Biomarkers measured in minimally invasive and repeatable ways can expedite the early diagnosis of disease, the measurement of therapeutic toxicity, the indication of disease prognosis, and the discovery of new drug targets for therapy. Traditional serum proteomic profiling mainly uses low-resolution mass spectrometry (MS) to identify high-mass protein products. However, the low-molecular weight (LMW) end of the spectrum—that tends to contain more biomarkers—is missed. High-resolution mass spectrometry, such as MALDI-TOF (matrix-assisted laser desorption/ionization orthogonal time-of-flight), is required to analyze rich LMW biomarkers.

In this talk, I will present high-resolution MS data analysis methods for multi-class proteomic biomarker selection and classification. These analysis methods use MALDI-TOF or other high-resolution input-data, which are superior to low-resolution spectra in terms of sensitivity and specificity. First, a multi-class classifier based on a re-designed ECOC (error-correcting output code) and pairwise SVM (support vector machine) ensemble will be presented.

Currently, the major obstacle in analyzing high-resolution MS data has been that the number of data points can easily go beyond one million for a single sample. A subspace extended Markov blanket feature (biomarker, mass/charge ratio) selection algorithm using a wrapper framework will then be shown, which solves this difficulty. Unique biomarker patterns will be found differentiating different phenotypes.

Finally, to identify the molecular formulae of the markers, a two-way parallel searching method will be described, which applies a tandem mass spectrometry (MS/MS) technique. I will conclude the talk with an overview of my other projects including protein microarray for aging signaling pathway identification, medical imaging for sub-cellular structure mobility study, and system biology modeling of biomaterial-mediated foreign body responses.