Automated Inference of Molecular Mechanisms of Disease from Amino Acid Substitutions

*Biao Li, **Vidhya G. Krishnan**, Kishore K. Kamati, Matthew E. Mort, Fuxiao Xin, David N. Cooper, Sean D. Mooney, and Predrag Radivojac*

Buck Institute for Age Research, Novato, California and School of Informatics, Indiana University, Bloomington, Indiana

Email : smooney@buckinstitute.org; predrag@indiana.edu

Aim: *We present our method called 'MutPred' that helps to understand the molecular mechanism underlying disease / phenotypic altering amino acid substitutions.*

Methods: *We collected five datasets of human amino acid substitutions from online databases and literature. These include substitutions data from cancer, kinase, HGMD and Swiss-Prot disease and putatively neutral. We generated broad range of attributes using protein sequence based on predicted protein structure, dynamics, functional properties, evolutionary information. To distinguish disease-causing substitutions from neutral we applied Random Forest (RF) classifiers.*

Results: *MutPred, is based on protein sequence, and which models changes of structural features and functional sites between wild-type and mutant sequences. These changes, expressed as probabilities of gain or loss of structure and function, can provide insight into a particular molecular mechanism responsible for the disease state. MutPred uses the established SIFT method but offers improved classification accuracy with respect to human disease mutations.*

Conclusion: *Given conservative thresholds on the predicted disruption of molecular function, we propose that MutPred can generate accurate and reliable hypotheses on the molecular basis of disease for ~11% of known inherited disease-causing mutations. We also note that the proportion of changes of functionally relevant residues in the sets of cancer-associated somatic mutations is higher than for the inherited lesions in the Human Gene Mutation Database which are instead predicted to be characterized by disruptions of protein structure.*
The tool MutPred is available at http://mutdb.org/mutpred.

Acknowledgments

We would like to thank Prateek Kumar and Pauline Ng from the J. Craig Venter Institute for their help with the SIFT tool. We also thank the anonymous reviewers on their comments and suggestions that improved the quality of this work. Funding: NIH R01LM009722 to SDM and NSF DBI-0644017 to PR.