

SUMO ligase challenge : Mooney-Radivojac group - submission 4

Submitters: Vikas Pejaver, Chen Cui, Predrag Radivojac, Sean D. Mooney

A data set of all missense mutations was first created by taking the union of all mutations from all three subsets. MutPred¹, an algorithm for the prediction of pathogenicity of missense mutations, was then run on this data set to obtain scores between zero and one. The underlying assumption of this approach is that scores of pathogenicity are expected to be correlated with the growth assay scores. A MutPred2 score of zero indicates a benign mutation (similar to wild-type) and a score of one indicates a pathogenic mutation (similar to null). The raw prediction scores were then simply “flipped” by subtracting each score from one. No further attempt was made to transform this distribution to the experimental distribution. The motivation behind the use of raw prediction scores is to evaluate the direct utility of pathogenicity prediction scores to prediction problems like the SUMO ligase challenge. One potential drawback of this approach is that, by definition, gain-of-function predictions will never be made. Nonetheless, it appears that these are rare events in the experimental distribution and may still be captured if ranking measures are used for evaluation.

References:

1. Li, Biao, et al. "Automated inference of molecular mechanisms of disease from amino acid substitutions." *Bioinformatics* 25.21 (2009): 2744-2750.