Proteomics with Galaxy

Ira Cooke
Protein ID by Tandem Mass Spectrometry

Proteins → Digestion & ionisation → Peptides → Fragmentation → Peptide Fragments → MS → MS/MS → Analysis → Protein ID's

Protein ID's:
- B3AT_HUMAN
- DEMA_HUMAN
- EPB42_HUMAN
Protein ID by Tandem Mass Spectrometry
The diagram illustrates a workflow for protein analysis. The process begins with raw data from Bruker and Thermo, which are subjected to data analysis. This analysis proceeds through tools such as CompassExport and mSConvert, which process files like mtf and mzML. Further analysis includes searches using Mascot, OMSAA, and Tandem Search, each generating outputs like pepXML. Data is then subsetted and indexed within the database. Annotations are fetched, and protein information is organized in a table:

<table>
<thead>
<tr>
<th>Protein</th>
<th>Probability</th>
<th>Protein Info</th>
</tr>
</thead>
<tbody>
<tr>
<td>B3AT_HUMAN</td>
<td>1.0</td>
<td>...</td>
</tr>
<tr>
<td>DEMA_HUMAN</td>
<td>0.99</td>
<td>...</td>
</tr>
<tr>
<td>EPB42_HUMAN</td>
<td>0.99</td>
<td>...</td>
</tr>
</tbody>
</table>
Mascot Search

OMSAA Search

Tandem Search

mgt → mzXML

Search Parameters

mascot2XML

cleanup pepXML

tandem2XML

Peptide Prophet

iProphet

Protein Prophet

protXML

Metadata Fetch

<table>
<thead>
<tr>
<th>Protein</th>
<th>Probability</th>
<th>Protein Info</th>
</tr>
</thead>
<tbody>
<tr>
<td>B3AT_HUMAN</td>
<td>1.0</td>
<td>...</td>
</tr>
<tr>
<td>DEMA_HUMAN</td>
<td>0.99</td>
<td>...</td>
</tr>
<tr>
<td>EPB42_HUMAN</td>
<td>0.99</td>
<td>...</td>
</tr>
</tbody>
</table>
Proteomics Toolkit
“protk”
Proteomics Toolkit
“protk”
Proteomics Toolkit
“protk”

TPP | Open MS | Search Engines | Databases
Proteomics Toolkit
“protk”

TPP-Galaxy

TPP
Open MS
Search Engines
Databases
Proteomics Toolkit
“protk”
<table>
<thead>
<tr>
<th>Protein</th>
<th>Peptides</th>
<th>Coverage</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>41_HUMAN</td>
<td>HSNLMLELDLKQSEEIKK 2 more</td>
<td>5.7%</td>
<td>1.0</td>
</tr>
<tr>
<td>B3AT_HUMAN</td>
<td>GTIVLDIQETSLAGVANQLDLR 1 more</td>
<td>4.5%</td>
<td>1.0</td>
</tr>
<tr>
<td>SPTA1_HUMAN</td>
<td></td>
<td>12.5%</td>
<td>1.0</td>
</tr>
<tr>
<td>ACTB_HUMAN</td>
<td>TGTIVMDSDGDSGVTHVTPYEGYALPHAILR 1 more</td>
<td>10.7%</td>
<td>1.0</td>
</tr>
<tr>
<td>ANK1_HUMAN</td>
<td>GTFPLYMACENLHLEVK 5 more</td>
<td>6.8%</td>
<td>1.0</td>
</tr>
<tr>
<td>SPTB1_HUMAN</td>
<td>ALEDLOAELEKENYIDOKR 14 more</td>
<td>13.9%</td>
<td>1.0</td>
</tr>
<tr>
<td>DEMA_HUMAN</td>
<td>HLSAEFSR Show coverage</td>
<td>2.2%</td>
<td>0.993</td>
</tr>
<tr>
<td>HBB_HUMAN</td>
<td>LHWDPENFR Show coverage</td>
<td>6.1%</td>
<td>0.9841</td>
</tr>
<tr>
<td>GSTA1_HUMAN</td>
<td>EATLRTFOETK/GGEFARTEFEEFSG Show coverage</td>
<td>10.0%</td>
<td>0.9338</td>
</tr>
</tbody>
</table>
Genomics and Proteomics
Genomics and Proteomics

- Proteomics can be used to improve genome annotation. “Proteo-genomics”
Genomics and Proteomics

- Proteomics can be used to improve genome annotation. “Proteo-genomics”

Verify Expression
Genomics and Proteomics

- Proteomics can be used to improve genome annotation. “Proteo-genomics”
  Verify Expression
  Correct gene predictions
Genomics and Proteomics

- Proteomics can be used to improve genome annotation. “Proteo-genomics”
  Verify Expression
  Correct gene predictions
  Find new genes
Genomics and Proteomics

• Proteomics can be used to improve genome annotation. “Proteo-genomics”
  Verify Expression
  Correct gene predictions
  Find new genes

• Genomics workflows are used to make protein databases
Genomics and Proteomics

• Proteomics can be used to improve genome annotation. “Proteo-genomics”
  Verify Expression
  Correct gene predictions
  Find new genes

• Genomics workflows are used to make protein databases
  Emerging model organisms
Genomics and Proteomics

• Proteomics can be used to improve genome annotation. "Proteo-genomics"
  Verify Expression
  Correct gene predictions
  Find new genes

• Genomics workflows are used to make protein databases
  Emerging model organisms
  Variant analysis
Where to get the tools

- Under “Proteomics” on the main toolshed
- [https://bitbucket.org/iracooke/protk](https://bitbucket.org/iracooke/protk)
- [https://bitbucket.org/Andrew_Brock/proteomics-visualise](https://bitbucket.org/Andrew_Brock/proteomics-visualise)
- [https://bitbucket.org/jmchilton/galaxyp-vm-launcher](https://bitbucket.org/jmchilton/galaxyp-vm-launcher)
Acknowledgements

★ Mass Spec & Biochemistry

Fiona Durand, Pierre Faou, Vita Levina, Gert Talbo, Nick Hoogenraad, Hoogenraad Lab Members

★ Programming

Andrew Brock (Protvis), John Chilton

★ Funding

VLSCI, e-Research, CRC for Biomarker translation