Protein Validation (Statistical Inference) and Protein Quantification
## Terminology

- Peptide Spectrum Match
- Target / Decoy
- False discovery rate
- Shared peptide
- Parsimony
- One hit wonders
SPECTRUM

Relative Abundance

$m/z$
Tandem mass spectrometry and peptide sequence

- An MS/MS spectrum contains a mixture of b and y ions
PEPTIDE SPECTRUM MATCH

Mass spectrum

Reference Protein Database from genomic annotation

Peptide Spectral Match
PEPTIDE SPECTRUM MATCH

- Visual inspection
- Scoring function $f(S,P)$ measures the quality of the match between spectrum $S$ and peptide sequence $P$

- Look at fragment ions
  - predicted mass
  - expected intensity
  - compare to next best
  - presence of immonium ions
  - etc.

PEPTIDE SPECTRAL MATCH

Nesvizhskii et al
TARGET-DECOY SEARCH

REVERSE DATABASE SEARCH

Mass spectrum

> IPI:IPI00205563.1|Gene_Symbol=Tmsbl1 thymosin beta-like protein
MSDKPDLSEVFDFDSSKLKNTNTEKNTLPSKETIQQEKEYNQRS

> IPI:REV_IPI00205563.1|Gene_Symbol=Tmsbl1 thymosin beta-like protein
SRQNYKEEQQITKESPLTKNEETNKKTKLKSDFTEVESLDKPDSM
TARGET-DECOY SEARCH

FALSE DISCOVERY RATE ANALYSIS


Inputs: PEAKLISTS and SEARCH db  SEARCHGUI  PEPTIDESHAKER
PROTEIN INFERENCE: FROM PEPTIDES TO PROTEINS

Slide from Alexey Nesvizhskii talk at http://www.scivee.tv/node/12671
Shared Peptides

**Peptides identified:**

<table>
<thead>
<tr>
<th></th>
<th>Peptide Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>TIGGGDDSNTFFSETGAGK</td>
</tr>
<tr>
<td>2</td>
<td>AVFVDLEPTVIDEVRE</td>
</tr>
<tr>
<td>3</td>
<td>QLFHPEQLITGKEDANNYAR</td>
</tr>
<tr>
<td>4</td>
<td>NLDIERPVTYTNLNR</td>
</tr>
<tr>
<td>5</td>
<td>IHFPATYAPVISAEK</td>
</tr>
<tr>
<td>6</td>
<td>AYHEQLSVAEITNAFCFEPANQMVK</td>
</tr>
<tr>
<td>7</td>
<td>YMACCLLYR</td>
</tr>
<tr>
<td>8</td>
<td>SIQFVDWCPPTGFK</td>
</tr>
<tr>
<td>9</td>
<td>VGINYQPPTVVPGDLAK</td>
</tr>
<tr>
<td>10</td>
<td>AVCMLSNTTAIAEAWAR</td>
</tr>
<tr>
<td>11</td>
<td>LDHKFDLMYAK</td>
</tr>
</tbody>
</table>

**Assignment of peptides to proteins:**

<table>
<thead>
<tr>
<th>Proteins</th>
<th>Peptides</th>
</tr>
</thead>
<tbody>
<tr>
<td>P05209</td>
<td>alpha-1</td>
</tr>
<tr>
<td>Q13748-1</td>
<td>alpha-2</td>
</tr>
<tr>
<td>Q13748-2</td>
<td>alpha-2</td>
</tr>
<tr>
<td>NP_006000</td>
<td>alpha-3</td>
</tr>
<tr>
<td>P05215</td>
<td>alpha-4</td>
</tr>
<tr>
<td>Q9BQIE3</td>
<td>alpha-6</td>
</tr>
<tr>
<td>Q9NY65</td>
<td>alpha-8</td>
</tr>
</tbody>
</table>

Nesvizhskii, A. I.; Aebersold, R.
Shared Peptides

Parsimony

SIQFVDWCPTGFK

YMACCLLYR

Tubulin alpha 3

Tubulin alpha 4

Tubulin alpha 6

85%
From Peptides to Proteins

- AEPTIR: 85%
- IDVCIVLLQHK: 65%
- NTGDR: 25%

Protein: (4%)

0.15 * 0.35 * 0.75 = 0.04

One Hit Wonders

IDVCIVLLQHK → 95% Protein

(5%)
One Hit Wonders

• Quantify as a score:
  If different peptides agree: Good!
  If peptides are one-hit-wonders: Bad!

• Protein Prophet, etc.

• Then can use FDR at protein level
Peptide / Protein Quantification
### Terminology

- Absolute Quantification
- Relative Quantification
- Label-free
- Normalization
- Labelled
- iTRAQ
- Reporter ions
Peptide / Protein Quantification

- **Absolute** – Estimate the molar amount of protein / peptide in the biological sample
  - PTMs
  - Validation

- **Relative** – Fold change / statistically significant difference between 2 biological states
  - Biological variation
  - Biomarker studies
Label-free

- Area Under Curve
  - MS1
  - Integrate XIC

- Spectral Counting
  - MS2
  - High abundant proteins

MS Quantification

- MS not inherently quantitative
- Physiochemical properties invoke different MS responses
- MS only samples a small percentage of total peptides
- Bias and variability
Normalization

• Remove bias and variability *between* runs
• Global – commonly used
  – Median scale
  – Total ion current (TIC)
• Local – very recent development
  – Proximity-based intensity normalization (PIN)
Labeled Quantification

- Run samples simultaneously on in a single run
- Add label to samples
- Mix samples together
- Compute ratios / statistically significant diffs.
Labeled

- **Isobaric**
  - MS2, iTRAQ
  - Number of samples
- **Synthetic Peptides**
  - MS1
  - Absolute (AQUA)
- **Metabolic**
  - MS1, SILAC
  - Not higher life forms

iTRAQ® 8-Plex Reagent Chemical Structure

Isobaric Tag
Total mass = 305

Reporter Group
113 – 119, 121 m/z

Balance Group (?)
Mass 184, 186 – 192 m/z

Amine specific peptide reactive group (NHS)
N-hydroxysuccinimide

Applied Biosystems has granted permission to use this slide.
iTRAQ Experiment

Obtain protein-containing sample, extract protein

Proteolytic Digestion

Label peptides with iTRAQ® Reagents

iTRAQ Experiment MS2 Spectrum

Peptide $\text{VAIVVGAPR}$

$\text{MW}_{\text{mono}}$ 1024.62

Protein Match Platelet membrane glycoprotein $\text{11b}$

“Reporter Ion Mass Tags” from which quantitation is calculated

Peptide match is made from product ions, e.g., b- and y-ion series

Peptide = VAIVVGAPR

MW = 1024.62

Protein ID = platelet membrane glycoprotein $\text{11b}$

"Reporter Ion Mass Tags" from which quantitation is calculated
iTRAQ Results

• Reporter ion intensities reflect relative peptide amounts

No change disease:control
iTRAQ Results

• What fold changes are significant?

• Do they represent biological relevance as opposed to experimental variability?

Increase disease:control

Decrease disease:control
<table>
<thead>
<tr>
<th>Label-free quantification</th>
<th>Label-based quantification</th>
</tr>
</thead>
<tbody>
<tr>
<td>counting MS/MS spectra</td>
<td>isobaric peptide labeling (ITRAQ)</td>
</tr>
<tr>
<td>peak integration</td>
<td>labeled synthetic peptides</td>
</tr>
<tr>
<td>lysis and fractionation</td>
<td>metabolic labeling (SILAC)</td>
</tr>
<tr>
<td>digestion</td>
<td></td>
</tr>
</tbody>
</table>