Proteomic data analysis for Orbitrap datasets using Resources available at MSI.

September 28th 2011
Pratik Jagtap

The Minnesota Supercomputing Institute for Advanced Computational Research

http://www.mass.msi.umn.edu/
Proteomics workflow

1. Trypsin
2. Peptides
3. RP HPLC MS/MS
4. Ion Chromatogram
6. Search Against Database
7. Protein Identification

© 2009 Regents of the University of Minnesota. All rights reserved.
Proteomic data analysis for Orbitrap datasets using Resources available at MSI.

- Orbitrap and precision Proteomics.
- Data formats.
- Data processing and effects
- **Proteomics workflow**
  - Search algorithms
  - **Statistical validation of protein identification**
  - Visualization
  - Descriptive Statistics: PDST (ProteinPilot outputs)
  - Pathway analysis
- Applications of workflows: *Horses for Courses.*
**Precision Proteomics**

<table>
<thead>
<tr>
<th>Low Resolution</th>
<th>Medium Resolution</th>
<th>High-Resolution</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 – 0.1 Da accuracy</td>
<td>0.1-0.01 Da accuracy</td>
<td>0.01-0.001 Da accuracy</td>
</tr>
<tr>
<td>Ion Traps, Quadrupoles, triple quadrupoles</td>
<td>Time-of-Flight,* hybrids with quadrupoles</td>
<td>FT ICR MS, FT-Orbitraps, hybrids with ion traps</td>
</tr>
</tbody>
</table>

Bette certainty of protein identifications
Ability to detect polymorphisms, post-translational modifications

“Precision proteomics: The case for high resolution and high mass accuracy”; PNAS (2008) Mann and Kelleher.

Need for newer approaches to assign peptides by taking advantage of increase in mass resolution and accuracy of new mass spectrometers.
Accuracy vs Precision
Accuracy and Precision
Importance of high mass accuracy measurements

“Low” Mass Accuracy: 1500 ppm (LTQ)

631 ± 1

630.0
630.1
630.2
630.3

197 proteins have a peptide in the range of 630 - 632

Upon searching the Mus database:

“High” Mass Accuracy: 10 ppm (Orbitrap)

631.345 ± 0.006

631.339
630.341
630.343

3 proteins have a peptide in the range of 631.339 - 631.351
Orbital Trap Mass Analyzer

- **Axial Frequency is Independent of Initial Energy, Angle, and Position of Ions**

- **Frequency is Similar to a Simple Harmonic Oscillator** — \( \omega \propto (m/z)^{1/2} \)

University of Minnesota Center for Proteomics and Mass Spectrometry currently has:

1 LTQ-OrbitrapXL (ThermoScientific)

It will be upgraded to a Velos-Orbitrap (ThermoScientific) within a few months.
• LTQ-Orbitrap has proven to be a tremendous advance for shotgun proteomics, combining high resolving power, mass accuracy, and reliability in a relatively compact form.

• The high-resolution part of the instrument consists of an Orbitrap analyzer, which is a new type of mass spectrometer developed by Alexander Makarov.

• The LTQ offers high sequencing speed, high sensitivity, and very robust performance but low resolving power.

• The use of LTQ-Orbitrap has lead to higher quality datasets and reduced false positive peptide identifications.
The Orbitrap is capable of very high mass accuracy because of the axial motion and oscillation of ions along the central spindle. Peptide fragmentation is mostly performed in low-resolution but very sensitive and fast linear ion traps.
Schematic of the LTQ Orbitrap Velos MS instrument.

A, the stacked ring ion guide (S-Lens) increases the ion flux from the electrospray ion source into the instrument by a factor 5–10.

B, the dual linear ion trap design enables efficient trapping and activation in the high-pressure cell (left) and fast scanning and detection in the low pressure cell (right).

C, the combo C-trap and HCD collision cell with an applied axial field with improved fragment ion extraction and trapping capabilities.
Conclusions

1. **High Mass Accuracy Measurements Facilitate Database Matching and Increase Peptide Identification Confidence**

2. **Orbitrap Mass Spectrometers Have Capability of Performing Robust and Accurate Quantitative Proteomics Experiments**
Proteomics workflow

Orbitrap

Mass spectral data. (.raw)

Processing

Search Algorithm

Statistical validation of Protein Identification

Visualization

Descriptive Statistics Pathway Analysis

© 2009 Regents of the University of Minnesota. All rights reserved.
<table>
<thead>
<tr>
<th>Mass Spectrometers &amp; Data Formats</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Thermofinnigan</strong></td>
</tr>
<tr>
<td><strong>AB Sciex</strong></td>
</tr>
<tr>
<td><strong>Waters</strong></td>
</tr>
<tr>
<td><strong>Bruker</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><strong>Sequest</strong></td>
</tr>
<tr>
<td><strong>ProteinPilot</strong></td>
</tr>
</tbody>
</table>
**Orbitrap : Data formats**

- **mzXML**: XML (eXtensible Markup Language) based common file format for proteomics mass spectrometric data.

- **mzML**: Joint format developed from mzData and mzXML formats.

- **Sequest (DTA) Files**: Convert .RAW to .dta file using extract_msn.exe utility. Name: Myoglobin_digest.0012.0015.2.dta  
  *File starts with: 1999 2*

- **Mascot Generic Format (MGF)**: ASCII format used as an input for Mascot searches (also OMSSA, X!tandem, Phenyx, ProteinPilot, Myrimatch, TagIdent, etc.). The precursor peptide mass is an observed m/z value, from which relative molecular mass is calculated using the prevailing charge state.  
  \[
  \text{PEPMASS}=1000 \\
  \text{CHARGE}=2+ \\
  \text{Relative molecular mass } M_r \text{ is } 1998. \]
Proteomic data analysis for Orbitrap datasets using Resources available at MSI.

- Data formats.
- Data processing and effects
  - Search algorithms
  - De novo analysis: Peaks
  - Statistical validation of protein identification
  - Visualization
  - Descriptive Statistics: PDST (ProteinPilot outputs)
  - Pathway analysis
- Applications of workflows: Horses for Courses.
### Raw Data

Peaks with wider distribution (Area under curve)

Additional information (Noise Peaks)

### PeakList (Processed)

Peaks (centroid) with intensity values.

Removal of noise peaks.

#### Rat9 SILAC (MaxQuant generated mgf)

<table>
<thead>
<tr>
<th>Mass (Da)</th>
<th>Intensity</th>
</tr>
</thead>
<tbody>
<tr>
<td>216.17106</td>
<td>24726</td>
</tr>
<tr>
<td>232.16430</td>
<td>584018</td>
</tr>
<tr>
<td>238.15761</td>
<td>464704</td>
</tr>
<tr>
<td>337.22811</td>
<td>1.66E-15</td>
</tr>
<tr>
<td>238.06562</td>
<td>6.59E-03</td>
</tr>
<tr>
<td>311.03233</td>
<td>1.13E-03</td>
</tr>
<tr>
<td>344.02660</td>
<td>1.38E-03</td>
</tr>
<tr>
<td>370.26761</td>
<td>2.59E-03</td>
</tr>
<tr>
<td>373.26174</td>
<td>2.03E-03</td>
</tr>
<tr>
<td>374.26425</td>
<td>3.87E-03</td>
</tr>
<tr>
<td>416.22447</td>
<td>2.19E-03</td>
</tr>
<tr>
<td>460.32152</td>
<td>9.88E-03</td>
</tr>
<tr>
<td>461.25520</td>
<td>1.19E-03</td>
</tr>
<tr>
<td>494.40381</td>
<td>3.80E-03</td>
</tr>
<tr>
<td>496.37443</td>
<td>1.93E-03</td>
</tr>
<tr>
<td>502.34441</td>
<td>5.26E-04</td>
</tr>
<tr>
<td>503.46895</td>
<td>1.63E-04</td>
</tr>
<tr>
<td>506.76793</td>
<td>5.02E-04</td>
</tr>
<tr>
<td>571.18562</td>
<td>2.18E-04</td>
</tr>
<tr>
<td>572.48300</td>
<td>1.64E-03</td>
</tr>
<tr>
<td>152.02855</td>
<td></td>
</tr>
</tbody>
</table>

© 2009 Regents of the University of Minnesota. All rights reserved.
Average ppm and Standard deviation improves when MaxQuant processed files are used.
Orbitrap: Processing and effects

- Peak detection
- Noise removal
- Baseline correction
- Monoisotoping
- Charge state

Spectrum #: 1.1.1.11461
Theor. m/z: 547.8194
Precursor m/z: 547.8195
Charge: 2+
Delta Mass: -0.001
Best Sequence: NLAVSQVHVK
Modification:
Conf: 64  Sc: 10
Protein Rank: 711

Spectrum #: 1.1.1.9961
Theor. m/z: 547.8194
Precursor m/z: 547.8193
Charge: 2+
Delta Mass: -0.004
Best Sequence: NLAVSQVHVK
Modification:
Conf: 99  Sc: 12
Protein Rank: 240

© 2009 Regents of the University of Minnesota. All rights reserved.
RAW DATA CONVERSION TOOLS

XRawfile library from ThermoFinnigan Xcalibur software.

- ReAdW
- MSConvert
- Raw2MSM

ProteoWizard

mzXML

http://proteowizard.sourceforge.net/
http://tinyurl.com/Raw2MSM
Proteomic data analysis for Orbitrap datasets using Resources available at MSI.

- **Data formats.**
- **Data processing and effects**
  - **Search algorithms**
  - **Statistical validation of protein identification**
  - **Visualization**
  - **Descriptive Statistics : PDST (ProteinPilot outputs)**
  - **Pathway analysis**
- **Applications of workflows : Horses for Courses.**
Proteomics workflow

Orbitrap Mass spectral data (.raw)

Processing Search Algorithm
SEARCH ALGORITHM

Peptide

© 2009 Regents of the University of Minnesota. All rights reserved.
Mass Spectrometry

Search algorithm

- Sequest
- X!tandem
- OMSSA
- MaxQuant
- ProteinPilot
- Mascot

RDC : sdvlapp32, PC2 and PC3 in SDVL.

CGL 138.

https://sequest7.msi.umn.edu/mascot
Protip / TINT

https://tropix.msi.umn.edu/

John Chilton  Mark Nelson
LTQ-iQuant: A freely available software pipeline for automated and accurate protein quantification of isobaric tagged peptide data from LTQ instruments

Getiria Onsongo¹, Matthew D. Stone², Susan K. Van Riper³, John Chilton⁴, Baolin Wu⁵, LeeAnn Higgins², Troy C. Lund⁶, John V. Carlis¹ and Timothy J. Griffin²

Getiria Onsongo
Protip

Raw Data from Orbitrap

14 node SEQUEST cluster

msconvert

mzxml format

mgf format

dta format

X!TANDEM search
OMSSA search
SEQUEST search

Scaffold Analysis

Scaffold Viewer

Powered by Tropix
Protip

https://tropix.msi.umn.edu/
MAXQUANT

1. Raw Data
2. MAXQUANT QUANT module
3. MSM
4. Mascot .dat file
5. Protein Summary with information that can be used for Pathway Analysis, Functional Enrichment, Localization prediction etc.
6. Protein Identification List, Peptide or MS/MS Summary
IMPROVING MASS ACCURACY

© 2009 Regents of the University of Minnesota. All rights reserved.
References / Weblinks


“Precision proteomics: the case for high resolution and high mass accuracy.”

• http://www.maxquant.org
• http://groups.google.com/group/maxquant-list
• http://tinyurl.com/maxquant-intro
• http://tinyurl.com/maxquant-tutorial

© 2009 Regents of the University of Minnesota. All rights reserved.
MaxQuant : Current availability

• Currently MaxQuant v1.0.13.13 is available on PC1 and PC3
• You can login into these machines using your MSI password.
• Please store your raw files in U: drive
• Transfer your .par files to the C: drive after Quant analysis.
PROTEINPILOT

• MGF peaklist generated from an Orbitrap dataset can be used as an input for ProteinPilot searches.
• Performs normalization and quantification of datasets acquired in HCD mode.
• Multiple features such as modifications, semi-tryptic and non-enzymatic digestion are detected simultaneously.
• Robust FDR estimation.

https://netfiles.umn.edu/users/pjagtap/Proteomics_Webinars_at_MSI/ProteinPilot/Web_Modules.htm
Introduction to ProteinPilot (LeeAnn Higgins and Pratik Jagtap).
https://netfiles.umn.edu/users/pjagtap/Proteomics_Webinars_at_MSI/ProteinPilot_Web_Modules.htm

- **Module One : Introduction to ProteinPilot (LeeAnn Higgins)**
  [http://mediamill.cla.umn.edu/mediamill/embed/66243](http://mediamill.cla.umn.edu/mediamill/embed/66243)
  Time : 23 minutes

- **Module Two : ProteinPilot Outputs**
  [http://mediamill.cla.umn.edu/mediamill/embed/68879](http://mediamill.cla.umn.edu/mediamill/embed/68879)
  Time : 27 minutes

- **Module Three : ProteinPilot Peptide Summary (LeeAnn Higgins)**
  [http://mediamill.cla.umn.edu/mediamill/embed/68701](http://mediamill.cla.umn.edu/mediamill/embed/68701)
  Time : 21 minutes

- **Module Four : ProteinPilot Protein Summary and Quantitation.**
  [http://mediamill.cla.umn.edu/mediamill/embed/68840](http://mediamill.cla.umn.edu/mediamill/embed/68840)
  Time : 15 minutes

- **Module Five : False Discovery Rate Analysis and ProteinPilot FDR output.**
  [http://mediamill.cla.umn.edu/mediamill/embed/68865](http://mediamill.cla.umn.edu/mediamill/embed/68865)
  Time : 25 minutes

- **Module Six : Beyond ProteinPilot. (LeeAnn Higgins)**
  [http://mediamill.cla.umn.edu/mediamill/embed/68866](http://mediamill.cla.umn.edu/mediamill/embed/68866)
  Time : 15 minutes
# ACQUISITION MODE AND DATA ANALYSIS.

<table>
<thead>
<tr>
<th>Data Acquisition Mode</th>
<th>Conversion Tool</th>
<th>Optimal use</th>
<th>Software</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Centroid Mode</strong></td>
<td>Any</td>
<td>Identification</td>
<td>Any</td>
</tr>
<tr>
<td><strong>Profile Mode</strong></td>
<td>Raw2MSM yields better results</td>
<td>MS-based quantification</td>
<td>MaxQuant</td>
</tr>
<tr>
<td><strong>CID</strong></td>
<td>Any</td>
<td>Identification</td>
<td>Any</td>
</tr>
<tr>
<td><strong>HCD</strong></td>
<td>Any</td>
<td>iTRAQ Quantification</td>
<td>ProteinPilot and LTQ-iQuant</td>
</tr>
<tr>
<td><strong>PQD</strong></td>
<td>Any</td>
<td>iTRAQ Quantification</td>
<td>LTQ-iQuant</td>
</tr>
<tr>
<td><strong>ETD</strong></td>
<td>Any</td>
<td>Modifications and c and z ions</td>
<td>OMSSA and COMPASS</td>
</tr>
</tbody>
</table>
Proteomic data analysis for Orbitrap datasets using Resources available at MSI.

- Orbitrap and precision Proteomics.
- Data formats.
- Data processing and effects
- Proteomics workflow
  - Search algorithms
  - Statistical validation of protein identification
  - Visualization
  - Descriptive Statistics: PDST (ProteinPilot outputs)
  - Pathway analysis
- Applications of workflows: Horses for Courses.
Statistical Validation of Peptide and Protein Identification

Orbitrap → Mass spectral data (.raw) → Processing → Search Algorithm → Statistical validation of Protein Identification
False Discovery Rate Analysis

Statistical validation of peptide and protein identifications.

Sennels et al BMC Bioinformatics 2009, 10:179

© 2009 Regents of the University of Minnesota. All rights reserved.
Slide Courtesy: Brian Searle (Proteome Software)

Statistical validation of peptide and protein identifications.

https://www.msi.umn.edu/sw/scaffold-for-pro
Combining results from multiple search algorithms increases the confidence and number of peptide and protein identifications.

HUMAN DATASET

<table>
<thead>
<tr>
<th>Algorithm Combination</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sequest</td>
<td>5522</td>
</tr>
<tr>
<td>X! tandem</td>
<td>5137</td>
</tr>
<tr>
<td>Mascot</td>
<td>5486</td>
</tr>
<tr>
<td>All Together</td>
<td>8162</td>
</tr>
<tr>
<td>Sequest + Mascot</td>
<td>6554</td>
</tr>
<tr>
<td>Sequest + X! tandem</td>
<td>6962</td>
</tr>
<tr>
<td>X! tandem + Mascot</td>
<td>7443</td>
</tr>
</tbody>
</table>
Scaffold : Current availability

• Currently Scaffold (v 3.03) is available on app1.msi.umn.edu
• You can Remote login into app1.msi.umn.edu using your MSI password.
• Please store your files on U: drive
• Scaffold v 3.03 is also available on ProTIP.
<table>
<thead>
<tr>
<th>Feature</th>
<th>ProteoIQ</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>LABEL FREE QUANTITATION</strong></td>
<td></td>
</tr>
<tr>
<td>Normalize between biological samples</td>
<td>✔️</td>
</tr>
<tr>
<td>Support for replicate analyses</td>
<td>✔️</td>
</tr>
<tr>
<td>Calculate statistics on replicates</td>
<td>✔️</td>
</tr>
<tr>
<td>View protein expression plots</td>
<td>✔️</td>
</tr>
<tr>
<td><strong>COMPARATIVE PROTEOMICS</strong></td>
<td></td>
</tr>
<tr>
<td>Compare by difference, intersection and union</td>
<td>✔️</td>
</tr>
<tr>
<td>Apply functions to two or more proteomes</td>
<td>✔️</td>
</tr>
<tr>
<td><strong>STATISTICAL VALIDATION</strong></td>
<td></td>
</tr>
<tr>
<td>Peptide and Protein Probability</td>
<td>✔️</td>
</tr>
<tr>
<td>Protein False Discovery Rate</td>
<td>✔️</td>
</tr>
<tr>
<td>Peptide False Discovery Rate</td>
<td>✔️</td>
</tr>
<tr>
<td><strong>COMPLEX SEARCH QUERIES</strong></td>
<td></td>
</tr>
<tr>
<td>Search by protein or peptide properties</td>
<td>✔️</td>
</tr>
<tr>
<td>Build new protein sets from returned results</td>
<td>✔️</td>
</tr>
</tbody>
</table>

Statistical validation of peptide and protein identifications.
ProteoIQ: Current Availability

- Currently ProteoIQ (v 2.1.03) is available on app1.msi.umn.edu
- You can Remote login into app1.msi.umn.edu using your MSI password.
- Please store your files on U: drive
Proteomic data analysis for Orbitrap datasets using Resources available at MSI.

- Orbitrap and precision Proteomics.
- Data formats.
- Data processing and effects
- Proteomics workflow
  - Search algorithms
  - Statistical validation of protein identification
  - Visualization
  - Descriptive Statistics: PDST (ProteinPilot outputs)
  - Pathway analysis
- Applications of workflows: Horses for Courses.
Mass Spectrometry

Mass spectral data.

Search algorithm

Statistical validation of peptide and protein identifications.

Visualization Descriptive Statistics Pathway analysis
Spectral Visualization

Scaffold Viewer . MaxQuant Viewer

ProteoIQ . ProteinPilot . SeeMS . Spectra Viewer

© 2009 Regents of the University of Minnesota. All rights reserved.
The ProteinPilot™ Descriptive Statistics Template (PDST) Tool

• Description:
  • A tool for users of ProteinPilot™ Software that provides a large amount of additional analyses not provided directly by the software.

• Features:
  • Describe and assure the quality of identification and quantitation results.
  • Optimize acquisition effectiveness with detailed metrics on acquisition redundancy, chromatography, mass accuracy, etc.
  • Characterize digestion, quantitation labeling efficiency, minor artifact modifications, etc.
  • Detect when important features are missing from search space.
  • Provide basic estimation of quantitation false discovery rates for workflows using iTRAQ® reagents.

*Slide content provided by Sean L. Seymour, AB Sciex*
PDST : Descriptive Statistics

- Quality Control of Identification and Quantitation.
- Detailed metrics on acquisition redundancy, chromatography, mass accuracy, etc.
- Characterize digestion, labeling efficiency, minor artifact modifications, etc.
- Provide basic estimation of quantitation false discovery rates for workflows using iTRAQ® reagents.
Metrics of Proteolytic Digestion

The OUTPUT–Characterize Peptides worksheet gives two tables that give aggregate metrics of digestion. Tracking these values can be a very useful way of assuring consistent digestion – for example, in the study of a series of similar samples. The percent of peptides with 1 missed cleavage and the percent semitryptics are particularly useful numbers to watch.

Expected Digestion (Over-cleavage)

<table>
<thead>
<tr>
<th>Peptide Type</th>
<th>Spectra in Selected Set</th>
<th>% of Selected Set</th>
</tr>
</thead>
<tbody>
<tr>
<td>Expected termini</td>
<td>13053</td>
<td>94.9%</td>
</tr>
<tr>
<td>Semi-specific (only one expected terminus)</td>
<td>573</td>
<td>4.2%</td>
</tr>
<tr>
<td>Non-specific (neither terminus expected)</td>
<td>128</td>
<td>0.9%</td>
</tr>
</tbody>
</table>

Missed Cleavages (Under-cleavage)

<table>
<thead>
<tr>
<th>Missed Cleavages</th>
<th>Spectra in Selected Set</th>
<th>% of Selected Set</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>13129</td>
<td>95.5%</td>
</tr>
<tr>
<td>1</td>
<td>623</td>
<td>4.5%</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>0.0%</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>0.0%</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>0.0%</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>0.0%</td>
</tr>
</tbody>
</table>

Slide content provided by Sean L. Seymour, AB Sciex

© 2009 Regents of the University of Minnesota. All rights reserved.
**PDST: Descriptive Statistics**

**Most Frequent Modifications and Substitutions**

- The **OUTPUT—Most Frequent Deltas** worksheet gives three tables summarizing the most frequent modification and substitution features several ways. The left table views features by nominal delta, which means that isobaric and nearly isobaric modifications are all counted as one delta. The middle table breaks this apart by specific feature and exact delta mass. The right table pertains to finding missing features and is discussed in the search space optimization and QC section.

### Most Frequent Nominal Delta Masses

<table>
<thead>
<tr>
<th>Overall Rank</th>
<th>Nominal Delta</th>
<th>Total Count</th>
<th>Modifications</th>
<th>Substitutions</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>304</td>
<td>17945</td>
<td>7762</td>
<td>2154</td>
</tr>
<tr>
<td>2</td>
<td>46</td>
<td>2721</td>
<td>2524</td>
<td>363</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>2313</td>
<td>268</td>
<td>25</td>
</tr>
<tr>
<td>4</td>
<td>16</td>
<td>2609</td>
<td>357</td>
<td>13</td>
</tr>
<tr>
<td>5</td>
<td>-1</td>
<td>2771</td>
<td>215</td>
<td>29</td>
</tr>
<tr>
<td>6</td>
<td>18</td>
<td>705</td>
<td>107</td>
<td>19</td>
</tr>
<tr>
<td>7</td>
<td>25</td>
<td>547</td>
<td>115</td>
<td>19</td>
</tr>
<tr>
<td>8</td>
<td>-17</td>
<td>2259</td>
<td>204</td>
<td>27</td>
</tr>
<tr>
<td>9</td>
<td>6</td>
<td>170</td>
<td>155</td>
<td>27</td>
</tr>
<tr>
<td>10</td>
<td>42</td>
<td>163</td>
<td>139</td>
<td>27</td>
</tr>
<tr>
<td>11</td>
<td>98</td>
<td>156</td>
<td>124</td>
<td>27</td>
</tr>
<tr>
<td>12</td>
<td>-2</td>
<td>109</td>
<td>187</td>
<td>37</td>
</tr>
<tr>
<td>13</td>
<td>-48</td>
<td>89</td>
<td>27</td>
<td>27</td>
</tr>
<tr>
<td>14</td>
<td>15</td>
<td>84</td>
<td>22</td>
<td>27</td>
</tr>
<tr>
<td>15</td>
<td>-14</td>
<td>57</td>
<td>57</td>
<td>27</td>
</tr>
<tr>
<td>16</td>
<td>28</td>
<td>49</td>
<td>6</td>
<td>27</td>
</tr>
<tr>
<td>17</td>
<td>27</td>
<td>42</td>
<td>20</td>
<td>27</td>
</tr>
<tr>
<td>18</td>
<td>48</td>
<td>41</td>
<td>22</td>
<td>20</td>
</tr>
<tr>
<td>19</td>
<td>-28</td>
<td>36</td>
<td>26</td>
<td>26</td>
</tr>
<tr>
<td>20</td>
<td>2</td>
<td>35</td>
<td>2</td>
<td>21</td>
</tr>
<tr>
<td>21</td>
<td>15</td>
<td>33</td>
<td>2</td>
<td>21</td>
</tr>
<tr>
<td>22</td>
<td>32</td>
<td>25</td>
<td>6</td>
<td>17</td>
</tr>
<tr>
<td>23</td>
<td>27</td>
<td>29</td>
<td>27</td>
<td>17</td>
</tr>
<tr>
<td>24</td>
<td>-20</td>
<td>26</td>
<td>27</td>
<td>17</td>
</tr>
<tr>
<td>25</td>
<td>-16</td>
<td>25</td>
<td>27</td>
<td>17</td>
</tr>
</tbody>
</table>

### Most Frequent Single Features

<table>
<thead>
<tr>
<th>Rank</th>
<th>Feature</th>
<th>Exact Delta</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>TRA@N-term</td>
<td>304 2054</td>
</tr>
<tr>
<td>2</td>
<td>TRA@N-term</td>
<td>2313 2054</td>
</tr>
<tr>
<td>3</td>
<td>MethylHcyC</td>
<td>45 99872</td>
</tr>
<tr>
<td>4</td>
<td>O-Methylated</td>
<td>2609 2054</td>
</tr>
<tr>
<td>5</td>
<td>TRA@N-term</td>
<td>304 2054</td>
</tr>
<tr>
<td>6</td>
<td>O-Methylated</td>
<td>15 99940</td>
</tr>
<tr>
<td>7</td>
<td>TRA@N-term</td>
<td>2609 2054</td>
</tr>
<tr>
<td>8</td>
<td>O-Methylated</td>
<td>2609 2054</td>
</tr>
<tr>
<td>9</td>
<td>Dehydrated</td>
<td>18 99355</td>
</tr>
<tr>
<td>10</td>
<td>O-Methylated</td>
<td>2609 2054</td>
</tr>
<tr>
<td>11</td>
<td>O-Methylated</td>
<td>2609 2054</td>
</tr>
<tr>
<td>12</td>
<td>O-Methylated</td>
<td>2609 2054</td>
</tr>
<tr>
<td>13</td>
<td>O-Methylated</td>
<td>2609 2054</td>
</tr>
<tr>
<td>14</td>
<td>O-Methylated</td>
<td>2609 2054</td>
</tr>
<tr>
<td>15</td>
<td>O-Methylated</td>
<td>2609 2054</td>
</tr>
<tr>
<td>16</td>
<td>O-Methylated</td>
<td>2609 2054</td>
</tr>
<tr>
<td>17</td>
<td>O-Methylated</td>
<td>2609 2054</td>
</tr>
<tr>
<td>18</td>
<td>O-Methylated</td>
<td>2609 2054</td>
</tr>
<tr>
<td>19</td>
<td>O-Methylated</td>
<td>2609 2054</td>
</tr>
<tr>
<td>20</td>
<td>O-Methylated</td>
<td>2609 2054</td>
</tr>
<tr>
<td>21</td>
<td>O-Methylated</td>
<td>2609 2054</td>
</tr>
<tr>
<td>22</td>
<td>O-Methylated</td>
<td>2609 2054</td>
</tr>
<tr>
<td>23</td>
<td>O-Methylated</td>
<td>2609 2054</td>
</tr>
<tr>
<td>24</td>
<td>O-Methylated</td>
<td>2609 2054</td>
</tr>
<tr>
<td>25</td>
<td>O-Methylated</td>
<td>2609 2054</td>
</tr>
</tbody>
</table>

### Looking for Surrogate Deltas via Co-Occurrence Frequency (Distinct Peptide-Centric)

<table>
<thead>
<tr>
<th>Rank</th>
<th>Nominal Delta</th>
<th>Actual Count</th>
<th>Expected Count</th>
<th>Enrichment Rate</th>
<th>In Single Delta List</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>690</td>
<td>7258</td>
<td>2257</td>
<td>1.01</td>
<td>Yes</td>
</tr>
<tr>
<td>2</td>
<td>25</td>
<td>2402</td>
<td>4427</td>
<td>1.01</td>
<td>Yes</td>
</tr>
<tr>
<td>3</td>
<td>305</td>
<td>2368</td>
<td>5503</td>
<td>0.99</td>
<td>Yes</td>
</tr>
<tr>
<td>4</td>
<td>305</td>
<td>2368</td>
<td>5503</td>
<td>1.01</td>
<td>Yes</td>
</tr>
<tr>
<td>5</td>
<td>305</td>
<td>2368</td>
<td>5503</td>
<td>0.99</td>
<td>Yes</td>
</tr>
<tr>
<td>6</td>
<td>286</td>
<td>725</td>
<td>808</td>
<td>0.90</td>
<td>No</td>
</tr>
<tr>
<td>7</td>
<td>304</td>
<td>2339</td>
<td>1800</td>
<td>1.25</td>
<td>Yes</td>
</tr>
<tr>
<td>8</td>
<td>47</td>
<td>192</td>
<td>182</td>
<td>1.05</td>
<td>Yes</td>
</tr>
<tr>
<td>9</td>
<td>287</td>
<td>173</td>
<td>136</td>
<td>1.05</td>
<td>Yes</td>
</tr>
<tr>
<td>10</td>
<td>306</td>
<td>1030</td>
<td>670</td>
<td>1.56</td>
<td>Yes</td>
</tr>
<tr>
<td>11</td>
<td>256</td>
<td>98</td>
<td>310</td>
<td>0.88</td>
<td>No</td>
</tr>
<tr>
<td>12</td>
<td>342</td>
<td>118</td>
<td>65</td>
<td>1.02</td>
<td>No</td>
</tr>
<tr>
<td>13</td>
<td>45</td>
<td>85</td>
<td>65</td>
<td>1.33</td>
<td>No</td>
</tr>
<tr>
<td>14</td>
<td>45</td>
<td>85</td>
<td>65</td>
<td>1.33</td>
<td>No</td>
</tr>
<tr>
<td>15</td>
<td>508</td>
<td>67</td>
<td>175</td>
<td>0.87</td>
<td>Yes</td>
</tr>
<tr>
<td>16</td>
<td>302</td>
<td>63</td>
<td>66</td>
<td>0.92</td>
<td>Yes</td>
</tr>
<tr>
<td>17</td>
<td>62</td>
<td>54</td>
<td>77</td>
<td>0.99</td>
<td>Yes</td>
</tr>
<tr>
<td>18</td>
<td>135</td>
<td>48</td>
<td>96</td>
<td>2.50</td>
<td>Yes</td>
</tr>
<tr>
<td>19</td>
<td>72</td>
<td>40</td>
<td>91</td>
<td>0.78</td>
<td>No</td>
</tr>
<tr>
<td>20</td>
<td>94</td>
<td>39</td>
<td>51</td>
<td>1.22</td>
<td>Yes</td>
</tr>
<tr>
<td>21</td>
<td>2</td>
<td>38</td>
<td>45</td>
<td>0.87</td>
<td>Yes</td>
</tr>
<tr>
<td>22</td>
<td>27</td>
<td>36</td>
<td>25</td>
<td>1.01</td>
<td>Yes</td>
</tr>
<tr>
<td>23</td>
<td>27</td>
<td>36</td>
<td>25</td>
<td>1.01</td>
<td>Yes</td>
</tr>
<tr>
<td>24</td>
<td>27</td>
<td>36</td>
<td>25</td>
<td>1.01</td>
<td>Yes</td>
</tr>
<tr>
<td>25</td>
<td>2</td>
<td>23</td>
<td>33</td>
<td>0.70</td>
<td>Yes</td>
</tr>
</tbody>
</table>

---

*Slide content provided by Sean L. Seymour, AB Sciex*

© 2009 Regents of the University of Minnesota. All rights reserved.
Pathway Analysis

Software: Ingenuity Pathway Analysis

Supercomputing Institute
for Advanced Computational Research
Pathway Analysis

Database for Annotation, Visualization and Integrated Discovery (DAVID)

Welcome to DAVID 6.7

2003 - 2010

The Database for Annotation, Visualization and Integrated Discovery (DAVID) v6.7 is an update to the sixth version of our original web-accessible program. DAVID now provides a comprehensive set of functional annotation tools for investigators to understand biological meaning behind large list of genes. For any given gene list, DAVID tools are able to:

- Identify enriched biological themes, particularly GO terms
- Discover enriched functional-related gene groups
- Cluster redundant annotation terms
- Visualize genes on BioCarta & KEGG pathway maps
- Display related many-genes-to-many-terms on 2-D view
- Search for other functionally related genes not in the list
- List interacting proteins
- Explore gene names in batch
- Link gene-disease associations
- Highlight protein functional domains and motifs
- Redirect to related literatures
- Convert gene identifiers from one type to another
- And more

http://david.abcc.ncifcrf.gov/

© 2009 Regents of the University of Minnesota. All rights reserved.
Proteomic data analysis for Orbitrap datasets using Resources available at MSI.

- Orbitrap and precision Proteomics.
- Data formats.
- Data processing and effects
- Proteomics workflow
  - Search algorithms
  - Statistical validation of protein identification
  - Visualization
  - Descriptive Statistics : PDST (ProteinPilot outputs)
  - Pathway analysis
- Applications of workflows : Horses for Courses.
Choosing the workflow outputs to maximise identifications: Horses for courses

Mascot

Protein Pilot

Phenyx

Sequest

OMSSA

Xi tandem

ARTIST: G. RENEE GUZLAS
Choosing the workflow outputs to maximise identifications: Horses for courses

**ORKITRAP DATASET**

- Raw Data
  - MAXQUANT
    - QUANT module
    - MSM
      - mgf
      - Mascot
        - .dat file
          - IDENTIFY module
            - Protein Summary with information that can be used for Pathway Analysis, Functional Enrichment, Localization prediction etc.
            - Protein Identification List, Peptide or MS/MS Summary

*ProteinPilot*

- .group file
  - ProteinPilot Descriptive Statistical Template.
  - Predicted PTMs.
  - Thorough FDR Analysis.
  - Protein Summary
  - Peptide Summary

© 2009 Regents of the University of Minnesota. All rights reserved.
Orbitrap: Processing and Effects

![Graphs and charts showing precursor ppm and mz error scatter plots.](image)
Protein identification: Orbitrap datasets
Conclusions

• Multiple search algorithms and statistical validation tools are available for analysis. (including Quantitative Analysis Tools)
• Various visualization, descriptive statistics and pathway tools are available.
• Outputs from various workflows can be creatively merged to harness the best features of each workflow.