

A Computational Method to Classify Variants of Uncertain Significance Using Functional Assay Data with Application to *BRCA1*

Edwin S. Iversen, Jr.¹, Fergus J. Couch², David E. Goldgar³, Sean V. Tavtigian⁴, and Alvaro N.A. Monteiro⁵

Abstract

Background: Besides revealing cancer predisposition variants or the absence of any changes, genetic testing for cancer predisposition genes can also identify variants of uncertain clinical significance (VUS). Classifying VUSs is a pressing problem, as ever more patients seek genetic testing for disease syndromes and receive noninformative results from those tests. In cases such as the breast and ovarian cancer syndrome in which prophylactic options can be severe and life changing, having information on the disease relevance of the VUS that a patient harbors can be critical.

Methods: We describe a computational approach for inferring the disease relevance of VUSs in disease genes from data derived from an *in vitro* functional assay. It is based on a Bayesian hierarchical model that accounts for sources of experimental heterogeneity.

Results: The functional data correlate well with the pathogenicity of *BRCA1* BRCT VUSs, thus providing evidence regarding pathogenicity when family and genetic data are absent or uninformative.

Conclusions: We show the utility of the model by using it to classify 76 VUSs located in the BRCT region of *BRCA1*. The approach is both sensitive and specific when evaluated on variants previously classified using independent sources of data. Although the functional data are very informative, they will need to be combined with other forms of data to meet the more stringent requirements of clinical application.

Impact: Our work will lead to improved classification of VUSs and will aid in the clinical decision making of their carriers. *Cancer Epidemiol Biomarkers Prev*; 20(6); 1078–88. ©2011 AACR.

Introduction

Variants of uncertain significance (VUS) are genetic variants, often missense mutations, of unknown effect on protein function and whose disease relevance is undefined. Because genetic testing for disease syndromes becomes increasingly common in clinical practice, the number of VUSs will continue to rise and the need for methods allowing determination of the disease relevance of the VUSs will increase. In cases such as the breast ovarian cancer syndrome, that is associated with inactivating mutations in the *BRCA1* and *BRCA2* genes, many individuals have been found to carry VUSs. In particular, approximately 10% of variants found in Caucasians, up to 35% of variants found in Hispanics, and up to 50% of

variants found in African Americans have this designation (1–3). Given that more rigorous screening for cancer and/or prophylactic surgery is useful in lowering morbidity and mortality in women who carry a high-risk pathogenic mutation in these genes (4), a relatively large number of women can benefit from the classification (determination of the disease relevance) of VUSs in these genes as neutral/low clinical significance or pathogenic.

Various classification schemes have been proposed to address the important clinical problem of defining which VUSs are pathogenic and which are neutral. In particular, a modified linkage approach based on cosegregation of mutations with disease in families that calculates the odds or likelihood of causality has been utilized for classification of *BRCA1* and *BRCA2* VUSs (5, 6). The likelihood is evaluated under the hypothesis that the variant under consideration has the same penetrance as the "average" known pathogenic mutation, compared with the hypothesis that the variant segregates independently of disease in the pedigree(s) under study. Thus, by collecting pedigrees for families with specific *BRCA1* or *BRCA2* VUSs and by obtaining genotyping data on as many individuals as possible in these pedigrees, it is possible to estimate the odds in favor of particular VUSs being disease-associated or disease-neutral. However, classification of VUSs is difficult because the statistical

Authors Affiliations: ¹Department of Statistical Science, Duke University, Durham, North Carolina; ²Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, Minnesota; Departments of ³Dermatology and ⁴Oncological Sciences, University of Utah School of Medicine, Salt Lake City, Utah; and ⁵Risk Assessment, Detection, and Intervention Program, H. Lee Moffitt Cancer Center & Research Institute, Tampa, Florida.

Corresponding Author: Edwin Iversen, Duke University, 122B Old Chemistry Building, Durham, NC 27708. E-mail: iversen@STAT.duke.edu

doi: 10.1158/1055-9965.EPI-10-1214

©2011 American Association for Cancer Research.

power of classification methods that depend on segregation analysis and other family-based genetic methods is typically limited because of the rarity (<1:10,000) of individual VUSs. As a result, it is necessary to incorporate other sources of data, such as protein function, that are related to the likelihood of a VUS being pathogenic and to employ models and methods that simultaneously classify many variants, gaining strength from common patterns in the data.

To this end, we describe a method for classifying a set of VUSs that have undergone evaluation by a particular functional assay. We focus on data from an *in vitro* assay designed to determine whether a *BRCA1* variant affects the transcriptional activity of the C-terminal BRCT domains of *BRCA1*. The assay provides a direct measure of structural integrity and an indirect measure of function (7, 8).

Materials and Methods

Terminology

We define VUSs as genetic variants of unknown effect on protein function whose disease relevance is undefined. As we note above, uncertainty in the disposition of a VUS may be addressed by several forms of data. Our focus here is on functional assays designed to determine whether a variant is damaging or benign to protein function. In contrast, *in silico* results, currently based largely on alignments and amino acid chemistry, address whether a variant is (evolutionarily) deleterious or benign; and family-based genetic methods address whether a variant is disease-associated or disease-neutral. When multiple forms of data are combined to address the disease relevance of a variant, we adopt the terms pathogenic and neutral to describe the variant's relationship with disease.

The terms VUS and unclassified variant (UV) are often used interchangeably in the literature. This can be a source of confusion as the former suggests a variant whose disease relevance is uncertain, whereas the latter suggests one that has not yet been studied and, hence, has not been classified. The confusion arises because classification schemes such as that advocated for in the work of Plon and colleagues (9) may include an uncertain/indeterminate category. Hence, a variant may simultaneously be classified and of uncertain significance. In this article, we refer to all variants that meet the definition above *prior to their evaluation for pathogenicity* as VUSs regardless of whether such an evaluation has been made or not. This is in keeping with the probabilistic framework we take in this article: *A posteriori*, there will always remain uncertainty, perhaps vanishingly small, about the true status of such a variant.

The functional assay

The transcriptional assay is based on the ability of *BRCA1* constructs fused to a heterologous DNA-binding domain to activate transcription of a reporter gene in

yeast and mammalian cells (10). Constructs containing variants that decrease or abrogate transcriptional activation are likely to impact protein function (and be damaging), whereas variants with activity similar to the wild-type construct are not likely to impact protein function and to be benign (11 and references therein). Further details related to the conceptual basis of the assay can be found in the work of Carvalho and colleagues (8).

Although the assay has been implemented in both yeast and mammalian cell lines, we focus here on results from the mammalian assay. The raw data that result from the assay are in arbitrary luciferase units. This is a ratio of firefly luciferase counts to *Renilla* luciferase counts. The firefly luciferase counts serve as a reporter of the activity of the *BRCA1* variant in question, whereas the *Renilla* luciferase counts serve as a constitutive reporter used to control for efficiency of transfection in individual samples.

These data are typically obtained in triplicate and are relatively well behaved (SD < 15%). However, they are variable from experiment to experiment and can be different by an order of magnitude. Sources of heterogeneity include the number of cells, the efficiency of transfection, the transfection ratio of the effector and reporter plasmids, the buffer volume in which the cells were lysed, the volume of sample that was used to make the reading, and the luciferase detection reagent.

Details of the *BRCA1* BRCT domain data set

The data set we analyze here comprises 893 individual ratio measurements from 30 individual experiments conducted over a period of 7 years. Each experiment measures the functional activity of a batch of 2 to 26 variants, 1 of which is always the wild-type (WT) control. Each variant is typically replicated 3 or 4 times within an experiment. Variants other than the WT control may be included in multiple experimental batches: About 25% of the variants appear in only 1 batch, about 66% appear in 2 to 5 batches, and the remaining appear in 6 to 23 batches.

In total, 76 VUSs were assessed for functional effects. In addition, 4 protein-truncating, pathogenic variants were included as "negative controls" and the *BRCA1* WT was included as the "positive control". Of the 76 VUSs, 9 were previously classified neutral and 6 pathogenic using a classification model not involving the functional data analyzed here. Table 1 provides evidence of pathogenicity and references for these variants. The 6 variants classified as pathogenic were estimated to be so with a posterior probability of at least 0.99 using a model incorporating family-based likelihood estimates and probabilities of causality estimated on the basis of a sequence conservation analysis. The variants previously classified as neutral had estimated posterior probabilities of being pathogenic of 0.0001 or less. The 15 previously classified variants were included in the analysis *blinded to their prior classification* for purposes of model evaluation. A complete list of the variants analyzed can be found in Table 2.

Two distinct experimental "contexts" were employed in the experiments represented here. Contexts are defined

Table 1. Summary of evidence excluding the functional data in favor of or against the pathogenicity of each of the 15 VUSs included in the analysis ‘blinded’ for purposes of model evaluation

Variant	Prior		Likelihood ratio		Posterior probability
	Probability	Reference	Combined LR	Reference	
G1706E	0.81	Tavtigian (26)	589	Easton (27)	0.9996
G1738R	0.81	Tavtigian (26)	115	Easton (27)	0.998
L1764P	0.29	Tavtigian (26)	347	Easton (27)	0.993
M1775R	0.66	Tavtigian (26)	229 ^a	Miki (28)	0.998
T1685I	0.81	Tavtigian (26)	138	Easton (27)	0.998
V1688del	0.35	Easton (27)	269	Easton (27)	0.993
F1662S	0.01	Tavtigian (26)	5.13e-03	Easton (27)	0.00005
P1614S	0.01	Tavtigian (26)	2.30e-09	Easton (27)	<0.00001
P1859R	0.01	Tavtigian (26)	3e-05	Easton (27)	<0.00001
R1751Q	0.01	Tavtigian (26)	6.76e-03	Easton (27)	0.00007
S1512I	0.01	Tavtigian (26)	<1e-10	Tavtigian (29)	<0.00001
S1613G	0.01	Tavtigian (26)	<1e-10	Tavtigian (29)	<0.00001
T1720A	0.01	Tavtigian (26)	<1e-10	Easton (27)	<0.00001
V1534M	0.01	Tavtigian (26)	4.8e-03	Chenevix-Trench (30)	0.00005
V1804D	0.01	Tavtigian (26)	1.2e-03	Easton (27)	0.00001

NOTE: Columns 2 and 3 provide the sequence conservation-based prior probability and literature source; columns 4 and 5 provide the family and genetic data-based likelihood ratios and literature source; and column 5 provides the resulting estimated posterior probability of pathogenicity.

^aCalculated using the LOD score of kindred 2099, the pedigree in which M1775R was first observed and found to segregate.

in terms of the range of amino acid (AA) residues or, alternately, exons in the gene that form the basis for the experiment. Five of the earlier experiments were carried out in the context of exons 16 to 24 (AAs 1,560 to 1,863), whereas the remaining 25 were done in the context of exons 13 to 24 (AAs 1,396 to 1,863). In general, the same variant will behave in very much the same way (as a percentage of WT function as measured in the same context) when done in both contexts (data not shown). The absolute activities, however, are different. All but 4 (M1775R, R1699W, Y1853X, and WT) of the variants were studied in one context or the other but not both.

The classification model

The model we propose for classifying VUSs is a 2-component mixture model in which 1 component is a model for the distribution of the data—here the log ratio obtained from the functional assay—for damaging variants and the other is a model for the distribution of the data for benign variants. Each variant, v , is accorded a binary variable, D_v , indicating whether it is damaging ($D_v = 1$) or not ($D_v = 0$). This variable is treated as missing for the VUSs and the focus of the analysis is estimation of $\Pr(D_v = 1 | \text{Data})$ for these variants. For purposes of the present analysis, we assume that VUSs fall into 2 distinct risk classes. It may be that some variants predispose to intermediate levels of risk and it is conceptually simple to extend the model to allow for categorical or continuous versions of D_v .

Let c index AA context with $c = 1$ denoting AAs 1,560 to 1,863 and $c = 2$ denoting AAs 1,396 to 1,863; $b = 1, \dots, 30$ index experiments (i.e., batches); $v = 1, \dots, 81$ index variants; and $r_v = 1, \dots, R_v$ index replicates of variant v . Despite the fact that batch and context are aliased, with batches 1 through 5 corresponding to $c = 1$ and batches 6 through 30 to $c = 2$, we find it convenient to retain both indices. If we denote by f_{vrbc} the log of the ratio of luciferase counts r_{vrbc} for replicate r of variant v from batch b using context c and by θ , the model's parameters, the model can be written as follows:

$$\prod_v \Pr(f_v | D_v, X_v, \theta) = \prod_{\{v: D_v=0\}} \Pr(f_v | D_v = 0, X_v, \theta) \prod_{\{v: D_v=1\}} \Pr(f_v | D_v = 1, X_v, \theta) \prod_{\{v: D_v \text{ is Unknown}\}} [\pi \Pr(f_v | D_v = 1, X_v, \theta) + (1 - \pi) \Pr(f_v | D_v = 0, X_v, \theta)], \quad (\text{A})$$

where f_v represents all measurements for variant v , X_v identifies the replicate, batch, and context for each, and π is the prior probability that a VUS in the sample is damaging.

This basic approach, as applied to family history data, is outlined in the work of Zhou and colleagues (12). In that setting, this formulation has the advantage that it accounts for ascertainment (sampling of families) without conditioning on the family history data, provided

Table 2. Table of variant-specific summaries: variant name (columns 1 and 5), prior classification (columns 2 and 6), posterior probability that the variant is protein damaging (columns 3 and 7); and natural logarithm of the posterior odds in favor of the variant being protein damaging (columns 4 and 8)

Variant	Classification	Posterior probability damaging	log _e posterior odds ^a	Variant	Classification	Posterior probability damaging	log _e posterior odds ^a
5673insC	Pathogenic ^b	1.0000		M1783L	VUS	0.0013	6.63
del17/18	Pathogenic	1.0000		M1783T	VUS	0.0346	3.33
W1837X	Pathogenic	1.0000		P1614S	VUS _{0.0001}	0.0011	6.82
WT	Neutral ^c	0.0000		P1771L	VUS	0.0066	5.01
Y1853X	Pathogenic	1.0000		P1806A	VUS	0.0016	6.45
A1669S	VUS	0.0625	2.71	P1812A	VUS	0.0050	5.29
A1708E	VUS	1.0000	17.67	P1859R	VUS _{0.0001}	0.0013	6.64
A1752P	VUS	1.0000	18.02	Q1785H	VUS	0.0012	6.77
A1823T	VUS	0.9999	8.89	Q1826H	VUS	0.0031	5.77
A1830T	VUS	0.0015	6.53	R1443G	VUS	0.0223	3.78
C1697R	VUS	0.9999	9.50	R1495M	VUS	0.0019	6.24
D1546N	VUS	0.0015	6.51	R1699L	VUS	1.0000	17.50
D1692N	VUS	0.0948	2.26	R1699Q	VUS	0.9987	6.65
E1644G	VUS	0.0016	6.45	R1699W	VUS	0.8795	1.99
E1794D	VUS	0.0013	6.62	R1726G	VUS	0.0013	6.66
F1662S	VUS _{0.0001} ^d	0.0017	6.40	R1751P	VUS	1.0000	17.54
F1695L	VUS	0.0013	6.63	R1751Q	VUS _{0.0001}	0.0022	6.11
G1706A	VUS	1.0000	11.85	R1753T	VUS	1.0000	17.57
G1706E	VUS _{0.999} ^e	1.0000	16.61	R1835Q	VUS	0.0021	6.15
G1738E	VUS	0.9952	5.34	S1512I	VUS _{0.0001}	0.0013	6.62
G1738R	VUS _{0.99} ^f	1.0000	17.39	S1613C	VUS	0.0012	6.69
G1788D	VUS	0.0147	4.21	S1613G	VUS _{0.0001}	0.0009	7.03
G1788V	VUS	1.0000	17.80	S1655F	VUS	0.9995	7.56
H1402Y	VUS	0.0032	5.73	S1715N	VUS	0.9994	7.49
H1421Y	VUS	0.0017	6.37	S1715R	VUS	0.9997	8.16
H1746N	VUS	0.0231	3.74	T1685I	VUS _{0.99}	1.0000	18.59
I1766S	VUS	1.0000	16.01	T1700A	VUS	0.4291	0.29
K1487R	VUS	0.0045	5.39	T1720A	VUS _{0.0001}	0.0014	6.59
K1847R	VUS	0.0025	5.98	V1534M	VUS _{0.0001}	0.0011	6.81
L1407P	VUS	0.4052	0.38	V1665M	VUS	0.0263	3.61
L1564P	VUS	0.0058	5.14	V1688del	VUS _{0.99}	0.9995	7.60
L1664P	VUS	0.0017	6.38	V1713A	VUS	1.0000	14.52
L1764P	VUS _{0.99}	1.0000	16.85	V1736A	VUS	0.9998	8.69
L1844R	VUS	0.0013	6.62	V1741G	VUS	0.7891	1.32
M1628T	VUS	0.0031	5.77	V1804D	VUS _{0.0001}	0.0057	5.17
M1628V	VUS	0.0031	5.78	V1809F	VUS	1.0000	17.74
M1628V+S1613G	VUS	0.0026	5.97	V1833M	VUS	0.0017	6.36
M1652I	VUS	0.0013	6.63	W1837R	VUS	1.0000	16.96
M1689R	VUS	1.0000	16.00	WTdelta14	VUS	0.0013	6.67
M1775K	VUS	1.0000	13.37	WTdelta14+S1616G	VUS	0.0012	6.74
M1775R	VUS _{0.99}	1.0000	17.24				

^aThe posterior odds are equivalent to the Bayes factor, $\Pr(\text{Data} | \text{Damaging}) / \Pr(\text{Data} | \text{Benign})$, because the prior distribution is uniform, and could be combined in this form with other conditionally independent data types.

^bKnown pathogenic variant ("negative control").

^cWild-type, known benign variant ("positive control").

^dVariant of unknown significance (VUS) previously classified as neutral with an estimated posterior probability of pathogenicity less than 0.0001 but blinded for purposes of model evaluation.

^eVUS previously classified as pathogenic with an estimated posterior probability greater than 0.999 but blinded for purposes of model evaluation.

^fVUS previously classified as pathogenic with an estimated posterior probability of between 0.99 and 0.999 but blinded for purposes of model evaluation.

that data from each of the classes of variants is sampled in the same way. A less statistically efficient alternative is to condition on the family history data (13); that formulation is outlined in the work of Goldgar and colleagues (14). The ultimate goal is to combine multiple sources of data, including functional assay and family history data, into a comprehensive 2 component mixture model.

Incorporating the functional assay data

Previous analyses of these data (7, 8, 15, 16) have adjusted for experiment-to-experiment variability by normalizing each ratio measurement in a batch to the mean of the WT ratios in that batch, focusing analysis on

$$g_{vbc} = \frac{r_{vbc}}{\bar{r}_{WT,bc}}$$

This suggests that the raw ratio data are scaled by an arbitrary multiplicative constant from batch to batch. The model we describe here extends this idea by treating the batch effects as parameters and by allowing all variants assayed in multiple batches, not just the WT, to inform their estimation. This approach will lead to more accurate estimates of these effects and inferences regarding the models other parameters will explicitly account for their uncertainty.

We assume that the average (over replicates) log ratio measurements are conditionally independent given batch, variant, and context; hence, that

$$\Pr(f_{v^*} | D_{v^*}, X_{v^*}, \theta) = \prod_{\{(v,b,c) : v=v^*\}} \Pr(\bar{f}_{vbc} | D_v, b, c, \theta). \quad (B)$$

We model the \bar{f}_{vbc} using a Bayesian random effects regression model with batch-, context-, and variant-specific random effects and constant error variance ψ^2 . The batch random effects distribution is centered at a value, κ_c , that depends on AA context, $c = 1$ or 2 . The core component of the model is the variant random effects distribution. It is a 2-component normal mixture model with component membership indexed by D_v , the indicator of a variant's status as damaging or benign to protein function, and having component-specific mean and variance terms (η_{WT}, ρ^2) , when $D_v = 0$, and $(\eta_{WT} + \delta, \gamma^2)$, when $D_v = 1$; δ is constrained to be positive. The parameter η_{WT} is the WT effect; for purposes of model identifiability we set this parameter to 3.8, the average WT log ratio measurement. This value is arbitrary; inference regarding the status of variants is not sensitive to its choice. The hierarchical model specification is made complete with uniform distributions on the location hyperparameters κ_1 and κ_2 and half-Cauchy distributions on the scale hyperparameters (ψ , λ , γ , and ρ). The half-Cauchy has been shown to be superior to other commonly used, weakly informative prior distributions for variance component parameters (17).

In particular, the right-hand side term in Equation (B) is specified by the following hierarchical model.

$$\begin{aligned} \bar{f}_{vbc} &\sim \text{Normal}(\beta_{bc} + \eta_v, \psi^2) \\ \beta_{bc} &\sim \text{Normal}(\kappa_c, \lambda^2) \quad b = 1, \dots, 30 \\ \eta_{WT} &= 3.8 \\ \eta_v | D_v = 0 &\sim \text{Normal}(\eta_{WT}, \rho^2) \quad \{v : D_v = 0 \wedge v \neq \text{WT}\} \\ \eta_v | D_v = 1 &\sim \text{Normal}(\eta_{WT} + \delta, \gamma^2) \quad \{v : D_v = 1\} \\ \delta &\sim \text{Uniform}(0, 10) \\ \kappa_1, \kappa_2 &\stackrel{\text{iid}}{\sim} \text{Uniform}(-10, 10) \\ \psi, \lambda, \gamma, \rho &\stackrel{\text{iid}}{\sim} \text{HalfCauchy}(0, a) \\ D_v &\sim \text{Bernoulli}(\pi) \quad \{v : D_v \text{ is Unknown}\} \\ \pi &\sim \text{Uniform}(0, 1) \end{aligned} \quad (C)$$

In the above, $\text{Normal}(m, s^2)$ denotes the normal distribution with mean m and variance s^2 ; $\text{Uniform}(c, d)$ is the uniform distribution on the interval $[c, d]$; and $X \sim \text{HalfCauchy}(0, a)$ indicates that X is Cauchy distributed with mode 0 and scale parameter \sqrt{a} , but constrained to be positive. In our analysis, we set $a = 1$ but also carry out sensitivity analyses with a more ($a = 10$) and a less ($a = 0.1$) diffuse prior specification for the model's variance parameters. The ranges of the uniform distributions placed on the model's mean parameters were chosen to comfortably exceed the range of values expected for those parameters. Finally, as a further gauge on the influence of these modeling decisions, we present an empirical analysis of the data and contrast summaries of that analysis with comparable summaries of the Bayesian analysis.

Inference

Inference for the above model is based on summaries of a single long run of a Markov chain Monte Carlo algorithm. The algorithm was coded in the WinBugs language (18) and the remaining aspects of the data analysis were carried out in the R programming environment (19). WinBugs constructs and executes a Gibbs sampler from a user-supplied description of the model written in metacode. Our WinBugs code is available on request. We ran the sampler for an initial 200,000 "burn-in" iterations followed by another 2.3 million iterations. We discarded the burn-in iterations and retained every 100th of the subsequent samples, leaving 23,000 samples for inference. We estimated marginal posterior means (PM), SD, and intervals of the model's parameters by their sample-based counterparts and formed "Rao-Blackwellized" estimates (20) of the posterior probabilities of each of the variants being protein damaging, $\Pr(D_v | \text{Data})$. The chain passed the Raftery and Lewis (21) convergence diagnostic implemented in the R package CODA (22) at its default settings, indicating that the chain's burn-in length was sufficiently long and that its remaining samples were more than sufficient to estimate the 0.025th quantile of the marginal posterior distribution of each parameter to a precision of ± 0.005 with probability 0.95.

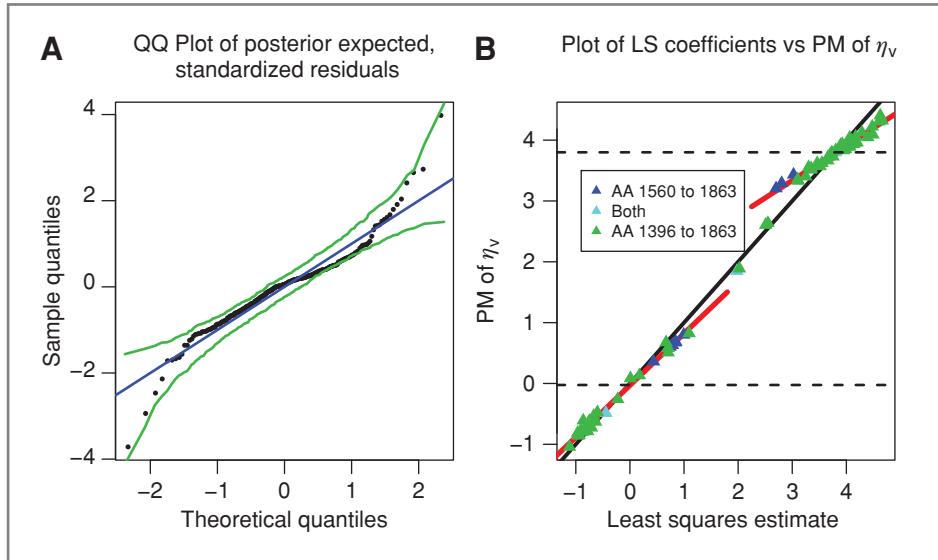


Figure 1. Model goodness of fit. A, normal quantile–quantile (QQ) plot of standardized residuals from our fit of the functional assay data. Standardized residuals were averaged over posterior uncertainty in the model's parameters. The green lines denote a simultaneous 95 interval estimate under the normal model. B, variant effects, η_v versus least squares (LS) estimates of the variant effect adjusted for batch. The black line is the diagonal. Points are highlighted according to whether the variant was studied in AA context 1396 to 1863 only (green), in AA context 1560 to 1863 only (blue) or both (teal). Bayesian shrinkage to the component-specific means (black dashed line) is indicated by the 2 red lines.

Classification and classification accuracy

Plon and colleagues (9) suggest summarizing the posterior probability in favor of a variant's pathogenicity on a scale of 1 to 5 where class 1, "not pathogenic," corresponds to $\Pr(\text{Pathogenic}|\text{Data}) < 0.001$; class 2, "likely not pathogenic," corresponds to $0.001 \leq \Pr(\text{Pathogenic}|\text{Data}) < 0.05$; class 3, "uncertain," corresponds to $0.05 \leq \Pr(\text{Pathogenic}|\text{Data}) < 0.95$; class 4, "likely pathogenic," corresponds to $0.95 \leq \Pr(\text{Pathogenic}|\text{Data}) < 0.99$; and class 5, "definitely pathogenic," corresponds to $0.99 \leq \Pr(\text{Pathogenic}|\text{Data})$. We consider 2 classification schemes when evaluating the model: S_1 , in which classes 1 and 2 are declared "neutral," class 3 "uncertain," and classes 4 and 5 "pathogenic;" and S_2 , in which class 1 variants are declared "neutral," classes 2, 3, and 4 "uncertain" and class 5 "pathogenic". We define the sensitivity of a classification scheme as the probability that a pathogenic variant is classified as protein damaging and its specificity as the probability that a neutral variant is classified as benign. We estimate these probabilities by their PM under a binomial sampling model and Jeffreys' prior and we provide 95% posterior credible intervals (CI), defined as the 2.5 th and 97.5 th percentage points of the posterior distributions.

Results

The process by which we arrived at the model outlined above involved an iterative procedure of model checking and evaluation. Our understanding of the biology coupled with results of an exploratory analysis (results

not reported here) suggested that the log ratio measurements depend on batch, context, and variant. We began with a model for the individual replicates, f_{vrb} but found that it was sufficient to analyze the mean log ratio measurement for each variant within a batch, \bar{f}_{vbc} . This greatly simplifies the model but does not affect our substantive conclusions [the correlation between estimates of $\Pr(D_v|\text{Data})$ derived from the 2 models is 0.9998]. The batch and variant effects β_{bc} and η_v are not identifiable through the model's likelihood but can be made so either by specifying a prior distribution that is informative for one of their means or by setting one effect to a constant. We experimented with several such choices and found that model fit was improved by fixing the WT effect. Figure 1A is a normal quantile–quantile plot of standardized residuals from our final model, averaged over posterior parameter uncertainty. A simultaneous 95% interval estimate for the empirical quantiles (green lines) includes the observed quantiles, indicating that the error structure of the model accurately describes residual variability in the data.

To provide a basis for comparison with the Bayesian model, we fit a fixed-effects regression of f_{vrb} on the factors variant and batch. This analysis provided us with least squares estimates of the variant effects adjusted for batch. Figure 1B, plots PM of the variant-level effects from the Bayesian hierarchical model against their corresponding least squares estimates. In this plot, the black line is the diagonal and points are highlighted according to whether the variant was studied in AA context 1,396 to 1,863 only (green), in AA context 1,560 to 1,863 only (blue)

Table 3. Summaries of the marginal posterior (columns 2–4) and marginal prior (columns 5–6) distributions of the models parameters excluding the batch- and variant-specific random effects and classification indicators

Parameter	Posterior		95 HPD Region	Prior	
	Mean	SD		Mean	SD
ψ	0.451	0.025	(0.403, 0.499)	∞	∞
λ	1.381	0.195	(1.029, 1.769)	∞	∞
ρ	0.412	0.131	(0.167, 0.667)	∞	∞
γ	0.948	0.241	(0.485, 1.413)	∞	∞
κ_1	0.063	0.286	(0.497, 0.621)	0.000	5.774
κ_2	0.160	0.639	(1.396, 1.120)	0.000	5.774
δ	3.825	0.226	(4.252, 3.371)	5.000	2.887
π	0.396	0.060	(0.277, 0.510)	0.500	0.289

NOTE: Columns 2 and 3 report estimates of PM and SD; Bayesian 95 posterior high density regions are provided in column 4. Prior means and SD appear in columns 5 and 6. Note that that half-Cauchy distribution has no finite moments.

or both (teal). A modest amount of Bayesian shrinkage to the component-specific means (depicted by a black dashed line) is evident. This effect is highlighted by the red lines which plot the least squares fits among variants with a PM greater than 3.0 (likely members of the protein neutral component) and less than 1.2 (likely members of the protein damaging component). Despite the shrinkage, the Bayesian estimates remain highly correlated (0.9964) with their frequentist fixed-effects counterparts, suggesting that influence of the models higher level parameters is negligible and that the model is not overfitting the data.

Table 3 summarizes the marginal prior and posterior distributions of the models higher level parameters (i.e., those not indexed by batch or variant). In all cases, the prior distributions are diffuse and relatively noninformative, whereas the marginal posterior distributions are concentrated. Inferences about the pathogenicity of the 76 VUSs are driven by the 2-component normal mixture model for the variant-specific random effects, η_v . With the benign variants centered at the WT effect (assumed to be 3.8 for purposes of model identifiability), we estimate that the variants damaging to protein function have effects centered at 0.025 with a SD of 0.226 adjusting for batch and context. There is considerable batch-to-batch variability in the average level of the log ratio data: We estimate the SD of the batch random effects distribution, λ , to be 1.381. Estimates of these parameters were insensitive to the scaling factor a used in the prior distributions: In both cases the PM estimated for these parameters in the sensitivity analyses had correlation greater than 0.999 with those reported in Table 3.

Figure 2 provides a graphical summary of the variant-specific effects and of their mixture model marginal distribution. In this plot, each variant is represented by a boxplot summarizing the marginal posterior distribution of its random effect, η_v . The WT variant is plotted in

green, the known pathogenic variants are plotted in red, the VUSs are plotted in blue, and the variants previously classified as pathogenic/benign and included in the model blinded to this information are plotted in orange/yellow. A point estimate of the mixture model is plotted in the right margin. Its left (top) component corresponds to benign variants, whereas its right (bottom) component corresponds to damaging variants; estimates of the means of the components are plotted as green and red dotted lines, respectively. The WT is the only assumed known benign variant in this analysis and, as such, serves to anchor its component. Its variant-specific effect is fixed at 3.8 as described above. There are 4 known pathogenic variants that serve to anchor the component representing protein-damaging variants. These have variant-specific effects ranging from 0.76 (5673insC) up to 0.68 (W1837X). As the plot suggests, the damaging variant effects are more variable than their benign counterparts (the SD of their component is estimated to be 2.30 times as large as the SD of the benign variant component).

Table 2 provides variant-specific summaries from the analysis. Each variant is labeled according to the classification assumed for it prior to the analysis: pathogenic, neutral, VUS, VUS previously classified as "neutral" with an estimated posterior probability of pathogenicity less than 0.0001, VUS previously classified as pathogenic with an estimated posterior probability greater than 0.999, and VUS previously classified as pathogenic with an estimated posterior probability of between 0.99 and 0.999. The last 3 classes of VUS were included in the analysis "blinded" to their prior classification for purposes of model evaluation. In addition, the table reports the posterior probabilities and the natural logarithm of the posterior odds for each variant being damaging. Estimates of these posterior probabilities were also insensitive to the scaling factor a used in the prior distributions: In both

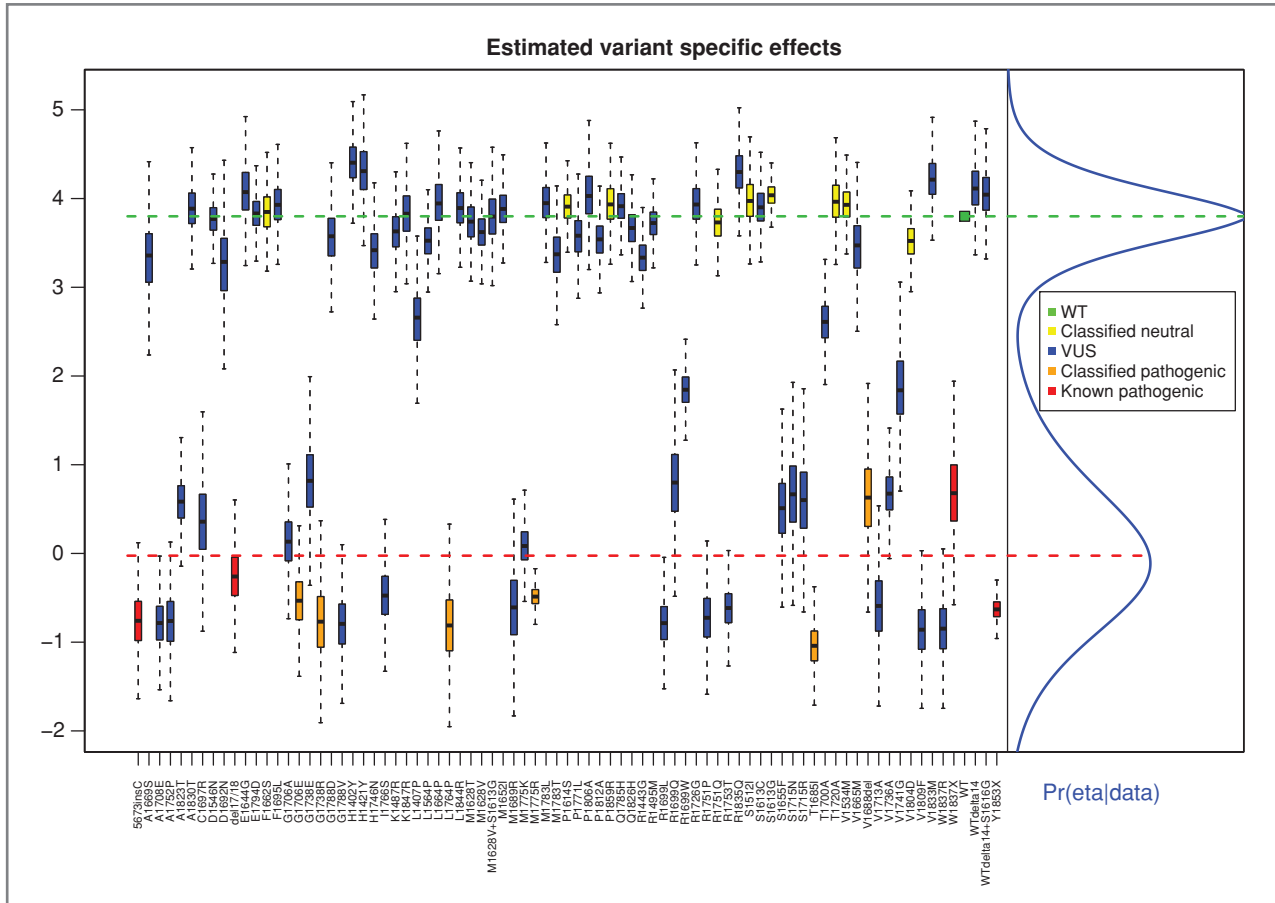


Figure 2. Boxplots summarizing the marginal posterior distributions of the variant-specific effect parameters, η_v . The WT variant is indicated in green, the known pathogenic variants in red, and the VUSs in blue. VUSs previously classified as neutral using other forms of data and included in the analysis for purposes of model evaluation are indicated in yellow, whereas those previously classified pathogenic and included for the same reason are depicted in orange. The η_v are modeled as arising from a 2-component mixture of normal distributions. The posterior expected value of this curve is plotted on the right margin. Its left (top) component corresponds to variants that are benign to protein function, whereas its right (bottom) component corresponds to variants that are damaging to protein function. The mean of the benign/damaging component is plotted as a green/red dotted horizontal line on the plot.

cases the estimates of these probabilities in the sensitivity analyses had correlation greater than 0.999 with those reported in Table 2.

For purposes of this analysis, we assumed that the prior probability of a variant being protein damaging for each VUS is 0.5 (see Equation C). This is a common "noninformative" choice when there is no prior data to inform the distribution of a binary variable. Our choice to isolate this analysis from the influence of other variables allows us to clearly summarize the utility of the functional data as a form of evidence for classifying VUSs. It also allows one to interpret the posterior odds of variant being damaging as a Bayes factor (BF; ref. 23), a Bayesian integrated likelihood ratio statistic that measures the degree to which the data—in this case, the functional annotation measurements—support the hypothesis that a variant is protein damaging. Jeffreys (24) provides a commonly cited metric for judging BFs in which BFs between 1 and 3.2 (0 and 1.16 on the log_e scale used in

columns 4 and 8 in Table 2) provide "indeterminate" evidence in favor of the hypothesis, those between 3.2 and 10 (1.16 and 2.30) provide "positive" evidence of the hypothesis, those between 10 and 31.2 (2.30 and 3.44) provide "strong" evidence of the hypothesis, those between 31.2 and 100 (3.44 and 4.61) provide "very strong" evidence of the hypothesis, and those above 100 (4.61) provide "decisive" evidence of the hypothesis. The BF is symmetric: Its multiplicative inverse provides a comparable measure of evidence *against* the hypothesis. Hence, for example, a BF between 1/100 and 1/31.2 (4.61 and 3.44) provides "very strong" evidence against the hypothesis on Jeffreys' scale of evidence.

The primary goal of this analysis is to determine how informative and accurate the functional data are for classifying VUSs before including it in a composite model that incorporates family, cosegregation, sequence alignment, and other forms of data. Our analysis suggests that the functional data are very informative. Indeed, 65 of 76

VUSs have BFs, either in favor of or against the variant being protein damaging, in the "decisive" range. Under classification scheme S_1 , 27 variants would be classified as damaging, 43 as neutral, and 6 would remain unclassified, whereas under the much more stringent scheme S_2 , 27 variants would be classified as damaging, 1 as neutral, and 48 would be left unclassified.

Furthermore, results for the blinded samples suggest that classification probabilities derived from the model are both sensitive and specific. Our estimates of the posterior probabilities (PP) of the variants being protein damaging for the VUSs previously classified as neutral range from 0.00089 (S1613G) up to 0.0057 (V1804D); all have BFs falling in the "decisive" category of evidence *against* the variants being protein damaging. Our PP estimates for the VUSs previously classified as pathogenic by other methods were 0.999 (V1688del), 1.00 (L1764P), 1.00 (T1685I), 1.00 (G1706E), 1.00 (M1775R), and 1.00 (G1738R); their BFs were all in the "decisive" range in *favor* of being protein damaging. All of these variants are classified correctly (i.e., the same as they had been previously) using classification scheme S_1 . In particular, our approach is both sensitive [observed relative frequency, RF = 6/6, PM = 0.93, CI = (0.67–1.00)] and specific [RF = 9/9, PM = 0.95, CI = (0.76–1.00)] when the variants are classified using this rule. In contrast, while sensitivity remains high [RF = 6/6, PM = 0.93, CI = (0.67–1.00)], specificity [RF = 1/9, PM = 0.15, CI = (0.01–0.41)] declines markedly under the more stringent scheme S_2 . This suggests that, while the functional data are very informative, they will need to be combined with other data, such as *in silico* and family-based results, to meet the more stringent requirements of clinical application.

In principle, a classification scheme with similar operating characteristics could be constructed by applying empirically determined thresholds to estimates of the variant-specific coefficients in the batch-adjusted linear model. Classification against thresholds of 1 and 3 (x -axis, Fig. 1B), for example, would correctly call all blinded variants and would correspond to a procedure with estimated sensitivity of 0.93 and specificity of 0.95. Overall, 26 of the 76 variants would be classified as pathogenic and 43 as neutral, whereas 7 would remain unclassified. While the predictive utility of the functional data is evident from these simple summaries of the data, the 2-component mixture model establishes the formal framework required for inference in this vexing problem. It provides a natural yardstick for classification (the posterior probability) and a modeling approach that can readily be extended to include the other forms of data relevant to classification of VUSs.

Discussion

In this article we have described a computational approach for inferring the disease relevance (pathogenicity) of VUSs in disease genes from data derived from a laboratory-based functional assay. We also show the

utility of the model by applying it to the classification of 76 *BRCA1* VUSs located in the *BRCA1* BRCT region. The approach is shown to be both sensitive and specific when evaluated on variants that had been previously classified using different sources of data.

Our results are very encouraging but caution is warranted when interpreting the results of functional assays in general, and of transcriptional assays in particular. False negatives—variants that confer increased disease risk but behave as benign in the assay—can occur when domain-specific protein activity assays do not evaluate a particular function important for tumor suppression. For example, the protease sensitivity assay (11, 25), which detects variants that cause large effects on protein folding, cannot assess the role of surface variants that disrupt binding sites without affecting the general fold. False positives are also possible, as a variant may disrupt a function that is not required for tumor suppression. This is a particularly vexing problem for multifunctional proteins, such as *BRCA1*, for which it is still unclear which of many functions are required for the normal phenotype. Therefore, in light of the severe nature of the clinical prophylactic options, we strongly advise against using the results of functional assays *in isolation* to classify a variants pathogenicity.

The model presented here is one step toward an integrative classification model based on Equation A that incorporates multiple sources of data, including segregation; personal and family history of disease; sequence based on position in the protein, and deviation of missense residues from that range of variation; co-occurrence data; and tumor grade and immunohistochemistry, to yield more definitive predictions. This will dramatically improve the model's ability to account for the sensitivity and specificity of each individual measurement, provided that the joint distribution of the variables and D_v is carefully modeled. Measurements that occasionally conflict with the others, especially the more accurate sources (e.g., the family history-based variables) will be down weighted by the model. As an example, suppose that an assay tests for only part of the gene's relevant functional activity and that a second assay tests for the rest. The second assay—in combination with the *in silico* and family-based variables—will help identify pathogenic variants that have WT level values from the first assay (i.e., those whose effect is through the alternate activity). These variants will now play a much larger role in estimation of the pathogenic component of the mixture model than when the first assay was in the model alone. This will have the effect of drawing the 2 components closer to one another and increasing the variance of the pathogenic component on the axis representing the first assay, thereby reducing the influence of that assay on classification. Presence of the first assay in the model will similarly balance the second assay.

For this to work effectively, the multivariate relationships among the data sources will need to be modeled carefully and estimated accurately. This will require a

coanalysis of a large, diverse set of VUSs representative of what would be expected in clinical application of the model along with large numbers of known pathogenic variants and WT families. All classes of variant should include as many examples with informative family history and genetic data as are available. A large, diverse set of variants will improve representation across the various classes characterized by the specific functional mechanisms potentially tied to disease predisposition. Ultimately, it will be important for health care providers to be made aware of the model's limitations and their implications for decision making. This can be achieved by increased education and training coupled with clear recommendations from the research community.

The model we present is a Bayesian hierarchical model that accounts for experimental heterogeneity at the batch, context, and variant levels. The prior distribution was chosen so as to be minimally informative but proper. In particular, it does not incorporate other forms of existing data on the VUSs and does not unduly restrict the effective range of the model's parameters. The posterior distribution from the analysis of BRCA1 BRCT domain variants is highly concentrated relative to the prior distribution suggesting that the data are informative for the effects captured by the parameters. Furthermore, we found that inferences regarding both the model's parameters and the estimated classification probabilities were insensitive to the prior scale, a , assumed for the models variance parameters. In subsequent applications of the model to data generated using the same protocols, the prior distribution may be made more informative for parameters, such as the batch and variant random effects, related to experimental heterogeneity.

For reasons described above, we placed a uniform prior distribution on π , the probability of a VUS being protein damaging. However, in settings where something is known *a priori* about the probability of pathogenicity of the population of observed and/or possible sequence variants under study, we suggest that this distribution be changed to a beta distribution with suitably chosen parameters. For example, when the goal is classification

of BRCA1 and/or BRCA2 VUSs, the model may be made conditional on the variant-specific sequence conservation classes described in (26). The model described in Equation C is easily adapted to incorporate the probabilities derived in (27) and (26) and sometimes referred to as "prior probabilities" for having been calculated prior to variant level family and genetic data, as prior probabilities for those classes. Using our notation, (26) estimates $\Pr(D_v = 1 | SCC_v = C65) = 0.81$ and reports (0.61, 0.95) as the 95% interval estimate for this quantity, where SSC_v denotes the sequence conservation class of variant v and C65 is a particular class. This suggests that we assume, for example,

$$D_v | SCC_v = C65 \sim \text{Bernoulli}(\pi_{C65})$$

$$\pi_{C65} \sim \text{beta}(14.58, 3.42),$$

as the beta distribution with parameters 14.58 and 3.42 has mean 0.81, 2.5th percentile 0.61 and 97.5th percentile 0.95.

VUSs are difficult to classify, but the clinical imperative to do so is clear: Patients tested for disease genes that are found to harbor a VUS are left with little data on which to make difficult decisions about prophylactic treatments. The current analysis and the program of research that it is a part of aim to improve this situation.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Acknowledgments

The authors thank 3 anonymous referees and the editor for their thoughtful review.

Grant Support

This work was funded in part by NIH grants P50-CA116201, R01-CA116167, U19-CA148112, and U56-CA11809.

Received November 19, 2010; revised February 14, 2011; accepted March 20, 2011; published OnlineFirst March 29, 2011.

References

1. Kurian AW. BRCA1 and BRCA2 mutations across race and ethnicity: distribution and clinical implications. *Curr Opin Obstet Gynecol* 2010;22:72–8.
2. Weitzel JN, Lagos V, Blazer KR, Nelson R, Ricker C, Herzog J, et al. Prevalence of BRCA mutations and founder effect in high-risk Hispanic families. *Cancer Epidemiol Biomarkers Prev* 2005;14:1666–71.
3. Nanda R, Schumm LP, Cummings S, Fackenthal JD, Sveen L, Ademuyiwa F, et al. Genetic testing in an ethnically diverse cohort of high-risk women: a comparative analysis of BRCA1 and BRCA2 mutations in American families of European and African ancestry. *JAMA* 2005;294:1925–33.
4. Domchek SM, Friebel TM, Singer CF, Evans DG, Lynch HT, Isaacs C, et al. Association of risk-reducing surgery in BRCA1 or BRCA2 mutation carriers with cancer risk and mortality. *JAMA* 2010;304:967–75.
5. Thompson D, Easton DF, Goldgar DE. A full-likelihood method for the evaluation of causality of sequence variants from family data. *Am J Hum Genet* 2003;73:652–5.
6. Goldgar DE, Easton DF, Deffenbaugh AM, Monteiro AN, Tavtigian SV, Couch FJ. Integrated evaluation of DNA sequence variants of unknown clinical significance: application to BRCA1 and BRCA2. *Am J Hum Genet* 2004;75:535–44.
7. Carvalho MA, Marsillac SM, Karchin R, Manoukian S, Grist S, Swaby RF, et al. Determination of cancer risk associated with germ line BRCA1 missense variants by functional analysis. *Cancer Res* 2007;67:1494–501.
8. Carvalho M, Pino MA, Karchin R, Beddor J, Godinho-Netto M, Mesquita RD, et al. Analysis of a set of missense, frameshift and in-frame deletion variants of BRCA1. *Mutat Res* 2009;660:1–11.

9. Plon SE, Eccles DM, Easton D, Foulkes WD, Genuardi M, Greenblatt MS, et al. Sequence variant classification and reporting: recommendations for improving the interpretation of cancer susceptibility genetic test results. *Hum Mutat* 2008;29:1282–91.
10. Carvalho MA, Couch FJ, Monteiro AN. Functional assays for BRCA1 and BRCA2. *Int J Biochem Cell Biol* 2007;39:298–310.
11. Lee MS, Green R, Marsillac SM, Coquelle N, Williams RS, Yeung T, et al. Comprehensive analysis of missense variations in the BRCT domain of BRCA1 by structural and functional assays. *Cancer Res* 2010;70:4880–90.
12. Zhou Z, Iversen ES, Parmigiani G. Classification of missense mutations of disease genes. *J Am Stat Assoc* 2005;100:51–60.
13. Iversen ES, Chen S. Population-calibrated gene characterization: estimating distributions of age at onset associated with breast cancer susceptibility genes. *J Am Stat Assoc* 2005;100:399–409.
14. Goldgar DE, Easton DF, Byrnes GB, Spurdle AB, Iversen ES, Greenblatt MS, et al. Genetic evidence and integration of various data sources for classifying uncertain variants into a single model. *Hum Mutat* 2008;29:1265–72.
15. Vallon-Christersson J, Cayanan C, Haraldsson K, Loman N, Bergthorsson JT, Brøndum-Nielsen K, et al. Functional analysis of BRCA1 C-terminal missense mutations identified in breast and ovarian cancer families. *Hum Mol Genet* 2001;10:353–60.
16. Phelan CM, Dapic V, Tice B, Favis R, Kwan E, Barany F, et al. Classification of BRCA1 missense variants of unknown clinical significance. *J Med Genet* 2005;42:138–46.
17. Gelman A. Prior distributions for variance parameters in hierarchical models. *Bayesian Anal* 2006;1:515–33.
18. Lunn DJ, Thomas A, Best N, Spiegelhalter D. WinBUGS – a Bayesian modelling framework: concepts, structure, and extensibility. *Stat Comput* 2000;10:325–37.
19. R Development Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2009. ISBN 3-900051-07-0.
20. Gelfand A, Smith AFM. Sampling-based approaches to calculating marginal densities. *J Am Stat Assoc* 1990;85:398–409.
21. Raftery AE, Lewis SM. Implementing MCMC. In: Gilks WR, Richardson S, Spiegelhalter DJ, editors. *Markov Chain Monte Carlo in Practice*. London, UK: Chapman & Hall; 1996. p. 115–27.
22. Plummer M, Best N, Cowles K, Vines K. CODA: Convergence diagnosis and output analysis for MCMC. *R News*; 2006;6:7–11. Available from: <http://CRAN.R-project.org/doc/Rnews/>, http://CRAN.R-project.org/doc/Rnews/Rnews_2006-1.pdf
23. Kass RE, Raftery AE. Bayes factors. *J Am Stat Assoc* 1995;90:773–95.
24. Jeffreys H. *Theory of Probability*. Oxford: Oxford Univ. Press; 1961.
25. Williams RS, Chasman DI, Hau DD, Hui B, Lau AY, Glover JNM. Detection of protein folding defects caused by BRCA1-BRCT truncation and missense mutations. *J Biol Chem* 2003;278:53007–16.
26. Tavtigian SV, Byrnes GB, Goldgar DE, Thomas A. Classification of rare missense substitutions, using risk surfaces, with genetic- and molecular-epidemiology applications. *Hum Mutat* 2008;29:1342–54.
27. Easton D, Deffenbaugh A, Pruss D, Frye C, Wenstrup RJ, Allen-Brady K, et al. A systematic genetic assessment of 1,433 sequence variants of unknown clinical significance in the BRCA1 and BRCA2 breast cancer predisposition genes. *Am J Hum Genet* 2007;81:873–83.
28. Miki Y, Swensen J, Shattuck-Eidens DJ, Futreal PA, Harshman K, Tavtigian S, et al. A strong candidate for the breast and ovarian cancer susceptibility: gene *brca1*. *Science* 1994;266:66–71.
29. Tavtigian SV, Deffenbaugh AM, Yin L, Judkins T, Scholl T, Samollow PB, et al. Comprehensive statistical study of 452 BRCA1 missense substitutions with classification of eight recurrent substitutions as neutral. *J Med Genet* 2006;43:295–305.
30. Chenevix-Trench G, Healey S, Lakhani S, Waring P, Cummings M, Brinkworth R, et al. Genetic and histopathologic evaluation of BRCA1 and BRCA2 DNA sequence variants of unknown clinical significance. *Cancer Res* 2006;66:2019–27.