

*Mānoa Seminar Series in Machine Learning and  
Computational Neuroscience presents:*

*Friday, April 13, 3.10 pm, HL 3F (Hamilton library basement)*

# **Protein Identification in Shotgun Proteomics Experiments**

**Dr. Marina Spivak**

**University of Washington  
Genome Sciences**

Mass spectrometry is a powerful technique in analytical chemistry that was originally designed to determine the composition of small molecules in terms of their constituent elements. In the last several decades, it has begun to be used for more complex tasks, including the detailed analysis of the amino acid sequence that makes up an unknown protein and even the identification of multiple proteins present in a complex mixture. I will discuss the problems of the computational analysis of proteomic data derived from mass spectrometry experiments. The algorithmic procedures used for the mass spectrometry data are closely tied to the specifics of the design of proteomics experiments. Proteins in the mixture are first broken up into smaller pieces (peptides) to minimize the effects of complicated protein chemistry. Then, the analysis of each peptide using the mass spectrometer yields a spectrum that represents a large portion of the full peptide sequence. The challenge of the data analysis consists in reconstructing original proteins from the spectra. The overall analysis of the mass spectrometry data involves several sub-tasks: (1) matching the spectra to candidate peptides, (2) evaluating the quality of these matches and (3) inferring the original proteins from the resulting peptides. In most current approaches, each sub-task is solved by a stand-alone procedure. By doing so, the existing algorithms ignore higher-level information available at the protein level. We integrate these separate stages into a single optimization problem. In the process, we show that the three tasks are cooperative and that solving them in parallel achieves superior results, justifying our approach.

This is collaborative work with collaboration with Jason Weston at Google Research, Bill Noble and Michael MacCoss at the University of Washington.