iPRG 2011:


ABRF 2011, San Antonio, TX

2/20/11
INTRODUCTION:
CID AND ETD FROM 30K FEET
Collision Induced Dissociation (CID) relies on a series of bimolecular events (collisions) to provide the peptide precursor with sufficient energy to fragment (ergodic process). CID typically causes backbone fragmentation. \( y \) and \( b \) ions are by far the most prevalent fragment types.

Electron Transfer Dissociation (ETD) relies on the transfer of a single electron to a peptide precursor. This transfer likely creates a radical that very quickly decays into ion fragments (a \textit{non-ergodic} process). Like CID, ETD typically causes backbone fragmentation, but mostly resulting in \( c \) and \( z \) ions.
CID and ETD spectra - example
iPRG 2011 STUDY: CONCEPT
Study Goals

• **Primary:** Evaluate the ability of participants to identify ETD spectra

• **Secondary:** Find out why result sets might differ between participants

• **Tertiary:** Produce a benchmark dataset, along with a spectral library and an analysis resource
Study Design

• Use a common, rich dataset
• Use a common sequence database
• Allow participants to use the bioinformatic tools and methods of their choosing
• Use a common reporting template
• Report results at an estimated 1% FDR (at the spectrum level)
• Ignore modification localization
• Ignore protein inference
The sample

NIST yeast lysate (six vials of RM8323), 228µg protein, LysC digest separated on SCX column

Sample prep by
Robert Chalkley,
UCSF

SCX by
Jinal Patel,
The Broad Institute
Choosing a fraction for the study

- Precursors selected for CID / ETD
- QAD identifications from CID / ETD
Study Materials (i)

- 1 LTQ-Orbitrap XL dataset (eq. 1 RAW file)
  - RAW, mzML, mzXML, MGF, dta – conversions by ProteoWizard 2.1.2051
- 1 fasta file (UniProtKB/SwissProt S. cerevisiae from Sept. 2010)
- 1 spectral library in SpectraST format (contributed by Henry Lam)
- 1 template (Excel)
- 1 on-line survey (Survey Monkey)
**Study Materials (ii) – additional data**

<table>
<thead>
<tr>
<th>Frac</th>
<th>Instrument</th>
<th>Fragmentation</th>
<th>MSMS Res/Acc</th>
<th>Spike?</th>
</tr>
</thead>
<tbody>
<tr>
<td>9</td>
<td>D100914_yeast_SCX09_rak_ft8DT_pc_01.RAW</td>
<td>XL</td>
<td>DT (C or E)</td>
<td>LL</td>
</tr>
<tr>
<td></td>
<td>K100923_Yeast_SCX09_ft16DT_pcc_01.raw</td>
<td>V</td>
<td>DT (C or E)</td>
<td>LL</td>
</tr>
<tr>
<td></td>
<td>K100923_Yeast_SCX09_ff6f6HE_pcc_01.raw</td>
<td>V</td>
<td>H+E</td>
<td>HH</td>
</tr>
<tr>
<td></td>
<td>D100915_yeast_SCX09S_rak_ft8E_pc_02.RAW</td>
<td>XL</td>
<td>E</td>
<td>L</td>
</tr>
<tr>
<td></td>
<td>V20100923-23</td>
<td>V</td>
<td>C+E</td>
<td>LL</td>
</tr>
<tr>
<td>10</td>
<td>D100914_yeast_SCX10_rak_ft8DT_pc_01.RAW</td>
<td>XL</td>
<td>DT (C or E)</td>
<td>LL</td>
</tr>
<tr>
<td></td>
<td>K100923_Yeast_SCX10_ft16DT_pcc_01.raw</td>
<td>V</td>
<td>DT (C or E)</td>
<td>LL</td>
</tr>
<tr>
<td></td>
<td>K100923_Yeast_SCX10_ff6f6HE_pcc_01.raw</td>
<td>V</td>
<td>H+E</td>
<td>HH</td>
</tr>
<tr>
<td></td>
<td>D100917_yeast_SCX10_rak_ft8E_pc_01.RAW</td>
<td>XL</td>
<td>E</td>
<td>L</td>
</tr>
<tr>
<td></td>
<td>D100930_yeast_SCX10S_rak_ft8E_pc_01.RAW</td>
<td>XL</td>
<td>E</td>
<td>L</td>
</tr>
<tr>
<td></td>
<td>V20100923-24</td>
<td>V</td>
<td>C+E</td>
<td>LL</td>
</tr>
<tr>
<td></td>
<td>V20100923-29</td>
<td>V</td>
<td>DT (C or E)</td>
<td>LL</td>
</tr>
<tr>
<td></td>
<td>V20100923-31</td>
<td>V</td>
<td>E</td>
<td>L</td>
</tr>
<tr>
<td></td>
<td>V20100923-32</td>
<td>V</td>
<td>DT (H or E)</td>
<td>HH</td>
</tr>
<tr>
<td>11</td>
<td>D100914_yeast_SCX11_rak_ft8DT_pc_01.RAW</td>
<td>XL</td>
<td>DT (C or E)</td>
<td>LL</td>
</tr>
<tr>
<td></td>
<td>K100923_Yeast_SCX11_ft16DT_pcc_01.raw</td>
<td>V</td>
<td>DT (C or E)</td>
<td>LL</td>
</tr>
<tr>
<td></td>
<td>K100923_Yeast_SCX11_ff6f6HE_pcc_01.raw</td>
<td>V</td>
<td>H+E</td>
<td>HH</td>
</tr>
<tr>
<td></td>
<td>D100917_yeast_SCX11S_rak_ft8E_pc_02.RAW</td>
<td>XL</td>
<td>E</td>
<td>L</td>
</tr>
<tr>
<td></td>
<td>V20100923-25</td>
<td>V</td>
<td>C+E</td>
<td>LL</td>
</tr>
<tr>
<td></td>
<td>V20100923-30</td>
<td>V</td>
<td>DT (C or E)</td>
<td>LL</td>
</tr>
<tr>
<td></td>
<td>V20100923-33</td>
<td>V</td>
<td>DT (H or E)</td>
<td>HH</td>
</tr>
<tr>
<td>12</td>
<td>D100914_yeast_SCX12_rak_ft8DT_pc_01.RAW</td>
<td>XL</td>
<td>DT (C or E)</td>
<td>LL</td>
</tr>
<tr>
<td></td>
<td>K100923_Yeast_SCX12_ft16DT_pcc_01.raw</td>
<td>V</td>
<td>DT (C or E)</td>
<td>LL</td>
</tr>
<tr>
<td></td>
<td>K100923_Yeast_SCX12_ff6f6HE_pcc_01.raw</td>
<td>V</td>
<td>H+E</td>
<td>HH</td>
</tr>
<tr>
<td></td>
<td>V20100923-26</td>
<td>V</td>
<td>C+E</td>
<td>LL</td>
</tr>
<tr>
<td></td>
<td>V20100923-27</td>
<td>V</td>
<td>E</td>
<td>LL</td>
</tr>
<tr>
<td>S48</td>
<td>D100914_Sigma48_ft4t4_pcc_01.RAW</td>
<td>XL</td>
<td>C+E</td>
<td>LL</td>
</tr>
<tr>
<td></td>
<td>K100922_Sigma48_ft16DT_pcc_01.raw</td>
<td>V</td>
<td>DT (C or E)</td>
<td>LL</td>
</tr>
<tr>
<td></td>
<td>K100923_Sigma48_ff6f6HE_pcc_01.raw</td>
<td>V</td>
<td>H+E</td>
<td>HH</td>
</tr>
<tr>
<td></td>
<td>V20100923-28</td>
<td>V</td>
<td>E</td>
<td>L</td>
</tr>
</tbody>
</table>
Instructions to Participants

1. Retrieve and analyze the data file in the format of your choosing, with the method(s) of your choosing

2. Report the peptide to spectrum matches in the provided template

3. Fill out the survey

4. Attach a 1-2 page description of the methodology employed
## ABRF iPRG 2011 Study Template: ETD Data Analysis

**Instructions:** Please fill in all fields required fields (marked with *). After deleting the example rows, create a new row for each peptide spectrum match. Indicate whether each match is better than a 1% FDR on the spectrum-level. Include identifications above and below threshold. Results should be sorted by 'Search Engine Score' from most to least confident. Additional instructions can be found above each field header. Results should be emailed to 'anonymous.iprg2011@gmail.com' no later than Dec. 10, 2010. Please make sure to fill out the REQUIRED survey (URL).

### Identifiers should be unique scan numbers from data file. Retention times and spectrum indices (e.g., from the MGF file) are also acceptable if described in the 1-2 page methods report.

<table>
<thead>
<tr>
<th>Spectrum Identifier*</th>
<th>Precursor m/z</th>
<th>Mass Error*</th>
<th>Peptide Sequence*</th>
<th>Modifications</th>
<th>Protein Accession(s)</th>
<th>Search Engine Score*</th>
<th>Better than 1% FDR threshold?*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scan:1753</td>
<td>563.7818</td>
<td>-0.0004</td>
<td>mvGvNRyleK</td>
<td>1,M:oxidation;4,N:deamidated</td>
<td>Q12672</td>
<td>0.999999 Y</td>
<td></td>
</tr>
<tr>
<td>Scan:4659</td>
<td>842.4291</td>
<td>0.0004</td>
<td>FFgFTPEgVAEQK</td>
<td></td>
<td>P23254</td>
<td>0.999999 Y</td>
<td></td>
</tr>
<tr>
<td>Scan:2156</td>
<td>673.3224</td>
<td>-0.0009</td>
<td>TSgiyADFaeFK</td>
<td></td>
<td>P22146</td>
<td>0.999999 Y</td>
<td></td>
</tr>
<tr>
<td>Scan:1571</td>
<td>414.8557</td>
<td>0.0007</td>
<td>nSTKhHSLVVK</td>
<td>1,N:deamidated;6,N:deamidated</td>
<td>P41540</td>
<td>0.999999 Y</td>
<td></td>
</tr>
<tr>
<td>Scan:6017</td>
<td>838.5427</td>
<td>-0.0031</td>
<td>qGVlPTrKLLtK</td>
<td>1,Gln-&gt;pyroglu</td>
<td>P02365</td>
<td>0.999999 Y</td>
<td></td>
</tr>
<tr>
<td>Scan:4212</td>
<td>617.3125</td>
<td>0.0008</td>
<td>YKgFTEFK</td>
<td>4,N:deamidated</td>
<td>P16521</td>
<td>0.999999 Y</td>
<td></td>
</tr>
<tr>
<td>Scan:1587</td>
<td>658.3452</td>
<td>-0.0188</td>
<td>HAcTNVAKDK</td>
<td></td>
<td>Q12447</td>
<td>0.999999 Y</td>
<td></td>
</tr>
<tr>
<td>Scan:7333</td>
<td>917.7254</td>
<td>-0.0013</td>
<td>GLVSDPAGSALNVLYKFDYN28,C,carboxidemethyl</td>
<td>P54639</td>
<td>0.999999 Y</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
iPRG 2011 STUDY:
PARTICIPATION
Soliciting Participants and Logistics

Study advertised on the ABRF website and listserv, Molecular and Cellular Proteomics blogsite, ECD/ETD conference attendants, GenomeWeb and by direct invitation from iPRG members

1. Email participation request to ‘iPRGxxxx@gmail.com’

2. Send official study letter with instructions

3. All further communication (e.g., questions, submission) through ‘iPRGxxx.anonymous@gmail.com’

“Anonymizer”
Participants (i) – overall numbers

• 40 requests / 35 submissions (‘88% return’)
  – Some participants submitted two result sets

• 9 initialed iPRG member submissions (with appended ‘i’)

• 7 vendor submissions (identifiable by appended ‘v’)

16
Participants (ii) - demographics

**Membership**
- 59% Member
- 41% Non member

**Location**
- North America (specify below)
- Asia (specify below)
- Europe (specify below)
- Australia/NZ (specify below)

**Primary Job Function**
- 66% Director/Manager
- 18% Bioinformatician/Software Developer
- 11% Mass Spectrometrist
- 5% Lab Scientist

**Resource Lab Status**
- Conduct both core functions and non-core lab research
- Non-core research lab
- Core only
- Software development only (no research facility)

**Type of Lab**
- 85% Academic
- 8% Manufacturer/Vendor
- 5% Government
- 2% Other (please specify)

**ETD Data Interpretation Experience**
- Complete novice
- I have worked with a few data sets
- I have worked with several data sets
- I routinely analyze ETD data

**Years of Experience in Proteomics**
- < 1 year
- 1-2 years
- 3-4 years
- 5-10 years
- > 10 years
Participants (iii) – study opinions

Study Difficulty Level

- Easy: 50%
- Just right: 30%
- Challenging: 20%

Reporting Difficulty Level

- Easy: 37%
- Just right: 36%
- Challenging: 27%

Have you participated in previous ABRF studies?

- No: 36%
- Yes: 64%

Do you provide this service?

- No: 18%
- No but have plans to do so: 7%
- Yes: 48%
- Not applicable: 27%
Participants (iv) – methods (i)

Data Format Used

- MGF: 25 participants
- RAW: 20 participants
- mzXML: 5 participants
- DTA: 2 participants
- mzML: 2 participants
Participants (v) – methods (ii)

Spectral Pre-Processing

Other / in-house (public / non-public):
DTARefinery, DeconMSn, DTA Generator, Etdgenerator, RawExtractor, Hardklor, multiplierz, ReAdW, ByonicZ
Participants (vi) – methods (iii)

Peptide identification

Other / in-house (public / non-public):
pFind, ByOnic, ProteinScape, MS_LIMS, PVIEW, PepArML, Byonic2, Proteome Discoverer
Participants (vii) – *methods (iv)*

**Results filtering**

- In-house software (public)
- Excel
- Other
- Protein Prospector
- None
- IDPicker
- Mascot
- PEAKS
- Peptide/ProteinProphet
- TransProteomic Pipeline (TPP)
- R
- Spectrum Mill
- Xcalibur
- In-house software (not public)

**Other / in-house (public / non-public):**
- pBuild, ComByne, ProteinScape, Percolator, PVIEW, Epitomize, FDR Optimizer, MSblender, OmssaParser, MascotDatFile, multiplierz, ComputeFDR, Proteome discoverer
Participants (viii) – *time spent (hours)*

- Spectral pre-processing
- Reformattting FASTA DB
- Search time
- Filtering for 1% FDR
- Manual validation
- Preparing Excel template
- Total analysis time

Total analysis time: 23 hours
Participants (ix) – confidence

Confidence in processing ETD data

- No experience, never tried it before
- Not confident
- Confident
- Very confident
iPRG 2011 STUDY:
PRELIMINARY ANALYSIS
**Total identifications and methods**

### Proteome Informatics Research Group

**Total identifications**

- **# spectra Id Yes**: 3000, 3500, 4000, 4500
- **# unique Peptides UC ID Yes**: 1000, 1500, 2000, 2500

### Spectrum Pre-processing

- **OM**: OM_Spectrum X (OM)
- **PP**: pFind = pF
- **Xc**: Xcalibur = Xc
- **PK**: Per = Peptide/ProteinProphet = P/PP

### Peptide Identification

- **OM**: OM_Spectrum X (OM)
- **PP**: pFind = pF
- **Xc**: Xcalibur = Xc
- **PK**: Per = Peptide/ProteinProphet = P/PP

### Result Filtering

- **OM**: OM_Spectrum X (OM)
- **PP**: pFind = pF
- **Xc**: Xcalibur = Xc
- **PK**: Per = Peptide/ProteinProphet = P/PP

### Years Experience

- **OM**: OM_Spectrum X (OM)
- **PP**: pFind = pF
- **Xc**: Xcalibur = Xc
- **PK**: Per = Peptide/ProteinProphet = P/PP

---

**Bioworks** = Bw
**DTA Generator** = Dg
**Excel** = Ex
**Extract_msn** = Ems
**IDPicker** = idp
**Inspect** = Ins

**Xcalibur** = Xc
**XTandem** = XI
**Spectrum Mill** = SM
**SpectraST** = SpST
**OMSSA** = OM
**Other** = Other

**MS-GFDB** = M-G
**MSQuant** = MQ
**MyriMatch** = My
**Mascot** = Ma
**Distiller** = MaD
**PK** = PEAKS

**Percolator** = Per
**Phenyx** = Ph
**ProteoWizard** = PW
**ReAdW** = Re
**SEQUEST** = SQ

**pFind** = pF
**OMSSA** = OM
**Distiller** = MaD
**ProteoWizard** = PW
**ReAdW** = Re

**In-house software (freely available)** = IH(P)
**In-house software (not public)** = IH(np)

**Phenyx** = Ph
**ProteoWizard** = PW
**Protein Prospector** = PP

**TransProteomic Pipeline (TPP)** = TPP

---

**Experience**

- **Bioworks** = Bw
- **DTA Generator** = Dg
- **Excel** = Ex
- **Extract_msn** = Ems
- **IDPicker** = idp
- **Inspect** = Ins

---

26
Modifications

Proteome Informatics Research Group

- C
- ^q, ^c, m,n,q
- Carbamyl, Acetyl, Carbamidomethyl, Acetaldehyde
- adefghiklprstvwy

The table shows the number of spectra for different modifications:

- No Mods
- Common Mods
- N-term Mods
- Rare Mods

The graph illustrates the distribution of these modifications across various spectra.
Charge state distributions

Proteome Informatics
Research Group

*** Underperformed with Z2
Overlap of spectrum identifications

44 participants

Unanimous

Cumulative # spectra

# Participants Agreeing

0 5 10 15 20 25 30 35 40 45

# Participants Agreeing
Characteristics of consensus spectra

4455 spectra >=3 participants agreeing on sequence

# Spectra Containing Mod

- Nterm-Acetyl
- Nterm-Carbamyl
- Nterm-Carbamidomethyl
- Nterm-Acetaldehyde
- Nterm-Other
- PyroGlu Q
- PyroGlu E
- PyroCarbamidomethylCys
- n: Deamidation
- m: Oxidation
- q: Deamidation
- k: Carbamyl,Acetyl,Carbamidomethyl
- h: Carbamidomethyl
- w: Oxidation
- r: Carbamyl,Carbamidomethyl
- d: Carbamidomethyl
- No Variable Mods

Consensus requires agreement on Sequence, but not modification localization
Room for improvement in thresholding?

44 participants
3 for consensus
ESR and FDR

Extraordinary Skill Rate or High False Discovery Rate?

ESR + FDR = 100* (Y<3P+YD)/total ids

44 participants
3 for consensus
**Resource for inspecting ID overlap**

**YS:** Y – identification, and top sequence same as consensus

**NS:** N – identification, but top sequence same as consensus

**YD:** Y – identification, and top sequence different than consensus

<table>
<thead>
<tr>
<th>spectruneSequence</th>
<th>nTerm</th>
<th>cursor mz</th>
<th>Charge</th>
<th>ED14941i</th>
<th>RC94158i</th>
<th>HL37133i</th>
<th>20965v</th>
<th>32747</th>
<th>25125</th>
<th>51076</th>
<th>NB21194i</th>
<th>86010</th>
<th>KC87034i</th>
<th>numY</th>
<th>DinRov</th>
<th>numYSinRov</th>
<th>numDifferent</th>
</tr>
</thead>
<tbody>
<tr>
<td>6712</td>
<td>GVSAAVAK</td>
<td>771.4227</td>
<td>4</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>0</td>
<td>34</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>7661</td>
<td>PGVLVPFEL</td>
<td>892.4985</td>
<td>3</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>ND</td>
<td>ND</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>0</td>
<td>18</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1801</td>
<td>EMHHECLELQ</td>
<td>456.2673</td>
<td>2</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>0</td>
<td>0</td>
<td>20</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>5006</td>
<td>NYVKAETR</td>
<td>376.3682</td>
<td>2</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>0</td>
<td>0</td>
<td>22</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>3074</td>
<td>GTGWNLAY</td>
<td>792.4985</td>
<td>3</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>0</td>
<td>0</td>
<td>26</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>3509</td>
<td>YEARK-VWV</td>
<td>428.3682</td>
<td>2</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>1</td>
<td>1</td>
<td>37</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>7652</td>
<td>SRLAYTV</td>
<td>117.3682</td>
<td>2</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>3163</td>
<td>QAPKDTVI</td>
<td>582.4985</td>
<td>3</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>1</td>
<td>1</td>
<td>40</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>4863</td>
<td>KFRTVSPS</td>
<td>606.3682</td>
<td>3</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>1</td>
<td>1</td>
<td>22</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1178</td>
<td>LEATH</td>
<td>549.3682</td>
<td>2</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>0</td>
<td>0</td>
<td>30</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>4944</td>
<td>TFYEDRLW</td>
<td>620.3682</td>
<td>2</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>2</td>
<td>2</td>
<td>12</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>4361</td>
<td>KEELAKLVR</td>
<td>552.3682</td>
<td>2</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>0</td>
<td>0</td>
<td>5</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>5499</td>
<td>SNNNYNK</td>
<td>549.3682</td>
<td>2</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>0</td>
<td>0</td>
<td>26</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>1250</td>
<td>SVDLRRH</td>
<td>434.3682</td>
<td>2</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>1</td>
<td>1</td>
<td>21</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>6642</td>
<td>HTZKESPN</td>
<td>581.3682</td>
<td>2</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>1</td>
<td>1</td>
<td>14</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>5810</td>
<td>GFLGKQ</td>
<td>581.3682</td>
<td>2</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>4</td>
<td>4</td>
<td>19</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1630</td>
<td>VRLS</td>
<td>581.3682</td>
<td>2</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>0</td>
<td>0</td>
<td>23</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>3491</td>
<td>EKLTPAGY</td>
<td>581.3682</td>
<td>2</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>0</td>
<td>0</td>
<td>21</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>4340</td>
<td>PNYETLEG</td>
<td>581.3682</td>
<td>2</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>0</td>
<td>0</td>
<td>40</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>3185</td>
<td>YQHNYH</td>
<td>581.3682</td>
<td>2</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>0</td>
<td>0</td>
<td>32</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>3109</td>
<td>KQLY</td>
<td>581.3682</td>
<td>2</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>0</td>
<td>0</td>
<td>38</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>7022</td>
<td>DFADYNFSK</td>
<td>581.3682</td>
<td>2</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>1</td>
<td>1</td>
<td>7</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>728</td>
<td>P3R3DADL</td>
<td>581.3682</td>
<td>2</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>0</td>
<td>0</td>
<td>29</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1638</td>
<td>DLLQDN</td>
<td>581.3682</td>
<td>2</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>1</td>
<td>1</td>
<td>17</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>5060</td>
<td>GTVRWCK</td>
<td>581.3682</td>
<td>2</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>YS</td>
<td>YD</td>
<td>YS</td>
<td>YS</td>
<td>2</td>
<td>2</td>
<td>26</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>
iPRG 2011 STUDY:
TWO SURPRISES AT THE END
Surprise № 1: Sigma-48 spike-in

Sigma 48 digested separately and not subjected to SCX. Spike-in at a level to yield ~10% of ID peptides. *Biases against non-iPRG DB users, and SCX prediction users.*
Sigma-48 as TP estimator

Proxy for FD rate

Sigma48 IDs - (Sigma48 IDs * YD rate)

true positives
**Orig. yeast sequences**

>sp|P47002|SSY5_YEAST SPS-sensor serine protease component

SSY5 OS=Saccharomyces cerevisiae GN=SSY5 PE=1 SV=2

MVRFFGLNKKNEKTDLPADNEQNAETSSNVSGNEDERIDNNTIPENANNDDA STTFSSISQSSLIFSRRGMYTYGTGASSMATSEMRSHSSGSHSISRNLQGFDVQKFL RAVSFLSPVREKESQQTQNLVDSLSSSTSTATSENARENPSHGRSITLHEIKSSEL EKVIVMDGCIQNDVIVISGKAVIRAIYKPEFLPTTRQEPVRKISKLEKSSIRAKNIVL HFLDNLVSDANFSNRRSILLRRFPYFLKLKLNDLDDDLGEPVCPLVSCPICSNNCLP SMDKLGMILDELTMDKSSISDEQEGAFiAPiLGiTiPSSSTILMFQLPNLQHMYKIK VLYSLPFPVHNVCYDKIKKAAASYPSHTATIDtDIAPTQFSPPYAYQSNFPLELP IMSLSTETSSAKTITGTLGGYLPQGSDKKFSQFASCSFATCHVLSKEQDYPVMVF SNVLTQSYKVLTKEFDPRGDSVKEFAPFLEVEQRYQDQNNQNKINFKQFGQVVGEARIV HLSDFAIKVNSFSKCQNTLNGKSPFDPTRFRQFNLVHRRKTPFKKPGMKVFPIGASt G1TSVSMISTKLWYAGKQLQSEFVAVSTPLFASAGDSQWLTIKLEDRLGLLGVMQL HCYGEQROGQLTPTYGDMLRHYAIVNQNDPGQPLCG

>sp|P38788|SSZ1_YEAST Ribosome-associated complex subunit SSZ1 OS=Saccharomyces cerevisiae GN=SSZ1 PE=1 SV=2

MSSFVIGITFGNSTTSSIAIYNFKDNDVIANPDGERAIPSALSYVGEDEYHQQALQLQLI RNPKNTINIFDGFIPDFDKCDVSKCNGAPAVEVDVGHVFRSREGGEKEELLTDVVS RHLNRKLLAAEDYISGAVKAVEAVTLPTNFSEQKXKTKLAKASAKIQLGVQIVFNEPSAALL AHAAGPQFPEKDVVAVDFGIISRDAIAVNRGFTTIALAHDSLGLGDLTVLTVYFEV ASEPQKYQANPKNARLAKLKNSITKTILSNAGS1I3DSLADGFDYHASNMRM YELVKVFAQCSSFSVDSVIAKALPDLPIDAVLTLTGSQFKTILTNRNLYLTPESEVIL GPQWKNASNNPNELASGALQLRARLISYDADELAALYPVFINHSLK5P1GIGAKGE FHVLLAETSFSFVQKLTQKAKDGFLGVYEGDDHIEEKLTLPEIKEPAENAADESEWS DDEPEVREKLYTLTGLTKLMGLIKNANGVEIFNINKDALRVTARDLdTAVKGE

**Previously unidentified yeast protein**

**Fusion junction (underlined)**

>sp|P47002|SSY5_YEAST SPS-sensor serine protease component

SSY5 OS=Saccharomyces cerevisiae GN=SSY5 PE=1 SV=2

MGIPMGKSMLVLLTLAFASCIAAYRPSETLCCGEILVTLQFQVGCGBKGFYFSPASRVS RRSGRIEVECCFRSCDALLLETYCATPAKSERDVSTPPFP4NFFFRPVKGFQYQDYTWK QSTQRLRGRPLALLRARRHGVHLEAFREAKHRPLIALTPDFAHGGAPFEMASNRK

**Final fusion in FASTA**

>sp|P47002|SSY5_YEAST SPS-sensor serine protease component

SSY5 OS=Saccharomyces cerevisiae GN=SSY5 PE=1 SV=2

MGIPMGKSMLVLLTLAFASCIAAYRPSETLCCGEILVTLQFQVGCGBKGFYFSPASRVS RRSGRIEVECCFRSCDALLLETYCATPAKSERDVSTPPFP4NFFFRPVKGFQYQDYTWK QSTQRLRGRPLALLRARRHGVHLEAFREAKHRPLIALTPDFAHGGAPFEMASNRK

**Next protein in FASTA**

>sp|P01344|IGF2_HUMAN Insulin-like growth factor II OS=Homo sapiens GN=IGF2 PE=1 SV=1

MGIPMGKSMLVLLTLAFASCIAAYRPSETLCCGEILVTLQFQVGCGBKGFYFSPASRVS RRSGRIEVECCFRSCDALLLETYCATPAKSERDVSTPPFP4NFFFRPVKGFQYQDYTWK QSTQRLRGRPLALLRARRHGVHLEAFREAKHRPLIALTPDFAHGGAPFEMASNRK

**Sigma48 protein**
Identification of fusion peptides

Five participants reported the peptide:
KLVAASQAAALGLMNYLETQLNKK

C-terminus of Human Serum Albumin - N-terminus of Pachytene arrest protein SAE3

Consensus Answer:
Acetyl-SRSGVAVADESLTAFLKLGK(Carbamidomethyl)K^{4+}

rare mod
Conclusions

• Study went well and had global participation, with quite a few first-timers joining in

• Earlier software problems with interpreting doubly-charged precursors have been largely cleared up

• Experience with software is probably a better measure of performance than the actual tool used

• People are generally over-optimistic about how reliable their results are (FDR underestimation)

• However, false negatives (NS) are generally much higher than false positives, so there is room for improvement there
What did the participants think?

"I work in an environment without a group of peers doing proteomics. So having a study like this definitely gives me a chance to compare notes with others and benchmark my abilities. It also gives me a piece of evidence to prove my ability to people that are not in the field, such as users, administrators, advisory committee. I believe all core only facilities should participate in these studies."

100% of participants found the study useful

"The results of this study will be good to show how much room for improvement there is for the popular identification tools in ETD analysis. It's a good opportunity for lesser known and more open software to make a significant impact in the field."
THANK YOU TO ALL STUDY PARTICIPANTS!

iPRG
- Manor Askenazi
- Nuno Bandeira
- Robert Chalkley
- Karl Clauser
- Eric Deutsch
- Henry Lam *(ad hoc member)*
- Paul Rudnick
- Tom Neubert *(EB liaison)*
- Hayes McDonald *(chair elect)*
- Lennart Martens *(chair)*
- John Cottrell *(member elect)*
- Matt Chambers *(member elect)*
- Ruixiang Sun *(member elect)*
- Eugene Kapp *(member elect)*

Indispensable help came from:
- Jinal Patel, The Broad Institute
- Namrata Udeshi, The Broad Institute
- Jeremy Carver, UCSD