TWO-DAY COURSE, Saturday and Sunday
01 Bioinformatics for Protein Identification

Instructors

Nuno Bandeira
University of California
San Diego

Alexey Nesvizhskii
University of Michigan
Health System

David Tabb
Stellenbosch University

Bioinformatics tools are routinely used to identify proteins from “shotgun” LC-MS/MS data sets. These tools, however, are often poorly understood by their users. This course seeks to familiarize proteomics researchers with the inner workings of the software that enables this field. The course will emphasize the key steps in protein identification: peptide-spectrum matching, error rate estimation, and protein assembly. It will build on these topics by discussing advanced techniques for MS/MS identification through spectral libraries and algorithms that leverage *de novo* sequence inference. Live demonstrations of the Trans-Proteomic Pipeline, ProteoSAFe, and BumberDash / IDPicker tools will familiarize participants with the decisions that produce reliable results from these systems. Participants will emerge from the course with a solid understanding of the underlying algorithms that enable protein identification as well as the software available to accomplish this task.

Day 1:

- Database search algorithms: structure, configuration, and implementations
- Error rate estimation: target/decoy searching and distributional analysis
- Protein parsimony: fielding shared peptides and database selection
- HANDS-ON: The Trans-Proteomic Pipeline, from Seattle Proteome Center (ISB)

Day 2:

- Spectral Library Search: Principles and Resources
- HANDS-ON: BumberDash and IDPicker, from the Tabb Laboratory (Vanderbilt)
- Post-translational modifications: sequence tagging and the challenge of localization
- HANDS-ON: ProteoSAFe, from the Center for Computational Mass Spectrometry (UCSD)