The 2012 PRG study: Assessing longitudinal variability in routine peptide LC-MS/MS analysis
Thanks!

2012 PRG
Maureen K. Bunger (Chair) - North Carolina Central University
Tracy M Andacht - Centers for Disease Control and Prevention
Keiryn Bennett - Center for Molecular Medicine of the Austrian Academy of Sciences
Cory Bystrom - Quest Diagnostics
Larry Dangott - Texas A&M University
John Leszyk - University of Massachusetts Medical School
Henrik Molina - The Rockefeller University
Robert L Moritz - Institute for Systems biology
Brett S Phinney (EB liaison) - Proteomics Core UC Davis Genome Center

To Our Participants!

PRG adhoc members and other supporters
J. Will Thompson – Duke University
Matthew Chambers- Vanderbilt University
David Tabb – Vanderbilt University
Paul Rudnick – National Institutes of Standards and Technologies

RG LC-MS/MS
Dr Rich Eigenheer - UC Davis Genome Center
Christian Knoll - CeMM,

Companies that contributed

BRUKER – MICHRM
bioproximity
Key Goals

- Measure intra-laboratory variation in LC-MS performance over time
- Survey types of QC procedures implemented in proteomics facilities
- Survey elements of system design/setup that correlate with variability
- Educate and Inform
Study Design Principles

- Provide labs with purified, digested, aliquotted protein mixture
- Expect one run per month for 9 months
- Expect fresh aliquot to be used each time
- Expect user to implement same settings for each run, but settings they determine based on their typical setup/comfort
- User only needs to upload raw data
Other study designs that were considered

- Providing labs with both complex (biological organism) and simple protein mixtures
  - Too time consuming and expensive
  - Labs may not want a complex biological on system
  - Sample availability limited

- Requiring targeted MS1 level approach (e.g. monitoring 5-10 peptides)
  - May be out of comfort zone for many
  - Uncertain how results would translate to a more standard DDA identification analysis
Bovine 6 protein mix (Michrom)

• Tryptic digest of the following Proteins:
  – glutamate dehydrogenase 1, mitochondrial precursor
  – serum albumin precursor
  – beta-lactoglobulin precursor
  – carbonic anhydrase 2
  – lactoperoxidase precursor
  – alpha-S1-casein precursor

• 1 pmol (total material) equimolar mix provided in each aliquot (approx 166 fmol of each protein)
Timeline

2012

- Invitation
- Final Participant List
- Samples Shipped

Data collection

2013

Data processing

ABRF 2013
Data pipeline

Participant → Bioproximity “Proteome Cluster” → Vanderbilt University

- LC-MS/MS raw data
  - RAW, WIFF, YEP, etc.
  - ProteoWizard msConvert
  - mz5 files
    - centroided data
    - MyriMatch search engine
    - mzIdentML files
      - unfiltered identifications
      - IDPicker 3.0 protein assembler
      - idpDB files
        - filtered identifications
        - QuaMeter metric computer

PRG → metrics for LC-MS/MS components
QuaMeter Metrics

- MS2.4C Fraction of total MS2 scans ID'd in the third quartile of peptides sorted by MS1 max intensity
- MS2.4A Fraction of total MS2 scans ID'd in the first quartile of peptides sorted by MS1 max intensity
- MS2.4D Fraction of total MS2 scans ID'd in the fourth quartile of peptides sorted by MS1 max intensity
- MS2.4B Fraction of total MS2 scans ID'd in the second quartile of peptides sorted by MS1 max intensity
- C.1A Fraction of all peptides ID'd at least 4 minutes earlier than max MS1 for ID
- C.1B Fraction of all peptides ID'd at least 4 minutes later than max MS1 for ID
- IS.3A Number of 1+ peptides over 2+
- IS.3C Number of 4+ peptides over 2+
- IS.3B Number of 3+ peptides over 2+
- MS2.1 MS2 ion injection time in ms (multiplied by 0.01)
- MS1.1 MS1 ion injection time in ms
- IS.1B Number of times where MS1 signal greatly increased between adjacent scans more than 10 fold
- IS.1A Number of times where MS1 signal greatly decreased between adjacent scans more than 10 fold
- MS1.2A Median signal-to-noise value (ratio of maximum to median peak height) for MS1 spectra
- MS1.2B Median TIC value for ID'd peptides up to and including C.2A (time period)
- MS1.3A Ratio of 95th over 5th percentile MS1 maximum intensity values for ID'd peptides
- MS1.3B Median maximum MS1 value for ID'd peptides
- DS.3A Ratio of MS1 maximum to MS1 value at sampling for median decile of peptides by MS1 maximum intensity
- DS.3B Ratio of MS1 maximum to MS1 value at sampling for bottom 50% of peptides by MS1 maximum intensity
- DS.2B Number of MS2 scans taken of C.2A (times 0.001)
- C.2B Ratio of peptides per minute ID'd during C.2A
- DS.2A Number of MS1 scans taken over C.2A (times 0.01)
- C.2A Time period over which 50% of peptides were ID'd (times 0.1)
- C.3B Measure of the distribution of the peak widths
- C.3A Median peak widths for all ID'd unique peptide(s)
- P.3 Ratio of semi/fully tryptic peptide IDs (times 100)
- IS.2 Median m/z for all ID'd peptides(unique ions) times 0.001
- DS.1A Ratio of peptides ID'd by one spectrum to number ID'd by two spectra
- DS.1B Ratio of peptides ID'd by two spectra to number ID'd by three spectra
- P.2C Number of unique tryptic peptide sequences ID'd
- P.2B Number of tryptic peptide ions ID'd; ions differing by charge state and/or modification are counted separately
- P.2A Spectral count

## Metrics part II: The Survey

<table>
<thead>
<tr>
<th>Variables</th>
<th>Maintenance</th>
<th>Chromatography</th>
<th>Mass Spectrometry</th>
<th>Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Calibration</td>
<td>Gradient</td>
<td>DDA settings</td>
<td>Load amount</td>
</tr>
<tr>
<td></td>
<td>Tuning</td>
<td>Flow rate</td>
<td>Exclusion list settings</td>
<td>Dissolution solution</td>
</tr>
<tr>
<td></td>
<td>Preventative Maintenance</td>
<td>Column Specifics</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Column change</td>
<td>Mobile Phases</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- Mobile Phases
  - Mass Spectrometry
  - Sample
  - Load amount
  - Dissolution solution
Testing Data example: IQ RT period for IDs

RG participant 1
Velos-Orbitrap

RG participant 2
LTQ
Testing Data Example: Median Peak widths

RG participant 1
Velos-Orbitrap

RG participant 2
LTQ
Contaminants

- lactotransferrin precursor [Bos taurus]
- alpha-S2-casein precursor [Bos taurus]
- superoxide dismutase [Cu-Zn] [Bos taurus]
- cationic trypsin precursor [Bos taurus]
Number of Participants

• Started with approximately 100 requests
• Scared about 30 people away by sending out an e-mail that required a response before we would send out the samples.
• Approximately 70 remaining
Participant Demographics

Other (Australia, Turkey, Singapore, Israel, South Africa, India) - 9.1%
Europe - 40.3%
North America - 50.6%
Instrument diversity

- Shimadzu: 1
- Unknown: 1
- Agilent: 3
- Bruker: 6
- Waters: 17
- AB Sciex: 19
- Thermo-Fisher: 61
Instrument Diversity part II

Pie charts showing the distribution of different types of instruments:

- **Thermo FTMS**: 4% (left chart)
- **Thermo Ion Trap**: 2% (left chart)
- **Thermo Orbitrap**: 7% (left chart)
- **Thermo Q-Exactive**: 18% (left chart)
- **Thermo Unknown**: 69% (left chart)

- **AB Sciex Q-TOF**: 19% (right chart)
- **AB Sciex Q-Trap**: 33% (right chart)
- **AB Sciex TOF/TOF**: 5% (right chart)
- **AB Sciex Triple-TOF**: 43% (right chart)

- **Waters Q-TOF**: 6% (bottom chart)
- **Waters QTOF Premier**: 6% (bottom chart)
- **Waters Synapt G2**: 12% (bottom chart)
- **Waters Synapt G2 HDMS**: 6% (bottom chart)
- **Waters Synapt HD**: 6% (bottom chart)
- **Waters Synapt HDMS**: 12% (bottom chart)
- **Waters Ultima Q-tof**: 29% (bottom chart)
Initial upload info: 42 Survey entries

75% of participants have more than 5 years experience
QC variability among participants

Did you perform a system suitability or QC test prior to running the PRG sample?

- Yes: 99.1%
- No: 0.9%

How often do you perform the QC test?

- Several times per day: 30%
- Once a day: 20%
- Once every few days: 15%
- Once a week: 10%
- Once a month: 5%
- Other (please specify): 5%

Describe your QC/System suitability method:

- Direct infusion/acquisition: 99.1%
- LC-MS/MS with Data Dependent Data Acquisition (DEA): 2.6%
- LC-MS/MS with targeted acquisition: 0%
- LC-MS only: 0%

Your primary QC/System suitability sample is:

- Digested single protein: 73.7%
- A known peptide or mix of known peptides: 13.2%
- Digested complex protein mix: 13.2%
- Molecule different from peptide: 0%
Other Variables

- Flow rate: range 100nL/min – 2uL/Min (one reported 150uL/min)
- Amount injected: range 40fmol – 1pmol
- Resolving gradient: range 11 min – 100 min
Goals for next year:

- Collect Data
- Harass participants
- Decide what Metrics vary most across labs
- Try to determine what metrics correlate with each other
- Correlate survey data with variability data
- Try to figure out a long term solution to make sure the raw data is available for future studies or different tools (1-2TB)