MS\textsuperscript{E} based workflow for quantitative proteomics

Kerman Aloria
Proteomics Core Facility-SGIKER, UPV/EHU
Proteomics Core Facility

Quantitation results
Proteomics Core Facility

Quantification workflow:
- Sample preparation
- Liquid Chromatography
- Mass Spectrometry
- Bioinformatic analysis

Standard, flexible and robust workflow
**MS<sup>E</sup> acquisition mode**

- **MS<sup>2</sup> is a SERIAL process**
- **MS<sup>E</sup> is a PARALLEL process**

**Retention time**

![Graph of retention time vs. intensity for MS and MS<sup>E</sup>](image)

**Low collision energy MS**

![Graph of retention time vs. intensity for low collision energy MS](image)
MS\textsuperscript{E} based protein identification and quantification

- **MS\textsuperscript{E} data**
  - Protein identification
    - **Absolute quantification**: Based on an added standard protein
    - **Relative quantification**: Based on absolute values and peptide intensities

All information in one run
Estimated absolute protein quantification

“the average MS signal response for the three most intense tryptic peptides per mole of protein is constant within a coefficient of variation of less than ±10%”


Comparative analysis of ion currents

Probabilistic algorithm, where relative quantification of all peptides is performed


About the accuracy of the absolute abundance measurement: “...provided an estimated average error rate of 1.8-fold for the extracted ion intensities, and ~threefold for the spectral counting.”

Quantification workflow:

Sample preparation:
- Label-free approach
- 2-D Clean-Up
- RapiGest
- Tryptic digestion

LC: NanoAcquity UPLC System
MS: SYNAPT HDMS
MS^5 acquisition mode

Data analysis:
- ProteinLynx Global Server
- PAnalyzer
- Excel (MS office)
Label-free quantification

Independent and parallel sample analysis

↓

Reproducible manipulation along the whole workflow
Sample preparation:

Sample cleaning and concentration: 2-D Clean-Up

Protein denaturation: Rapigest

Standard in solution tryptic digestion

- Escherichia coli
- Vibrio harveyi: different growth temperatures
- Brucella abortus: ± erythritol treatment
- Rhodococcus sp.: different culture media
- Synechococcus sp.: different culture media
- Saccharomyces cerevisiae
- Botrytis cinerea (plant pathogenic fungus):
  - mycelium of different strains
  - culture media

* Betula pendula: pollen extract
* Phleum pratense: pollen extract

Human:
- Tear (control and disease)
- Depleted plasma
- MDA-MB-468 breast cancer cells (± EGF treatment)
- HEK 293T cells (± sh RNA treatment)
- Adipose Stem Cells: exosomes

Mouse:
- T-lymphocytes: wt and KO
- NIH 3T3 cells: - nuclear proteins (± doxycycline treatment)
  - chromatin bound proteins
Sample preparation

Digestion efficiency: percentage of peptides with missed cleavages

Digestion efficiency independent of the sample
Sample preparation

BPI chromatogram of 6 different sample preparations of protein extracts form human HEK293T cells

Reproducible sample preparation
Overlap in protein identification is not affected by sample preparation
Liquid chromatography

LC reproducibility: Standard deviation of the retention time of replicate analysis

Reproducible LC performance independently of sample type and preparation

*E. coli*: commercial sample (no sample preparation). LC replicates.
*Botrytis* (laboratory sample): LC replicates
*HEK 293T* (laboratory sample): biological replicates
Approximately 70% of the proteins are identified in >1 replicate.

E. coli: commercial sample (no sample preparation). LC-MS\(^E\) replicates.
Botrytis, T-Lymphocytes and NIH 3T3 (laboratory samples) LC-MS\(^E\) replicates
E. coli: commercial sample (no sample preparation). LC-MS<sup>E</sup> replicates. Botrytis, T-Lymphocytes and NIH 3T3 (laboratory samples) LC-MS<sup>E</sup> replicates

~75% of the quantified proteins have a CV < 25%
Bioinformatics

Protein identification

Absolute quantification
- Based on an added standard protein

Relative quantification
- Based on absolute values
- Based on peptide intensities

PAnalyzer
Bioinformatics

**Protein identification**

**Absolute quantification**
- Based on an added standard protein

**Relative quantification**
- Based on absolute values
- Based on peptide intensities
Protein inference: the task of assembling the sequences of identified peptides to infer the protein content of the sample

Identifications in PLGS 2.4:
- Hit: a group of proteins (can have only one member) in which one of them has an exclusive peptide
- Protein: proteins that have at least one identified peptide

Proteins are grouped in hits where one of the proteins has exclusive peptide(s) and the other members have only shared peptide(s)

1 hit and 2 proteins: protein A names the hit
No experimental evidence to select “Polyubiquitin B” to name the hit
2 hits: - Protein A  
- Protein C  
In which hit should be protein B?

One protein assigned to more than one group and counted several times (2 hits, 8 proteins)
Bioinformatics

2 hits: - Protein A
- Protein C
In which hit should be protein B?

ProteinLynx Global Server 2.4: 8 hits, 32 proteins

Totally identified proteins: 11
Conclusive proteins: they have at least one distinct/unique peptide.

Protein group: All peptides are shared and several combinations of proteins can explain the presence of all peptides.

Indistinguishable proteins: all the peptides are shared and at least one of them is distinct/unique for this group.

Nonconclusive proteins: All peptides are shared and their presence is explained by other proteins that must be present in the sample.

Interpretation of Shotgun Proteomic Data. The Protein Inference Problem

**Bioinformatics**

**PAAnalyzer:** A software tool that classifies peptide and protein identifications from ProteinLynx Global Server based on the proposal of Nesvizhskii and Aebersold (with minor modifications).
PAnalyzer reads the results .xml file from ProteinLynx Global Server and classifies the identified proteins based on the defined criteria.

Possibility to filter peptides by score thresholds
Comparison between Panalyzer and PLGS 2.4

<table>
<thead>
<tr>
<th></th>
<th>Panalyzer</th>
<th>PLGS 2.4</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>High confidence</td>
<td>Medium confidence</td>
</tr>
<tr>
<td>Conclusive</td>
<td>149</td>
<td>194</td>
</tr>
<tr>
<td>Indistinguishable</td>
<td>81 (34 groups)</td>
<td>122 (49 groups)</td>
</tr>
<tr>
<td>Group</td>
<td>13 (1 group)</td>
<td>0</td>
</tr>
<tr>
<td>Non conclusive</td>
<td>92</td>
<td>62</td>
</tr>
<tr>
<td>Filtered</td>
<td>54</td>
<td>11</td>
</tr>
<tr>
<td>Total</td>
<td>389</td>
<td>389</td>
</tr>
</tbody>
</table>

PAnalyzer reports more precise and clear results than PLGS 2.4
estimated absolute protein quantification
“the average MS signal response for the three most intense tryptic peptides per mole of protein is constant within a coefficient of variation of less than ±10%”


Quantification in PLGS 2.4:
Absolute quantification of proteins identified with only 1 or 2 peptides is performed

Export data to excel and filter
Results

Commercial samples
(LC-MS\(^E\) replicates only)

**ECOLI_1**
0.5 µg *E. coli* digest

- ADH1_YEAST: 1
- ALBU_BOVIN: 1
- ENO1_YEAST: 1
- PYGM_RABIT: 1

**ECOLI_2**
0.5 µg *E. coli* digest

- ADH1_YEAST: 1
- ALBU_BOVIN: 8
- ENO1_YEAST: 2
- PYGM_RABIT: 0.5

**Identified proteins**

<table>
<thead>
<tr>
<th>Replicate</th>
<th>ECOLI_1</th>
<th>ECOLI_2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>132</td>
<td>180</td>
</tr>
<tr>
<td>2</td>
<td>134</td>
<td>197</td>
</tr>
<tr>
<td>3</td>
<td>32</td>
<td>45</td>
</tr>
<tr>
<td>4</td>
<td>29</td>
<td>35</td>
</tr>
<tr>
<td>5</td>
<td>180</td>
<td>197</td>
</tr>
</tbody>
</table>

**Quantified proteins:**
- > 2 replicates
- > 2 peptides

**Coefficient of variation of the quantified proteins**

![Graph showing coefficient of variation (CV) of quantified proteins]
# Results

<table>
<thead>
<tr>
<th>Protein</th>
<th>Expected ratio</th>
<th>Ratio based on ion current</th>
<th>Ratio based on absolute quantification</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALBU_BOVIN</td>
<td>8.0</td>
<td>7.61</td>
<td>6.95</td>
</tr>
<tr>
<td>ENO1_YEAST</td>
<td>2.0</td>
<td>2.04</td>
<td>1.91</td>
</tr>
<tr>
<td>PYGM_RABIT</td>
<td>0.5</td>
<td>0.55</td>
<td>0.54</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Protein</th>
<th>Expected ratio</th>
<th>Ratio based on ion current</th>
<th>Ratio based on absolute quantification</th>
</tr>
</thead>
<tbody>
<tr>
<td>HSLV_ECOLI</td>
<td>1</td>
<td></td>
<td>0.62</td>
</tr>
<tr>
<td>RL13_ECOLI</td>
<td>1</td>
<td></td>
<td>2.10</td>
</tr>
<tr>
<td>TDCE_ECOLI</td>
<td>1</td>
<td>0.65</td>
<td></td>
</tr>
</tbody>
</table>

False positive quantifications:
- Ion current: 0.05>p>0.95
- 0.66>ratio>1.5
- Absolute quantification: t-test p<0.05
- 0.66>ratio>1.5

Commercial samples (LC-MS$^5$ replicates only):
- **ECOLI_1** 0.5 µg *E. coli* digest
  - ADH1_YEAST: 1
  - ALBU_BOVIN: 1
  - ENO1_YEAST: 1
  - PYGM_RABIT: 1

- **ECOLI_2** 0.5 µg *E. coli* digest
  - ADH1_YEAST: 1
  - ALBU_BOVIN: 8
  - ENO1_YEAST: 2
  - PYGM_RABIT: 0.5

False positive quantifications:
- Ion current: 0.05>p>0.95
- 0.66>ratio>1.5
- Absolute quantification: t-test p<0.05
- 0.66>ratio>1.5
## Results

<table>
<thead>
<tr>
<th>Protein</th>
<th>Theoretical ratio (B/A)</th>
<th>Ratio based on absolute quantification (B/A)</th>
<th>CV Sample A</th>
<th>CV Sample B</th>
<th>p (t-test)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CYC_HORSE</td>
<td>1.50</td>
<td>1.61</td>
<td>1.0</td>
<td>7.2</td>
<td>8.3e-4</td>
</tr>
<tr>
<td>MYG_RABIT</td>
<td>0.38</td>
<td>0.41</td>
<td>7.9</td>
<td>9.5</td>
<td>3.0e-4</td>
</tr>
<tr>
<td>ALDOA_RABIT</td>
<td>0.50</td>
<td>0.4</td>
<td>4.0</td>
<td>2.9</td>
<td>8.7e-5</td>
</tr>
<tr>
<td>ALBU_BOVIN</td>
<td>5.0</td>
<td>2.08</td>
<td>15.9</td>
<td>6.1</td>
<td>7.5e-4</td>
</tr>
</tbody>
</table>

---

**False positive quantifications:**
- t-test p<0.05
- 0.8>ratio>1.25

<table>
<thead>
<tr>
<th>Protein</th>
<th>Theoretical ratio (B/A)</th>
<th>Ratio based on absolute quantification (B/A)</th>
<th>CV Sample A</th>
<th>CV Sample B</th>
<th>p (t-test)</th>
</tr>
</thead>
<tbody>
<tr>
<td>FETP_ECOLI</td>
<td>1</td>
<td>0.2</td>
<td>5.5</td>
<td>7.9</td>
<td>2.3e-5</td>
</tr>
</tbody>
</table>

0.65%  
(1 out of 153)
## Results

### Sample

<table>
<thead>
<tr>
<th>Protein</th>
<th>Sample A</th>
<th>Sample B</th>
<th>Sample C</th>
<th>Sample D</th>
<th>Sample E</th>
</tr>
</thead>
<tbody>
<tr>
<td>CAH1_HUMAN</td>
<td>1</td>
<td>2</td>
<td>5</td>
<td>10</td>
<td>50</td>
</tr>
<tr>
<td>CAH2_HUMAN</td>
<td>1</td>
<td>2</td>
<td>5</td>
<td>10</td>
<td>50</td>
</tr>
<tr>
<td>CRP_HUMAN</td>
<td>1</td>
<td>2</td>
<td>5</td>
<td>10</td>
<td>50</td>
</tr>
</tbody>
</table>

### Protein Levels

<table>
<thead>
<tr>
<th>Protein</th>
<th>E/A</th>
<th>E/B</th>
<th>E/C</th>
<th>E/D</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Theor</td>
<td>Exp</td>
<td>Theor</td>
<td>Exp</td>
</tr>
<tr>
<td>CAH1</td>
<td>50</td>
<td>42.4</td>
<td>25</td>
<td>31.1</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>10.1</td>
<td>5</td>
<td>4.9</td>
</tr>
<tr>
<td>CAH2</td>
<td>50</td>
<td>42.8</td>
<td>25</td>
<td>18.9</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>10.7</td>
<td>5</td>
<td>3.9</td>
</tr>
<tr>
<td>CRP</td>
<td>50</td>
<td>54.2</td>
<td>25</td>
<td>17.7</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>6.4</td>
<td>5</td>
<td>7.1</td>
</tr>
</tbody>
</table>

### Graph

- CAH1
- CAH2
- CRP

<table>
<thead>
<tr>
<th>Sample</th>
<th>CAH1</th>
<th>CAH2</th>
<th>CRP</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>84.9</td>
<td>85.6</td>
<td>108.4</td>
</tr>
<tr>
<td>B</td>
<td>124.5</td>
<td>107.1</td>
<td>98.2</td>
</tr>
<tr>
<td>C</td>
<td>100.7</td>
<td>64.4</td>
<td>79</td>
</tr>
<tr>
<td>D</td>
<td>142.8</td>
<td>100.7</td>
<td>75.6</td>
</tr>
</tbody>
</table>

**S. cerevisiae** background +
**Results**

- **WT**
  - T-lymphocyte purification
  - protein extraction
    - Sample cleaning and digestion
      - LC-MS\(^5\) 5 replicates

- **KO**

**Identified**

- WT: 333 proteins
- E2F2\(^{-/-}\): 331 proteins

**Quantified**

- WT: 288 proteins
- E2F2\(^{-/-}\): 331 proteins

**Deregulated**

- Based on ion current: 41 proteins
- Based on absolute estimation: 44 proteins

* Identified: proteins identified in > 1 replicate
* Quantified: proteins identified in >2 replicates with > 2 peptides
* Deregulated: - ion current: 0.05<p>0.95 ratio>1.25
  - absolute: t-test: p<0.05 ratio>1.25
Results

NIH 3T3
- treatment
+ treatment

Nuclear protein extraction

Cation exchange chromatography pH=8.5

Sample cleaning and digestion

LC-MS\textsuperscript{E} 5 replicates

Identified proteins
quantified

Based on ion current
96

Based on absolute estimation
97

84
deregulated

- Treatment 252
+ Treatment 207

* Identified: proteins identified in > 1 replicate
* Quantified: proteins identified in >2 replicates with > 2 peptides
* Deregulated: - ion current: 0.05>p>0.95 ratio>1.25
        - absolute: t-test: p<0.05 ratio>1.25
Results

Bacteria: 3 growing conditions

3 biological replicates

Extraction of membrane proteins

Sample cleaning and digestion

<table>
<thead>
<tr>
<th>Identified A</th>
<th>Identified B</th>
<th>Quantified</th>
<th>Deregulated</th>
<th>Unique</th>
</tr>
</thead>
<tbody>
<tr>
<td>111</td>
<td>69</td>
<td>56</td>
<td>7</td>
<td>17</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Identified A</th>
<th>Identified C</th>
<th>Quantified</th>
<th>Deregulated</th>
<th>Unique</th>
</tr>
</thead>
<tbody>
<tr>
<td>111</td>
<td>94</td>
<td>62</td>
<td>12</td>
<td>13</td>
</tr>
</tbody>
</table>

* Identified: proteins identified in > 1 replicate
* Quantified: proteins identified in >1 replicate with > 2 peptides
* Deregulated: t-test: p<0.05 ratio>1.25
* Unique: proteins identified 3 times in one condition and 0 in the other
Conclusions

$\text{MS}^E$ is an accurate approach for label-free relative quantification in discovery type of experiments

$\text{MS}^E$ approach can be applied to a wide variety of protein samples
Proteomics Core Facility - SGIKER

Kerman Aloria

Biochemistry and Molecular Biology

Jabier Beaskoetxea
Miren Josu Omaetxebarria
Jesus Mari Arizmendi

Genetics, Physical Anthropology and Animal Physiology

Asier Fullaondo

Electronics and Telecommunications

Gorka Prieto

www.ehu.es
www.ehu.es/sgiker
www.proteored.org