Title: Accounting for splicing effects in known missense variants improves in silico prediction of deleterious effect

Authors: D. Qian, J. Clifford, A. Tchourbanov, Y. Tian, M.H. Black, H.M. Lu, Z. Zhu, S. Li Institutions: Ambry Genetics, Aliso Viejo, California.

Abstract: To characterize missense variation across the human genome, various in silico predictors exist for assessing deleterious effects from a combination of sequence, structural and evolutionary features. We previously showed that an ensemble in silico variant prediction (IVP) score from a Bayesian logistic regression model had high degree of prediction accuracy. In this study, an ensemble score is first estimated in a subset of 15 in silico predictors (including SIFT, PolyPhen2, PhyloP, PROVEAN and others), and then combined with a splicing impact prediction to form a joint score. The splicing impact prediction score was obtained from a previously developed tool, SpliceScan II, which combines ab initio splicing prediction with evolutionary conservation in the context of known gene structures to assess variant impact on splicing. We derived an ensemble score without accounting for potential splicing (IVPns; IVP with no splicing predictor), and the score accounting for potential splicing (IVPws; IVP with splicing predictor). We then compared the IVPws and IVPns scores in known (likely) pathogenic and (likely) benign missense variants with ClinVar classifications, defined as the most supported category among six submitters providing assertion criteria. For evaluating and reporting prediction performance, we separated the missense variants in 338 genes into two variant sets based on SpliceScan II at a cutoff 0.05 determined by the distribution of classified variants: 1) 9,761 variants with SpliceScan II 0.05 for potential splicing impact. The prediction performance of IVPws and IVPns scores were compared within each of the two variant sets by area under the curve (AUC) statistic using direct cross validation. For variants in 1) with no predicted splicing impact, the AUCs of IVPws and IVPns were equivalent (0.9761 and 0.9756; p =0.58), showing that including splicing score in IVP model didn’t impair the prediction performance. For variants in 2) with potential splicing impact, the AUC of IVPws was 4.8% higher than that of IVPns (0.971 vs. 0.927; p =0.01). To conclude, for missense variants with potential splicing impact, ensemble scores that incorporate splicing information improve in silico prediction of deleterious effects. 6/6/2019 Abstracts | ASHG 2019 Annual Meeting