**Comparison of variant classification algorithms incorporating clinical and family history for breast and ovarian cancer**

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

Pathogenic mutations are more likely to occur in high-risk individuals while benign variants are unrelated to personal and family history. A proband's clinical and family history may be used to form a likelihood ratio (LR) of the probability that the individual is a mutation carrier vs. non-carrier. A variant LR is computed by estimating the joint probability of carrier status across all probands, which may be derived either from odds obtained from a logistic regression model (LRM) or observed proportions in a clinical population (history weighting algorithm (HWA)). While both methods are useful for variant classification, the relative merits of each under a variety of scenarios have not been directly compared. To demonstrate differences in performance and clinical utility of these approaches, we applied both methods to estimate the LRs of classified variants in *BRCA1* and *BRCA2*, using detailed clinical and family history data from 61,851 patients who underwent genetic testing at a single diagnostic laboratory in 2012-2015.To evaluate the performance of each approach under increasing numbers of independent variables, we tested gene-specific models with 6 and 20 binary variables, respectively. Discriminatory power for correctly classifying variants was then assessed using AUCs with 2-fold cross-validation. We also simulated synthetic variant LR distributions for a fixed number of mutation carriers/non-carriers; the number of carriers was incrementally increased to determine the minimum number of probands such that $\leq$5% of the variant LR distribution for mutation carriers overlapped with the distribution for non-carriers. We observed similar performance, as measured by AUC, for both methods with 6 variables for *BRCA1* (LRM: 0.71, HWA: 0.70) and *BRCA2* (LRM: 0.61, HWA: 0.61). When the number of independent variables increased, LRM had higher AUC (*BRCA1*: 0.83, *BRCA2*: 0.70) than HWA (*BRCA1*: 0.74, *BRCA2*: 0.64). For the 20 variable LRM, our simulations indicated that 40 probands for *BRCA2* and 12 for *BRCA1* were the minimum number necessary to separate 95% of the carrier and non-carrier LR distributions. Variant LR log(e) thresholds were found to predict synthetic variants as benign (*BRCA1/2*:-4.8) and pathogenic (*BRCA*1:5.7, *BRCA2*:5.5) with 95% accuracy. These data show that when the number of variables is large, the LRM may outperform HWA, which may have important implications for variant assessment given the increasing availability of clinical and family history data.

Characters with spaces: 2,500