

Project Title: Subset Clustering of Clinical Phenotypes based on Molecular Signatures

Company Contact:

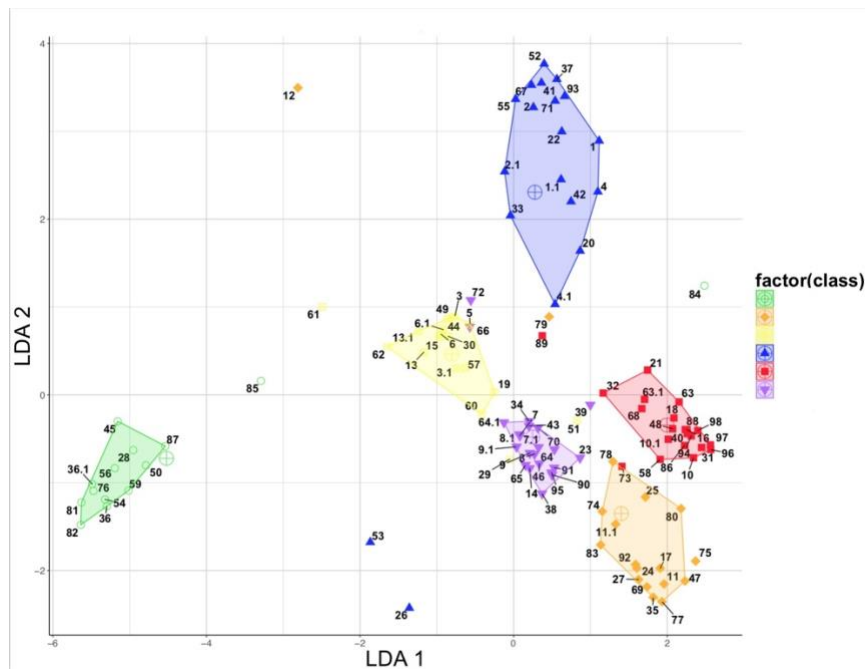
Wolfgang Rumpf

Wolfgang.rumpf@gmail.com

614-638-6797

Project Background Information:

We have collected protein samples from 112 individuals. Each individual falls into one of 6 classes: Control as well as 5 related disease conditions. An LDA-based clustering analysis was able to separate each class into unique clusters based on the full (~650 peptide) protein spectral counts obtained from MALDI-TOF. The first two LDA axes (LDA1 and LDA2) were used as X and Y coordinates in a scatter plot (see attached). Note that several classes have “outliers”, but overall clustering based on spectral counts agrees with clinical phenotype:



Nature of the Research & Future Plans:

This project has two end goals: 1) produce a simple diagnostic test (e.g. a test strip) that can distinguish between each of the six classes, and 2) determine target proteins for further investigation into the mechanism of the disease.

Project Goal:

Your mission is to find a way of reproducing the clustering analysis results (separating subjects into their assigned classes) using 20 or fewer peptides from the data set.

Specific Objectives:

1. Generate a cluster analysis that accurately reflects the patient classes, separating each individual into a cluster of commonly-diagnosed individuals, using 20 or fewer peptides
2. Visualize the cluster analysis and show statistics such as mean distance to centroid for each class, etc.
3. Output the list of proteins for further analysis