Spectrum Mill – A Powerful Tool for Proteomics

Karl Clauser
Broad Institute of MIT and Harvard
February 26, 2015
Abstract

Driven by new instrumentation and techniques actively used in my proteomics research, collaborative development with Agilent has evolved the tools and algorithms within the Spectrum Mill software package. In this webinar several features will be highlighted.

- Scoring of peptide MS/MS spectrum matches capitalizes on low ppm product ion mass accuracies with fragment-ion type models optimized for qToF-CID(HCD), ion trap CID, and ETD dissociation methods.
- Summary reports enable quantitative comparison across samples with organization not only by protein, distinct peptide, or peptide spectrum match, but also by PTM site.
- Sophisticated protein parsimony assembles distinct peptides into protein groups which enables protein-level quantitation that allows for peptides shared between multiple proteins to be either included or excluded. This is particularly empowering for human/mouse xenograft experiments in need of species-specific quantitation.
- Quantitative data analysis methods support iTRAQ, TMT-6/10, SILAC, and label-free paradigms.
- Quality Metrics enable troubleshooting mass spectrometry, chromatography, sample handling, and chemical labeling efficiency problems, as well as measuring effectiveness of MS/MS data acquisition method alterations.
- Automation on an SM server through a service request manager that maximizes use of all CPUs and queues processes from multiple users.
### Process Automation Tools

- **Workflows**: Automate data extraction, search, validation, and summary.
- **Request Queue**: View status of request queue for submitted requests, monitor results output.
- **Completion Log**: View log of completed extractions, searches, validations, and summaries.

### Mass Spectral Interpretation Tools

- **Data Extractor**: Prepare MS/MS data files for Spectrum Mill processing.
- **MS/MS Search**: Search database with MS/MS spectra.
- **Autovalidation**: Autovalidate MS/MS spectra, calculate false-positive rates.
- **Manual PMF Search**: Search database with MS spectra from Peptide Mass Fingerprint data.

### Result Summary Tools

- **Protein/Peptide Summary**: Summarize results from MS/MS searches.
- **Quality Metrics & FDR**: Summarize metrics for quality (MS performance, MS/MS interpretation, chromatography, sample handling) and FDR (spectra, peptide, protein).
- **Spectrum Summary**: Summarize characteristics of MS/MS spectra.

### Utilities

- **Tool Belt**: Collection of Spectrum Mill utility tools.
- **Protein Databases**: Manipulate FASTA sequence databases for use with Spectrum Mill software.
- **Archive Data**: Zip and Unzip dataset directories.
- **Build TIC**: Display a TIC or neutral loss chromatogram.
- **Peptide Selector**: Select peptides from protein digest likely to produce high quality MS/MS spectra.
- **MRM Selector**: Build/Export MRM transition lists based on experimental data in Spectrum Mill.
- **Multiple Sequence Aligner**: Run Clustal W and highlight non-identical AA's across the alignment.
- **MS Digest**: Predict peptide masses for enzymatic digestion of protein.
- **MS Edman**: Search database with text or partial peptide sequence.
- **Peptide String Match**: Search database with a list of peptide sequences.
- **Peptide List to Masses**: Convert a list of peptides to masses and formulas.
- **MS Product**: Predict product ion masses from peptide sequence.
- **MS Comp**: Calculate AA compositions fitting parent mass and ammonium ions.
- **MS Isotope**: Display isotope patterns of peptides.
Summary of New Features in Next Release (B.05.00)

- 64-bit extraction
  - Readily handle data files > 4Gb
  - Support for latest instruments
    - Agilent 6500 Series Q-TOF
    - ThermoFisher: QExactive Plus, QExactive HF, Fusion?
- 64-bit search
  - Handle sequence databases > 2Gb
- Data Extraction supports “multi-core” or “Maximize CPUs” (core per file)
- Workflow automation
  - Quality and FDR Metrics reports
  - Data Archival
- Additional Quality Metrics (isobaric label incorporation, LC gradient shape)
- Enhanced MassProfiler Professional export for Agilent data
- TMT-10 Extraction, Search, Quantitation support
- Peptide String Match Utility
  - Search database with a list of peptide sequences to aid configuring targeted MS/MS assays
- Protein-termini modifications support in MS/MS Search
- Tool to create a FASTA subset from accession numbers
Spectrum Mill Workflow

**LC-MS/MS Run(s)**

- Data Extractor
- precursor ion
- Reporter ion
- Peak Areas

**Extracted Filtered MS/MS Spectra**

- *d
- *.RAW
- *mzXML

**Candidate Peptide ID’s (PSM’s) VM site localizations**

**FDR Validated Peptide ID’s (PSM’s), VM site localizations**

**Identified, Quantitated Grouped Proteins**

- Localized, Identified, Quantitated Protein grouped VM sites

- Localized, Identified, Quantitated Peptides

- Localized, Identified, Quantitated Peptides or PSM’s

**PP Summary**

**Autovalidation**

- hitTable.1.tsv
- spectrumTable.1.tsv

**proteinProteinCentricColumnsExport.1.ssv**

**proteinPeptideComparisonColumnsExport.S.1.ssv**

**proteinPeptideComparisonColumnsExport.VM.1.ssv**

**proteinPeptideComparisonColumnsExport.1.ssv**

**peptideExport.CS.1.ssv**

**peptideExport.1.ssv**

**PSMexport.1.ssv**

**tagSummary.1.tsv**
Use Windows Explorer to Move data files to the SM server

At Broad Institute
- Nightly, data files backed up from instrument to archive server.
- Following extraction, raw data files automatically deleted from SM server, replaced by shortcut.
### Configuring an Extraction

**Spectrum Mill - Data Extractor**

**Extraction**
- **Extract**
  - Save As...
  - Load...
  - Remove all results
  - Delete data files after extraction

**Data Directories**
- **Select**
  - Chris/2014-11-04-TB-std

**Modifications**
- **Choose...**
  - Fixed: Carbamidomethylation (C)

**MS/MS Spectral Feature Filtering**
- **Precursor **
  - MH+: 600.0 → 6000 Da
  - Scan time range: 0 → 300 min
- **Sequence tag length** > 0 (For MALDI: Set tag length to -1 and merge secs to total run time.)
- **Ignore spectra with dissociation mode**: CID, ETD

**Merge nearby MS² scans with same precursor m/z**
- Retention time & m/z tolerance: ± ± 45 m/z
  - also used for calculating chromatographic peak area of precursor in MS scans
- General MS/MS Merging Constraints: Spectral Similarity & RT & m/z

**Precursor m/z & Charge Assignment**
- **Precursor Charge**: Find 128 Maximum (z): 6 Minimum MS1 S/N: 25
- **Find ¹³C precursor m/z**
- MS Noise threshold: 400 counts (only applies to Agilent Q-TOF)

#### Agilent Instruments

- Convert to .mzXML
- Filter out low quality MS/MS
- Calculate precursor ion XIC's
- Correct precursor monoisotopic m/z assignments
- Merge replicate MS/MS from same chromatographic peak

#### ThermoFisher Instruments

- Convert to .mzXML
- Filter out low quality MS/MS
- Calculate precursor ion XIC's
- Correct precursor monoisotopic m/z assignments
- Merge replicate MS/MS from same chromatographic peak
# Configuring an MS/MS Search

**Agilent Spectrum Mill - MS/MS Search**

### Search

- **Start Search**
- **Save As**
- **Load**
- **Remove all prior MS/MS Search results**
- **Maximize CPUs**

### Data Directories

- **Select Data Directories**
  - **Select...**
  - **07TCGA_A8_A067_A2_A0D1_A2_A0CM_20130401**

### Search Parameters

**Validation filter:** spectrum not marked - sequence not validated

- **Batch size:** 500
- **Search previous hits**
- **Max repeat matches:**

**Database:** RefSeq 20130727 - Human contains

**Digest:** Trypsin + K

**Species:** All

### Modifications

**Fixed:**
- Carbamidomethylation (C)
- ITRAQ Full Lys only
- Carbamyl IAA-mix (N-term,k)

**Variable:**
- Acetyl (Prot-N-term)
- Oxidized methionine (M)
- Pyroglutamic acid (N-termQ)
- Deamidated (N)
- Dimers Carbamidomethyl Cys (N-termC)

### Search Criteria

**Matching matched peak intensity:** 40%

- **Instrument:** ESI QExactive HCD
- **Masses are:** Monoisotopic
- **Precursor mass tolerance:** +/- 20 ppm
- **Product mass tolerance:** +/- 20 ppm
- **Maximum ambiguous precursor charge:** 3

**Spectral Quality**

- **Sequence tag length:** 3

**Search Mode**

- **Calculate reversed database scores**

- **Search mode:** Variable modifications
  - **Precursor mass shift range:** -13.0 to 64 Da

**Data Files**

- **Fragmentation mode:** All

- **Spectrum files (.mzXML):**
  - .mzXML

- **Spectrum files (.pk1):**
  - .pk1

---

Karl Clauser
Proteomics and Biomarker Discovery
Key Features of Spectrum Mill Database Searching with MS/MS Spectra

- Scoring specific to instrument types, dissociation modes
  - Ion types allowed and scoring weights
  - On board peak detection (de-isotoping, S/N thresholding, fragment z assignment).
- Product ion mass tolerance in ppm or Da.
- Spectra grouped in batches of similar precursor m/z, z for fast precursor mass filtering.
- Target-decoy FDR accomplished by on-the-fly peptide inner reversal
  - SAMPLER becomes SELPMAR
  - Search time is only ~1.5x as long as a target only database, instead of 2x for a concatenated forward/reverse database. 1x digestion of all proteins, 1x precursor mass filtering of all peptides, 2x MS/MS matching of every sequence (fwd & rev)
  - Search results do not comingle target and decoy hits.
    - For each PSM report top target hit, and delta Fwd – Rev score of top decoy hit
    - A false positive PSM had delta Fwd-Rev < 0. (Top Decoy hit has higher score)
- Multiple cycles for fixed/mix modifications or unknown charge, with combined single result.
- Variable modifications constrained by precursor mass shift, not by # of mods/peptide.
- Second pass search possible with leftover spectra from 1st pass.
- Homology modes available
  - Unassigned single mass gap
  - Allow single mutation per peptide
Fixed, Mix and Variable Modifications

**Fixed**
Redefine the wild type as...

**Variable**
Allow 2 possibilities for an AA. Allow both in 1 spectrum if more than one location/AA.

**Mix**
Search in 2 cycles
Cycle 1: all KR light
Cycle 2: all KR heavy
DO NOT allow both light and heavy in 1 spectrum

---

**Fixed/Mix Modifications**

- **Cysteine**
  - Carbamidomethylation (C)
- **N-terminus**
  - unmodified
- **C-terminus**
  - unmodified

**Other amino acids**
- D: unmodified
- E: unmodified
- K: unmodified
- M: unmodified

**Metabolic isotope labels**
- All: unmodified
- K: SILAC 2 (Arg 0-10Da Lys 0-8Da)-mix (R,K)
- L: unmodified
- M: unmodified
- R: SILAC 2 (Arg 0-10Da Lys 0-8Da)-mix (R,K)
- V: unmodified

**Variable Modifications**

- Acrylamide (C)
- Carboxymethylation (C)
- Carbamidomethylation (C)
- MMTS (C)
- Pyridylethylation (C)
- Cysteine sulfide (C)
- Acetyl (K)
- Guanidination (K)
- Carbamylated lysine (K)
- Oxidized methionine (M)
- Pyroglutamic acid (N-termQ)
- Deamidated (N)
- Phosphorylated S (S)
- Phosphorylated T (T)
- Phosphorylated Y (Y)
- Ile 6 C-13 (I)
- Ile 7 C-13 N-15 (I)
- Leu 6 C-13 (L)
- Leu 7 C-13 N-15 (L)
- Lys 6 C-13 (K)
PSM Scoring/ Phosphosite Localization
Spectrum Mill Scoring of MS/MS Interpretations

(R)I/S|F/D/E/F/I/K(I)

Peak Selection: De-Isotoping, S/N thresholding, Parent - neutral removal, Charge assignment
Match to Database Candidate Sequences

Score = Assignment Bonus (Ion Type Weighted) + Marker Ion Bonus (Ion Type Weighted) - Non-assignment Penalty (Intensity Weighted)

SPI (%) Scored Peak Intensity

PIP(%) Precursor Ion Purity

<table>
<thead>
<tr>
<th>Fragment Ion (m/z)</th>
<th>120.080</th>
<th>129.102</th>
<th>130.086</th>
<th>136.075</th>
<th>147.112</th>
<th>173.125</th>
<th>201.122</th>
<th>207.112</th>
<th>219.112</th>
<th>245.075</th>
<th>269.155</th>
<th>267.115</th>
<th>352.143</th>
<th>407.263</th>
<th>461.163</th>
<th>533.304</th>
<th>595.366</th>
<th>597.366</th>
<th>626.242</th>
<th>651.332</th>
<th>702.390</th>
<th>798.409</th>
<th>833.430</th>
<th>867.416</th>
<th>885.430</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frac. Int. (% of TIC)</td>
<td>-1.00</td>
<td>-1.00</td>
<td>-1.00</td>
<td>-1.00</td>
<td>-1.00</td>
<td>-1.00</td>
<td>-1.00</td>
<td>-1.00</td>
<td>-1.00</td>
<td>-1.00</td>
<td>-1.00</td>
<td>-1.00</td>
<td>-1.00</td>
<td>-1.00</td>
<td>-1.00</td>
<td>-1.00</td>
<td>-1.00</td>
<td>-1.00</td>
<td>-1.00</td>
<td>-1.00</td>
<td>-1.00</td>
<td>-1.00</td>
<td>-1.00</td>
<td>-1.00</td>
<td></td>
</tr>
<tr>
<td>Rel. Int. (% of BP)</td>
<td>0.53</td>
<td>0.22</td>
<td>3.03</td>
<td>11.38</td>
<td>16.93</td>
<td>16.93</td>
<td>16.93</td>
<td>4.54</td>
<td>11.12</td>
<td>5.41</td>
<td>6.07</td>
<td>16.93</td>
<td>5.97</td>
<td>3.61</td>
<td>11.65</td>
<td>5.91</td>
<td>4.32</td>
<td>10.00</td>
<td>4.32</td>
<td>10.00</td>
<td>15.16</td>
<td>20.21</td>
<td>50.16</td>
<td>4.73</td>
<td></td>
</tr>
<tr>
<td>Signal/Noise</td>
<td>17.9</td>
<td>6.4</td>
<td>2.7</td>
<td>11.1</td>
<td>7.4</td>
<td>10.7</td>
<td>4.4</td>
<td>10.7</td>
<td>4.4</td>
<td>10.7</td>
<td>4.4</td>
<td>10.7</td>
<td>4.4</td>
<td>10.7</td>
<td>4.4</td>
<td>10.7</td>
<td>4.4</td>
<td>10.7</td>
<td>4.4</td>
<td>10.7</td>
<td>4.4</td>
<td>10.7</td>
<td>4.4</td>
<td>10.7</td>
<td>4.4</td>
</tr>
<tr>
<td>Score</td>
<td>1.00</td>
<td>0.20</td>
<td>0.260</td>
<td>0.01</td>
<td>1.00</td>
<td>0.250</td>
<td>1.00</td>
<td>0.01</td>
<td>1.00</td>
<td>0.250</td>
<td>1.00</td>
<td>0.01</td>
<td>1.00</td>
<td>0.250</td>
<td>1.00</td>
<td>0.01</td>
<td>1.00</td>
<td>0.250</td>
<td>1.00</td>
<td>0.01</td>
<td>1.00</td>
<td>0.250</td>
<td>1.00</td>
<td>0.01</td>
<td>1.00</td>
</tr>
<tr>
<td>Ion-type</td>
<td>F</td>
<td>K</td>
<td>y-NH₃</td>
<td>y₁</td>
<td>y₂</td>
<td>y₂</td>
<td>b₂</td>
<td>b₂</td>
<td>b₂</td>
<td>b₂</td>
<td>b₂</td>
<td>b₂</td>
<td>b₂</td>
<td>b₂</td>
<td>b₂</td>
<td>b₂</td>
<td>b₂</td>
<td>b₂</td>
<td>b₂</td>
<td>b₂</td>
<td>b₂</td>
<td>b₂</td>
<td>b₂</td>
<td>b₂</td>
<td></td>
</tr>
<tr>
<td>Delta ppm</td>
<td>-4.1</td>
<td>-5.0</td>
<td>-4.1</td>
<td>-5.2</td>
<td>-5.5</td>
<td>-5.0</td>
<td>-5.0</td>
<td>-5.0</td>
<td>-5.0</td>
<td>-5.0</td>
<td>-5.0</td>
<td>-5.0</td>
<td>-5.0</td>
<td>-5.0</td>
<td>-5.0</td>
<td>-5.0</td>
<td>-5.0</td>
<td>-5.0</td>
<td>-5.0</td>
<td>-5.0</td>
<td>-5.0</td>
<td>-5.0</td>
<td>-5.0</td>
<td>-5.0</td>
<td></td>
</tr>
</tbody>
</table>

11.52
81.5%
80.3%
Localizing a Phosphorylation Site

same spectrum  
2 different interpretations

L/F|P/A/D|T/s/P/S T A\ T K

MSTag 1

1.06e+3
3.39%

T | A | T | S | P | S | T | D | A | P | F

167 101

L/F|P/A/D|t S/P/S T A\ T K

MSTag 2

1.06e+3
3.39%

T | A | T | S | P | S | T | D | A | P | F

D20090930_PM_K562_SCX-IMAC_fxn04.3801.3801.2.pkl

MH+: 1415.6421  m/z: 788.3247  z: 2
<table>
<thead>
<tr>
<th>Locations Tested</th>
<th>Conclusion</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>No possible ambiguity</strong></td>
<td></td>
</tr>
<tr>
<td>AVsEEQQPALK</td>
<td>AVS(1.0)EEQQPALK</td>
</tr>
<tr>
<td></td>
<td># PO₄ sites = # S,T, or Y</td>
</tr>
<tr>
<td><strong>Single Site</strong></td>
<td></td>
</tr>
<tr>
<td>APSLTDLVK *</td>
<td>APS(0.99)LT(0.0)DLVK</td>
</tr>
<tr>
<td>APSLtDLVK -</td>
<td></td>
</tr>
<tr>
<td>sSSAGPEGPQLDVPR *</td>
<td>S(0.50)S(0.50)S(0.0)AGPEGPQLDVPR</td>
</tr>
<tr>
<td>sSSAGPEGPQLDVPR *</td>
<td></td>
</tr>
<tr>
<td>SSsAGPEGPQLDVPR -</td>
<td></td>
</tr>
<tr>
<td><strong>Multiple Sites</strong></td>
<td></td>
</tr>
<tr>
<td>VTNDISPEssPGVGR *</td>
<td>VT(0.0)NDIS(0.99)PES(0.50)S(0.50)PGVGR</td>
</tr>
<tr>
<td>VTNDISPEssPGVGR *</td>
<td></td>
</tr>
<tr>
<td>VTNDISPEssPGVGR -</td>
<td></td>
</tr>
<tr>
<td>VtNDISPEssPGVGR -</td>
<td></td>
</tr>
<tr>
<td>VtNDISPEssPGVGR -</td>
<td></td>
</tr>
<tr>
<td>VtNDISPEssPGVGR -</td>
<td></td>
</tr>
</tbody>
</table>
PTM Site Localization – Confident Localization

(K)A/P*s/L/T D|L\V K(S)

APS(0.99)LT(0.0)DLVK

$y_6, y_7$ ions provide confident localization to the Ser
PTM Site Localization – Ambiguous Localization

\[(R)S\ s/S/A/G/P\ E/G/P\ Q\ L|D|V|P\ R(E)\]

S(0.50)S(0.50)S(0.0)AGPEGPQLDVPR

\[y_{13}^{++}\] ion excludes localization to Ser 3
PTM Site Localization – Ambiguous Localization
2 sites: 1 confident, 1 ambiguous

(R)V T N D I S/P E I S/P G V\G R(R)

VT(0.0)NDIS(0.99)PES(0.50)S(0.50)PGVGR

$y_9, y_{10}$ ions provide confident localization to the Ser -6
$y_9, y_{10}$ ions provide ambiguous localization to Ser-9, Ser-10
Spectrum Mill Variable Modification Localization Score

VML score = Difference in Score of same identified sequences with different variable modification localizations

VML score > 1.1 indicates confident localization

Why a threshold value of 1.1?
1 implies that there is a distinguishing ion of b or y ion type
0.1 means that when unassigned, the peak is 10% the intensity of the base peak
Phosphosite Localization Scoring - Ascore

- Typical score thresholds are equivalent to 2 peaks supporting localization
- Peak selection probability term requires Da tolerance instead of ppm


http://ascore.med.harvard.edu/
Supports Sequest results only, Linux only
## VML Score Features Reported

<table>
<thead>
<tr>
<th>#</th>
<th>Filename</th>
<th>z Score</th>
<th>Fwd-Rev Score</th>
<th>Rank1-Rank2 Score</th>
<th>SPI (%)</th>
<th>Score VML</th>
<th># STY Sites</th>
<th># CSTY Sites</th>
<th>Ambiguity Sites</th>
<th>Variable Sites</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>H20120518_JQ_CPTAC2_COMPREF4_IMAC_06.0578.0578.3</td>
<td>3</td>
<td>12.57</td>
<td>12.57</td>
<td>12.57</td>
<td>70.4</td>
<td>1.148</td>
<td>4</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>H20120518_JQ_CPTAC2_COMPREF4_IMAC_06.0900.0900.3</td>
<td>3</td>
<td>13.63</td>
<td>13.63</td>
<td>13.63</td>
<td>71.7</td>
<td>2.406</td>
<td>6</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>H20120518_JQ_CPTAC2_COMPREF4_IMAC_06.1127.1127.2</td>
<td>2</td>
<td>19.07</td>
<td>19.07</td>
<td>19.07</td>
<td>89.0</td>
<td>3.970</td>
<td>5</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>H20120518_JQ_CPTAC2_COMPREF4_IMAC_06.1413.1413.4</td>
<td>4</td>
<td>12.66</td>
<td>12.66</td>
<td>8.96</td>
<td>62.3</td>
<td>0.550</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>H20120518_JQ_CPTAC2_COMPREF4_IMAC_06.2362.2362.3</td>
<td>3</td>
<td>13.29</td>
<td>13.29</td>
<td>13.29</td>
<td>70.0</td>
<td>3.457</td>
<td>4</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>H20120518_JQ_CPTAC2_COMPREF4_IMAC_06.2323.2323.3</td>
<td>3</td>
<td>17.30</td>
<td>11.73</td>
<td>10.60</td>
<td>76.9</td>
<td>5.510</td>
<td>3</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>H20120518_JQ_CPTAC2_COMPREF4_IMAC_06.2493.2493.3</td>
<td>3</td>
<td>15.09</td>
<td>15.09</td>
<td>15.09</td>
<td>54.7</td>
<td>1.050</td>
<td>3</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>8</td>
<td>H20120518_JQ_CPTAC2_COMPREF4_IMAC_06.2479.2479.3</td>
<td>3</td>
<td>15.63</td>
<td>15.63</td>
<td>15.63</td>
<td>63.9</td>
<td>0.000</td>
<td>7</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>9</td>
<td>H20120518_JQ_CPTAC2_COMPREF4_IMAC_06.2737.2737.5</td>
<td>5</td>
<td>13.10</td>
<td>13.10</td>
<td>13.10</td>
<td>68.9</td>
<td>1.195</td>
<td>3</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>10</td>
<td>H20120518_JQ_CPTAC2_COMPREF4_IMAC_06.2846.2846.3</td>
<td>3</td>
<td>13.34</td>
<td>13.34</td>
<td>13.34</td>
<td>69.3</td>
<td>0.000</td>
<td>5</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>11</td>
<td>H20120518_JQ_CPTAC2_COMPREF4_IMAC_06.3057.3057.4</td>
<td>4</td>
<td>12.16</td>
<td>7.65</td>
<td>7.34</td>
<td>71.4</td>
<td>3.717</td>
<td>3</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

### Review Fields

- **Filename**
- **Score**
- **FDR (Discriminant)**
- **Rank 1-2 score**
- **SPI (%)**
- **Unmatched ions**
- **Var mod sites**
- **Prep Avg Chi**
- **Isol Pur**
- **VML score**
- **Accession #**
- **Protein name**

### Validation and Sorting Filter results by:

- **Valid**
- **Retention Time**
- **Fwd.Rev score**
- **Rank 1-2 score**
- **SPI (%)**
- **Unmatched ions**
- **Var mod sites**
- **Prep Avg Chi**
- **Isol Pur**
- **VML score**
- **Accession #**
- **Protein name**

### Summarize Results for Review

- **Save as...**
- **Load...**

### Data directories:

- **Select...**
- **CPTAC2/COMPREF4/IMAC_06.35**

### Modification names

- **N-term**
- **C-term**
- **Cysteines**

### Equation ratios invert

- **Control: 127N**
- **Control: 128N**

### Reporter 

- **TMT 10**
- **TMT 12**
- **TMT 14**

### Fragmentation mode

- **Category:**
- **Modification names**
- **N-term**
- **C-term**
- **Cysteines**

---

Karl Clauser
Proteomics and Biomarker Discovery
Autovalidation
Confident identification and False Discovery Rates (FDR) at the PSM, peptide, and protein levels.
Autovalidation – Peptide/PSM level

Optimize thresholds by directory
Useful for low frequency matches in complex samples
• PTM’s in unenriched samples
• Proteogenomic variants
• High charge states (>4+)

Min Sequence length filter
• Short peptides often contribute to multiple proteins, inclusion may skew protein quantitation.
• Extractor MH+ lower limit: 750 for iTRAQ excludes short Arg peps
• Autovalidation MSL filter excludes short Lys & Arg peps

Separate Score thresholds for each:
• LC-MS/MS run
• Precursor charge
Separate precursor mass error range for each LC-MS/MS run
• Median +/- 3 std dev
Protein polishing assembles protein groups from the autovalidated PSM’s passing the peptide-level autovalidation step for all 4 patients (directories), determines the maximum protein level score of a protein group that consists entirely of distinct peptides estimated to be false-positive identifications (PSM’s with negative delta forward-reverse scores). Then PSM’s are unvalidated if they contribute to protein groups that:

a) have protein scores at or below the larger of
   i. the minimum protein score
   ii. the max false-positive protein score
b) are derived from only a single patient.
Collapsing Peptide Spectrum Matches (PSM’s) for Quantitation at the Protein level and Phosphosite level
Global proteome and phosphoproteome discovery workflow for TCGA breast tumors

105 TCGA breast cancer samples

Cryo fracture → Protein extraction → LysC/Trypsin → iTRAQ label (4-plex)

Tumor 1 → Tumor 2 → Tumor 3

Combine iTRAQ-labeled peptides

Basic Reversed-Phase

Internal Reference: 40 tumors peptide mix

Basic RP fractions (pH 10)

Metal affinity; enrich 95% of each fraction

Proteome (5%)

Phosphopeptides (12 fractions)

All peptides (24 fractions)

LC-MS/MS (Q-Exactive, Top12, HCD)

Data analysis in Spectrum Mill

Clustering, Classification, Differential Analysis

1 mg total protein per tumor

Internal reference: equal representation of basal, Her2 and Luminal A/B subtypes
Hierarchical clustering of proteome yields 3 major groups: basal, luminal, and a new group

- Basal-Enriched
- Luminal-Enriched
- Uncharacterized Clusters
- Basal-Enriched

Row – protein
Cell - iTRAQ ratio
Column - patient

Ratio count >1;
Stdev >1.5
>15 tumors
4,657 proteins
Phosphoproteome data clusters tumors identically to the proteome data

stdev > 1.75
> 15 tumors
5,065 p-sites

• Row - phosphosite
• Cell - iTRAQ ratio
• Column - patient

Luminal-Enriched
Uncharacterized Cluster
Basal-Enriched
Protein Inference Tutorial
Nesvizhskii, *Mol Cell Proteomics*, 4, 1419-1440, **2005**.
### Collapsing PSM’s to Protein level

<table>
<thead>
<tr>
<th>Protein Name</th>
<th>Group #</th>
<th>Subgroup #</th>
</tr>
</thead>
<tbody>
<tr>
<td>poly(C) binding protein 1</td>
<td>525</td>
<td>525.1</td>
</tr>
<tr>
<td>poly(C)-binding protein 2 isofrom b variant</td>
<td>526</td>
<td>526.1</td>
</tr>
<tr>
<td>PDZ and LM domain 2 isofrom 2</td>
<td>526</td>
<td>526.1</td>
</tr>
<tr>
<td>protein phosphatase 1, catalytic subunit, beta isofrom 1</td>
<td>527</td>
<td>527.1</td>
</tr>
<tr>
<td>protein phosphatase 1, catalytic subunit, alpha isofrom 1</td>
<td>527</td>
<td>527.2</td>
</tr>
<tr>
<td>calumenin</td>
<td>527</td>
<td>527.3</td>
</tr>
<tr>
<td>protein phosphatase 1, catalytic subunit, gamma isofrom</td>
<td>528</td>
<td>528.1</td>
</tr>
<tr>
<td>RPL7</td>
<td>528</td>
<td>528.1</td>
</tr>
<tr>
<td>lactate dehydrogenase A</td>
<td>530</td>
<td>530.1</td>
</tr>
<tr>
<td>lactate dehydrogenase B</td>
<td>530</td>
<td>530.2</td>
</tr>
<tr>
<td>cAMP responsive element binding protein 1 isofrom B</td>
<td>531</td>
<td>531.1</td>
</tr>
</tbody>
</table>

Distinct Peptides must be length >8 to spawn a new subgroup.
<table>
<thead>
<tr>
<th>ID</th>
<th>Sub Group</th>
<th>Length</th>
<th>Identical Axs</th>
<th>%ID</th>
<th>Species</th>
<th>Protein Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>54696354</td>
<td>1</td>
<td>327</td>
<td>327</td>
<td>100.0</td>
<td>Homo sapiens</td>
<td>protein phosphatase 1, catalytic subunit, beta isoform</td>
</tr>
<tr>
<td>46249376</td>
<td>2</td>
<td>286</td>
<td>286</td>
<td>99.7</td>
<td>Homo sapiens</td>
<td>protein phosphatase 1, catalytic subunit, beta isoform 1</td>
</tr>
<tr>
<td>48527798</td>
<td>2</td>
<td>286</td>
<td>286</td>
<td>88.8</td>
<td>Homo sapiens</td>
<td>protein phosphatase 1, catalytic subunit, alpha isoform 2</td>
</tr>
<tr>
<td>4506003</td>
<td>2</td>
<td>330</td>
<td>291</td>
<td>88.2</td>
<td>Homo sapiens</td>
<td>protein phosphatase 1, catalytic subunit, alpha isoform 1</td>
</tr>
<tr>
<td>56790945</td>
<td>2</td>
<td>341</td>
<td>290</td>
<td>85.0</td>
<td>Homo sapiens</td>
<td>protein phosphatase 1, catalytic subunit, alpha isoform 3</td>
</tr>
<tr>
<td>5668560</td>
<td>3</td>
<td>294</td>
<td>269</td>
<td>91.5</td>
<td>Homo sapiens</td>
<td>serine/threonine phosphatase 1 gamma</td>
</tr>
<tr>
<td>4506007</td>
<td>3</td>
<td>323</td>
<td>285</td>
<td>88.2</td>
<td>Homo sapiens</td>
<td>protein phosphatase 1, catalytic subunit, gamma isoform</td>
</tr>
</tbody>
</table>

Highlighted identified peptides:

- **54696354 (1)**: MADG-ELNVDLSLTIQLLEVR---------GCPKPHYVQRTEABVRGLCIKHRSEIFSQPILEAPLKICDGDIHGYTDLLRLEFYYGFPPEANYLF
- **46249376 (1)**: MADG-ELNVDLSLTIQLLEVR---------GCPKPHYVQRTEABVRGLCIKHRSEIFSQPILEAPLKICDGDIHGYTDLLRLEFYYGFPPEANYLF
- **48527798 (2)**: MDSDEKLLNLSIIQLLEVR---------GCPKPHYVQRTEABVRGLCIKHRSEIFSQPILEAPLKICDGDIHGYTDLLRLEFYYGFPPEANYLF
- **4506003 (2)**: MDSDEKLLNLSIIQLLEVR---------GCPKPHYVQRTEABVRGLCIKHRSEIFSQPILEAPLKICDGDIHGYTDLLRLEFYYGFPPEANYLF
- **56790945 (3)**: LGDYVDRCQSLETICLLAYIKYPENFFLLIRCNHACSNIRYYGFDYDKRRFNNIKLMHFTDFCNCLPLAAVIDBKFICCHGLSDLQMEQIDRI
- **5668560 (3)**: LGDYVDRCQSLETICLLAYIKYPENFFLLIRCNHACSNIRYYGFDYDKRRFNNIKLMHFTDFCNCLPLAAVIDBKFICCHGLSDLQMEQIDRI
- **4506007 (3)**: LGDYVDRCQSLETICLLAYIKYPENFFLLIRCNHACSNIRYYGFDYDKRRFNNIKLMHFTDFCNCLPLAAVIDBKFICCHGLSDLQMEQIDRI
- **54696354 (1)**: MRPTDVPDGLLLCDLWSDLKDQVQWGENDRGVSFTGADUVSKFLNRHDLDDLICRAHVQVGEDYFFAKRQLVTLSAPNYCGEFDNAGCHMSVDTEKL
- **46249376 (1)**: MRPTDVPDGLLLCDLWSDLKDQVQWGENDRGVSFTGADUVSKFLNRHDLDDLICRAHVQVGEDYFFAKRQLVTLSAPNYCGEFDNAGCHMSVDTEKL
- **48527798 (2)**: MRPTDVPDQLLLCDLWSDLKDQVQWGENDRGVSFTGADUVSKFLNRHDLDDLICRAHVQVGEDYFFAKRQLVTLSAPNYCGEFDNAGCHMSVDTEKL
- **4506003 (2)**: MRPTDVPDQLLLCDLWSDLKDQVQWGENDRGVSFTGADUVSKFLNRHDLDDLICRAHVQVGEDYFFAKRQLVTLSAPNYCGEFDNAGCHMSVDTEKL
- **56790945 (3)**: MRPTDVPDQLLLCDLWSDLKDQVQWGENDRGVSFTGADUVSKFLNRHDLDDLICRAHVQVGEDYFFAKRQLVTLSAPNYCGEFDNAGCHMSVDTEKL
- **5668560 (3)**: MRPTDVPDQLLLCDLWSDLKDQVQWGENDRGVSFTGADUVSKFLNRHDLDDLICRAHVQVGEDYFFAKRQLVTLSAPNYCGEFDNAGCHMSVDTEKL
- **4506007 (3)**: MRPTDVPDQLLLCDLWSDLKDQVQWGENDRGVSFTGADUVSKFLNRHDLDDLICRAHVQVGEDYFFAKRQLVTLSAPNYCGEFDNAGCHMSVDTEKL
- **54696354 (1)**: MCFSQLIKPSEK-KAKY-QYGLNS-GRPVTPFR---ANPPKRR
- **46249376 (1)**: MCFSQLIKPSEK-KAKY-QYGLNS-GRPVTPFR---ANPPKRR
- **48527798 (2)**: MCFSQLIKPADNNNGKGYKGFNSLNGPGRPRIP---SAA---
- **4506003 (2)**: MCFSQLIKPADNNNGKGYKGFNSLNGPGRPRIP---SAA---
- **56790945 (3)**: MCFSQLIKPADNNNGKGYKGFNSLNGPGRPRIP---SAA---
- **5668560 (3)**: MCFQ---------------------------
- **4506007 (3)**: MCFSQLIKPAEKKPKNA--------TRPVTPPRGMITHQAKK-
Collapsing PSM’s to VM site level

Non-conflicting localization requires

<table>
<thead>
<tr>
<th>m/z</th>
<th>ppm</th>
<th>ID</th>
<th>Score</th>
<th>Filename</th>
</tr>
</thead>
<tbody>
<tr>
<td>552.960</td>
<td>5.1</td>
<td>13.83</td>
<td>0.00</td>
<td>0 1 0.96 0.08 1.27</td>
</tr>
<tr>
<td>643.689</td>
<td>-0.1</td>
<td>10.84</td>
<td>0.00</td>
<td>0 1 0.93 0.06 1.45</td>
</tr>
<tr>
<td>500.923</td>
<td>-1.4</td>
<td>8.06</td>
<td>0.00</td>
<td>0 1 1.48 0.31 0.88</td>
</tr>
<tr>
<td>750.882</td>
<td>0.8</td>
<td>7.23</td>
<td>0.00</td>
<td>0 1 1.84 0.16 0.99</td>
</tr>
<tr>
<td>750.881</td>
<td>-0.2</td>
<td>7.09</td>
<td>0.00</td>
<td>0 1 1.96 0.14 0.71</td>
</tr>
<tr>
<td>678.833</td>
<td>3.6</td>
<td>7.07</td>
<td>0.00</td>
<td>0 1 0.99 0.10 0.96</td>
</tr>
<tr>
<td>551.943</td>
<td>0.1</td>
<td>10.81</td>
<td>4.43</td>
<td>1 0 0.40 0.70 1.84</td>
</tr>
<tr>
<td>755.361</td>
<td>2.1</td>
<td>7.91</td>
<td>7.91</td>
<td>1 0 1.03 0.89 1.60</td>
</tr>
<tr>
<td>827.408</td>
<td>-2.0</td>
<td>7.73</td>
<td>7.73</td>
<td>1 0 0.57 0.90 1.63</td>
</tr>
<tr>
<td>721.043</td>
<td>0.7</td>
<td>16.86</td>
<td>2.29</td>
<td>1 0 0.48 0.98 1.77</td>
</tr>
<tr>
<td>645.813</td>
<td>1.0</td>
<td>12.89</td>
<td>2.66</td>
<td>1 0 0.74 0.08 1.45</td>
</tr>
<tr>
<td>541.035</td>
<td>2.5</td>
<td>12.36</td>
<td>0.00</td>
<td>0 1 0.42 1.08 1.74</td>
</tr>
<tr>
<td>1081.057</td>
<td>-2.6</td>
<td>13.71</td>
<td>1.21</td>
<td>1 0 0.44 0.76 1.56</td>
</tr>
<tr>
<td>541.032</td>
<td>-1.7</td>
<td>8.73</td>
<td>0.00</td>
<td>0 1 0.48 0.74 1.56</td>
</tr>
<tr>
<td>721.040</td>
<td>-2.5</td>
<td>7.37</td>
<td>0.00</td>
<td>0 1 0.35 0.75 1.52</td>
</tr>
<tr>
<td>430.878</td>
<td>0.6</td>
<td>4.40</td>
<td>0.00</td>
<td>0 1 0.65 0.06 1.52</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>#Am</th>
<th>m/z</th>
<th>ppm</th>
<th>ID</th>
<th>Score</th>
<th>Filename</th>
</tr>
</thead>
<tbody>
<tr>
<td>551.943</td>
<td>0.1</td>
<td>10.81</td>
<td>2.66</td>
<td>1 0 0.74 0.08 1.45</td>
<td></td>
</tr>
<tr>
<td>755.361</td>
<td>1.0</td>
<td>12.89</td>
<td>2.66</td>
<td>1 0 0.74 0.08 1.45</td>
<td></td>
</tr>
<tr>
<td>827.408</td>
<td>-2.0</td>
<td>7.73</td>
<td>7.73</td>
<td>1 0 0.57 0.90 1.63</td>
<td></td>
</tr>
<tr>
<td>721.043</td>
<td>0.7</td>
<td>16.86</td>
<td>2.29</td>
<td>1 0 0.48 0.98 1.77</td>
<td></td>
</tr>
<tr>
<td>645.813</td>
<td>1.0</td>
<td>12.89</td>
<td>2.66</td>
<td>1 0 0.74 0.08 1.45</td>
<td></td>
</tr>
<tr>
<td>541.035</td>
<td>2.5</td>
<td>12.36</td>
<td>0.00</td>
<td>0 1 0.42 1.08 1.74</td>
<td></td>
</tr>
<tr>
<td>1081.057</td>
<td>-2.6</td>
<td>13.71</td>
<td>1.21</td>
<td>1 0 0.44 0.76 1.56</td>
<td></td>
</tr>
<tr>
<td>541.032</td>
<td>-1.7</td>
<td>8.73</td>
<td>0.00</td>
<td>0 1 0.48 0.74 1.56</td>
<td></td>
</tr>
<tr>
<td>721.040</td>
<td>-2.5</td>
<td>7.37</td>
<td>0.00</td>
<td>0 1 0.35 0.75 1.52</td>
<td></td>
</tr>
<tr>
<td>430.878</td>
<td>0.6</td>
<td>4.40</td>
<td>0.00</td>
<td>0 1 0.65 0.06 1.52</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>m/z</th>
<th>ppm</th>
<th>ID</th>
<th>Score</th>
<th>Filename</th>
</tr>
</thead>
<tbody>
<tr>
<td>552.960</td>
<td>5.1</td>
<td>13.83</td>
<td>0.00</td>
<td>0 1 0.96 0.08 1.27</td>
</tr>
<tr>
<td>643.689</td>
<td>-0.1</td>
<td>10.84</td>
<td>0.00</td>
<td>0 1 0.93 0.06 1.45</td>
</tr>
<tr>
<td>500.923</td>
<td>-1.4</td>
<td>8.06</td>
<td>0.00</td>
<td>0 1 1.48 0.31 0.88</td>
</tr>
<tr>
<td>750.882</td>
<td>0.8</td>
<td>7.23</td>
<td>0.00</td>
<td>0 1 1.84 0.16 0.99</td>
</tr>
<tr>
<td>750.881</td>
<td>-0.2</td>
<td>7.09</td>
<td>0.00</td>
<td>0 1 1.96 0.14 0.71</td>
</tr>
<tr>
<td>678.833</td>
<td>3.6</td>
<td>7.07</td>
<td>0.00</td>
<td>0 1 0.99 0.10 0.96</td>
</tr>
<tr>
<td>551.943</td>
<td>0.1</td>
<td>10.81</td>
<td>4.43</td>
<td>1 0 0.40 0.70 1.84</td>
</tr>
<tr>
<td>755.361</td>
<td>2.1</td>
<td>7.91</td>
<td>7.91</td>
<td>1 0 1.03 0.89 1.60</td>
</tr>
<tr>
<td>827.408</td>
<td>-2.0</td>
<td>7.73</td>
<td>7.73</td>
<td>1 0 0.57 0.90 1.63</td>
</tr>
<tr>
<td>721.043</td>
<td>0.7</td>
<td>16.86</td>
<td>2.29</td>
<td>1 0 0.48 0.98 1.77</td>
</tr>
<tr>
<td>645.813</td>
<td>1.0</td>
<td>12.89</td>
<td>2.66</td>
<td>1 0 0.74 0.08 1.45</td>
</tr>
<tr>
<td>541.035</td>
<td>2.5</td>
<td>12.36</td>
<td>0.00</td>
<td>0 1 0.42 1.08 1.74</td>
</tr>
<tr>
<td>1081.057</td>
<td>-2.6</td>
<td>13.71</td>
<td>1.21</td>
<td>1 0 0.44 0.76 1.56</td>
</tr>
<tr>
<td>541.032</td>
<td>-1.7</td>
<td>8.73</td>
<td>0.00</td>
<td>0 1 0.48 0.74 1.56</td>
</tr>
<tr>
<td>721.040</td>
<td>-2.5</td>
<td>7.37</td>
<td>0.00</td>
<td>0 1 0.35 0.75 1.52</td>
</tr>
<tr>
<td>430.878</td>
<td>0.6</td>
<td>4.40</td>
<td>0.00</td>
<td>0 1 0.65 0.06 1.52</td>
</tr>
</tbody>
</table>
Key Features of Spectrum Mill Quantitation

- MS/MS reporter ion ratio based: iTRAQ, TMT
  - Median of all PSM ratios for each Protein, VM-site
  - Not sum of all PSM reporter ion intensities, then ratio for Protein, VM-site
  - Exclude ratios with PIP < 50%
  - Exclude ½ of false positive ID’s: Delta Fwd-Rev < 0
  - Exclude peptides with no label

- MS precursor ion ratio based: SILAC
  - Only 1 member of H/L pair H/M/L triplet needs to be triggered for MS/MS
  - Median of all H/L pair H/M/L triplet ratios for each Protein, VM-site
  - Exclude ratios from poor precursor ion isotope cluster shape

- Label – free
  - Only uses peptides identified by MS/MS
  - Sums up all precursor ion peak area for all PSM’s for a protein

- Peptides shared between protein subgroups
  - Use shared in each subgroup
  - Use only unshared: Subgroup specific
Automation

Service Request Manager (SRM) and Workflows
### SRM Request Queue Status

There are 12 requests in the queue. Available memory: 48.6 Gb of 68.7 Gb

<table>
<thead>
<tr>
<th>#</th>
<th>Task Id</th>
<th>Status</th>
<th>Task Type</th>
<th>Data Directory</th>
<th>Dependencies</th>
<th>User</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>150225164124.4521</td>
<td>Running</td>
<td>MS/MS Search</td>
<td>Jana/2015_OCBC_study/OCBC_phospho</td>
<td>150225164053.4520</td>
<td>Anonymous gp45a-f4</td>
</tr>
<tr>
<td>2</td>
<td>150225164144.4522</td>
<td>Queued</td>
<td>Autovalidation</td>
<td>Jana/2015_OCBC_study/OCBC_phospho</td>
<td>150225164053.4520</td>
<td>Anonymous gp45a-f4</td>
</tr>
<tr>
<td>3</td>
<td>150225164203.4523</td>
<td>Running</td>
<td>MS/MS Search</td>
<td>Jana/2015_OCBC_study/OCBC_proteome</td>
<td>150225164046.4519</td>
<td>Anonymous gp45a-f4</td>
</tr>
<tr>
<td>4</td>
<td>150225164208.4524</td>
<td>Queued</td>
<td>Autovalidation</td>
<td>Jana/2015_OCBC_study/OCBC_proteome</td>
<td>150225164046.4519</td>
<td>Anonymous gp45a-f4</td>
</tr>
<tr>
<td>5</td>
<td>150225221936.4528</td>
<td>Queued</td>
<td>Autovalidation</td>
<td>Kar/ECMatlas/MM/nCO_5555_5557_OGE</td>
<td>150225221933.4525</td>
<td>Anonymous gp556-812</td>
</tr>
<tr>
<td>6</td>
<td>150225221937.4529</td>
<td>Queued</td>
<td>Autovalidation</td>
<td>Kar/ECMatlas/MM/nCO_5555_5557_OGE</td>
<td>150225221933.4525</td>
<td>Anonymous gp556-812</td>
</tr>
<tr>
<td>7</td>
<td>150225221938.4530</td>
<td>Queued</td>
<td>Autovalidation</td>
<td>Kar/ECMatlas/MM/nCO_5555_5557_OGE</td>
<td>150225221933.4525</td>
<td>Anonymous gp556-812</td>
</tr>
<tr>
<td>8</td>
<td>150225222805.4534</td>
<td>Queued</td>
<td>Autovalidation</td>
<td>Kar/ECMatlas/MM/nLU_5555_5557_OGE</td>
<td>150225222802.4531</td>
<td>Anonymous gp556-812</td>
</tr>
<tr>
<td>9</td>
<td>150225222806.4741</td>
<td>Queued</td>
<td>Autovalidation</td>
<td>Kar/ECMatlas/MM/nLU_5555_5557_OGE</td>
<td>150225222802.4531</td>
<td>Anonymous gp556-812</td>
</tr>
<tr>
<td>10</td>
<td>150225222807.4742</td>
<td>Queued</td>
<td>Autovalidation</td>
<td>Kar/ECMatlas/MM/nLU_5555_5557_OGE</td>
<td>150225222802.4531</td>
<td>Anonymous gp556-812</td>
</tr>
<tr>
<td>11</td>
<td>150225221934.4526P</td>
<td>Running (181:206)</td>
<td>MS/MS Search</td>
<td>Kar/ECMatlas/MM/nCO_5555_5557_OGE</td>
<td>150225221933.4525</td>
<td>Anonymous gp556-812</td>
</tr>
<tr>
<td>12</td>
<td>150225222804.4532P</td>
<td>Running (120:424)</td>
<td>MS/MS Search</td>
<td>Kar/ECMatlas/MM/nLU_5555_5557_OGE</td>
<td>150225222802.4531</td>
<td>Anonymous gp556-812</td>
</tr>
</tbody>
</table>

SRM coordinates task execution on the queue to maximize CPU usage while maintaining workflow order dependencies.

Tasks in each user’s workflow execute in serial.

User A’s workflow executes in parallel with user B’s workflow.

Any user can add/remove tasks from queue.
### SRM Completion Log

If a task fails the SRM aborts all subsequent task in the workflows.

<table>
<thead>
<tr>
<th>Task Id</th>
<th>Task Type</th>
<th>Task Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>150205160444.9455</td>
<td>P/P Summary</td>
<td>Jana/TMTviTRAQ/2015_researchTMTviTRAQ/TMT10_phosphoproteome1</td>
</tr>
<tr>
<td>150205160402.9453</td>
<td>Autovalidation</td>
<td>Jana/TMTviTRAQ/2015_researchTMTviTRAQ/TMT10_phosphoproteome1</td>
</tr>
<tr>
<td>150205155152.9451</td>
<td>MS/MS Search</td>
<td>Jana/TMTviTRAQ/2015_researchTMTviTRAQ/TMT10_phosphoproteome1</td>
</tr>
<tr>
<td>150205160403.9454</td>
<td>Autovalidation</td>
<td>Jana/TMTviTRAQ/2015_researchTMTviTRAQ/TMT10_phosphoproteome2</td>
</tr>
<tr>
<td>150205155163.9452</td>
<td>MS/MS Search</td>
<td>Jana/TMTviTRAQ/2015_researchTMTviTRAQ/TMT10_phosphoproteome2</td>
</tr>
<tr>
<td>150205150433.9447</td>
<td>P/P Summary</td>
<td>Jana/TMTviTRAQ/2015_researchTMTviTRAQ/TMT10_proteome1</td>
</tr>
<tr>
<td>150207145805.11388</td>
<td>Quality Metrics</td>
<td>Fil/Xman/20150206_singleshot</td>
</tr>
<tr>
<td>150207145723.11387</td>
<td>P/P Summary</td>
<td>Fil/Xman/20150206_singleshot</td>
</tr>
<tr>
<td>150205145341.9446</td>
<td>Autovalidation</td>
<td>Jana/TMTviTRAQ/2015_researchTMTviTRAQ/TMT10_proteome1</td>
</tr>
<tr>
<td>150206214430.9473</td>
<td>Quality Metrics</td>
<td>Fil/Xman/20150206_singleshot</td>
</tr>
<tr>
<td>150206214341.9472</td>
<td>P/P Summary</td>
<td>Fil/Xman/20150206_singleshot</td>
</tr>
<tr>
<td>150206214244.9471</td>
<td>Autovalidation</td>
<td>Fil/Xman/20150206_singleshot</td>
</tr>
<tr>
<td>150206214230.9470</td>
<td>Autovalidation</td>
<td>Fil/Xman/20150206_singleshot</td>
</tr>
</tbody>
</table>
# Simple Workflow – Broad Institute QC sample

## Agilent Spectrum Mill - Workflows

<table>
<thead>
<tr>
<th>Spectrum Mill</th>
<th>Request Queue</th>
<th>Completion Log</th>
<th>Extractor</th>
<th>MS/MS Search</th>
<th>Autovalidation</th>
<th>Protein/Peptide Summary</th>
<th>Tool Belt</th>
<th>Help</th>
</tr>
</thead>
</table>

### Data Directories

- **JurkatQC/Jurkat_Hubble2/20150204**

## Workflow

Select and Execute a workflow of Data Extraction, MS/MS Searches, Autovalidation, and Generation of a Result Summary.

**Select a Task to view its parameters**

- **Workflow:**
  - JurkatDC_Dig_141110
  - HLA_NoEnzyme
  - Trypsin_Phospho
  - Jurkat/Jurkat QC
  - Jurkat/Jurkat QC QE EMARG
  - Jurkat/Jurkat QC QE EMARG
  - Jurkat/Jurkat QC QE EMARG
  - Jurkat/Jurkat QC QE EMARG

- **Tasks:**
  - Extraction - Jurkat/Default_Extraction
  - MS/MS Search - Jurkat/UniprotHuman14_20_20_AmMqc_SPI40_QE
  - Autovalidation - Jurkat/peptide_auto_25_ppm_MSL6_1.0
  - Archive Data - Jurkat/results
  - Quality Metrics - Jurkat/FDR_ppm_z_PIP_PAU_id_chrom_digend_mods

## Agilent Spectrum Mill - MS/MS Search - Jurkat/UniprotHuman14_20_20_AmMqc_SPI40_QE

### Search

- **Validation filter:** spectrum-not-marked-sequence-not-validated
- **Batch size:** 500
- **Search previous hits**
- **Max reported hits:** 5

#### Search Parameters

- **Database:** UniProt human 20141017 RNFiSnr 150 contains
- **Species:** All

#### Modifications

- **Fixed:** Carbamidomethylation (C)
- **Variable:** Acetyl (ProtN-term) Oxidized methionine (M) Pyroglutamic acid (N-termQ) Deamidated (N) Pyrrol Carbamidomethyl Cys (N-termC)

### Search Criteria

- **Matching Tolerances**
  - **Minimum matched peak intensity:** 40%
  - **Instrument:** ESI QExactive HCD
  - **Masses are:** Monoisotopic
  - **Precursor mass tolerance:** ±20 ppm
  - **Product mass tolerance:** ±20 ppm

- **Search Mode**
  - **Calculate reversed database scores**
  - **Search mode:** Variable modifications
  - **Precursor mass shift range:** -18.0 to 64 Da

### Data Files

- **Fragmentation mode:** All
Complex Workflow – Mixed CID/ETD Dataset

Agilent Spectrum Mill - Workflows

<table>
<thead>
<tr>
<th>Spectrum Mill</th>
<th>Request Queue</th>
<th>Completion Log</th>
<th>Extractor</th>
<th>MS/MS Search</th>
<th>Autovalidation</th>
<th>Protein/Peptide Summary</th>
<th>Tool Belt</th>
<th>Help</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Directories</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Select ...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

✓ KarlECMAtlas/Naba_2014_BMC_Cancer/D120217_CoK_2205A
✓ KarlECMAtlas/Naba_2014_BMC_Cancer/D120417_LiN_1HS1
✓ KarlECMAtlas/Naba_2014_BMC_Cancer/D1201010_CoN_1948A
✓ KarlECMAtlas/Naba_2014_BMC_Cancer/D1201012_LiK_2333A

Workflow

Select and Execute a workflow of Data Extraction, MS/MS Searches, Autovalidation, and Generation of a Result Summary
Select a Task to view its parameters

Execute

Max CPUs per search

Edit Workflow

Refresh

Agilent Spectrum Mill - MS/MS Autovalidation - Karl'peptide_byDir_z4_1_0_MSL7_CID

Validation Parameters

Strategy:  ○ Fixed thresholds  ○ Auto thresholds  ○ Auto thresholds - discriminant

Mode: Peptide  ▼  Auto determine using score, delta R1-R2 thresholds to reach a target FDR

Optimize score & R1-R2 score thresholds with max FDR:  1.0  % across each:  ○ LC run  ○ Directory

Precursor charge range:  4  to  4  Min Sequence Length:  7  ▼

Required AAs: any  ▼  Disallowed AAs: none  ▼

Filtering

None (ppm)

None (SC/pl)

Fixed range for all runs

○ Fixed precursor mass error  ○ Fixed Solution Charge  ○ Fixed peptide pl

Low -1.0  High 30.0  ppm

Low -2  High 6

Low 3.0  High 10.0
Edit an existing workflow to search a different database
Quality Metrics

Troubleshooting
- LC Gradient and column performance
- Measuring data acquisition strategy changes
- Mass calibration drift
- MS ion optics need cleaning
- iTRAQ labeling efficiency
- IMAC phospho enrichment
- Cysteine reduction/alkylation failure
# MS and Chromatography Metrics

Measure effect of changes in acquisition parameters and chromatography on recently installed QExactive Plus at Broad Institute

<table>
<thead>
<tr>
<th>Comment</th>
<th>Pep Match</th>
<th>MS/MS spectra collected C</th>
<th>MS/MS spectra collected MI m/z = 0.0</th>
<th>MS/MS spectra filtered F</th>
<th>MS/MS spectra valid V</th>
<th>MS/MS spectra valid MI m/z = 0.0</th>
<th>Mean Precursor Mass Error (ppm)</th>
<th>Start time mid 90 % run (min)</th>
<th>End time mid 90 % run (min)</th>
<th>Time span mid 90 % run (min)</th>
<th>Gradient Shape mid 90 % filtered spectra in run</th>
<th>Median ID Score</th>
<th>Mean Precursor Mass Error (ppm)</th>
<th>PIP bin 100-90% matched spectra (%)</th>
<th>PIP bin 50-90% matched spectra (%)</th>
<th>Distinct Peps CS Total (#)</th>
<th>FDR Spectra (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>LC-Franklin, New pump B</td>
<td>pref</td>
<td>59,229</td>
<td>15,118</td>
<td>54,812</td>
<td>34,687</td>
<td>5,706</td>
<td>11.57</td>
<td>1.8</td>
<td>20</td>
<td>98</td>
<td>78</td>
<td>9876532</td>
<td>12</td>
<td>3.32E+13</td>
<td>21.5</td>
<td>10.3</td>
<td>28,743</td>
</tr>
<tr>
<td></td>
<td>pref</td>
<td>59,003</td>
<td>14,832</td>
<td>54,249</td>
<td>35,140</td>
<td>5,750</td>
<td>11.67</td>
<td>0.3</td>
<td>21</td>
<td>98</td>
<td>78</td>
<td>9986643</td>
<td>12</td>
<td>3.64E+13</td>
<td>21.8</td>
<td>10.6</td>
<td>28,683</td>
</tr>
<tr>
<td></td>
<td>on</td>
<td>52,270</td>
<td>-</td>
<td>47,892</td>
<td>34,811</td>
<td>-</td>
<td>12.07</td>
<td>0.3</td>
<td>21</td>
<td>98</td>
<td>78</td>
<td>9997754</td>
<td>12</td>
<td>6.15E+13</td>
<td>22.1</td>
<td>9.7</td>
<td>27,929</td>
</tr>
<tr>
<td>Isolation width decreased 2.5 to 2.0 on</td>
<td></td>
<td>51,827</td>
<td>-</td>
<td>49,963</td>
<td>36,876</td>
<td>-</td>
<td>12.68</td>
<td>1.1</td>
<td>17</td>
<td>98</td>
<td>81</td>
<td>9987753</td>
<td>12</td>
<td>6.31E+13</td>
<td>29.7</td>
<td>6.4</td>
<td>29,598</td>
</tr>
<tr>
<td>LC-Klaus</td>
<td>on</td>
<td>49,015</td>
<td>-</td>
<td>47,140</td>
<td>32,487</td>
<td>-</td>
<td>11.73</td>
<td>-0.2</td>
<td>18</td>
<td>98</td>
<td>80</td>
<td>9999876</td>
<td>15</td>
<td>5.35E+13</td>
<td>26.1</td>
<td>9.4</td>
<td>26,176</td>
</tr>
<tr>
<td></td>
<td>pref</td>
<td>60,705</td>
<td>19,121</td>
<td>57,497</td>
<td>38,127</td>
<td>8,198</td>
<td>11.79</td>
<td>2.2</td>
<td>21</td>
<td>100</td>
<td>79</td>
<td>9897654</td>
<td>11</td>
<td>4.88E+13</td>
<td>27.0</td>
<td>7.6</td>
<td>30,759</td>
</tr>
<tr>
<td>16 min longer gradient, isolation width decreased 2.0 to 1.6 pref</td>
<td></td>
<td>68,003</td>
<td>23,946</td>
<td>63,633</td>
<td>42,998</td>
<td>10,670</td>
<td>12.36</td>
<td>-0.4</td>
<td>20</td>
<td>115</td>
<td>94</td>
<td>8898754</td>
<td>13</td>
<td>6.13E+13</td>
<td>35.0</td>
<td>4.8</td>
<td>34,719</td>
</tr>
</tbody>
</table>

**Replaced Pump B**

**Replaced LC**

**Lengthened Gradient**
# MS and Chromatography Metrics

Measure effect of changes in acquisition parameters and chromatography on recently installed QExactive Plus at Broad Institute

| Comment | Pep Match | MS/MS spectra collected C | MS/MS spectra collected MI m/z = 0.0 | MS/MS spectra filtered F | MS/MS spectra valid V | MS/MS spectra valid MI m/z = 0.0 | Mean Precursor Mass Error (ppm) | Start time mid 90 % run (min) | End time mid 90 % run (min) | Time span mid 90 % run (min) | Gradient Shape mid 90 % filtered spectra in run | Mediation MS1 peak width mid 90 % run (sec) | Total precursor XIC mid 90 % matched spectra in run | PIP bin 1 100-90 Spectra (%) | PIP bin 6 <50 Spectra (%) | Distinct Peps CS Total (#) | FDR Spectra (%) |
|---------|-----------|--------------------------|-------------------------------------|--------------------------|------------------------|---------------------------------|--------------------------------|------------------------------|----------------------------|-----------------------------|---------------------------------|-------------------------------|-----------------------------|----------------------|-------------------|---------------------|
| Isolation width decreased 2.5 to 2.0 on | 52,270 | - | 47,892 | - | 34,811 | - | 12.07 | 0.3 | 21 | 98 | 78 | 9997754 | 12 | 6.15E+13 | 22.1 | 9.7 | 27,929 | 0.72 |
| LC-Klaus on | 51,827 | - | 49,963 | - | 36,876 | - | 12.68 | 1.1 | 17 | 98 | 81 | 9987753 | 12 | 6.31E+13 | 29.7 | 6.4 | 29,598 | 0.73 |
| 16 min longer gradient, isolation width decreased 2.0 to 1.6 pref | 60,705 | 19,121 | 57,497 | 38,127 | 8,198 | 11.79 | 2.2 | 21 | 100 | 79 | 9897654 | 11 | 4.88E+13 | 27.0 | 7.6 | 30,759 | 0.87 |
| pref | 68,003 | 23,946 | 63,633 | 42,998 | 10,670 | 12.36 | -0.4 | 20 | 115 | 94 | 8898754 | 13 | 6.13E+13 | 35.0 | 4.8 | 34,719 | 0.96 |

**Pep Match (on/pref)**
Precursor isotope cluster

**Lengthened Gradient**

**Narrowed isolation width**
## Phosphorylation Quality Metrics

| Directory | Files | MS/MS spectra | Raw spectra (% | s|t|y Sites spectra (%) | s|t|y Sites Localized spectra (#) | s|t|y Sites Distinct Peptides (CI%) | s|t|y Sites Distinct Peptides FDR (CI) | s|t|y Sites Distinct Peptides FDR (CI%) | s|t|y Sites FDR (CI%) | s|t|y Sites FDR (CI#) |
|-----------|-------|----------------|----------------|-------------------|---------------------|-------------------------------|-------------------------------|-------------------------------|----------------|----------------|
| 01TCGA    | 13    | 77,152         | 94.5           | 72,930            | 56.4                | 32,179                        | 92.4                          | 0.47                          | 0.92          | 32,179         |
| 02TCGA    | 13    | 82,578         | 90.0           | 74,288            | 55.7                | 31,843                        | 85.9                          | 0.51                          | 1.00          | 31,843         |
| 03TCGA    | 13    | 91,894         | 77.8           | 71,460            | 57.0                | 31,193                        | 75.6                          | 0.48                          | 0.95          | 31,193         |
| 04TCGA    | 13    | 98,744         | 72.6           | 71,690            | 57.9                | 31,921                        | 68.1                          | 0.57                          | 1.12          | 31,921         |
| 05TCGA    | 13    | 90,028         | 83.4           | 75,068            | 57.6                | 31,748                        | 78.2                          | 0.49                          | 1.05          | 31,748         |
| 06TCGA    | 13    | 96,310         | 53.2           | 51,264            | 58.7                | 22,933                        | 51.4                          | 0.64                          | 1.18          | 22,933         |
| 07TCGA    | 13    | 83,351         | 73.0           | 60,858            | 58.4                | 26,668                        | 68.9                          | 0.54                          | 1.03          | 26,668         |
| 08TCGA    | 13    | 102,722        | 66.7           | 68,521            | 55.1                | 30,355                        | 66.3                          | 0.65                          | 1.23          | 30,355         |
| 09TCGA    | 13    | 87,016         | 67.1           | 58,360            | 55.5                | 27,513                        | 65.1                          | 0.62                          | 1.19          | 27,513         |
| 10TCGA    | 13    | 88,406         | 78.5           | 69,368            | 55.7                | 29,632                        | 74.7                          | 0.58                          | 1.15          | 29,632         |
| 11TCGA    | 13    | 60,593         | 92.5           | 56,029            | 49.4                | 25,075                        | 90.0                          | 0.52                          | 0.95          | 25,075         |
| 12TCGA    | 13    | 63,858         | 94.1           | 60,117            | 50.5                | 26,540                        | 92.5                          | 0.45                          | 0.81          | 26,540         |
| 13TCGA    | 13    | 81,810         | 85.2           | 69,701            | 49.5                | 28,186                        | 79.7                          | 0.51                          | 0.93          | 28,186         |
| 14TCGA    | 13    | 68,580         | 90.7           | 62,196            | 48.5                | 26,335                        | 87.6                          | 0.49                          | 0.92          | 26,335         |
| 15TCGA    | 13    | 80,970         | 95.0           | 76,933            | 57.2                | 31,203                        | 92.8                          | 0.47                          | 0.97          | 31,203         |
| 16TCGA    | 13    | 83,601         | 93.1           | 77,874            | 56.2                | 32,908                        | 90.0                          | 0.46                          | 0.91          | 32,908         |
| 17TCGA    | 13    | 71,216         | 94.5           | 67,305            | 56.7                | 29,337                        | 91.9                          | 0.48                          | 0.95          | 29,337         |
| 18TCGA    | 13    | 71,290         | 93.9           | 66,930            | 53.0                | 27,688                        | 91.6                          | 0.47                          | 0.91          | 27,688         |
| 19TCGA    | 13    | 85,670         | 92.5           | 79,252            | 55.9                | 35,012                        | 90.2                          | 0.5                           | 0.98          | 35,012         |
| 20TCGA    | 13    | 85,860         | 94.7           | 81,314            | 55.8                | 35,404                        | 93.1                          | 0.48                          | 0.96          | 35,404         |
| 21TCGA    | 13    | 90,324         | 96.6           | 87,277            | 55.3                | 35,215                        | 95.1                          | 0.47                          | 0.98          | 35,215         |
| 22TCGA    | 13    | 83,748         | 97.5           | 81,684            | 56.0                | 34,594                        | 96.3                          | 0.46                          | 0.97          | 34,594         |
| 23TCGA    | 13    | 84,628         | 94.6           | 80,032            | 55.2                | 31,800                        | 92.1                          | 0.43                          | 0.94          | 31,800         |
| 24TCGA    | 13    | 75,699         | 95.7           | 72,416            | 55.4                | 28,790                        | 93.8                          | 0.46                          | 0.97          | 28,790         |

- Track level of phospho enrichment across IMAC experiments
- Separate FDR calculation for phosphopeptides and all peptides
# iTRAQ Labeling Efficiency of ECM Preps

- Measure fully (Nterm, Lys) and partially labeled (Lys only)

<table>
<thead>
<tr>
<th>File</th>
<th>Isobaric Labeled Spectra (%)</th>
<th>Isobaric Fully Labeled Spectra (%)</th>
<th>Isobaric No Label Spectra (%)</th>
<th>Isobaric Only Nterm Label Spectra (%)</th>
<th>Isobaric Only Lys Label Spectra (%)</th>
<th>Isobaric Labeled Precursor Intensity (%)</th>
<th>Isobaric Fully Labeled Precursor Intensity (%)</th>
<th>Isobaric No Label Precursor Intensity (%)</th>
<th>Isobaric Only Nterm Precursor Intensity (%)</th>
<th>Isobaric Only Lys Precursor Intensity (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>K130613_WT2_06wk_117_rafkt_ff12_cc_01</td>
<td>88.1</td>
<td>75.3</td>
<td>11.9</td>
<td>0.5</td>
<td>12.3</td>
<td>93.5</td>
<td>88.4</td>
<td>6.5</td>
<td>0.1</td>
<td>5.0</td>
</tr>
<tr>
<td>K130613_RIPTag2_06wk_116_rafkt_ff12_cc_01</td>
<td>86.4</td>
<td>69.2</td>
<td>13.6</td>
<td>0.6</td>
<td>16.7</td>
<td>92.4</td>
<td>87.6</td>
<td>7.6</td>
<td>0.2</td>
<td>4.6</td>
</tr>
<tr>
<td>K130613_RIPTag2_09wk_115_rafkt_ff12_cc_01</td>
<td>93.5</td>
<td>86.7</td>
<td>6.5</td>
<td>1.1</td>
<td>5.6</td>
<td>97.8</td>
<td>96.3</td>
<td>2.2</td>
<td>0.5</td>
<td>1.0</td>
</tr>
<tr>
<td>K130613_RIPTag2_12wk_114_rafkt_ff12_cc_01</td>
<td>95.1</td>
<td>89.4</td>
<td>4.9</td>
<td>0.6</td>
<td>5.1</td>
<td>97.8</td>
<td>96.1</td>
<td>2.2</td>
<td>0.1</td>
<td>1.5</td>
</tr>
<tr>
<td>K130614_WT2_6wk_117_2_rafkt_ff12_cc_01</td>
<td>94.2</td>
<td>90.1</td>
<td>5.8</td>
<td>0.7</td>
<td>3.4</td>
<td>98.6</td>
<td>98.2</td>
<td>1.4</td>
<td>0.0</td>
<td>0.4</td>
</tr>
<tr>
<td>K130614_RIPTag1_06wk_116_2_rafkt_ff12_cc_01</td>
<td>97.3</td>
<td>93.6</td>
<td>2.7</td>
<td>0.5</td>
<td>3.2</td>
<td>99.4</td>
<td>98.7</td>
<td>0.6</td>
<td>0.3</td>
<td>0.4</td>
</tr>
<tr>
<td>K130614_RIPTag1_09wk_115_2_rafkt_ff12_cc_01</td>
<td>96.6</td>
<td>92.6</td>
<td>3.6</td>
<td>1.2</td>
<td>2.6</td>
<td>99.1</td>
<td>98.7</td>
<td>0.9</td>
<td>0.2</td>
<td>0.2</td>
</tr>
<tr>
<td>K130614_RIPTag1_12wk_114_2_rafkt_ff12_cc_01</td>
<td>96.6</td>
<td>92.8</td>
<td>3.4</td>
<td>0.9</td>
<td>2.8</td>
<td>99.3</td>
<td>98.7</td>
<td>0.7</td>
<td>0.2</td>
<td>0.4</td>
</tr>
<tr>
<td>K130618_WT1_06wk_117_rafkt_ff12_cc_01</td>
<td>86.4</td>
<td>65.7</td>
<td>13.6</td>
<td>0.7</td>
<td>20.1</td>
<td>82.8</td>
<td>61.7</td>
<td>17.2</td>
<td>0.1</td>
<td>21.0</td>
</tr>
<tr>
<td>K130618_RIPTag1_06wk_116_1st_rafkt_ff12_cc_01</td>
<td>94.9</td>
<td>87.7</td>
<td>5.1</td>
<td>0.8</td>
<td>6.4</td>
<td>95.8</td>
<td>91.4</td>
<td>4.2</td>
<td>0.2</td>
<td>4.2</td>
</tr>
<tr>
<td>K130618_RIPTag1_09wk_115_1st_rafkt_ff12_cc_01</td>
<td>93.0</td>
<td>81.7</td>
<td>7.0</td>
<td>1.0</td>
<td>10.3</td>
<td>94.6</td>
<td>85.1</td>
<td>5.4</td>
<td>1.0</td>
<td>8.5</td>
</tr>
<tr>
<td>K130618_RIPTag1_12wk_114_1st_rafkt_ff12_cc_01</td>
<td>93.0</td>
<td>82.3</td>
<td>7.0</td>
<td>0.7</td>
<td>10.0</td>
<td>94.2</td>
<td>84.8</td>
<td>5.8</td>
<td>0.5</td>
<td>8.9</td>
</tr>
<tr>
<td>K130619_WT1_06wk_117_rafkt_ff12_cc_01</td>
<td>95.3</td>
<td>88.2</td>
<td>4.7</td>
<td>0.7</td>
<td>6.4</td>
<td>96.3</td>
<td>91.7</td>
<td>3.7</td>
<td>0.2</td>
<td>4.4</td>
</tr>
<tr>
<td>K130619_RIPTag1_06wk_116_2nd_rafkt_ff12_cc_01</td>
<td>97.9</td>
<td>94.9</td>
<td>2.1</td>
<td>0.5</td>
<td>2.5</td>
<td>99.3</td>
<td>97.9</td>
<td>0.7</td>
<td>0.1</td>
<td>1.3</td>
</tr>
<tr>
<td>K130619_RIPTag1_09wk_115_2nd_rafkt_ff12_cc_01</td>
<td>98.5</td>
<td>95.6</td>
<td>1.5</td>
<td>0.8</td>
<td>2.0</td>
<td>99.1</td>
<td>97.2</td>
<td>0.9</td>
<td>0.8</td>
<td>1.0</td>
</tr>
<tr>
<td>K130619_RIPTag1_12wk_114_2nd_rafkt_ff12_cc_01</td>
<td>97.5</td>
<td>93.0</td>
<td>2.5</td>
<td>0.7</td>
<td>3.7</td>
<td>98.6</td>
<td>95.7</td>
<td>1.4</td>
<td>0.5</td>
<td>2.4</td>
</tr>
<tr>
<td>K130827_WT3_06wk_117_rafkt_ff12_cc_01</td>
<td>98.4</td>
<td>95.8</td>
<td>1.6</td>
<td>1.1</td>
<td>1.6</td>
<td>99.3</td>
<td>98.9</td>
<td>0.7</td>
<td>0.2</td>
<td>0.2</td>
</tr>
<tr>
<td>K130827_RIPTag3_06wk_116_rafkt_ff12_cc_01</td>
<td>97.6</td>
<td>94.1</td>
<td>2.4</td>
<td>1.1</td>
<td>2.4</td>
<td>99.0</td>
<td>98.4</td>
<td>1.0</td>
<td>0.2</td>
<td>0.4</td>
</tr>
<tr>
<td>K130827_RIPTag3_09wk_115_rafkt_ff12_cc_01</td>
<td>97.4</td>
<td>93.9</td>
<td>2.6</td>
<td>2.0</td>
<td>1.6</td>
<td>99.2</td>
<td>98.8</td>
<td>0.8</td>
<td>0.3</td>
<td>0.2</td>
</tr>
<tr>
<td>K130827_RIPTag3_12wk_114_rafkt_ff12_cc_01</td>
<td>96.6</td>
<td>90.6</td>
<td>4.0</td>
<td>2.3</td>
<td>3.2</td>
<td>98.9</td>
<td>98.2</td>
<td>1.1</td>
<td>0.3</td>
<td>0.5</td>
</tr>
</tbody>
</table>

- % of spectra
- % of precursor intensity
## Reduction/Alkylation Problem

<table>
<thead>
<tr>
<th>PSMs Containing no Var Mods (#)</th>
<th>Containing m (#)</th>
<th>Containing c (#)</th>
<th>Containing p (#)</th>
<th>Containing C (#)</th>
<th>PSM's Cys/Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>5101</td>
<td>285</td>
<td>133</td>
<td>918</td>
<td>1527</td>
<td>0.213</td>
</tr>
<tr>
<td>3857</td>
<td>219</td>
<td>61</td>
<td>1051</td>
<td>778</td>
<td>0.140</td>
</tr>
<tr>
<td>4147</td>
<td>225</td>
<td>77</td>
<td>1107</td>
<td>866</td>
<td>0.145</td>
</tr>
<tr>
<td>5852</td>
<td>309</td>
<td>129</td>
<td>1178</td>
<td>1580</td>
<td>0.193</td>
</tr>
<tr>
<td>5204</td>
<td>522</td>
<td>223</td>
<td>1247</td>
<td>1855</td>
<td>0.228</td>
</tr>
<tr>
<td>4785</td>
<td>470</td>
<td>185</td>
<td>1072</td>
<td>1475</td>
<td>0.204</td>
</tr>
<tr>
<td>4161</td>
<td>287</td>
<td>157</td>
<td>1212</td>
<td>1170</td>
<td>0.184</td>
</tr>
<tr>
<td>6395</td>
<td>424</td>
<td>26</td>
<td>1371</td>
<td>533</td>
<td>0.058</td>
</tr>
<tr>
<td>7617</td>
<td>1199</td>
<td>303</td>
<td>1752</td>
<td>2431</td>
<td>0.206</td>
</tr>
<tr>
<td>7408</td>
<td>1078</td>
<td>259</td>
<td>1469</td>
<td>2084</td>
<td>0.188</td>
</tr>
<tr>
<td>6685</td>
<td>764</td>
<td>212</td>
<td>1524</td>
<td>1613</td>
<td>0.162</td>
</tr>
<tr>
<td>14035</td>
<td>946</td>
<td>113</td>
<td>1770</td>
<td>2770</td>
<td>0.152</td>
</tr>
<tr>
<td>1184</td>
<td>200</td>
<td>2</td>
<td>1144</td>
<td>77</td>
<td>0.029</td>
</tr>
<tr>
<td>1928</td>
<td>258</td>
<td>3</td>
<td>1004</td>
<td>91</td>
<td>0.027</td>
</tr>
<tr>
<td>4483</td>
<td>580</td>
<td>117</td>
<td>1381</td>
<td>1495</td>
<td>0.205</td>
</tr>
<tr>
<td>3684</td>
<td>447</td>
<td>5</td>
<td>1323</td>
<td>177</td>
<td>0.030</td>
</tr>
<tr>
<td>4017</td>
<td>434</td>
<td>22</td>
<td>1571</td>
<td>1061</td>
<td>0.166</td>
</tr>
<tr>
<td>5540</td>
<td>578</td>
<td>28</td>
<td>1434</td>
<td>1273</td>
<td>0.157</td>
</tr>
<tr>
<td>12714</td>
<td>1058</td>
<td>234</td>
<td>2129</td>
<td>3127</td>
<td>0.180</td>
</tr>
<tr>
<td>11774</td>
<td>936</td>
<td>70</td>
<td>1934</td>
<td>2376</td>
<td>0.151</td>
</tr>
</tbody>
</table>

### Reduction/alkylation repeated due to low #’s of PSM’s Containing Cys

**PSM’s Cys/Total**

<table>
<thead>
<tr>
<th>PSM</th>
<th>Cys/Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>5_Ctrl_10444</td>
<td>0.213</td>
</tr>
<tr>
<td>early_12256</td>
<td>0.228</td>
</tr>
<tr>
<td>i_late_97500</td>
<td>0.204</td>
</tr>
<tr>
<td>PDAC_10832</td>
<td>0.058</td>
</tr>
<tr>
<td>7_Wt_10856</td>
<td>0.206</td>
</tr>
<tr>
<td>Early_13301</td>
<td>0.188</td>
</tr>
<tr>
<td>i_late_10269</td>
<td>0.162</td>
</tr>
<tr>
<td>PDAC_10864</td>
<td>0.152</td>
</tr>
<tr>
<td>120_WEIGHTS</td>
<td>0.029</td>
</tr>
<tr>
<td>0_PDAC10864</td>
<td>0.027</td>
</tr>
<tr>
<td>20_Late_10044</td>
<td>0.205</td>
</tr>
<tr>
<td>Early_17536</td>
<td>0.030</td>
</tr>
<tr>
<td>0_PDAC10328</td>
<td>0.166</td>
</tr>
</tbody>
</table>

**Reduction/Alkylation repeated due to low #’s of PSM’s Containing Cys**

- **1 Fine**
- **2 Before**
- **2 After**
- **3 Before**
- **3 After**
# Peptide String Match

An aid for configuring targeted MS/MS assays

## Peptide String Match

<table>
<thead>
<tr>
<th>Spectrum Mill</th>
<th>MS Edman</th>
<th>Peptide Selector</th>
<th>Peptide List to Masses</th>
<th>MS Digest</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Find Peptide**

**Database:** Uniprot.human.mtou.20141017.RNFSnr.150contains

**Subset of Accession #s to search:**
- [ ] Require Preceding Tryptic Site
- [ ] Show Species
- [ ] Show Link to ClustalW Multi-Aligner
- [ ] Show Link to Database Website

**Enter peptide sequences:** (regular expressions allowed)

- VTVGDFGEQGCSGR
- NIVNALSNCCK
- LIEEDNETAR
- VSYIDFFTAGEQYPQOP
- EMQPTFPDR
- EDLYVYQAK

---

<table>
<thead>
<tr>
<th>#</th>
<th>Peptide</th>
<th># Missed Cleavages (Trypsin)</th>
<th># Matches</th>
<th># Different Gene Symbols</th>
<th>Gene Symbols</th>
<th>Link to ClustalW Multi-Aligner</th>
<th>Link to UniProt</th>
<th># Different Species</th>
<th>Species</th>
<th>Start AA</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>VTVGDFGEQGCSGR</td>
<td>0</td>
<td>5</td>
<td>1</td>
<td>VTVGDFGEQGCSGR</td>
<td>P54764</td>
<td>1</td>
<td>HUMAN</td>
<td>MOUSE</td>
<td>576</td>
</tr>
<tr>
<td>2</td>
<td>NIVNALSNCCK</td>
<td>0</td>
<td>41</td>
<td>8</td>
<td>NIVNALSNCCK</td>
<td>P65733</td>
<td>1</td>
<td>HUMAN</td>
<td>MOUSE</td>
<td>434</td>
</tr>
<tr>
<td>3</td>
<td>LIEEDNETAR</td>
<td>0</td>
<td>18</td>
<td>4</td>
<td>LIEEDNETAR</td>
<td>P65733</td>
<td>1</td>
<td>HUMAN</td>
<td>MOUSE</td>
<td>434</td>
</tr>
<tr>
<td>4</td>
<td>VSYIDFFTAGEQYPQOP</td>
<td>0</td>
<td>3</td>
<td>1</td>
<td>VSYIDFFTAGEQYPQOP</td>
<td>P65733</td>
<td>1</td>
<td>HUMAN</td>
<td>MOUSE</td>
<td>434</td>
</tr>
<tr>
<td>5</td>
<td>EMQPTFPDR</td>
<td>0</td>
<td>12</td>
<td>3</td>
<td>EMQPTFPDR</td>
<td>P65733</td>
<td>1</td>
<td>HUMAN</td>
<td>MOUSE</td>
<td>434</td>
</tr>
<tr>
<td>6</td>
<td>EDLYVYQAK</td>
<td>0</td>
<td>6</td>
<td>1</td>
<td>EDLYVYQAK</td>
<td>P65733</td>
<td>1</td>
<td>HUMAN</td>
<td>MOUSE</td>
<td>434</td>
</tr>
</tbody>
</table>

---

**ClustalW Multi-Aligner**
Future Directions

• Personalized sequence databases
• 2-3x faster searches, better handle peptide redundancy in databases from isoforms
• Store protein grouping for faster report generation
• VM site level polishing autovalidation
• VM site level FDR, FLR
• Better integration with post-SM statistical manipulation of quantitation
  • Normalization
  • Significance testing
  • Clustering
• Spectral library searching
• Identify multiple peptides per MS/MS spectrum
Dataflow for Proteomics with Personalized Sequence Databases

**Global Proteomics**

**Phospho Proteomics**

**CPTAC**

**105**

**Tumors**

**TCGA**

**germline**

**Whole Exome sequencing**

**RNA-Seq**

Copy number arrays (Affy)

DNA methylation

Whole genome sequencing

mRNA arrays (Affy)

miRNA sequencing

RPPA

**MS/MS**

**RefSeq Reference database**

**.Raw**

**.fasta**

**QUILTS**

quilts.fenyolab.org

**Sequence Analysis**

(TopHat, BowTie, GATK)

**Variants**, **Alternative Splicing**, **Frameshifts**, **Novel Expression**, **Fusion Genes**, **Indels**

**Identified/Quantitated Peptides and Proteins**

+ tumor specific peptides

+ patient specific peptides

**Proteogenomic mapping and integration**

**Spectrum Mill**

**.fasta**

(4x105)
Serial Search Strategy with Personalized Databases

25,776,160 Spectra
(105 patients)
(36 iTRAQ experiments)
(25 LC-MS/MS runs / experiment)

• Concatenated FASTA files, 105 patients
• Removed redundant entries

RefSeq-Hs-7/2013: 31,852

> Canonical Protein
SIGNALLINGPATHWAYREGULATOR

11,113,249 Spectra
355,654 Peptides
15,402 Proteins
(43% of total)
(0.43% FDR)

Variants: 132,769

• Canonical – Variant Patient 1
SIGNALLINGPATHWAYHREGULATOR
• Canonical Protein – Variant Patient 2
SIGNALLINGPATHWAYREGULATOR

Alternative Splices: 68,434

• Canonical – Alternate splice Patient 1
SIGNALLINGREGULATOR
• Canonical – Alternate splice Patient 2
SIGNALLINGPATHWAYREGULATOR

Frameshifts: 187,059

• Canonical – Truncation Patient 1
SIGNALLINGPAT FRAMESHIFT
• Canonical – Novel Exon Insert Patient 2
SIGNALLINGPATHWAY INSERTREGULATOR
• Canonical – Partial Exon Deletion Patient 3
SIGNALLINGPATHWAYULATOR

3,330 Variants Matched
196 Splice Junctions Matched
476 Truncations Matched
12 Insertions Matched
68 Deletions Matched

Low confidence thresholds for Genome Transcriptome calls
• Variants: >2 QUAL score (phred-scaled)
• Alternative splices, frameshifts: >1 read

High confidence for Proteome IDs
• <0.1% FDR peptide spectrum match
Acknowledgements

Broad Institute Proteomics Platform Group – Steve Carr

Agilent Technologies Inc.
Joe Roark
Chris Miller