Published in final edited form as:

Breast Cancer Res Treat. 2014 February ; 144(1): 171–177. doi:10.1007/s10549-014-2845-6.

Estimating risks for variants of unknown significance according to their predicted pathogenicity classes with application to *BRCA1*

J. G. Dowty¹, E. Lee², R. McKean-Cowdin², B. E. Henderson², L. Bernstein³, G. Ursin², and J. L. Hopper^{1,4}

¹Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, The University of Melbourne, Parkville, Victoria, Australia

²Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, California, USA

³City of Hope Comprehensive Cancer Center and Beckman Research Institute, Duarte, California, USA

⁴Department of Epidemiology and Institute of Health and Environment, School of Public Health, Seoul National University, Seoul

Abstract

Sequence-based testing of disease-susceptibility genes has identified many variants of unknown significance (VUSs) whose pathogenicity is unknown at the time of their measurement. Female breast cancer cases aged 20–49 years at diagnosis and with VUSs in *BRCA1* and no mutations in *BRCA2* have previously been identified through the population-based Los Angeles County Cancer Surveillance Program. These nominal *BRCA1* VUSs were classified as ‘low’, ‘medium’ and ‘high’ risk by four classification methods: Align-GVGD, Polyphen, Grantham matrix scores and sequence conservation in mammalian species. Average hazard ratios (HRs) for classes of variants, i.e. the age-specific incidences of cancer for carriers of such variants divided by the population incidences, were estimated from the cancer family histories of first- and second-degree relatives of the index cases using modified segregation analysis. The study sample comprised 270 index cases and 4543 of their relatives. There was weak evidence that the risk of breast cancer increases with the degree of sequence conservation ($p=0.03$) and that missense variants at highly conserved sites are associated with a 5.6-fold (95% confidence interval 1.4–22.2; $p=0.05$) increased incidence of

Corresponding author: Prof John L. Hopper, Centre for Molecular, Environmental, Genetic and Analytic Epidemiology, The University of Melbourne, Level 3/207 Bouverie St., Carlton, 3053, Victoria, Australia. j.hopper@unimelb.edu.au.

Ethics statement

The Women’s Learning the Influence of Family and Environment Study was approved by the Institutional Review Board of the University of Southern California. All participants provided written informed consent. All experiments complied with the current laws of the US where they were performed.

Disclaimer

Authors had full responsibility for the design of the study, the collection of the data, the analysis and interpretation of the data, the decision to submit the manuscript for publication, and the writing of the manuscript. The ideas and opinions expressed herein are those of the authors, and no endorsement by the State of California, Department of Health Services is intended or should be inferred.

Disclosures

The authors declare that they have no conflicts of interest.

breast cancer. An upper bound of 2.3 is given for the average breast cancer HRs corresponding to variants classified as ‘low risk’ by any of the four VUS classification methods. In summary, we have given a method to estimate cancer risks for groups of VUSs by combining existing classification methods with traditional penetrance analyses. This analysis suggests that classification methods for *BRCA1* variants based on sequence conservation might be useful in a clinical setting. We have shown in principle that our method can be used to classify VUSs into clinically useful risk categories, but our specific findings should not be put into clinical practice unless confirmed by larger studies.

Keywords

Unclassified variants; variants of unknown significance; *BRCA1*; breast cancer; ovarian cancer

INTRODUCTION

Sequence-based testing of high-risk disease-susceptibility genes in clinical and research contexts usually identifies a large number of genetic variants which are neither clearly pathogenic nor clearly benign [1,2]. Carriers of such variants of unknown significance (VUSs) cannot be given definite information about their risks of disease, and the appropriate clinical management of such carriers is debatable.

Most VUSs are missense variants which result in the change of a single amino acid in the protein product of the gene [1,2]. The occurrence of a VUS in a series of multiple-case families does not imply that the VUS confers a high risk of disease unless it is known that the VUS occurs more rarely in comparable families without family histories of the disease. Some VUSs are sufficiently common that their associated risks of disease can be estimated but most VUSs are rare, so direct risk estimates are not possible [3–5].

A number of methods exist to predict the pathogenicity of individual VUSs based on the predicted effect of the variant on the encoded protein [6,7] and/or the level of sequence conservation at the variant’s homologous sites in non-human species [8,9]. These classification methods can be used to classify individual VUSs into nominally ‘low’, ‘medium’ or ‘high’ risk categories, but the actual disease risks conferred by variants in these categories are unknown.

In this paper, we use penetrance analysis to estimate the average risks of disease associated with certain classes of VUSs, such as the set of all VUSs classified as ‘high risk’ by a given method. We suggest that, in the absence of further information about an individual VUS, these average risks can be used for the genetic counseling and clinical management of a carrier of any VUS in the class. This is in line with current clinical practice for deleterious mutations, where even though disease risks might vary from mutation to mutation, the genetic counseling and clinical management of carriers is uniform across this class of variants, and is based on average risks obtained from pooled analyses of families with different mutations. We focus on the gene *BRCA1* (OMIM 113705) and the diseases breast and ovarian cancer in this paper but our approach can easily be applied to other genes and diseases.

MATERIALS AND METHODS

Subjects

Data collection methods used in conducting the Women's Learning the Influence of Family and Environment (LIFE) Study have previously been described [10]. In brief, female patients diagnosed with histologically-confirmed first primary invasive breast cancer were identified through the Los Angeles County Cancer Surveillance Program (CSP), a population-based Surveillance, Epidemiology and End Results registry supported by the State of California and the National Cancer Institute. Eligible cases were: (1) US-born and English speaking, (2) white (including Latina) or African-American, (3) aged 20–49 years at diagnosis, and (4) Los Angeles County residents at diagnosis. A total of 2,882 eligible cases were identified (2,534 whites and 348 African-Americans) between February, 1998 and May, 2003. Recruitment of African-Americans began after the initiation of the study with eligible African-American cases diagnosed from January, 2000 to May, 2003.

Among the 2,882 potentially eligible cases, 1,794 (62%) were interviewed (1,585 white, 209 African-American). Reasons for non-participation were patient refusal (n=428), no longer a resident of Los Angeles County (n=37), not located (n=88), death (n=38), serious illness or disability (n=18), physician refusal (n=50), or inability to schedule the interview within 18 months of diagnosis (n=429). The original study was approved by the Institutional Review Board of the University of Southern California (USC). All participants provided written informed consent.

Data and blood specimen collection

An in-person interview was completed using a modified version of the structured questionnaire used in the Women's Contraceptive and Reproductive Experiences (CARE) Study [11]. The questionnaire included detailed information on demographic characteristics, detailed family history of breast or ovarian cancer, ethnic origin, and environmental factors such as oral contraceptive use, reproductive history, alcohol use, smoking history, and radiation exposure. We obtained information up to the date of breast cancer diagnosis. Blood specimens were collected from 1519 participants (85%) and transported to the USC Norris Comprehensive Cancer Center Genetics Core Laboratory in Styrofoam containers on frozen ice packs. For the first 50 samples, buffy coat was immediately extracted and stored, and for the remaining samples we stored whole blood.

Sequencing of *BRCA1* and *BRCA2* and classification of variants

The sequencing of *BRCA1* and *BRCA2* (OMIM 600185) and the classification of the resulting variants have been described previously [9]. In brief, all *BRCA1* and *BRCA2* exons (except *BRCA1* exons 1 and 4 and *BRCA2* exon 1) as well as all exon-intron boundaries were sequenced. Exon 1 was not sequenced for either gene because it is located upstream of the translation start site in both genes. *BRCA1* exon 4 was not sequenced because it is not found in the normal *BRCA1* mRNA transcript. DNA extraction, amplification and sequencing were done in the USC Genomics Core Laboratory using a protocol similar to that previously described [12]. We sequenced *BRCA1* and *BRCA2* for 1469 out of 1519 blood specimens. We were unable to sequence the remaining 50 specimens due to

insufficient DNA. Thirty-three randomly selected, blinded samples were re-sequenced for quality control purposes. The discordance rate was 0.19%: 16 discordant sequencing results out of the total 8646 variant sites sequenced (262 variant sites for each of the 33 samples). In addition, 166 subjects who had non-informative sequencing results on one or more variant sites were re-sequenced or genotyped using TaqMan assay (for I2490T, N372H, and N991D) as previously described [13].

BRCA1 and *BRCA2* variants were first classified into one of the following categories: (1) definitely disease-causing variants (DDCVs), including frameshift mutations, nonsense mutations, splice variants that were previously reported to affect splicing or located at the exon/intron boundary, and missense variants that were previously shown to be deleterious; (2) VUSs, including in-frame deletion/insertions, intronic variants that might affect splicing by creating a splice donor/acceptor site, variants next to the exon/intron boundary, and most missense variants; or (3) benign polymorphic variants, including synonymous variants, intronic variants that are unlikely to affect splicing, and a few missense mutations that have been reported to be benign. See [9] for a list of all variants identified in this study with their classification and reasons and references for such classification.

Restriction to women carrying VUSs in *BRCA1* and not carrying mutations in *BRCA1* or *BRCA2*

Of the 1469 cases whose *BRCA1* and *BRCA2* genes were sequenced, 984 carried a VUS in *BRCA1* or *BRCA2*, as described previously [9]. In the present study, we excluded cases carrying only polymorphisms in *BRCA1* (n=627), any DDCV in *BRCA1* (n=41) or *BRCA2* (n=9) and any splice variant or in-frame deletion in *BRCA1* (n=6) or *BRCA2* (n=2). We further excluded cases with non-informative genetic test results at any DDCV or VUS site in *BRCA1* (n=16) or *BRCA2* (n=13), giving a final dataset of 270 cases and their families.

Classification methods for *BRCA1* VUSs

All *BRCA1* VUSs were classified with the following methods, as described previously [9] (with the exception of A-GVGD):

1. *Sequence conservation.* *BRCA1* homologues in ten mammalian species were used to measure the degree of sequence conservation of each site of *BRCA1*. Variants at sites which are conserved in 9–10, 7–8 or 6 or less of these 10 species were, respectively, considered to be the ‘high’, ‘medium’ or ‘low’ risk variants of this method.
2. *Grantham matrix score (GMS)* [6]. Using a criterion used previously to define neutral missense variants[14], those with GMS less than 60 were considered the ‘low risk’ variants of this method. All other variants were classified as ‘high risk’.
3. *A-GVGD* [8]. The ‘high’, ‘medium’ and ‘low’ risk variants of this method were (respectively) those classified by A-GVGD as C45 or higher, C15–C35, or less than C15.

4. *Polyphen* [7]. Variants classified as ‘probably damaging’, ‘possibly damaging’, and ‘benign’ by the PolyPhen algorithm were, respectively, considered to be ‘high’, ‘medium’ or ‘low’ risk.

Statistical methods

For various groups of VUSs, the corresponding average hazard ratios (HRs) for breast and ovarian cancer, i.e. the age-specific incidence of cancer for women with VUSs in the given class divided by the appropriate population rate, were estimated by modified segregation analysis [15,16]. Models were fitted by the method of maximum likelihood with the likelihood for each pedigree given by Ott’s form¹⁶ in which the penetrance function was based on a parametric survival model. More specifically, the (unconditional) likelihood of each pedigree was assumed to have the following form

$$L(\theta) = \sum_g P(g) \prod_i P(x_i | g_i)$$

where: θ is the vector of unknown log HRs to be estimated; g is a vector specifying genotypes for all family members; the sum is over all genotypes consistent with the index case’s genotype; $P(g)$ is the joint probability of the family’s genotypes g ; the product is over all individuals i in the family; x_i is a vector specifying all phenotypic data for person i , consisting of his or her affected status and an age t_i equal to the age of diagnosis (if affected) or last known age (if unaffected); g_i is the genotype of person i according to the vector g ; either $P(x_i | g_i) = S(t_i | g_i)$ if individual i is unaffected or $P(x_i | g_i) = \lambda(t_i | g_i) S(t_i | g_i)$ if individual i is affected; $S(t_i | g_i)$ is the survival function for genotype g_i evaluated at age t_i years, equal to $\exp(-\int \lambda(t | g_i) dt)$ where the terminals of integration are 0 and t_i ; $\lambda(t_i | g_i)$ is the hazard evaluated at age t_i years, equal to $\lambda(t_i | g_i) = \text{HR}(g_i) \lambda(t_i)$; and $\lambda(t_i)$ is the age-standardized and age-, sex- and race-specific population incidence rate.

By design, each index case was affected and carried a VUS, so in order to avoid ascertainment bias due to this study design, the likelihood of each family was conditioned on the index case’s *BRCA1* VUS genotype, the fact that she was a breast cancer case and her age at onset [17,18]. The underlying genetic model used by the segregation analysis was a mixed model which incorporated a polygenic background, implemented via a hypergeometric polygenic model [19,15] with 4 loci and a standard deviation of 1.29 (as estimated previously [20]), in addition to the *BRCA1* major gene locus. HRs were assumed to be constant with age and each individual within each family was assumed to either have a polymorphism at the *BRCA1* locus (with $\text{HR}=1$) or an identical-by-descent copy of the index case’s VUS. Population breast and ovarian cancer incidence rates were sex-, age- and race-specific (race being white/black) for the US population 1989–93 and were taken from Tables IV-2 and XX-2 of *SEER Cancer Statistics Review 1973–1993* [21].

In the main analyses, average HRs were estimated for all variants placed into particular risk groups by particular classification methods. More detailed analyses were performed based on the raw (non-dichotomized) GMS, the number of mammalian species (out of the ten considered) for which the site of the variant was conserved and the number of classification

methods (out of four) which classified the given VUS as ‘high risk’. Tests for trend between cancer risks and these variables assumed a linear dependence between the log HR and the given variable, with the log HR fixed to 0 in these analyses for (respectively) VUSs with GMS equal to 0, VUSs at sites conserved in 7 species or less and VUSs classified as ‘high risk’ by 0 models.

Affected subjects with missing ages at diagnosis (comprising 10% of breast cancer diagnoses among relatives of the index cases and 11% of ovarian cancer diagnoses) were included in the analysis by marginalizing over the missing ages and maximizing the resulting likelihood. Unaffected subjects with missing ages (comprising 15% of all unaffected females) were censored at birth, effectively removing them from the analysis, since their affected statuses were considered to be unreliable.

Standard errors were obtained by inverting the observed Fisher information matrix while p-values were calculated from likelihood ratio tests. Segregation analyses were performed with Mendel [22] version 3.2 and other analyses were performed with R [23] version 2.15.0.

RESULTS

Among 1469 index cases who were sequenced for *BRCA1* and *BRCA2*, 270 had a *BRCA1* VUS and did not have a *BRCA1* or *BRCA2* mutation; 29 (11%) of these cases were black and 241 (89%) were white (including Latina white). The first- and second-degree relatives of these index cases comprised 2258 (50%) females and 2285 (50%) males. Of the female relatives, 164 (7%) had been diagnosed with first primary breast cancer and 23 (1%) with ovarian cancer. Table 1 gives the number of breast and ovarian cancer diagnoses in the female relatives of the index cases, broken down by the risk class of the index case’s VUS according to the different classification methods.

For each VUS, we counted the number of methods which classified the variant as ‘high risk’, giving a whole number between 0 and 4 which summarizes the classifications of all methods. There was no evidence that the risk of cancer increased with this number for either breast ($p=0.09$) or ovarian ($p=0.4$) cancer.

For each classification method, Table 2 gives the average HR for all VUSs classified as ‘high’, ‘medium’ or ‘low’ risk by the given method. There was weak evidence that missense variants at highly conserved sites are associated with an increased risk of breast cancer ($p=0.05$), with breast cancer incidences for carriers of such VUSs estimated to be 5.6 (95% CI: 1.4–22.2) times the population rates. Otherwise, our data was consistent with the average HR for each class of VUS being 1 for breast cancer (all $p \geq 0.2$) and ovarian cancer (all $p \geq 0.3$). The 95% confidence intervals for these HR estimates show that the true incidences of breast cancer for VUSs in the ‘low risk’ category of any classification method are no larger than 2.3 times the population incidences, on average. The corresponding ovarian cancer HR estimates are far less precise than those for breast cancer so the corresponding confidence intervals do not give very useful upper bounds on the true HR.

Lastly, we investigated whether risk depended on the degree of sequence conservation at the site of the VUS or on the VUS’s Grantham matrix score. No evidence was observed that

breast or ovarian cancer risks varied with the Grantham matrix score, with each 50 point increase in the Grantham matrix score associated with an estimated increase in cancer incidence by a factor of 1.4 (95% CI 0.89–2.1; $p=0.2$) for breast and 1.3 (95% CI 0.60–3.0; $p=0.6$) for ovarian cancer. There was also no evidence that ovarian cancer risk depended on the VUS's level of sequence conservation ($p=0.5$) but the risk of breast cancer increased with the amount of sequence conservation ($p=0.03$). Table 3 gives the HRs for VUS with various levels of sequence conservation.

DISCUSSION

We have shown that it is possible to estimate cancer risks for groups of VUSs by combining existing classification methods with traditional penetrance analyses. The estimated HRs give the average risks for all variants placed into a given risk class by a given classification method, just as estimated HRs for deleterious mutations are the average across the class of all such variants. We suggest that, in the absence of information about an individual VUS other than its nominal predictive risk class, these average risks can be used for the genetic counseling and clinical management of a carrier of any VUS in the class. They constitute the best evidence available, just as an estimate of average risk for the group of protein-truncating mutations is the best evidence available for counseling those carriers, even though their specific mutation might never have been previously observed.

Application of this method to *BRCA1* and breast and ovarian cancer risks gave weak evidence that missense variants at highly conserved sites are associated with an increased risk of breast cancer ($p=0.05$). It also showed that the risk of breast cancer increases with the degree of sequence conservation ($p=0.03$). This analysis also suggested that variants in the 'high risk' category of A-GVGD have risks close to those of mutations, although the confidence interval for this estimate was too wide to draw definite conclusions, so these results should not be put into clinical practice unless confirmed by larger studies.

We also showed that female carriers of variants classified as 'low risk' by a given classification method will, on average, have incidences of breast cancer at most 2.3 times the population incidences (and possibly no higher than for the population). This upper bound provides some certainty and reassurance to the vast majority of female VUS carriers whose VUSs are classified as 'low risk' by a given classification method.

Our study is similar to that of Akbari *et al.* [24], who used Cox proportional hazards to estimate cancer risks for the first degree relatives of ovarian cancer cases (unselected for a family history of cancer) by comparing the relatives of VUS carriers to those of polymorphism carriers. Their method therefore estimates the relative risk associated with having a VUS-carrying first-degree relative, while our method estimates the risk associated with personally carrying a VUS. Since on average half of the first-degree relatives of VUS carriers will not carry a VUS, the HRs obtained by Akbari *et al.* should be approximately equal to $(1 + \theta)/2$, where θ is the HR associated with personally carrying the VUS, as estimated in this paper. Therefore their estimate of θ for breast, ovarian and fallopian tube cancers (combined) for 'high risk' A-GVGD VUSs is about $2 \times 5.24 - 1 = 9.5$. Encouragingly, this is very similar to our estimate of 8.3 (0.83–83.0) for breast cancer

(which should dominate their combined class of cancers, since it is by far the most common cancer), though the wide confidence intervals on our estimate and theirs means that this near agreement could be due to chance. Relative strengths of our method over that of Akbari *et al.* include our direct estimation of the main quantity of interest, θ , our ability to rigorously adjust for complex ascertainment schemes and our ability to reduce bias in penetrance estimates [18] by incorporating residual familial aggregation of cancer.

Our estimates quantify the effect of the VUS on top of an unmeasured polygenic background of breast cancer risk. Including a polygenic background was necessary because simulation has shown that failure to use a mixed model or to otherwise account for residual familial aggregation of risk can bias penetrance estimates [18]. Since the polygenic background can be thought of as capturing a woman's family history of cancer, her incidence of breast cancer should be calculated as the relevant HR for her VUS multiplied by both the population incidence and a factor reflecting her family history, e.g. a factor of approximately 2 if she has an affected first degree relative.

Strengths of our study include the population-based study design and the powerful and rigorous statistical methods used. Weaknesses include the fact that, for each VUS classification method, relatively few families had VUSs that were classed as 'high risk', leading to imprecision in the resulting HR estimates. This problem was exacerbated by the relative rarity of ovarian cancer, giving us little statistical power to detect increased risks for this disease.

In summary, we have given a method to estimate cancer risks for groups of VUSs by combining existing classification methods with traditional penetrance analyses. This analysis suggests that classification methods for *BRCA1* variants based on sequence conservation might be useful in a clinical setting and that, on average, variants classed as 'low risk' increase breast cancer risks by at most a moderate amount. We have shown in principle that our method can be used to classify VUSs into clinically useful risk categories, but our specific findings should not be put into clinical practice unless confirmed by larger studies.

Acknowledgments

The authors would like to thank Sean V. Tavtigian for his insightful comments and for his help with Align-GVGD. The authors are also very grateful to the women who participated in this study, the interviewers who collected the data, the phlebotomists who collected the blood samples and Ms Juliana Bamrick who managed all study activities. This study was supported by grants CA17054 and CA74847 from the National Cancer Institute, National Institutes of Health, by 4PB-0092 from the California Breast Cancer Research Program of the University of California, and in part through contract number N01-PC-35139 and T32 ES-013678 from the National Institute of Environmental Health Sciences, National Institute of Health. The collection of cancer incidence data used in this publication was supported by the California Department of Health Services as part of the statewide cancer reporting program mandated by California Health and Safety Code Section 103885.

References

1. Woods MO, Williams P, Careen A, Edwards L, Bartlett S, McLaughlin JR, Younghusband HB. A new variant database for mismatch repair genes associated with Lynch syndrome. *Hum Mutat.* 2007; 28 (7):669–673.10.1002/humu.20502 [PubMed: 17347989]

2. Szabo C, Masiello A, Ryan JF, Brody LC. The breast cancer information core: database design, structure, and scope. *Hum Mutat.* 2000; 16 (2):123–131.10.1002/1098-1004(200008)16:2<123::AID-HUMU4>3.0.CO;2-Y [PubMed: 10923033]
3. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS, Sunyaev SR. A method and server for predicting damaging missense mutations. *Nature methods.* 2010; 7 (4):248–249.10.1038/nmeth0410-248 [PubMed: 20354512]
4. Arnold S, Buchanan DD, Barker M, Jaskowski L, Walsh MD, Birney G, Woods MO, Hopper JL, Jenkins MA, Brown MA, Tavtigian SV, Goldgar DE, Young JP, Spurdle AB. Classifying MLH1 and MSH2 variants using bioinformatic prediction, splicing assays, segregation, and tumor characteristics. *Hum Mutat.* 2009; 30 (5):757–770.10.1002/humu.20936 [PubMed: 19267393]
5. Goldgar DE, Healey S, Dowty JG, Da Silva L, Chen X, Spurdle AB, Terry MB, Daly MJ, Buys SM, Southey MC, Andrulis I, John EM, Khanna KK, Hopper JL, Oefner PJ, Lakhani S, Chenevix-Trench G. Rare variants in the ATM gene and risk of breast cancer. *Breast Cancer Res.* 2011; 13 (4):R73. bcr2919 [pii]. 10.1186/bcr2919 [PubMed: 21787400]
6. Grantham R. Amino acid difference formula to help explain protein evolution. *Science.* 1974; 185 (4154):862–864. [PubMed: 4843792]
7. Ramensky V, Bork P, Sunyaev S. Human non-synonymous SNPs: server and survey. *Nucleic Acids Res.* 2002; 30 (17):3894–3900. [PubMed: 12202775]
8. Tavtigian SV, Deffenbaugh AM, Yin L, Judkins T, Scholl T, Samollow PB, de Silva D, Zharkikh A, Thomas A. Comprehensive statistical study of 452 BRCA1 missense substitutions with classification of eight recurrent substitutions as neutral. *J Med Genet.* 2006; 43 (4):295–305. jmg.2005.033878 [pii]. 10.1136/jmg.2005.033878 [PubMed: 16014699]
9. Lee E, McKean-Cowdin R, Ma H, Chen Z, Van Den Berg D, Henderson BE, Bernstein L, Ursin G. Evaluation of unclassified variants in the breast cancer susceptibility genes BRCA1 and BRCA2 using five methods: results from a population-based study of young breast cancer patients. *Breast Cancer Res.* 2008; 10 (1):R19. bcr1865 [pii]. 10.1186/bcr1865 [PubMed: 18284688]
10. Ma H, Bernstein L, Ross RK, Ursin G. Hormone-related risk factors for breast cancer in women under age 50 years by estrogen and progesterone receptor status: results from a case-control and a case-case comparison. *Breast Cancer Res.* 2006; 8 (4):R39. bcr1514 [pii]. 10.1186/bcr1514 [PubMed: 16846528]
11. Marchbanks PA, McDonald JA, Wilson HG, Folger SG, Mandel MG, Daling JR, Bernstein L, Malone KE, Ursin G, Strom BL, Norman SA, Wingo PA, Burkman RT, Berlin JA, Simon MS, Spirtas R, Weiss LK. Oral contraceptives and the risk of breast cancer. *N Engl J Med.* 2002; 346(26):2025–2032. [pii]. 10.1056/NEJMoa013202346/26/2025 [PubMed: 12087137]
12. McKean-Cowdin R, Spencer Feigelson H, Xia LY, Pearce CL, Thomas DC, Stram DO, Henderson BE. BRCA1 variants in a family study of African-American and Latina women. *Hum Genet.* 2005; 116 (6):497–506.10.1007/s00439-004-1240-5 [PubMed: 15726418]
13. Freedman ML, Penney KL, Stram DO, Le Marchand L, Hirschhorn JN, Kolonel LN, Altshuler D, Henderson BE, Haiman CA. Common variation in BRCA2 and breast cancer risk: a haplotype-based analysis in the Multiethnic Cohort. *Hum Mol Genet.* 2004; 13(20):2431–2441. ddh270 [pii]. 10.1093/hmg/ddh270 [PubMed: 15317758]
14. Abkevich V, Zharkikh A, Deffenbaugh AM, Frank D, Chen Y, Shattuck D, Skolnick MH, Gutin A, Tavtigian SV. Analysis of missense variation in human BRCA1 in the context of interspecific sequence variation. *J Med Genet.* 2004; 41 (7):492–507. [PubMed: 15235020]
15. Antoniou AC, Pharoah PD, McMullan G, Day NE, Ponder BA, Easton D. Evidence for further breast cancer susceptibility genes in addition to BRCA1 and BRCA2 in a population-based study. *Genet Epidemiol.* 2001; 21 (1):1–18.10.1002/gepi.1014 [PubMed: 11443730]
16. Lange, K. *Statistics for biology and health.* 2. Springer; New York: 2002. Mathematical and statistical methods for genetic analysis.
17. Kraft P, Thomas DC. Bias and efficiency in family-based gene-characterization studies: conditional, prospective, retrospective, and joint likelihoods. *Am J Hum Genet.* 2000; 66 (3): 1119–1131. S0002-9297(07)64032-4 [pii]. 10.1086/302808 [PubMed: 10712222]
18. Gong G, Hannon N, Whittemore AS. Estimating gene penetrance from family data. *Genet Epidemiol.* 2010; 34 (4):373–381.10.1002/gepi.20493 [PubMed: 20397150]

19. Cannings C, Thompson E, Skolnick M. Probability functions on complex pedigrees. *Adv Appl Prob.* 1978; 10 (1):26–61.
20. Antoniou AC, Pharoah PD, McMullan G, Day NE, Stratton MR, Peto J, Ponder BJ, Easton DF. A comprehensive model for familial breast cancer incorporating BRCA1, BRCA2 and other genes. *Br J Cancer.* 2002; 86 (1):76–83.10.1038/sj.bjc.6600008 [PubMed: 11857015]
21. Ries, LAGKC.; Hankey, BF.; Harras, A.; Miller, BA.; Edwards, BK. SEER Cancer Statistics Review, 1973–1993: Tables and Graphs. National Cancer Institute; Bethesda, Md: 1996.
22. Lange K, Weeks D, Boehnke M. Programs for Pedigree Analysis: MENDEL, FISHER, and dGENE. *Genet Epidemiol.* 1988; 5 (6):471–472.10.1002/gepi.1370050611 [PubMed: 3061869]
23. Team RDC. R: A language and environment for statistical computing. R Foundation for Statistical Computing; Vienna: 2010.
24. Akbari MR, Zhang S, Fan I, Royer R, Li S, Risch H, McLaughlin J, Rosen B, Sun P, Narod SA. Clinical impact of unclassified variants of the BRCA1 and BRCA2 genes. *J Med Genet.* 2011; 48 (11):783–786.10.1136/jmedgenet-2011-100305 [PubMed: 21965345]

Table 1

The number of breast and ovarian cancer diagnoses in female relatives of index cases by the class of the index case's *BRCA1* VUS according to each classification method, with row percentages in parentheses.

Classification method	Predicted risk class	Breast cancer	Ovarian cancer	Total number of female relatives
Sequence conservation	low	98 (7%)	13 (1%)	1425
	medium	50 (7%)	8 (1%)	684
	high	16 (11%)	2 (1%)	149
Grantham matrix score	low	143 (7%)	19 (1%)	1987
	high	21 (8%)	4 (1%)	271
A-GVGD	low	157 (7%)	21 (1%)	2149
	medium	2 (3%)	1 (1%)	70
	high	5 (13%)	1 (3%)	39
PolyPhen	low	146 (7%)	18 (1%)	1987
	medium	4 (3%)	2 (2%)	128
	high	14 (10%)	3 (2%)	143

Table 2

Average hazard ratios for females carrying *BRCA1* VUS in each risk class according to each classification method.

Classification method	Predicted risk class	Breast cancer		Ovarian cancer		Number of families
		HR (95% CI)	p	HR (95% CI)	p	
Sequence conservation	low	1.2 (0.62–2.3)	0.6	0.63 (0.06–6.4)	0.6	171
	medium	1.2 (0.44–3.1)	0.8	1.5 (0.33–7.3)	0.6	83
	high	5.6 (1.4–22.2)	0.05	1.4 (0.02–116)	0.9	16
Grantham matrix score	low	1.2 (0.71–2.1)	0.5	0.68 (0.10–4.4)	0.6	239
	high	1.9 (0.53–6.6)	0.4	2.9 (0.55–15.1)	0.3	31
	low	1.3 (0.81–2.2)	0.3	0.80 (0.17–3.7)	0.8	260
A-GVGD	medium	NE	NE	4.9 (0.35–69.6)	0.4	7
	high	8.3 (0.83–83.0)	0.2	3.4 (0.07–164)	0.6	3
	low	1.3 (0.79–2.3)	0.3	0.55 (0.06–5.0)	0.5	239
PolyPhen	medium	NE	NE	5.0 (0.62–40.8)	0.3	15
	high	3.2 (0.76–13.8)	0.2	3.8 (0.56–25.6)	0.3	16

Abbreviations: HR, hazard ratio; CI, confidence interval; NE, not estimable due to low numbers.

Table 3

Hazard ratios for female carriers of *BRCA1* VUSs by degree of sequence conservation.

Number of mammalian species (out of 10) for which the site of the variant is conserved	Number of families	HR (95% CI) for	
		Breast cancer	Ovarian cancer
0–7	220	1.1 (0.58–1.9)	0.82 (0.16–4.2)
8	34	2.3 (0.70–7.3)	1.9 (0.21–16.2)
9	11	5.7 (1.1–30.8)	NE
10	5	5.4 (0.51–57.8)	10.6 (1.0–107)

Abbreviations: HR, hazard ratio; CI, confidence interval; NE, not estimable due to low numbers.