THURSDAY, OCT. 24
12:15 PM IN 1315 CHEMISTRY

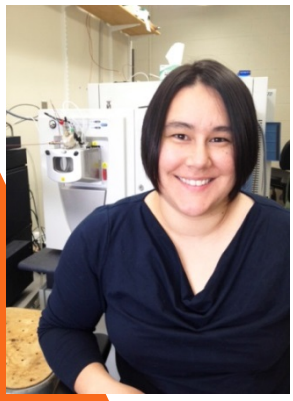
ANALYTICAL SEMINAR

PRESENTED BY

GLORIA SHEYNKMAN

2013 WINNER OF CASEY EXCELLENCE AWARD IN ANALYTICAL RESEARCH

SMITH RESEARCH GROUP



"Enabling the detection of protein variations through integration of mass spectrometry and RNA-Seq"

Protein reference databases required for mass spectrometry-based peptide identification are frequently updated and carefully curated, yet are still incomplete. Proteins exist in many different forms, termed proteoforms, which often are challenging to identify as they may not be present in generic databases. For example, distinct proteoforms can arise from differences in RNA alternative splicing, RNA editing, single nucleotide polymorphisms (SNPs), and insertions and deletions (indels) and these variations may not yet be populated in public databases. Fortunately, the unprecedented capabilities of next generation sequencing (NGS) and RNA-Seq enable the comprehensive measurement of every coding transcript in a sample. One can then empirically determine the mRNA sequences, translate them into protein sequences, and build a customized proteomic database that captures all sample-specific (i.e. specific to an individual) protein variations. This allows for the MS-based proteomics discovery of novel peptides corresponding to novel proteoforms. Using sample matched deep coverage RNA-Seq and proteomics data, we have developed several bioinformatic pipelines that enable the detection of novel peptides through the construction and use of customized proteomic databases. Currently, we are implementing these strategies into open-source workflows within the Galaxy-P framework.