA2 mutational information are considered in the model as covariates, \( h^2 \) is 46.7\%. In the final model, only BRCA1 exhibited a significant contribution. When the household effect was also considered, \( c^2 \) was not observed to be significant (data not shown).

**Nuclear pedigrees.** \( h^2 \) remained significant in the three models using nuclear pedigrees (model with no covariate, adjusted for sex/age and BRCA1/BRCA2) but reduced compared with the extended pedigrees. In particular, \( h^2 \) without any adjustment was 35.9, 39.4 and 34% with sex and age covariates and with BRCA1/BRCA2, respectively (Table V). In the final model, only BRCA1 demonstrated a significant contribution. The effect of siblings, or sibling-household effects, \( c^2 \), such as in extended pedigrees, were not observed to be significant (data not shown).

**Discussion**

A number of studies have been focused on the genetics of BC. This malignancy is ~twice as common in the first-degree relatives of women with the disease as in the general population, consistent with the variation in genetic susceptibility to the disease (29). Inherited mutations in BRCA1/2 genes lead to a
high risk of breast and other associated cancers (30). However, the majority of families with multiple patients with BC do not carry mutations in these genes (31). These observations have led to the proposal that BC susceptibility is largely polygenic, meaning that susceptibility is conferred by a large number of loci, each with a small effect on BC risk (32). This model is consistent with the observed patterns of familial aggregation of BC.

A base for counseling may be provided by mathematical models, which predict BRCA1/2 carrier probabilities and cancer risk. A number of models have been suggested in the literature (15-21). Several of these are logistic regression models that use descriptive measures of family history and may be implemented by adding other non-genetic risk factors. An alternative approach is to base predictions on a genetic model for the disease, for example the Claus (15-16) or Gail models (17,18,21). However, genetic susceptibility to BC is more complex than these models suggest. As has been previously mentioned, BRCA1/2 genes explain just a small proportion of hereditary BC. Different studies focused on the association between BC risk and 8 common alleles, each conferring a relative BC risk <1.5 (14,32-34). It is important to highlight that the attributable risk of these alleles is relatively high (13-16% for the allele of stronger effect) as they occur at a high frequency in the general population. However, the variants may only explain a proportion of BC when they perform a causal role (33,34).

The present study focused on the estimation of $h^2$ of BC susceptibility through 6 common variants used as instrumental variables in 52 Apulian nuclear families, with a relevant family history of BC. In addition, BC data analysis was performed for the first time by applying the ACE model on the liability scores, $Y^*$, of BC. This is defined by the liability of BC as predicted by SNPs that are not statistically associated with the
disease, in addition to the unobserved terms representing the variation in liability of BC, unexplained by SNPs.

Family information and genotypes were used to estimate $h^2$ in extended and nuclear pedigrees. Comparable results were obtained in the two cases (0.407±0.19 and 0.359±0.17, respectively). This first result indicates a fairly strong additive genetic effect of BC susceptibility scores, $Y^*$. The possibility to analyze extended and nuclear pedigrees allows more precise and detailed estimations of variances to be obtained. However, no statistically significance was observed with household and sib-household effects, while the additive genetic component ($h^2$) was fairly strong and significant. It is not possible to compare the results of the present study with the literature as no other similar study regarding BC $h^2$ has been performed previously.

Antoniou et al (35) developed a model using complex segregation analysis of breast and ovarian cancer occurrence. It allowed for the simultaneous effect of BRCA1 and BRCA2 and the effect of low-penetrance genes with multiplicative effects on BC risk. The results of the present study demonstrated that the presence of deleterious mutations in BRCA1 genes have to be considered in the estimation of polygenic risk. The largest limitation of the present study was the sample size. The number of pedigrees and other SNPs correlated with BC by GWAS studies needs to be considered to improve the power of the study. It is important to highlight that $h^2$ estimation is a snapshot of a moment in a limited geographical region and is not universally applicable.

Although the cohort size needs to be enlarged, the preliminary results are encouraging. The present study indicated the importance of the evaluation and consideration of the six SNPs, used as instrumental variables, in index patients in order to obtain improved genetic models useful for $h^2$ analysis.

**Acknowledgements**

The present study was partially funded by Regione Puglia (DIEF 2011).

**References**


---

### Table V. $h^2$ and covariate (none, sex/age and BCRA1/2) fixed effect estimates in (A) extended and (B) nuclear pedigrees.

**A, Extended families**

<table>
<thead>
<tr>
<th>Variable</th>
<th>No covariates estimate (SE)</th>
<th>P-value</th>
<th>Sex/age estimate (SE)</th>
<th>P-value</th>
<th>BRCA1/BRCA2 estimate (SE)</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$h^2$</td>
<td>0.407 (0.19)</td>
<td>0.005*</td>
<td>0.464 (0.204)</td>
<td>0.003*</td>
<td>0.465 (0.201)</td>
<td>0.003*</td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td></td>
<td>0.295 (0.197)</td>
<td></td>
<td>0.681</td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td></td>
<td></td>
<td>0.003 (0.006)</td>
<td></td>
<td>0.137</td>
<td></td>
</tr>
<tr>
<td>BRCA1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.405 (0.24)</td>
<td>0.092*</td>
</tr>
<tr>
<td>BRCA2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-0.122 (0.129)</td>
<td>0.631</td>
</tr>
</tbody>
</table>

**B, Nuclear families**

<table>
<thead>
<tr>
<th>Variable</th>
<th>No covariates estimate (SE)</th>
<th>P-value</th>
<th>Sex/age estimate (SE)</th>
<th>P-value</th>
<th>BRCA1/BRCA2 estimate (SE)</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$h^2$</td>
<td>0.359 (0.17)</td>
<td>0.009*</td>
<td>0.394 (0.18)</td>
<td>0.006*</td>
<td>0.339 (0.19)</td>
<td>0.022*</td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td></td>
<td>0.291 (0.18)</td>
<td></td>
<td>0.125</td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td></td>
<td></td>
<td>-0.003 (0.006)</td>
<td></td>
<td>0.631</td>
<td></td>
</tr>
<tr>
<td>BRCA1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.386 (0.211)</td>
<td>0.067*</td>
</tr>
<tr>
<td>BRCA2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-0.173 (0.21)</td>
<td>0.422</td>
</tr>
</tbody>
</table>

*P<0.05. BRCA1, breast cancer 1, early onset; BRCA2, breast cancer 2, early onset; SE, standard error; $h^2$, heritability.


