

The UCI Chao Family Comprehensive Cancer Center  
Biostatistics Shared Resource Seminar Series  
Statistical Issues in Clinical and Basic Science Research

# ***Identification of Targetable Pathways in Oral Cancers via Random Forest and Chemical Informatics***

Presented by

John P. Schomberg

Ph.D. student  
Department of Epidemiology  
University of California Irvine

Monday, April 10, 2017  
10 – 11 AM  
Sprague Hall Room 207

Differential gene expression (DGE) is one method used to identify gene signatures that stratify patients into groups that inform clinical treatment. An alternative to traditional differential gene expression is to apply a machine learning approach to identify those genes that act as the best classifiers. Identifying candidate drugs is a costly process that benefits from virtual screening methods that reduce the number of small molecules to test in a study of cell lines or murine models. By using only those genes with listed ligand structures for our random forest analysis we can apply a virtual screening method similar to linguistic analysis to ligand structure. This line of informatics will identify existing drugs that are more likely to act on those genes that are most influential in predicting treatment response. This study will also present a novel method applying machine learning to subsets of genes in oral cavity cancer patients listed in the cancer genome atlas. This approach aims to use machine learning and chemical informatics hand in hand to rapidly evaluate genes indicated in influencing treatment response and potential small molecules that may modify the effect of those genes.