

# **5<sup>th</sup> Annual Joint Bioinformatics Workshop**

**July 19, 2005**  
2229 Seamans Center  
University of Iowa

by

**Brian O'Leary**  
University of Iowa

## ***A Machine Learning Approach to Optimizing Annotation Parameters in the PAR Algorithm***

### ***ABSTRACT***

The process of screening for disease-causing sequence mutations typically requires the examination of sub-regions of genes. Normally the entire coding region of a gene is examined from the first exon continuing until the entire gene has been screened. Previous work has shown that screening a prioritized sub-region of a gene is more efficient than exhaustively screening the entire gene. The gene attributes used to prioritize these sub-regions of genes, combined with parameters assigned to each attribute, are used in a calculation called Prioritization of Annotated Regions (PAR). While we have shown this approach to be more efficient than previous methods, it has not been optimized based on available attributes. This research aims to improve the PAR calculation and accelerate mutation identification by using known mutations with sequence-based annotation and machine learning techniques to optimize the parameters assigned to each attribute. Our hypothesis is that there are optimal PAR parameter values that will increase the efficiency in using sequence attributes to find disease-related mutations.