PREDICTION OF DISEASE-ASSOCIATED REGULATORY MUTATIONS IN MONOGENIC AND COMPLEX DISEASE

Yiqiang Zhao, Sean D. Mooney*
Buck Institute for Age Research, 8001 Redwood Blvd, Novato, CA 94945 USA

*Email: smooney@buckinstitute.org
Email: yzhao@buckinstitute.org

Matthew Mort, David N. Cooper
Institute of Medical Genetics, School of Medicine, Cardiff University, Cardiff, United Kingdom

Wyatt Clark
School of Informatics and Computing, Indiana University, Bloomington, Indiana 47408

Understanding the underlying causes of human genetic disease at the DNA sequence level remains an important area for translational bioinformatics. A particular priority is to understand how single nucleotide polymorphisms (SNPs) and disease-causing mutations exert their effects on the expression of gene products. Recently, several studies have explored the relationship between regulatory polymorphisms and SNP-based features, identifying several features which appear to be sufficient to quantitatively discriminate true regulatory polymorphisms from neutral (decoy) polymorphisms. Here, we have extended this work by assessing the utility of both SNP-based and gene-based features for the identification of functional regulatory polymorphisms. Monogenic disease regulatory mutations were directly compared with complex disease regulatory mutations. Gene-based features were found to be much more important than SNP-based features in predicting known regulatory mutations although some SNP-based features still appear to have some utility. When gene-based features are included, regulatory SNPs can be predicted with high accuracy using a balanced approach with the best predictive features being based on gene expression, protein interaction and gene function. The best feature for predicting regulatory polymorphisms associated with monogenic disease genes was found to be the maximum expression level of the associated gene. By contrast, the most important feature for predicting regulatory polymorphisms associated with complex disease genes was found to be protein-protein interaction complexity.

Acknowledgments

This research is supported by R01LM009722 (PI: Mooney),