PERFORMANCE OF SNP PATHOGENICITY PREDICTION METHODS

J. Thusberg*

Buck Institute for Age Research, 8001 Redwood Blvd.
Novato, California 94945, U.S.A.
**Email: jthusberg@buckinstitute.org*

A. Olatubosun and M. Vihinen

Bioinformatics Group, Institute of Medical Technology
University of Tampere, FI-33014, Finland
Email: mauno.vihinen@uta.fi

Single nucleotide polymorphisms (SNPs) are the most common form of genetic variation in humans. The number of SNPs identified in the human genome is growing rapidly, but attaining experimental knowledge about the possible disease association of variants is laborious and time-consuming. Several computational methods have been developed for the classification of SNPs according to their predicted pathogenicity. In this study, we evaluated the performance of nine widely used SNP pathogenicity prediction methods available on the Internet. The evaluated methods were MutPred, nsSNPAnalyzer, Panther, PhD-SNP, Pmut, PolyPhen, SIFT, SNAP, and SNPs&GO. The methods were tested with a set of over 60 000 pathogenic and neutral variants. We also assessed whether the type of substituted or substituting amino acid residue, the structural class of the protein, or the structural environment of the amino acid substitution, had an effect on the prediction performance. The performances of the programs ranged from poor (Matthews correlation coefficient (MCC) 0.07) to reasonably good (MCC 0.41), and the results from the programs correlated poorly. The overall best performing methods in this study were PolyPhen, SNAP, and SNPs&GO, with accuracies reaching 0.70, 0.66, and 0.70, respectively. PolyPhen and SNPs&GO performed best when protein structural information was available; otherwise SNPs&GO was the best performing method.

Acknowledgments

The authors would like to thank Pier Luigi Martelli and Rita Casadio from the University of Bologna, Biao Li from Indiana University, and Sean Mooney from the Buck Institute for Age Research, for cooperation in running of data. The financial support from the Tampere Graduate School in Biomedicine and Biotechnology, the Sigrid Jusélius foundation, the Academy of Finland, and the Medical Research Fund of Tampere University Hospital is gratefully acknowledged.

* Corresponding author.