



The ProteoWizard Project: Origins and Tools Overview

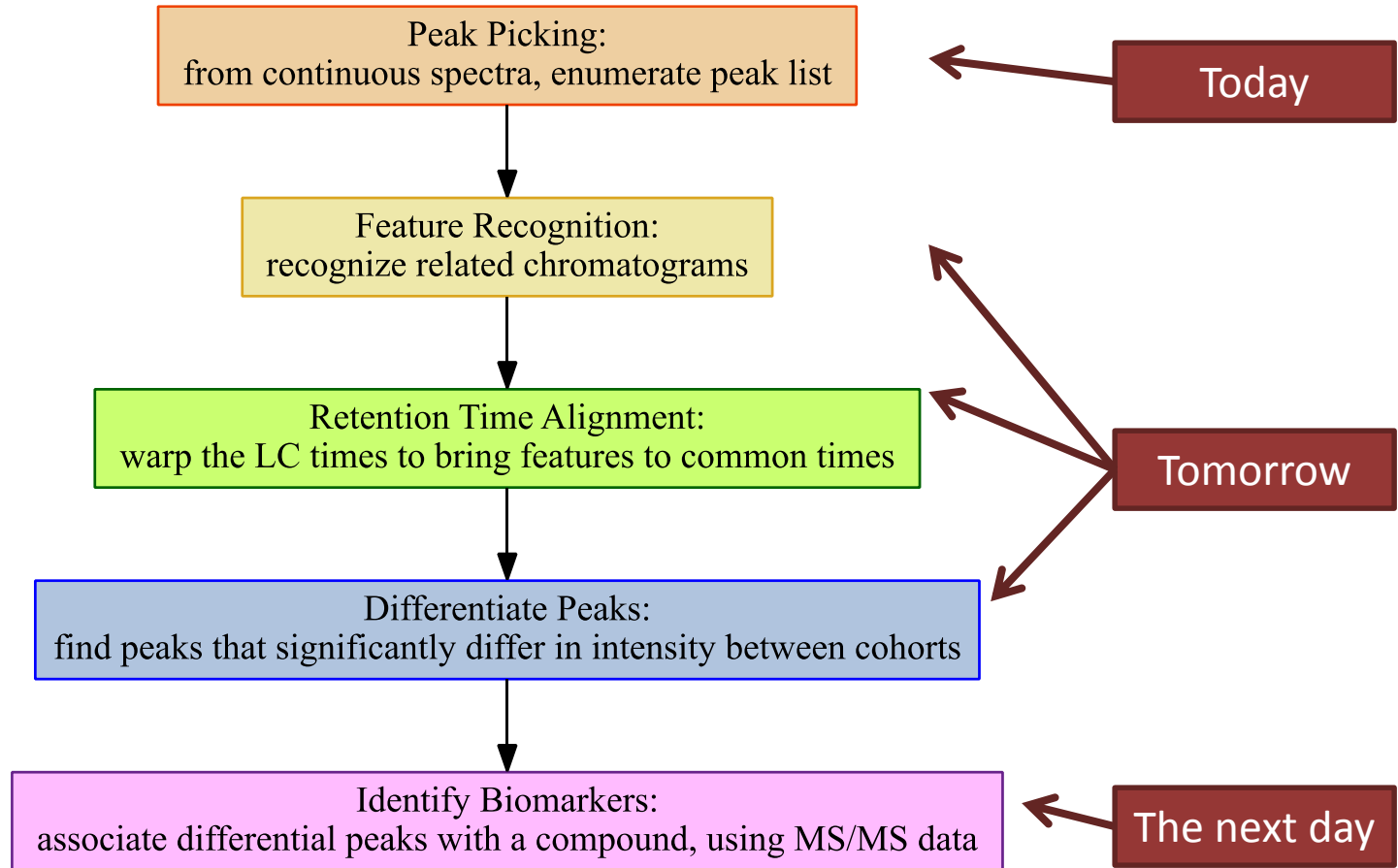
David L. Tabb, Ph.D.

dtabb1973@gmail.com

Overview

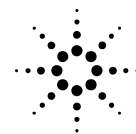
- Origins of ProteoWizard
- Structure of the library
- Who's who in ProteoWizard
- Tools distributed in the library

Metabolome informatics



A chaos of LC-MS/MS formats

- Vendor formats
- Open standard formats
 - mzXML
 - mzML
- Identification-only formats
 - MGF
 - DTA
 - PKL
 - MS2



Agilent Technologies



ThermoFisher
S C I E N T I F I C



Waters
THE SCIENCE OF
WHAT'S POSSIBLE.™

mzML metadata

```
<spectrum index="194" id="controllerType=0 controllerNumber=1 scan=195" defaultArrayLength="17">
  <cvParam cvRef="MS" accession="MS:1000580" name="MSn spectrum" value=""/>
  <cvParam cvRef="MS" accession="MS:1000511" name="ms level" value="2"/>
  <cvParam cvRef="MS" accession="MS:1000130" name="positive scan" value=""/>
  <cvParam cvRef="MS" accession="MS:1000127" name="centroid spectrum" value=""/>
  <cvParam cvRef="MS" accession="MS:1000504" name="base peak m/z" value="403.931457519531" unitCvRef="MS" unitAccession="MS:1000040" unitName="m/z"/>
  <cvParam cvRef="MS" accession="MS:1000505" name="base peak intensity" value="9.118535041809" unitCvRef="MS" unitAccession="MS:1000131" unitName="number of detector counts"/>
  <cvParam cvRef="MS" accession="MS:1000285" name="total ion current" value="72.539352416992"/>
  <cvParam cvRef="MS" accession="MS:1000528" name="lowest observed m/z" value="147.067260742188" unitCvRef="MS" unitAccession="MS:1000040" unitName="m/z"/>
  <cvParam cvRef="MS" accession="MS:1000527" name="highest observed m/z" value="591.20751953125" unitCvRef="MS" unitAccession="MS:1000040" unitName="m/z"/>
  <scanList count="1">
    <cvParam cvRef="MS" accession="MS:1000795" name="no combination" value=""/>
    <scan instrumentConfigurationRef="IC2">
      <cvParam cvRef="MS" accession="MS:1000016" name="scan start time" value="20.20338333333" unitCvRef="UO" unitAccession="UO:0000031" unitName="minute"/>
      <cvParam cvRef="MS" accession="MS:1000512" name="filter string" value="ITMS + c NSI d Full ms2 421.83@cid35.00 [105.00-855.00]"/>
      <cvParam cvRef="MS" accession="MS:1000616" name="preset scan configuration" value="9"/>
      <cvParam cvRef="MS" accession="MS:1000927" name="ion injection time" value="100.0" unitCvRef="UO" unitAccession="UO:0000028" unitName="millisecond"/>
      <userParam name="[Thermo Trailer Extra]Monoisotopic M/Z:" value="421.8348388671875" type="xsd:float"/>
      <scanWindowList count="1">
        <scanWindow>
          <cvParam cvRef="MS" accession="MS:1000501" name="scan window lower limit" value="105.0" unitCvRef="MS" unitAccession="MS:1000040" unitName="m/z"/>
          <cvParam cvRef="MS" accession="MS:1000500" name="scan window upper limit" value="855.0" unitCvRef="MS" unitAccession="MS:1000040" unitName="m/z"/>
        </scanWindow>
      </scanWindowList>
    </scan>
  </scanList>
  <precursorList count="1">
    <precursor spectrumRef="controllerType=0 controllerNumber=1 scan=194">
      <isolationWindow>
        <cvParam cvRef="MS" accession="MS:1000827" name="isolation window target m/z" value="421.83" unitCvRef="MS" unitAccession="MS:1000040" unitName="m/z"/>
        <cvParam cvRef="MS" accession="MS:1000828" name="isolation window lower offset" value="1.0" unitCvRef="MS" unitAccession="MS:1000040" unitName="m/z"/>
        <cvParam cvRef="MS" accession="MS:1000829" name="isolation window upper offset" value="1.0" unitCvRef="MS" unitAccession="MS:1000040" unitName="m/z"/>
      </isolationWindow>
      <selectedIonList count="1">
        <selectedIon>
          <cvParam cvRef="MS" accession="MS:1000744" name="selected ion m/z" value="421.834838867188" unitCvRef="MS" unitAccession="MS:1000040" unitName="m/z"/>
          <cvParam cvRef="MS" accession="MS:1000041" name="charge state" value="2"/>
          <cvParam cvRef="MS" accession="MS:1000042" name="peak intensity" value="240.620056152344" unitCvRef="MS" unitAccession="MS:1000131" unitName="number of detector counts"/>
        </selectedIon>
      </selectedIonList>
      <activation>
        <cvParam cvRef="MS" accession="MS:1000133" name="collision-induced dissociation" value=""/>
        <cvParam cvRef="MS" accession="MS:1000045" name="collision energy" value="35.0" unitCvRef="UO" unitAccession="UO:0000266" unitName="electronvolt"/>
      </activation>
    </precursor>
  </precursorList>
```

Versus MGF

BEGIN IONS

PEPMASS=421.834839 240.6201

CHARGE=2+

TITLE=A0218_1B-WHIM2-P5-11_20120725_r_klc_X_A1_t1_fr05_120814103313.00195.00195.2

147.067261 8.778127

149.202332 2.320779

195.249405 5.232951

197.068146 3.649321

267.800507 2.094611

295.291138 3.251245

321.154388 6.537436

325.938507 2.862236

330.879028 5.401159

332.079712 2.962690

334.918304 5.255876

352.198822 1.756394

380.961700 1.586437

395.014496 6.547868

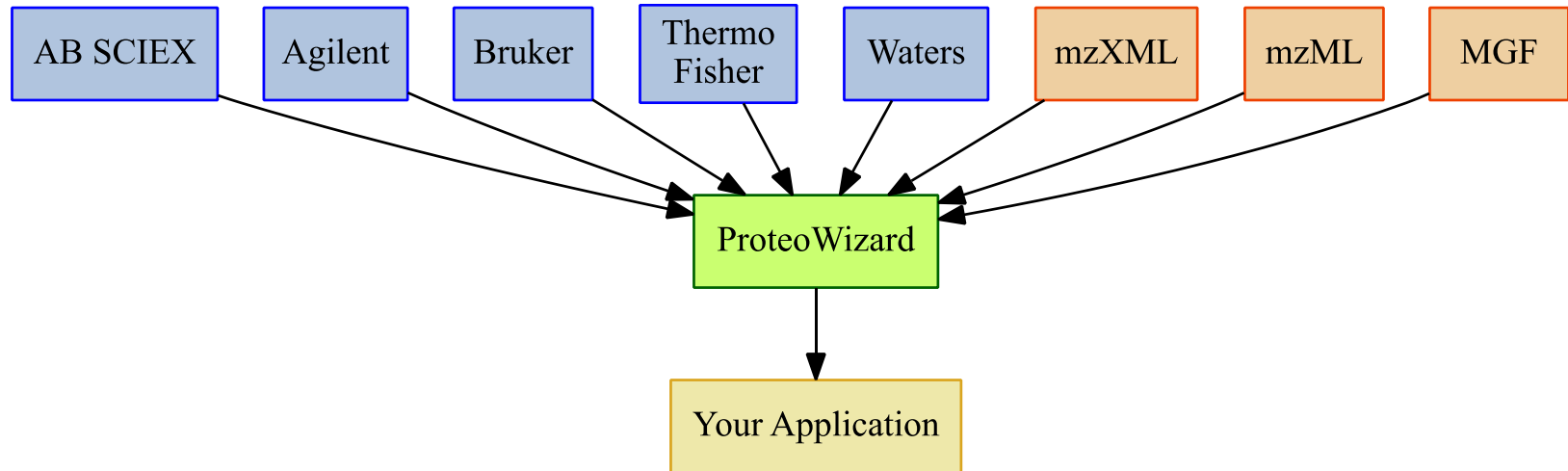
403.067566 3.779420

403.931458 9.118535

591.207520 1.404271

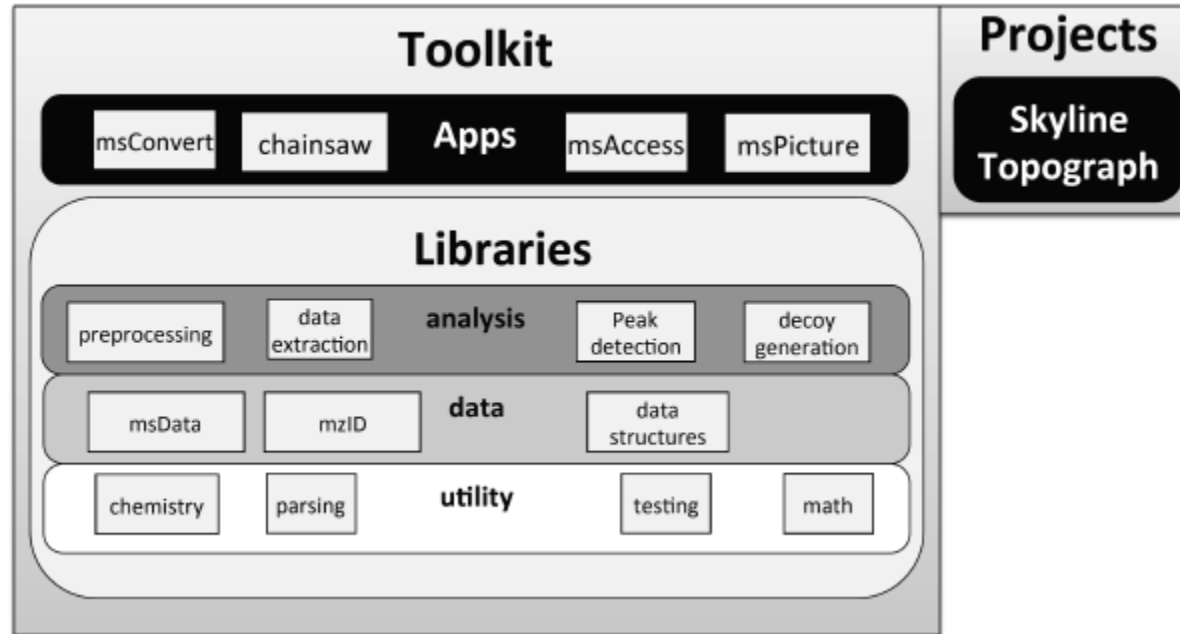
END IONS

PWiz: a Rosetta Stone



- The PWiz msData layer connects to mass spectrometry vendor libraries.
- Once the data are seen by msData, the format from which they came does not matter.

PWiz Structure



- Libraries embed data models for MS data and the identifications derived from them.
- Utility functions handle common activities such as protein digestion, peak integration, etc.

The team

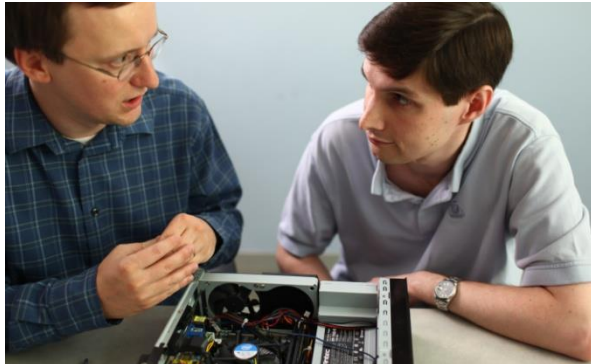
- Parag Mallick
- Darren Kessner



- Michael J. MacCoss
- Brendan Maclean

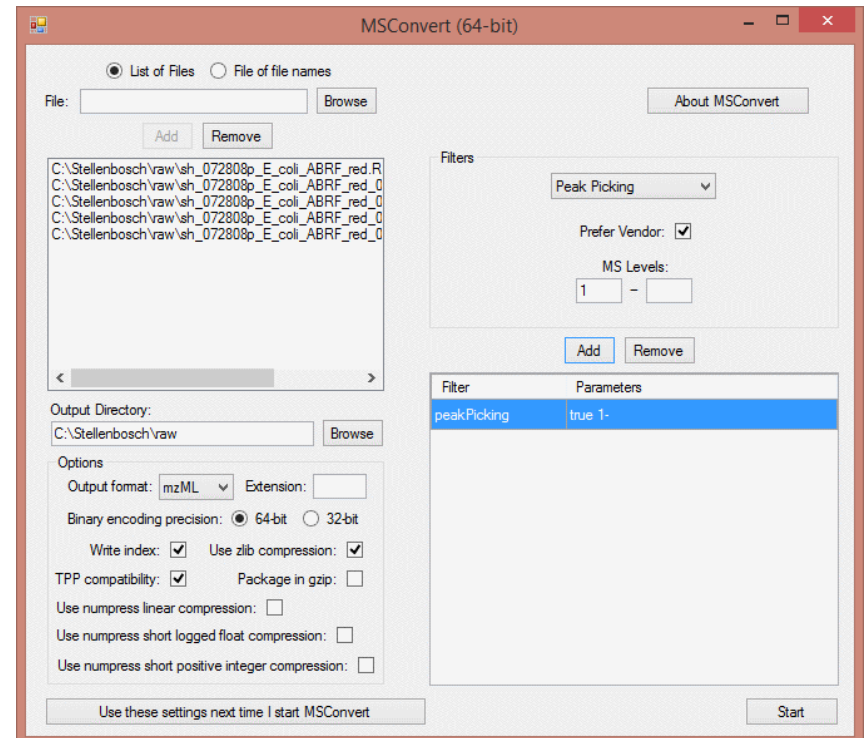


- David L. Tabb
- Matthew C. Chambers

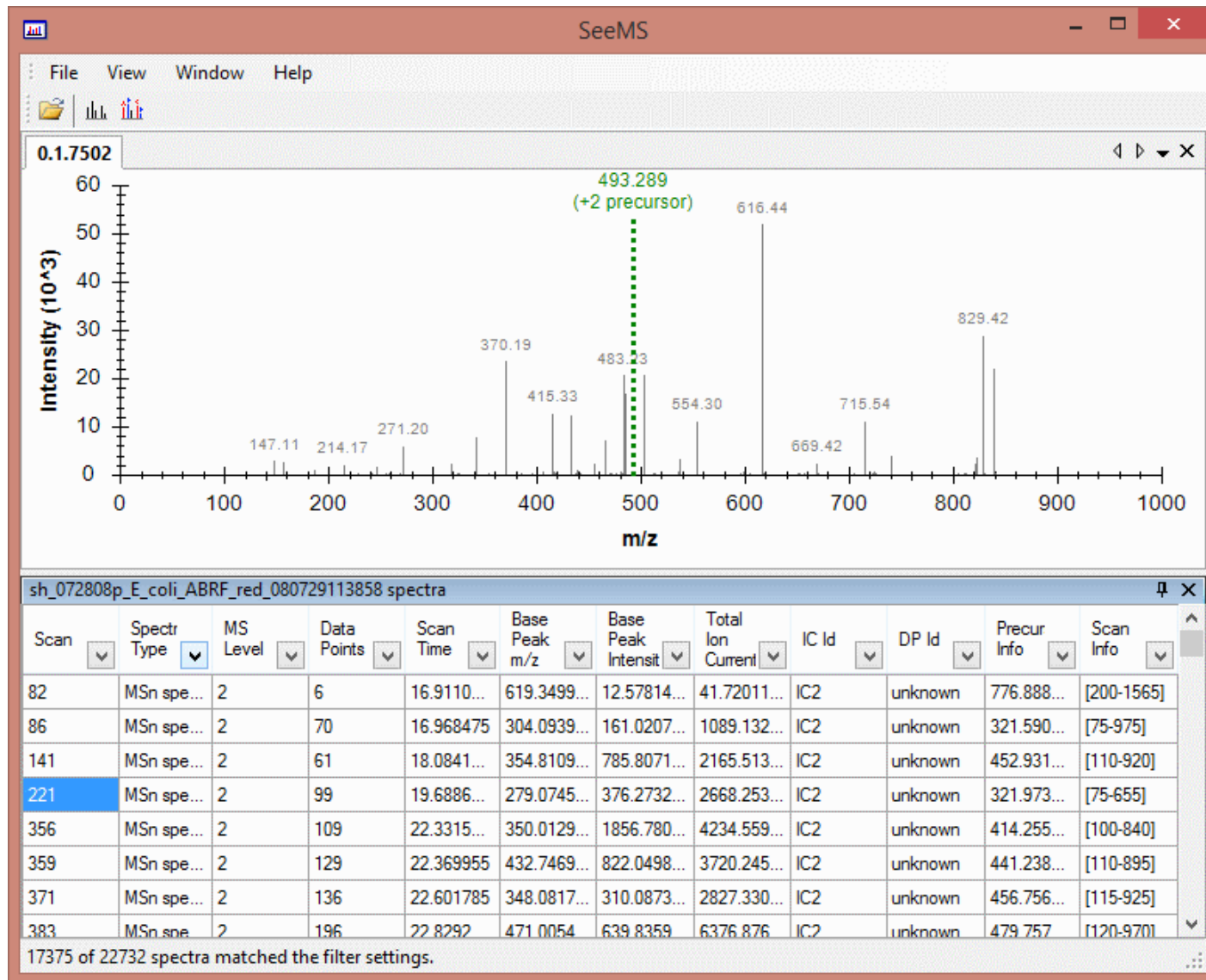


msConvert

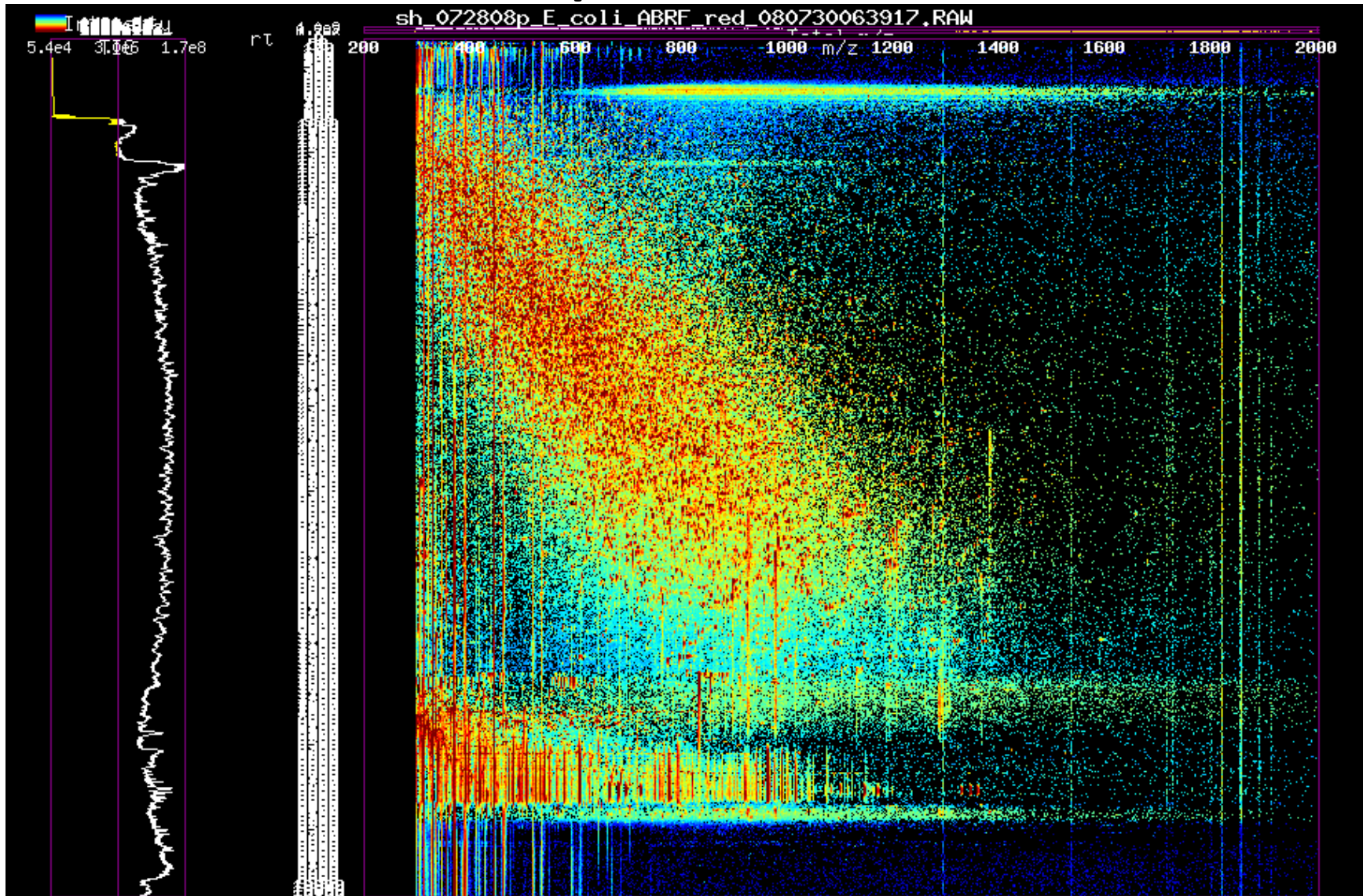
- The tool reads from one MS data format and writes to another.
- Both GUI and command-line interfaces are available.
- We will discuss it in detail later.



SeeMS spectrum and chromatogram viewer



msPicture experiment visualizer



QuaMeter “IDFree” metrics

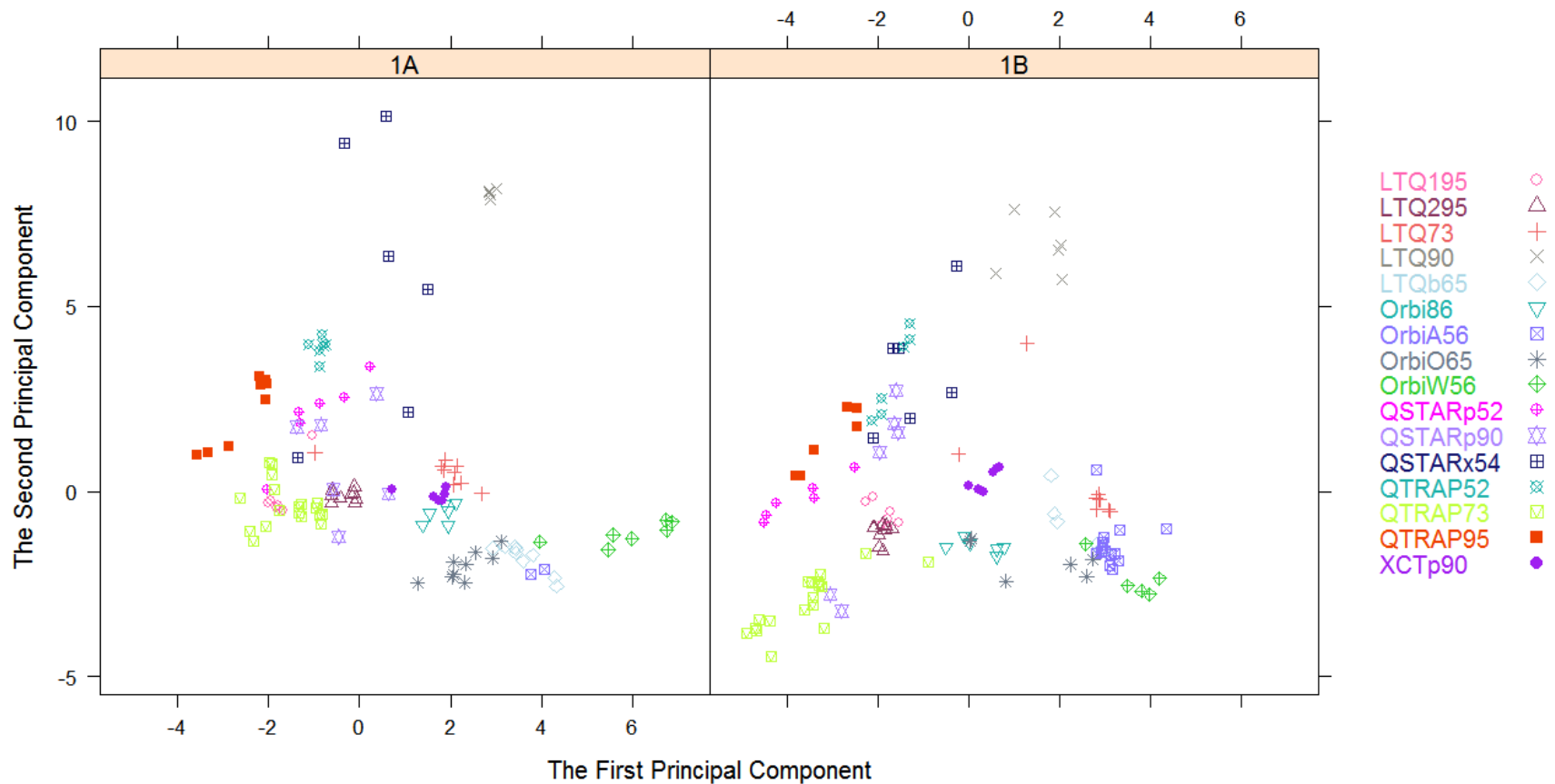
- 40 metrics categorized into extracted ion chromatograms, retention times, mass spectrometry, and tandem mass spectrometry
- Metrics emphasize quartiles and log ratios to give robust behavior in extreme variation
- Example fields for rapid characterization:
 - XIC-FWHM-Q2: median of peak widths in time
 - RT-TIC-Qx: relative duration for TIC integration
 - MS1-Density-Q2: median of peak counts for MS
 - MS2-Freq-Max: Fastest rate of MS/MS acquisition

```
quameter.exe *.raw -MetricsType idfree -OutputFilepath metrics.tsv  
ChromatogramMzLowerOffset = 10ppm  
ChromatogramMzUpperOffset = 10ppm
```

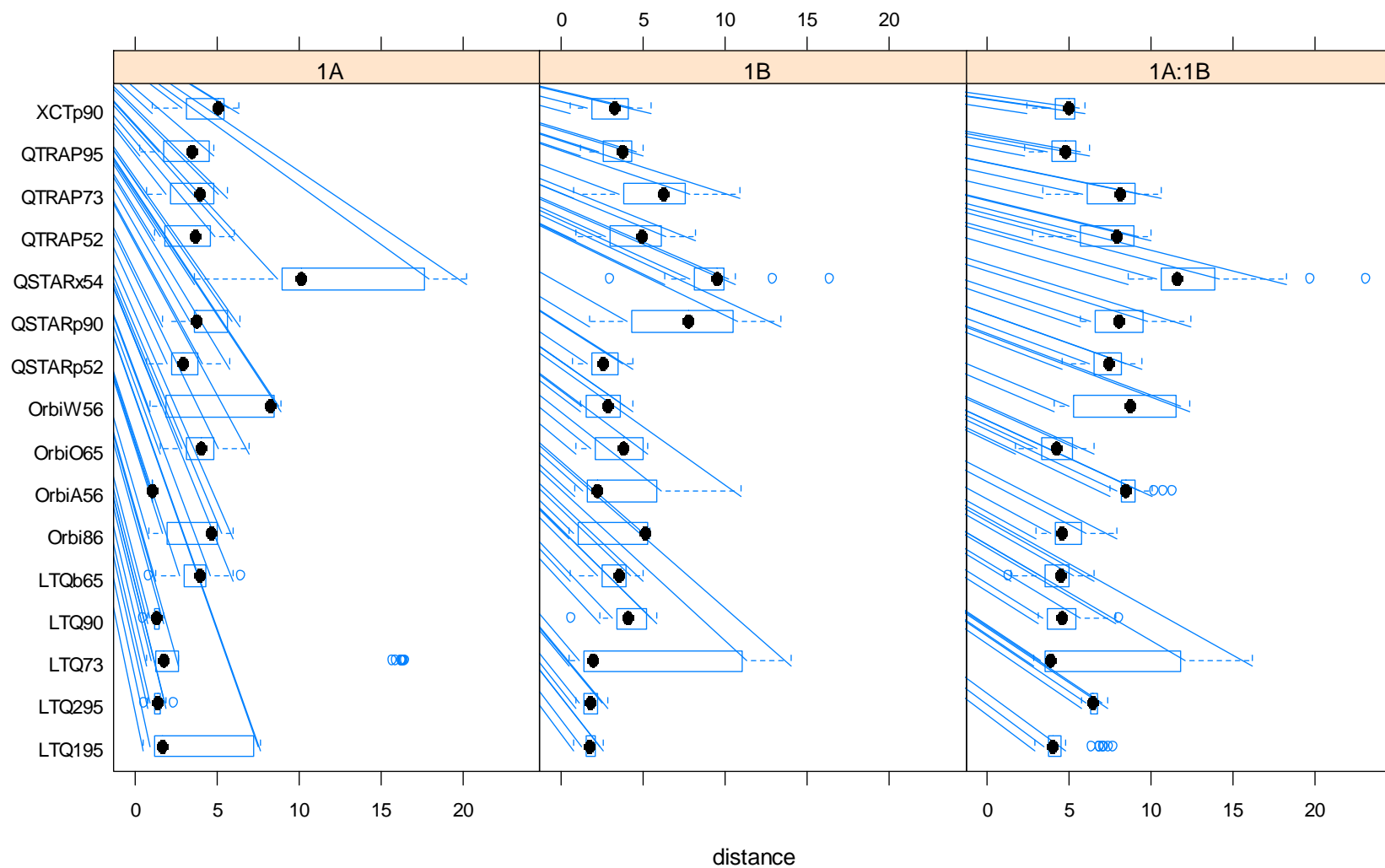
Ma et al. *Anal Chem* (2012) 84: 5845

Wang et al. *Anal Chem* (2014) 86: 2497

Robust PCA de-correlates metrics



PC distances compare files



Takeaway messages

- ProteoWizard reads lots of formats so your software doesn't have to do the work.
- ProteoWizard can be of service for metabolomics researchers, as well.
- Basic tools in ProteoWizard or built using ProteoWizard have advanced applications.