Title: Expanding and automating the calibration of computational tools for clinical variant classification

Authors & Affiliations: Vikas Pejaver (Icahn School of Medicine at Mount Sinai), Himanshu Sharma (Icahn School of Medicine at Mount Sinai), Timothy Bergquist (Icahn School of Medicine at Mount Sinai), Sarah L. Stenton (Broad Institute of MIT and Harvard, Boston Children's Hospital), Emily A. W. Nadeau (University of Vermont Larner College of Medicine), Alicia B. Byrne (Broad Institute of MIT and Harvard), Marc S. Greenblatt (University of Vermont Larner College of Medicine), Steven M. Harrison (Broad Institute of MIT and Harvard, Ambry Genetics), Sean V. Tavtigian (University of Utah School of Medicine), Anne O'Donnell-Luria (Broad Institute of MIT and Harvard, Boston Children's Hospital), Leslie G. Biesecker (National Institutes of Health), Predrag Radivojac (Northeastern University, Steven E. Brenner (University of California Berkeley)

Abstract: : The 2015 American College of Medical Genetics and Genomics (ACMG) and Association for Molecular Pathology (AMP) variant classification guidelines accorded Supporting (the weakest) evidence to scores from genetic variant impact or effect predictors (VIPs or VEPs). We developed and applied a local posterior probability-based calibration algorithm and showed that, for variants above certain score thresholds, some VIPs can provide stronger evidence for variant pathogenicity/benignity than Supporting. With the ClinGen Sequence Variant Interpretation Working Group, we proposed updated clinical recommendations for the reliable use of VIPs to provide evidence strength up to Strong. Subsequently, new tools using distinctive approaches have been released, and these methods must be independently calibrated for clinical application.

We expanded our calibration to three new VIPs (AlphaMissense, ESM1b, and VARITY), with revised evidence strength levels in alignment with the nascent ACMG/AMP/CAP/ClinGen Sequence Variant Classification v4 (SVC4) point-based system. The newer tools reached the Strong level of evidence for variant pathogenicity and Moderate for benignity, although sometimes for few variants. Compared with four previously recommended tools, these yielded at best only modest improvements in the trade-offs between evidence strength and false-positive predictions. This calibration and recommendation broadens the scope of computational tools for application in clinical variant classification and places these new VIPs on par with previously recommended predictors (and with functional assays for some variants).

Because calibrated outputs from VIPs can now provide more reliable, and sometimes stronger evidence, computational methods are expected to play more prominent roles in the clinical classification of variants and the resolution of variants of uncertain significance (VUS). Our approach is open to new VIPs, different variant types, specific gene(s), and

specific disease(s). To make our calibration approach and its results more repeatable, generalizable, and scalable, we developed a Python tool to calibrate VIPs for clinical variant classification in various contexts, and convert scores from already-calibrated tools to SVC4 clinical evidence strength levels and points, with publication-ready summaries and visualizations. Our work empowers VIP developers, as well as clinical laboratories in assessing and integrating predictive evidence into variant classification