Comparison of Variant Classification Algorithms Incorporating Clinical and Family History for Breast and Ovarian Cancer

Jacob Clifford, PhD; Mary Helen Black, PhD; Aaron Elliott, PhD; Hsiao-Mei Lu, PhD; Shuwei Li, PhD

Ambry Genetics, Aliso Viejo, CA

Background

- Pathogenic mutations are more likely to occur in high-risk individuals while benign variants are unrelated to personal and family history.
- Two approaches have been developed to assess variant pathogenicity based on personal and family history: **logistic regression model (LRM)[1,2] and history weighting algorithm (HWA)[3]**.
- To assess differences in performance and clinical utility, **we compared the two methods for classification of variants in genes associated with breast and/or ovarian cancer**.
- Trained models can be used to predict pathogenicity or provide evidence for variant assessment, especially for reclassifying Variant of Uncertain Significance (VUS).

Methods

1. Data Collection

- We collected genetic sequence data, and personal (phx) and family history (fhx) for 94,562 Caucasian patients referred for genetic testing between 2014 and 2016.
- For this analysis, we included10,750 probands carrying only one reportable (pathogenic or VUS) variant and 59,831 probands with no reportable variants.

2. Statistical Approaches

HWA: $$P(G=1|\boldsymbol{X}) = \frac{(n_x + P_{prior})}{((n_x + P_{prior}) + (n_y + (1 - P_{prior}))}$$

$n_x = \#$ mutation carriers satisfying X

$n_y = \#$ non-mut. carriers satisfying X

$G =$ mutation carrier status

$P_{prior} =$ mutation prevalence

LRM: $$\log \frac{P(G=1|\boldsymbol{X})}{1 - P(G=1|\boldsymbol{X})} = \beta_0 + \boldsymbol{\beta} \cdot \boldsymbol{X}$$

3. Comparing performance of LRM and HWA

- Explanatory variables were breast cancer diagnosis age, ovarian cancer diagnosis age, male breast cancer, weighted counts of affected family members, and a severity-weighted fhx score.
- Assessed several scenarios using a variety of phx and fhx variables.
- Applied 2-fold and 10-fold stratified cross-validation for each scenario.

4. Variant Assessment

Proband Odds Ratio (OR) of carrying a pathogenic mutation:

$$OR = \left(\frac{P(G=1|X)}{1 - P(G=1|X)}\right) \Big/ \left(\frac{P_{prior}}{1 - P_{prior}}\right)$$

Variant likelihood ratio of being a pathogenic mutation: LRv(*n*), where the variant is carried by *n* probands.

$$LRv(n) = OR_1 * OR_2 ... OR_n$$

Figure 1. Illustration of variant-specific thresholds for VUS reclassification from distributions of pathogenic and non-pathogenic variants based on random sampling from P (X|G)

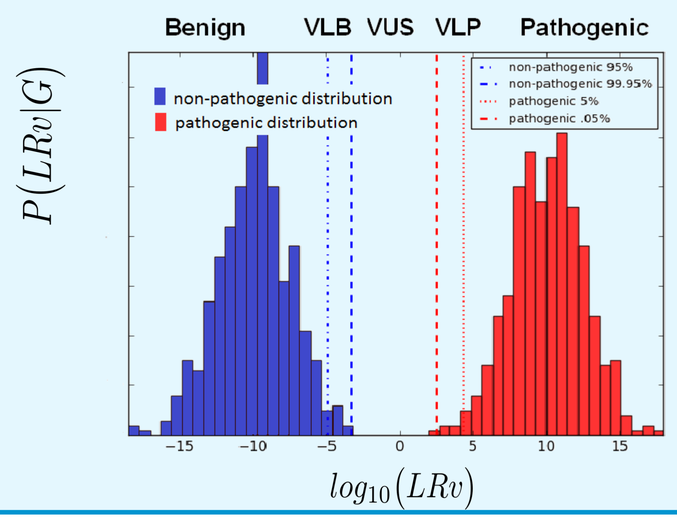
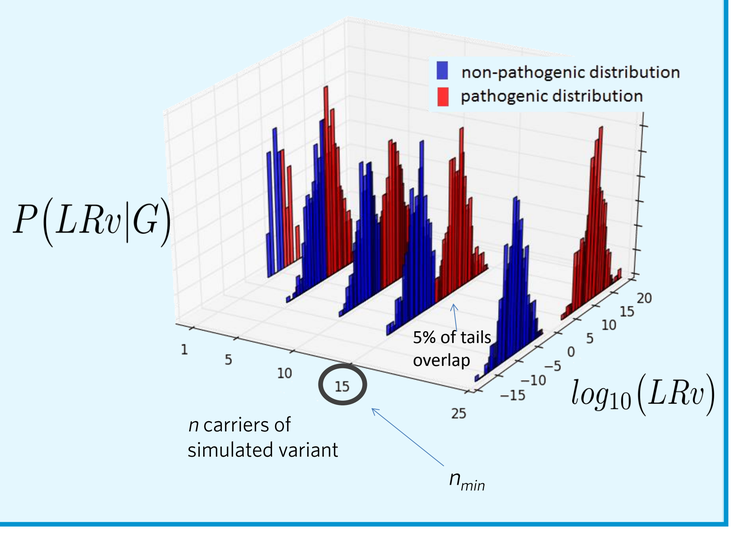


Figure 2. Illustration of minimum number (n_{min}) of carriers required for VUS reclassification per gene. Minimum number occurs when there is a 5% overlap between each conditional distribution.



Results

- For models including phx variables only, both HWA and LRM performed similarly (AUC range: 0.57-0.78 and 0.58-0.79; data not shown).
- For models including phx and severity-weighted fhx score (5-variable model), LRM outperformed HWA for some genes (Table 1).
- For models including phx and breast and ovarian cancer-specific fhx (7-variable model), LRM outperformed HWA for most genes (Table 1).
- High penetrance genes consistently performed better than low penetrance genes regardless of model.
- 158 VUSs were predicted to be benign/VLB based on variant-specific thresholds (Fig 1) given the 7-variable LRM model. As shown in Table 2,15 of the VUSs also meet or exceed the n_{min} required (Fig 2).

Table 1. Model Performance for 5- and 7-VariableScenarios with 2-Fold Cross-Validation

	5-Variable Model[a]			*7-Variable Model*[b]		
Gene	HWA AUC	LRM AUC	P-value[c]	HWA AUC	LRM AUC	P-value[c]
			High-penetrance genes			
BRCA1	0.80	0.81	0.33	0.72	0.84	9.8×10^{-11}
BRCA2	0.72	0.77	7.0×10^{-6}	0.65	0.77	8.2×10^{-10}
CDH1	0.65	0.67	0.47	0.70	0.69	0.86
TP53	0.64	0.73	0.03	0.69	0.77	0.04
PALB2	0.69	0.70	0.55	0.54	0.66	8.4×10^{-4}
PTEN	0.75	0.72	0.45	0.75	0.77	0.71
			Moderate-penetrance genes			
NBN	0.62	0.65	0.37	0.68	0.61	0.17
ATM	0.62	0.62	0.99	0.64	0.60	0.16
CHEK2	0.57	0.66	3.0×10^{-3}	0.70	0.66	0.52

[a] 5-variable model included breast cancer diagnosis age (dxage), ovarian cancer dxage, male breast cancer, multiple primaries of breast cancer, and severity-weighted fhx score.

[b] 7-variable model included breast cancer dxage, ovarian cancer dxage, male breast cancer, multiple primaries of breast cancer and 3 fhx variables: weighted number of relatives with breast cancer dxage<50, breast cancer dxage >50, and affected with ovarian cancer.

[c] P-values were derived from two-sided Delong AUC comparison test.

Table 2. VUSs predicted to be Benign or VLB[a] based on the 7-variable LRM

Genes	Variant	VUS	#Carriers of VUS	n_{min}[b]	LRM predicted class	ClinVar Classification[c]		
		Log_{10}LRv				Benign/VLB	Uncertain	Patho/VLP
PTEN	c.-835C>T	-38.19	37	9	Benign	1	2	0
PTEN	c.-1243G>A	-27.56	21	9	Benign	0	1	0
PTEN	c.-1171C>T	-21.49	30	9	Benign	0	4	0
PTEN	c.1061C>A	-20.97	17	9	Benign	0	4	0
PTEN	c.-976G>A	-13.62	24	9	Benign	0	1	0
PTEN	c.235G>A	-9.19	27	9	Benign	0	5	0
PTEN	c.-844T>C	-8.79	10	9	Benign	0	1	0
PTEN	c.-869G>C	-7.69	9	9	Benign	0	2	0
PTEN	c.-904_-883dup22	-7.09	9	9	Benign	0	1	0
PTEN	c.-944C>T	-1.14	9	9	Benign	0	2	0
PTEN	c.-1196_-1185del12	-1.46	30	9	VLB	0	2	0
TP53	c.847C>T	-12.73	24	13	Benign	0	6	0
TP53	c.869G>A	-7.64	24	13	Benign	0	5	1
NBN	c.511A>G	-4.62	100	38	VLB	1	5	0

[a] VLB: variant likely benign; VLP: variant likely pathogenic

[b] n_{min}: the number of proband carriers required (Fig 2).

[c] ClinVar classifications are based on ClinVar submitters that provided assertion criteria for a given variant class.

Conclusions

- These data demonstrate the performance of each model under a variety of data availability scenarios; LRM outperformed HWA for most scenarios across most genes.
- We combined the strength of LRM and variant-specific thresholds to reclassify VUSs.
- This may have important implications for variant assessment.

References

1. Easton D.F. *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**, 873-83 (2007).
2. Pruss D. *et al.* Development and validation of a new algorithm for the reclassification of genetic variants identified in the *BRCA1* and *BRCA2* genes. *Breast Cancer Res Treat* **147**, 119-32 (2014).
3. Goldgar D.E. *et al.* Integrated Evaluation of DNA Sequence Variants of Unknown Clinical Significance: Application to *BRCA1* and *BRCA2*. *The American Journal of Human Genetics* **75**, 535-544 (2004).