A Model Averaging Approach for Improved *in silico* Variant Prediction in Cancer Genes

Yuan Tian, MA; Shuwei Li, PhD; Mary Helen Black, PhD; Aaron Elliott, PhD; Hsiao-Mei Lu, PhD; Dajun Qian, PhD
Ambry Genetics, Aliso Viejo, CA

Motivation

- A variety of *in silico* scores are currently used as evidence of pathogenicity in variant assessment.
- Bayesian integrated evaluation is a commonly used method to leverage the information from these scores, in which *in silico* predictors are initially selected manually or by stepwise logistic regression (SLR).
- However, such methods often result in using uninformative predictors and do not account their gene-specific information. An accurate and robust gene-specific model is needed to calibrate information from multiple *in silico* tools.

Methods

- A collection of 1,090 classified missense variants among 8 genes (*BRCA1, BRCA2, CDH1, TP53, MLH1, MSH2, MSH6* and *PSM2*), with consensus classifications derived from ClinVar's expert reviewed or those supported from majority submitters, was obtained.
- Gene-specific model averaging logistic regression (MALR) models were trained on known classified variants using scores from 16 standalone publicly-available *in silico* tools as predictors (i.e., Grantham, GERP++, phastCons (vertebrate vs. mammalian), AGVGD, SIFT, MutPred, SiPhy, LRT, phyloP (vertebrate vs. mammalian), Polyphen2 (built on HumVar and HumDiv datasets), MutationAssessor, PROVEAN, FATHMM).
- For each variant, the pathogenicity prediction was estimated based on a Bayesian sampling from Polya-Gamma distributions over the gene-specific MALR models.
- The variation of MALR model coefficients was evaluated by a stratified bootstrapping resampling procedure.
- The prediction performance of MALR models were compared to those of SLR models.

Results

- MALR models showed improved prediction performance and model robustness when compared to SLR models by 5-fold cross-validations. Overall, 77.4% of benign/likely benign variants and 54.9% of pathogenic/likely pathogenic variants were correctly categorized by MALR models (Figure 1).
- When comparing prediction outcomes of MALR models to those of SLR models, the area under the curve (AUC) increased from 0.91 to 0.96, the proportion of predicted variants of uncertain significance (VUS rate) decreased from 36.0% to 28.7%, and positive and negative prediction values (PPV, NPV) increased by 2.3% and 7.2% respectively (Table 1).
- Essentially, MALR stabilized models by shrinking the variation of predictor coefficients when different samples were employed, especially for genes with small number of classified variants (Figure 2).

Figure 1. Prediction outcomes of MALR and SLR models in 8 genes

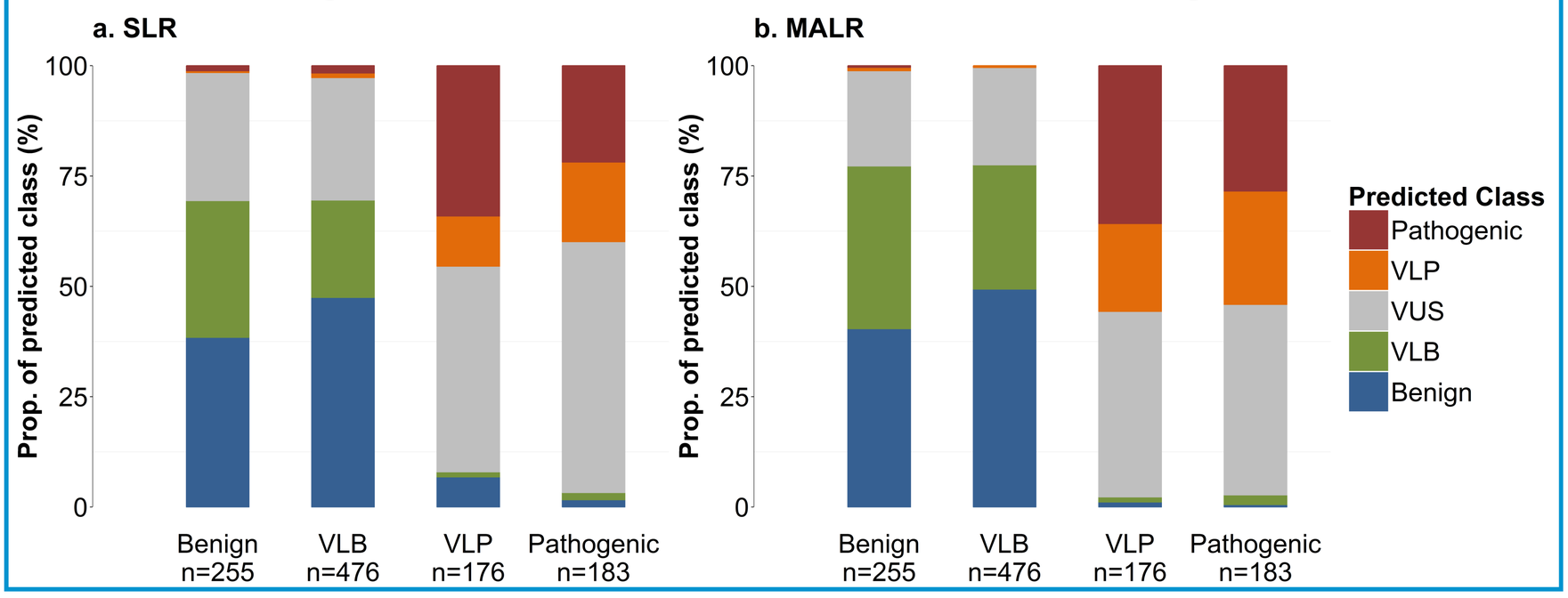

Figure 2. Variation of model coefficients in 1,000 bootstrap resampling analyses

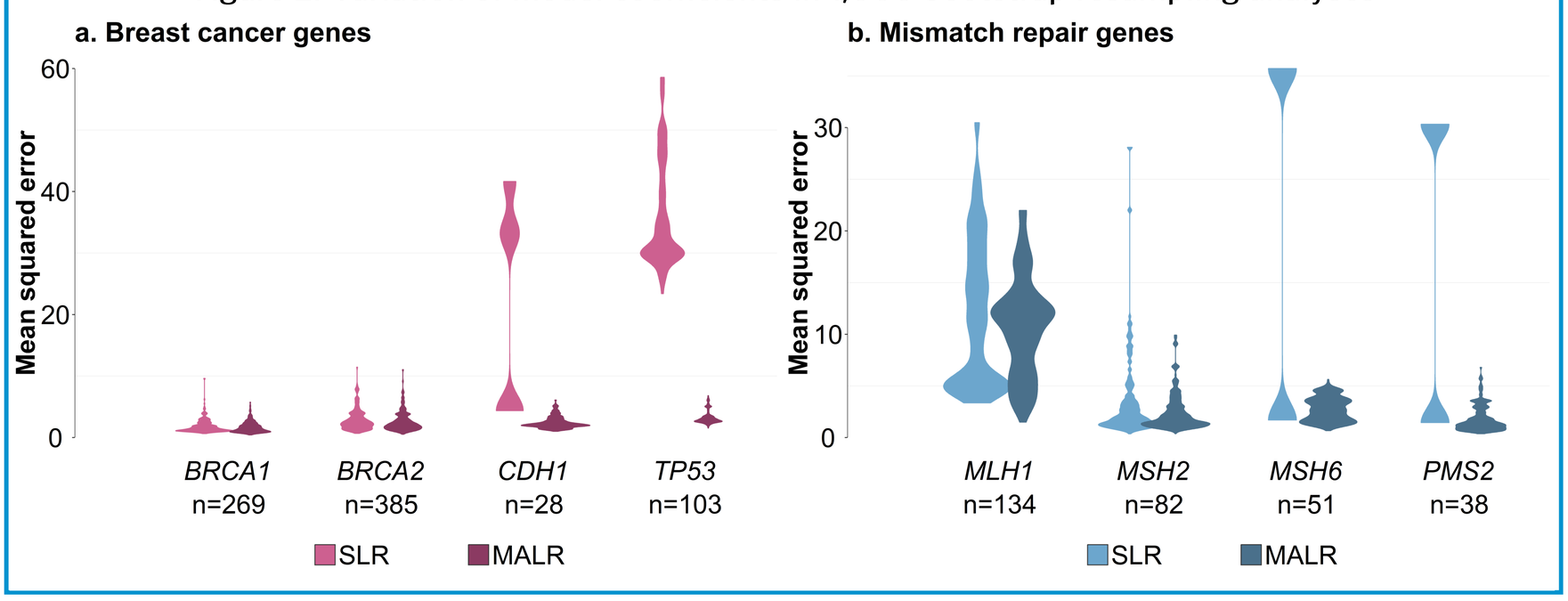

Table 1. Gene-specific performance comparison of MALR vs. SLR

Genes	MALR PPV	SLR PPV	MALR NPV	SLR NPV
BRCA1	99.2%	100.0%	91.3%	82.4%
BRCA2	100.0%	100.0%	0.0%	0.0%
CDH1	94.7%	82.6%	50.0%	0.0%
TP53	100.0%	92.3%	100.0%	98.9%
MLH1	100.0%	100.0%	95.0%	60.0%
MSH2	100.0%	100.0%	100.0%	100.0%
MSH6	97.0%	88.6%	100.0%	66.7%
PMS2	100.0%	93.9%	100.0%	50.0%
Total	99.1%	96.8%	95.6%	88.4%

Genes	MALR VUS rate	SLR VUS rate	MALR AUC	SLR AUC
BRCA1	46.5%	55.8%	0.946	0.922
BRCA2	28.3%	34.0%	0.942	0.917
CDH1	25.0%	7.1%	0.770	0.578
TP53	0.0%	2.9%	1.000	0.964
MLH1	77.6%	91.0%	0.850	0.802
MSH2	85.4%	89.0%	0.878	0.833
MSH6	9.8%	2.0%	0.989	0.838
PMS2	21.1%	2.6%	0.985	0.706
Total	28.7%	36.0%	0.963	0.913

Summary

- Compared to SLR, gene-specific MALR improved model robustness and prediction performance for classification of missense variants using multiple *in silico* scores. This approach may be useful in future variant classification efforts.