An integrative model for the comprehensive classification of \textit{BRCA1} and \textit{BRCA2} variants of uncertain clinical significance

Edwin S. Iversen Jr.\textsuperscript{1,} \textsuperscript{2,}, Gary Lipton\textsuperscript{1}, Steven N. Hart\textsuperscript{2}, Kun Y. Lee\textsuperscript{3}, Chunling Hu\textsuperscript{3}, Eric C. Polley\textsuperscript{3}, Tina Pesaran\textsuperscript{4}, Amal Yussuff\textsuperscript{4}, Holly LaDuca\textsuperscript{4}, Elizabeth Chao\textsuperscript{4}, Rachid Karam\textsuperscript{4}, David E. Goldgar\textsuperscript{5}, Fergus J. Couch\textsuperscript{1,7} and Alvaro N. A. Monteiro\textsuperscript{6,7}\textsuperscript{5}

Loss-of-function variants in the \textit{BRCA1} and \textit{BRCA2} susceptibility genes predispose carriers to breast and/or ovarian cancer. The use of germline testing panels containing these genes has grown dramatically, but the interpretation of the results has been complicated by the identification of many sequence variants of undefined cancer relevance, termed “Variants of Uncertain Significance (VUS).” We have developed functional assays and a statistical model called VarCall for classifying \textit{BRCA1} and \textit{BRCA2} VUS. Here we describe a multifactorial extension of VarCall, called VarCall XT, that allows for co-analysis of multiple forms of genetic evidence. We evaluated the accuracy of models defined by the combinations of functional, in silico protein predictors, and family data for VUS classification. VarCall XT classified variants of known pathogenicity status with high sensitivity and specificity, with the functional assays contributing the greatest predictive power. This approach could be used to identify more patients that would benefit from personalized cancer risk assessment and management.

\textit{npj Genomic Medicine} (2022) 7:35; https://doi.org/10.1038/s41525-022-00302-3

\textbf{INTRODUCTION}

The recent widespread utilization of clinical DNA sequencing has facilitated the implementation of precision medicine where clinical decisions, treatments, and procedures are tailored to each patient based on the identification of DNA variants. However, while a genetic variant may be identified in individuals with the disease, the likelihood of pathogenicity of the variant may not be sufficient to justify an intervention. Thus, genetic variants of uncertain clinical significance (VUS) present a critical challenge to the effective use of clinical genetic information\textsuperscript{1}.

\textit{BRCA1} and \textit{BRCA2} are among the genes most frequently subjected to clinical genetic testing because of the clinical importance of pathogenic variants. Carriers of pathogenic variants in \textit{BRCA1} or \textit{BRCA2} are at significantly increased risk (Relative Risk (RR) \textgreater{} 4) for breast and ovarian cancer and this knowledge can guide screening for early detection and the use of risk-reducing prophylactic surgeries. Importantly, the discovery of pathogenic variants also identifies patients likely to benefit from PARP inhibitors, a class of therapeutic agents that target cancers with homologous recombination DNA repair deficiency. Thus, the determination of the clinical relevance of VUS in \textit{BRCA1} and \textit{BRCA2} has significant implications for clinical care.

Currently, several non-mutually exclusive approaches exist for classifying \textit{BRCA1} and \textit{BRCA2} VUS. For example, loss-of-function variants in these tumor suppressors, for which the effect on protein function can be inferred from the genetic code (frameshift, nonsense, or consensus splicing acceptor/donor sites), are generally classified as pathogenic. A multifactorial model that includes family history, co-segregation, and co-occurrence with another known pathogenic variant in the same gene (biallelic inactivation is embryonic lethal and in the case of \textit{BRCA2} can cause Fanconi anemia) has been used for the remaining variant types (inframe indels, synonymous and missense in coding changes; intronic and regulatory)\textsuperscript{2,3}. It is also possible to classify variants based on the ACMG/AMP (The American College of Medical Genetics and Genomics; Association for Molecular Pathology) rule-based guidelines, in which pathogenicity may be predicted using multiple sources of the available evidence, including results from reproducible and clinically validated functional assays\textsuperscript{4,5}. However, current \textit{BRCA1} and \textit{BRCA2} variant classification models either do not incorporate functional data or use these data-only as supporting evidence, for or against pathogenicity\textsuperscript{6,7}. The potential for the application of functional data has not been fully realized because there is not an adequate statistical framework with which to (a) optimize the discrimination between pathogenic and non-pathogenic variants using functional data; and (b) integrate functional data with family history and other sources of data to obtain a global likelihood of pathogenicity.

Here, we present VarCall XT (VarCall eXtended), a Bayesian statistical framework that incorporates data from established functional assays with family data and sequence-based in silico prediction models. This model was used to determine the likelihood of pathogenicity for individual missense variants in \textit{BRCA1} and \textit{BRCA2} in the context of hereditary breast and ovarian cancer syndrome (HBOC), and to examine the effect on predictive accuracy realized by adding the available family or in silico data to the functional data.

\textbf{RESULTS}

Comparative results—classification accuracy

VarCall models make a binary prediction of pathogenic versus benign classification for each variant. In Fig. 1, we trace the estimated log-odds of pathogenicity for each variant based on the two progression series for \textit{BRCA1} or \textit{BRCA2} described above. While...
VarCall analyses of the functional data used here have been previously reported\(^7,8\), we refit these models using our VarCall XT software and summarize these fits here for the sake of consistency and to provide a basis of comparison for the multifactorial extensions of these models.

We initially considered the progression from the null to full models. Classification probabilities for both genes largely remain in the no-call range (VUS) based on the family data alone (Fig. 1). Only three of 96 variants, all in BRCA1, had enough family data for classification (all three were correctly classified as benign; Supplementary Table 1). The addition of the protein predictor data provided increased resolution and a greater degree of separation between the known pathogenic and benign variants, with 45 variants classified as benign or pathogenic (Fig. 1). In all cases, the classifications were correct, but benign variants were twice as likely to reach the threshold of classification as pathogenic variants. With the addition of the Align-GVGD score to the model with family and in silico data, six more variants were classified. However, three BRCA1 benign variants were incorrectly called pathogenic. The addition of the functional data, resulting in the “full model”, provided the clearest improvement in predictive accuracy, with 89 of the 96 variants correctly classified (Fig. 1).

Next, we took the original function-based VarCall model as the starting point. This model achieved 98% accuracy, with 94 of the 96 variants correctly classified and one BRCA1 benign and one BRCA2 pathogenic variant remaining as VUS. Progressively adding the family and protein predictor data diminished overall accuracy with four variants left unclassified. Progressing to the full model with the addition of the Align-GVGD score further diminishes accuracy with seven variants remaining unclassified (Fig. 1).

Table 1 displays operating characteristics estimates for the full set of fifteen predictive models when applied to the BRCA1 and BRCA2 data sets and computed using the \(n = 63\) BRCA1 and \(n = 33\) BRCA2 variants of known status. It is clear that the functional assays contributed the greatest predictive power to the VarCall XT model. The models that included the functional data all had sensitivity and specificity estimates greater than 0.65 and frequently exceeded 0.9. Models that did not use functional data were all lower than 0.65, because of the modest predictive power of the protein predictor data, Align-GVGD score, and family data.

---

Fig. 1 Trace plots of individual variant log-odds (y-axis) of pathogenicity across models (x-axis) for BRCA1 (top) and BRCA2 (bottom). The legend indicates the data elements included in the model with filled circles; the value plotted for the ‘Null’ model (i.e., the model with no data) is the marginal probability of pathogenicity estimated by the model including all data elements (labelled ‘Full’). Variants known a priori to be pathogenic are plotted in red, while those known to be benign are plotted in green; the VUS are plotted in gray. The light red region corresponds to posterior probabilities of pathogenicity of 99% or higher (IARC class 5), while the light green region corresponds to posterior probabilities of pathogenicity of 5% or lower (IARC classes 1 and 2); values in the white band fall within the ‘no-call’ range.

---

Table 1 displays operating characteristics estimates for the full set of fifteen predictive models when applied to the BRCA1 and BRCA2 data sets and computed using the \(n = 63\) BRCA1 and \(n = 33\) BRCA2 variants of known status. It is clear that the functional assays contributed the greatest predictive power to the VarCall XT model. The models that included the functional data all had sensitivity and specificity estimates greater than 0.65 and frequently exceeded 0.9. Models that did not use functional data were all lower than 0.65, because of the modest predictive power of the protein predictor data, Align-GVGD score, and family data.
The original functional assay-based VarCall model was the top model based on overall accuracy for both genes and, based on the scaled Brier score, was the top \textit{BRCA} 2 model and the second-ranked \textit{BRCA} 1 model (Supplementary Table 1). Furthermore, its accuracy was nearly as great when only the positive and negative controls where labeled: the average absolute difference in classification probabilities between the full and reduced training set fits was 0.00091 for the \textit{BRCA} 1 model and 0.0027 for the \textit{BRCA} 2 model (see Supplementary Fig. 3 and the discussion in Supplemental Results).

Because resolution to estimate predictive accuracy was limited by the number of variants of known status it was difficult to compare VarCall XT models. However, the contrast between models with and without the functional data was clear. The 95% CI limits for sensitivity among the models that excluded the functional data than for those models excluding the functional data (\textit{BRCA} 1 minimum = 93.7%; \textit{BRCA} 2 minimum = 87.9%) were uniformly higher than for those models excluding the functional data (\textit{BRCA} 1 maximum = 58.7%; \textit{BRCA} 2 maximum = 39.4%). Finally, Brier scores and the no-call rates were systematically lower among the models with function data than those without. Thus, when considering sensitivity and specificity jointly, the least accurate model with functional data was significantly more accurate than the most accurate model without functional data.

Predictive contribution of the family data
Family history data were available for 86 (25%) of the \textit{BRCA} 1 variants and 112 (45%) of the \textit{BRCA} 2 variants. Among the classified variants there were 17 \textit{BRCA} 1 (seven benign and 10 pathogenic) and 12 \textit{BRCA} 2 variants (two benign and 10 pathogenic) with family history summary data. When available, the number of families contributing to these data was often small (see Supplemental Data; Variant-level data tab, columns AT and AU). This highlights the fact that even when testing is widespread and prevalent, there will be no available family data on a large fraction of VUS that arise and the data that are available will often provide modest information overall.

We computed ratios of the estimated odds (ORs) in favor of pathogenicity, given both the function and family data, to the estimated odds given the functional data alone. The estimated mean OR among benign variants across both genes was <1.0 as expected (0.31, 95% CI 0.17–0.57) and was >1.0 among known
pathogenic variants (OR 1.04, 95% CI 0.39, 2.81). On average, the
effect of adding the family data moved the associated classification
probabilities marginally in the correct direction; however, this
pattern was clearer and more consistent among the known benign variants than the known pathogenic variants, as shown by
the wider interval estimates. Note that the inclusion of the family
data changes the posterior odds of pathogenicity of all variants,
not just those with the data. Estimated classifications of the
variants with those data shift causing estimates of the population
(of variants) level parameters that specify the Gaussian
classification model to adjust in response, thereby altering the
mapping, applicable to all variants, between functional effects estimates and their associated likelihoods of pathogenicity.
The classification status of the majority of variants was
unchanged with the addition of the family data. However, four
BRCA1 no-calls became benign and one pathogenic call became a
no-call (Fig. 2A). All of these variants are currently VUS. Scatter
plots of the log posterior odds of VUS pathogenicity using the
functional data-only (x-axis) versus the combined functional and
family data provide more resolution (Supplementary Fig. 2). There

---

**Fig. 2**  Changes in variant calls upon addition of family or protein predictor data. Circos plots illustrating changes in pathogenic calls (red ribbons), benign calls (blue ribbons), and no-calls (gray ribbons) with functional data-only (left half of the plot) and with the addition of family (A, B) or protein predictor data (C, D) for BRCA1 (A, C) and for BRCA2 (B, D). Outermost rings in each segment on the left side (starting calls) of the plot depicts the percentage of variants in function-only calls that remained in the same category after addition of family or protein predictor data. Outermost rings in each segment on the right side (final calls) depicts the contribution of variants that started as calls in the same category.
are six differences for BRCA2: two no-calls became benign and one became pathogenic and three pathogenic calls became no-calls (Fig. 2B). Four of these are VUS, two are known pathogenic variants. A comparable number of BRCA1 (n = 5) and BRCA2 variants (n = 6) shift into or out of a no-call region (Fig. 2A, B). Adjusted for the ‘size’ of the no-call region (there are about twice as many BRCA2 variants in the function-only model no-call region (n = 35) than in the corresponding BRCA1 region (n = 19)), the fraction of BRCA1 variants to shift classifications is larger than for BRCA2 (but is still small as a fraction of the total number of variants). No variants changed from pathogenic to benign or benign to pathogenic (Fig. 2A, B). Most of the changes observed were cases very near a classification threshold (these regions are highlighted in Supplementary Fig. 2). Among variants predicted by the original VarCall to be benign, 74% (37/50) of BRCA1 and 97% (59/61) of BRCA2 variants had decreased odds of pathogenicity with the addition of the family data. Likewise, among variants predicted to be pathogenic, 22% (7/32) of BRCA1 and 74% (26/35) BRCA2 variants had increased odds of pathogenicity. These patterns are reflected in the plotted loess regression curves in Supplementary Fig. 2.

**Contribution of protein predictor data**

While the classification status of the majority of variants remained unchanged with the addition of the protein predictor data as with the family data, differences were more frequent and pronounced, especially for BRCA2 (Fig. 2C, D). Among BRCA1 variants seven benign calls including a known benign variant became no-call VUS, and two no-call VUS became pathogenic (Fig. 2C). Similarly, among BRCA2 variants ten benign calls including a known benign variant became no-call VUS, and ten no-call VUS became pathogenic (Fig. 2D). While the number of BRCA2 variants to shift into or out of the no-call region is larger than for BRCA1, as a fraction of the variants in the gene’s function-only no-call region these values are comparable. The shift in classification categories for BRCA2 variants reflected an overall increase in the posterior log-odds of pathogenicity that resulted from the addition of the protein predictor data, which strengthened the original VarCall estimates for likely benign BRCA1 VUS, but had little effect on the variants called as likely pathogenic. The log-odds of pathogenicity of 150 of 229 variants called likely benign by the original function-only model decreased with the addition of the protein predictor data, while only 37 of 99 variants called likely pathogenic experienced an increase. In contrast, BRCA2 variants of all classifications experienced an increase in the log-odds of pathogenicity, with 146 of 154 variants called likely benign and 56 of 59 variants called likely pathogenic by the original VarCall displaying increased odds of pathogenicity with the addition of the protein predictor data.

Differences in probabilities of pathogenicity estimated using only the functional assay data and those estimated using only the protein predictor data are shown in Fig. 3. The probabilities of 42% of BRCA1 VUS changed by more than 0.1, 27% changed by more than 0.5, and 13% by more than 0.9. These figures were 63%, 29%, and 4%, respectively, for BRCA2. The pathogenicity probabilities estimated using only the in silico data were higher, on average

**Fig. 3** Differences in estimated probabilities of pathogenicity for BRCA1 (top) and BRCA2 (bottom) based on the protein predictor data alone and the functional assay data alone. The data for each variant are depicted as an arrow from the estimated probability given only the functional assay data to the estimated probability given only the protein predictor summary data. Variants are plotted on the x-axis in order of the location of the altered residue. Functional regions are highlighted and labeled.
than those estimated using only the functional assay data for both BRCA1 (mean difference = 0.22; sd = 0.40) and BRCA2 (mean difference = 0.13; sd = 0.41). In BRCA1, the mean difference was 0.22 (SE = 0.050) in the BRCT1 domain, 0.28 (SE = 0.049) in the BRCT2 domain, 0.13 (SE = 0.10) in the Coiled Coil domain, 0.42 (SE = 0.043) in the Linker region and 0.073 (SE = 0.034) in the remaining, unannotated regions. In BRCA2, the mean difference was 0.17 (SE = 0.048) in the Helical region of the DNA Binding Domain, 0.087 (SE = 0.059) in OB1, 0.052 (SE = 0.053) in OB2, 0.34 (SE = 0.082) in OB3 and 0.015 (SE = 0.16) in the unlabelled regions. These results, coupled with the poor operating characteristics of the protein predictor model (Table 1), suggest that the discrepancies are driven by a lack of calibration and refinement in the protein predictors.

The effects of the Align-GVGD score were similar to those associated with the in silico protein predictor data. The effects on BRCA1 classifications were modest. Sensitivity and specificity estimates were unchanged (Table 1) with only eight benign calls changing to no-call VUS and four changing from no-call VUS to pathogenic. Differences in log-odds of pathogenicity (see Supplementary Fig. 2, left panel) were small. The effects of adding Align-GVGD scores on BRCA2 classifications were more striking, with 44 variants called benign from functional data alone being reclassified as no-call VUS. Furthermore, 23 variants classified as no-call VUS reclassified as pathogenic. As a result, the estimated sensitivity decreased, the estimated specificity increased and differences in log-odds of pathogenicity were large (Supplementary Fig. 3). The systematic shift evident in this plot can be traced to a discrepancy between the Align-GVGD scores and the variant-level probabilities based on the functional data alone; across all variants, the Align-GVGD probabilities were larger by 0.155 (95% CI 0.099 to 0.212; p = 1.30 × 10⁻⁵); among the VUS by 0.179 (95% CI 0.116 to 0.242; p = 6.64 × 10⁻⁶ and, in contrast, among BRCA1 VUS by −0.00272 (95% CI −0.0465 to 0.0410; p = 0.903).

The Align-GVGD score also strengthened the original VarCall estimates for likely benign BRCA1 VUS, but had little effect on the likely pathogenic variants or on BRCA2 variants. The results we observe here may overstate Align-GVGD’s potential to improve predictive accuracy since it was also used as a prior probability in the multifactorial model used to establish the IARC classifications of the variants we use to assess accuracy.

DISCUSSION

We have extended the VarCall model to incorporate additional forms of data related to the pathogenicity of VUS and have characterized the predictive performance based on two functional data-only and 13 unique combinations of functional assay data, in silico protein predictions, Align-GVGD score, and a summary of cancer family history. The resulting model, referred to as VarCall XT, is structured to facilitate the addition of further forms of evidence, including data from additional functional assays, genetic analyses, and tumor pathology characteristics. VarCall XT provides a prediction of the probability that each VUS is pathogenic based on data from multiple sources, thus providing a measure for assessing the strength of evidence in favor of or against pathogenicity and for inclusion in VUS classification schema.

Available data for variant classification are expected to become richer over time as more individuals are tested, new models for in silico prediction of protein function are developed, other functional assays are developed and standardized, assays are extended to more VUS, and other forms of evidence become available. VarCall XT provides a flexible platform for updating predictions of pathogenicity as these new data arise. VarCall XT assumes that VUS have binary risk profiles, behaving like the wild-type sequence or pathogenic, protein-truncating variants. However, there are likely to be hypomorphic variants and distribution of penetrance values intermediate to these two extremes. The VarCall model can be easily adapted to allow for a different distribution of risk profiles.

As currently configured, VarCall identifies variants of known pathogenicity with high sensitivity and specificity with functional assays contributing the greatest predictive power to the model. Indeed, considering the functional data alone, 98.4% of BRCA1 variants of known disposition were correctly called (one known benign was a no-call VUS) and 97% of BRCA2 variants were correctly called (one known pathogenic variant was a no-call VUS). This level of accuracy is not surprising as these functional assays are well-calibrated. While it is difficult to identify a single best model, the contribution of the functional data is clear. The least accurate model that included the functional data was (much) more accurate than the most accurate model that did not include functional data. This can be explained, at least in part, as an artifact of statistical power: the variants that we focus on each gene have multiple measurements made by the functional assays. The available family data is much weaker: because many variants are seen rarely, truly informative family data is the exception, not the rule. The utility of in silico protein predictors and Align-GVGD score for the prediction of individual VUS is far less clear. Coverage of the protein predictor data is complete, but its ability to separate pathogenic from neutral BRCA1 and BRCA2 variants is imperfect, and may be because the protein predictor models have been developed to identify broader classes of likely functional variation.

Further, the family data corroborated the functional assays despite their limited extent and the lack of co-segregation data. With continued growth in testing, there is the potential for family history summaries to become much more informative, but this outcome will depend on the availability and quality of these data. Our results highlight the fact that, despite the current shortcomings of these sources of data, accurate classification of VUS can rely heavily on functional assay data. This is and will continue to be especially important for rare VUS that lack informative family data.

Importantly, the functional assays described above cannot be used to classify splice variants. Thus, specific RNA-based splicing assays or functional assays that allow for endogenous splicing within BRCA1 and BRCA2 may be needed for the classification of potential splice variants. Importantly, splicing data can be incorporated into the VarCall XT model as they become available.

An inherent limitation in the VarCall XT approach is the historical definition of pathogenic missense variants in the BRCA1 and BRCA2 genes. Both the BRCA1 transcriptional assay and the BRCA2 homologous recombination assay were clinically calibrated using controls with properties that were well established in the literature. It is therefore not surprising that the most informative predictive model contains functional data, and hence the lowest Brier scores since that score reflects the calibratedness of the predictor. Family history and protein prediction scores contribute some amount of measurement error and bias to classification schemes. However, these effects will diminish as testing-based family data increases in scope and coverage and as protein prediction models increase in accuracy.

DNA testing is a powerful tool used to tailor medical care based on individual cancer risks. However, DNA testing can produce inconclusive results when a VUS is identified. In this common scenario, healthcare providers might not have the information needed to recommend appropriate preventive and early detection steps, or certain therapeutic treatments, such as PARP inhibitors, and family members may not be referred for genetic testing or enhanced screening. Our results indicate that combining functional assay with other forms of data regarding BRCA1 and BRCA2 missense variants can overcome this limitation and lead to a more accurate determination of whether a variant results in increased cancer risk.
METHODS

**BRCA1 transcriptional assay (TA) data**

The BRCA1 transcriptional assay dataset has been described in detail elsewhere. Briefly, the assay is based on the ability of the BRCA1 C-terminal region to activate the transcription of a reporter gene when fused to a heterologous transcription factor DNA binding domain (Gal4 or LexA). This validated assay has been shown previously to have high predictive sensitivity and specificity.

In this dataset, 281 BRCA1 VUS, 42 IARC class 1 or 2 (benign and likely benign) by the IARC sequence variant classification scheme and 21 IARC class 4 or 5 variants (likely pathogenic and pathogenic), and two controls were assayed for transcriptional activity in 126 experimental batches (3,453 individual ratio measurements). Each batch included the wild-type (WT) as a positive control and p.(Met1775Arg) (IARC Class 5) as a negative control. Typically, three replicates of each variant were made within each batch. As previously, we averaged the within-batch replicate log-ratio values and analyzed the resulting variant- and batch-level mean values.

**BRCA2 homologous recombination (HR) assay data**

The BRCA2 homologous recombination assay and the associated experimental protocol have been described previously. The BRCA2 HR dataset analyzed here comprised 2233 individual log-ratio measurements made on a total of 248 variants. Of these, 22 were IARC class 1 or 2 variants that, together with the wild-type (WT) control, were labeled as benign in the analysis. In addition to the loss-of-function (LOF) control p.(Asp2723His), 11 IARC class 4 and 5 likely pathogenic and pathogenic variants were labeled as damaging to protein function. The remaining 166 variants were VUS and were unlabeled in the analysis. The data were collected in 162 experimental batches. The WT and LOF variants were replicated twice in every batch, while all other variants were replicated at least twice in at least one batch. As with the BRCA1 assay data, we averaged the within-batch replicate log-ratio values for each variant and analyzed the 984 resulting within-batch means.

**In silico protein prediction data**

In silico predictions from 27 protein prediction models for all assayed variants in BRCA1 and BRCA2 were obtained from the dbNSFP 

Using these estimated classification probabilities, we refer to these categories as called benign, called pathogenic and not called, respectively.

For each gene, we compared each of the 15 VarCall models on basis of the classification probabilities assigned to the variants of known disposition in the context of a leave-one-variant-out test. We systematically left each variant unlabeled, treating it as if it were a VUS, re-fit the model, and recorded the variant’s estimated classification probability. Based on this value and thresholds established in the literature, we classified the variant as ‘benign’ if $Pr(D = 1 | Data) < 0.05$, as ‘pathogenic’ if $Pr(D = 1 | Data) > 0.99$ and left it unclassified otherwise; to distinguish these from the known classifications, we refer to these categories as ‘called benign’, ‘called pathogenic’ and ‘not called’, respectively.

Using these estimated classifications, we compute and report model- and empirical, count-based estimates of the operating characteristics of the VarCall models. The empirical predictive summary includes tabulations of the estimated classifications conditional on known pathogenicity status; raw predictive accuracy rates that treat variants in the no-call category as incorrectly classified; and the scaled Brier score, which summarizes the discordance between the binary classifications, $D_0$ and the predicted probabilities, $Pr(D | Data)$ that result from a given model. We also compute Bayesian estimates of the sensitivity, specificity, false-positive, false-negative, and no-call rates associated with each VarCall model. In particular, for each model, we estimate the conditional distributions over the estimated classification categories given known disposition based on a Dirichlet-multinomial model for the observed counts assuming Jeffrey’s prior on the multinomial probability vector. We compute and report the posterior mean and 95% Bayesian high-density region (HDR) interval estimates for each operating characteristic based on the fitted model.
Supplemental data
The supplemental data file includes a detailed description of the models; results of an analysis of sensitivity to the fraction of labeled variants; a table of empirical summaries of the 15 VarCall models for BRCA1 and BRCA2; the protein predictor data; a figure of density plots of the top ten principal components of the protein predictor data; a figure of scatter plots of the log-odds of pathogenicity computed using the function data alone versus the function data combined with the Align-GVGD score; and a log-odds of pathogenicity computed using the function data alone versus density estimation. J. Am. Stat. Assoc. 97, 611–631 (2002).

DATA AVAILABILITY
The data are available from the corresponding author upon request.

CODE AVAILABILITY
The analysis scripts are available from the corresponding author upon request.

REFERENCES
1. Eccles, D. M. et al. BRCA1 and BRCA2 genetic testing—pitfalls and recommendations for managing variants of uncertain clinical significance. Ann. Oncol. 26, 2057–2065 (2015).
2. Goldgar, D. E. et al. Integrated evaluation of DNA sequence variants of unknown clinical significance: application to BRCA1 and BRCA2. Am. J. Hum. Genet. 75, 535–544 (2004).
3. Easton, D. 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. 81, 873–883 (2007).
4. Richards, S. et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and genomics and the Association for Molecular Pathology. Genet. Med. 17, 405–424 (2015).
5. Binhch, S. E. et al. Recommendations for application of the functional evidence PS3/BS3 criterion using the ACMG/AMP sequence variant interpretation framework. Genome Med. 12, 1 (2019).
6. Kanay, D. M. et al. Comparative analysis of functional assay evidence use by ClinGen variant curation expert panels. Genome Med. 11, 77 (2019).
7. Fernandes, V. et al. Impact of amino acid substitutions at secondary structures in the BRCT domains of the tumor suppressor BRCA1: Implications for clinical annotation. J. Biol. Chem. 294, 5980–5992 (2019).
8. Guidugli, L. et al. Assessment of the clinical relevance of BRCA2 missense variants by functional and computational approaches. Am. J. Hum. Genet. 102, 233–248 (2018).
9. Woods, N. T. et al. Functional assays provide a robust tool for the clinical annotation of genetic variants of uncertain significance. npj Genom. Med. 1, 16001 (2016).
10. Monteiro, A. N. A., August, A. & Hanafusa, H. Evidence for a transcriptional activation function of BRCA1 C-terminal region. Proc. Natl Acad. Sci. USA 93, 13595–13599 (1996).
11. Plon, S. E. et al. Sequence variant classification and reporting: recommendations for improving the interpretation of cancer susceptibility genetic test results. Hum. Mutat. 29, 1282–1291 (2008).
12. Iversen, E. S., Couch, F. J., Goldgar, D. E., Tavtigian, S. V. & Monteiro, A. N. A. A computational method to classify variants of uncertain significance using functional data with application to BRCA1. Cancer Epidemiol. Biomark. Prev. 20, 1078–1088 (2011).
13. Liu, X., Wu, C., Li, C. & Boerwinkle, E. dbNSFP v3.0: a one-stop database of functional predictions and annotations for human nonsynonymous and splice-site snvs. Hum. Mutat. 37, 235–241 (2016).
14. Kocher, J.-P. A. et al. The Biological Reference Repository (BioR): a rapid and flexible system for genomics annotation. Bioinformatics 30, 1920–1922 (2014).
15. Hart, S. N. et al. Comprehensive annotation of BRCA1 and BRCA2 missense variants by functionally validated sequence–based computational prediction models. Genet. Med. 21, 71–80 (2018).
16. Tavtigian, S. V., Byrnes, G. B., Goldgar, D. E. & Thomas, A. Classification of rare missense substitutions, using risk surfaces, with genetic– and molecular–epidemiology applications. Hum. Mutat. 29, 1342–1354 (2008).
17. Li, H. et al. Classification of variants of uncertain significance in BRCA1 and BRCA2 using personal and family history of cancer from individuals in a large hereditary cancer multigene panel testing cohort. Genet. Med. 22, 701–708 (2020).
18. Biswas, K. et al. A computational model for classification of BRCA2 variants using mouse embryonic stem cell–based functional assays.npj Genom. Med. 5, 52, https://doi.org/10.1038/s41525-020-00158-5 (2020).
19. Dean, N., Murphy, T. B. & Downey, G. Using unlabelled data to update classification rules with applications to food authenticity studies. J. R. Stat. Soc. Ser. C Appl. Stat. 55, 1–14 (2006).
20. McNicholas, P. D. Mixture Model-Based Classification. (Taylor & Francis, 2016).
21. Fraley, C. & Raftery, A. E. Model-based clustering, discriminant analysis, and density estimation. J. Am. Stat. Assoc. 97, 611–631 (2002).
22. O’Neill, T. J. Normal discrimination with unclassified observations. J. Am. Stat. Assoc. 73, 821–826 (1978).
23. Ganesalingam, S. & McLachlan, G. J. The efficiency of a linear discriminant function based on unclassified initial samples. Biometrika 65, 658–662 (1978).
24. Zhou, Z., Iversen, E. S. & Parmigiani, G. Classification of missense mutations of disease genes. J. Am. Stat. Assoc. 100, 51–60 (2005).
25. Parsons, T. et al. Large scale multifactorial likelihood quantitative analysis of BRCA1 and BRCA2 variants: an enigma resource to support clinical variant classification. Hum. Mutat. 40, 1557–1578 (2019).
26. Steyerberg, E. W. et al. Assessing the performance of prediction models: a framework for some common and novel measures. Epidemiology 21, 128–138 (2010).

ACKNOWLEDGEMENTS
This work was funded by the National Institutes of Health via grants numbered R01CA116167, R01CA192393, R33CA253187, P50CA116201, and R01CA225662, as well as by the Breast Cancer Research Foundation and the Moffitt Foundation.

AUTHOR CONTRIBUTIONS
E.S.I., F.J.C., A.N.A.M., and D.E.G. designed the study; F.J.C., K.Y.L., C.H., and A.N.A.M. provided functional assay data; T.P., A.Y., H.L., E.C., and R.K. provided family-based data; S.N.H. and D.E.G. provided in silico prediction data; E.S.I., G.L., E.C.P., and D.E.G. conducted statistical analyses; all authors contributed to manuscript writing.

COMPETING INTERESTS
T.P., A.Y., H.L., E.C., and R.K. were employees of Ambry Genetics during the time this study was conducted. The remaining authors declare no competing interests.

ADDITIONAL INFORMATION
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41525-022-00302-3.

Correspondence and requests for materials should be addressed to Edwin S. Iversen Jr. or Alvaro N. A. Monteiro.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

npj Genomic Medicine (2022) 35

© The Author(s) 2022

Published in partnership with CEGMR, King Abdulaziz University