# Label Free Quantitative Proteomics

Mi-Youn Brusniak, Ph.D.

NHLBI PROTEOMICS

INSTITUTE FOR Systems Biology

# Outline

- Principles of quantitative proteomics
- Labeling vs. non-labeling approaches for quantitative proteomics
- SpecArray: software tool for quantitative proteomics without isotopic labeling
- Corra: Framework to generate candidate biomarkers using non labeling methods

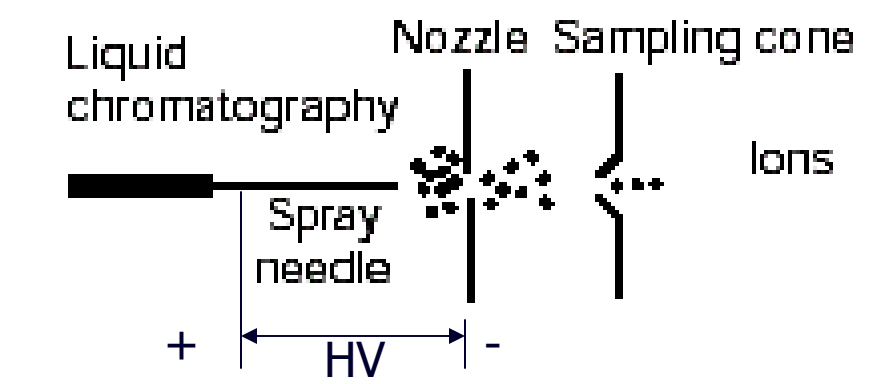
# Summary of LC-ESI-MS/MS

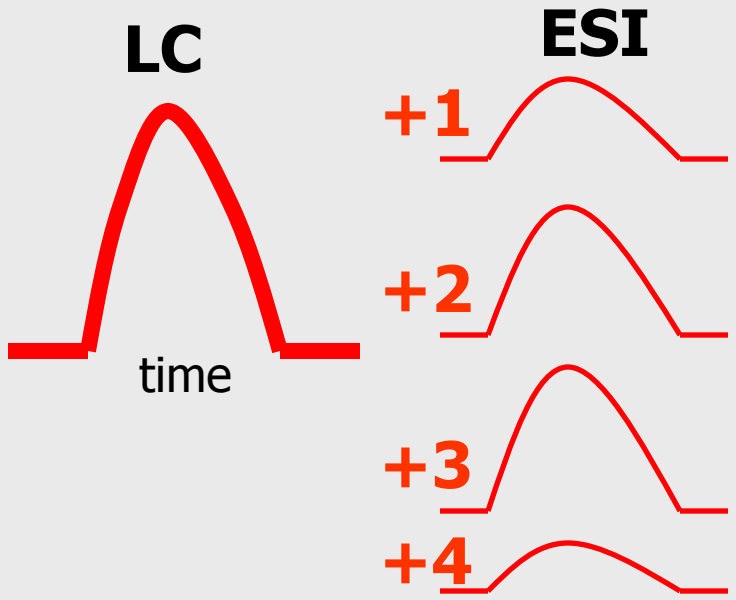
- Protein mixtures are digested into peptides
- Peptides are concentrated and fractionated by separation technologies such as SCX, IEF, RP, etc.
- While eluting from RP column, peptides are ionized by ESI and analyzed by MS/MS
- Peptides are identified from CID spectra
- Peptides are mostly quantified from MS signatures except in the case of iTRAQ

# Electrospray Ionization

- Multiple charge states: from +1 to +4

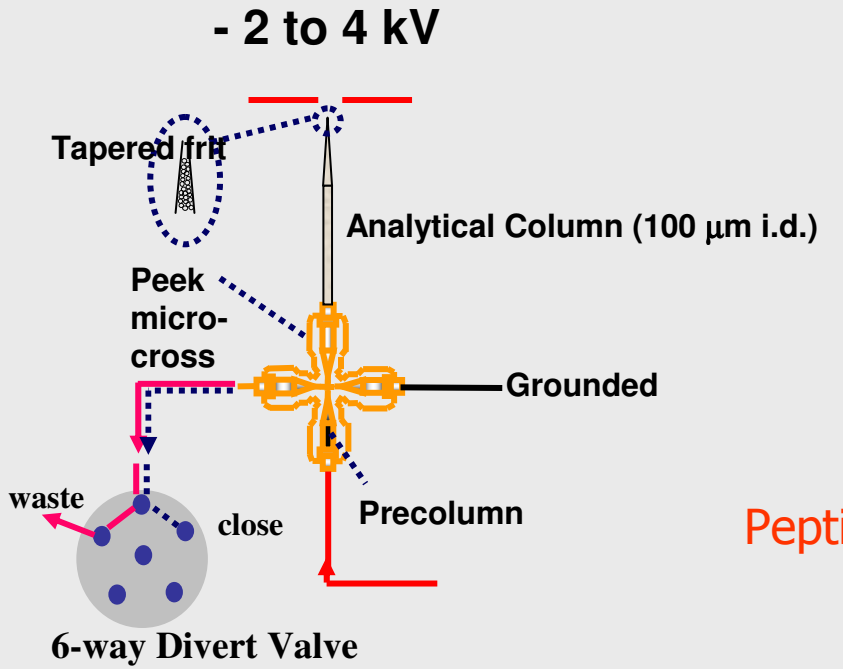
$M + z\ H^+ = M(H^+)_z$ $\qquad$ $m/z = (M+z^*H)/z$

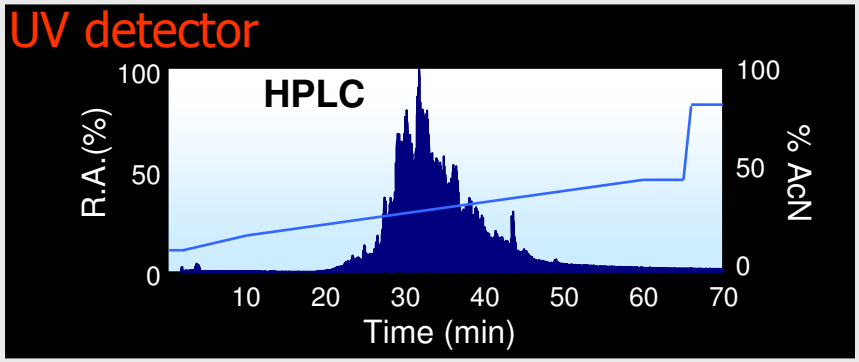


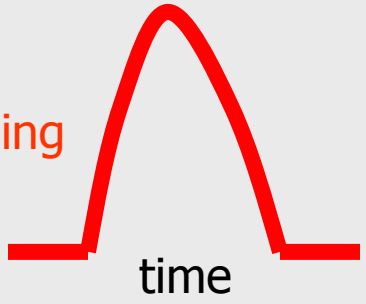
# Reversed-Phase Chromatography

- Separate peptides by hydrophobicity
- Reproducible
- Automated, coupled online with MS



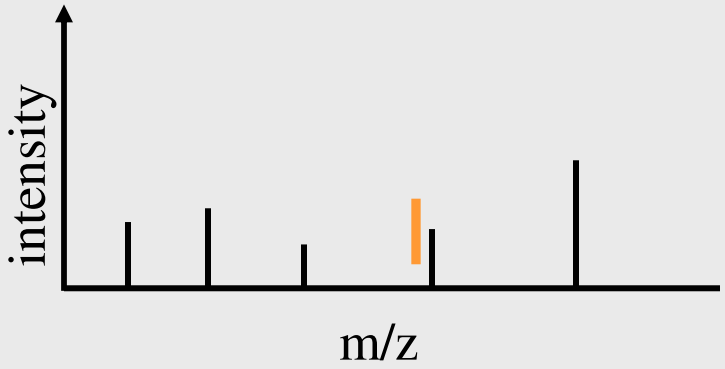




Peptide eluting profiling

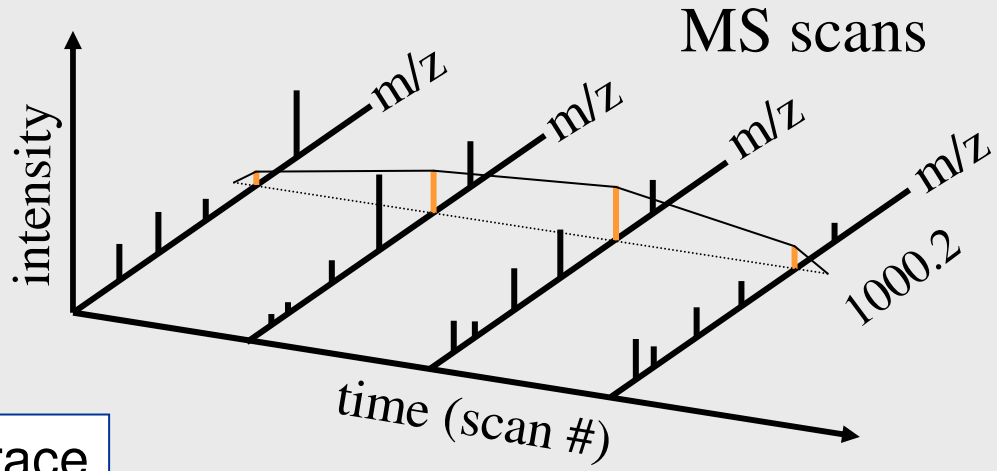


# Single Ion Chromatogram

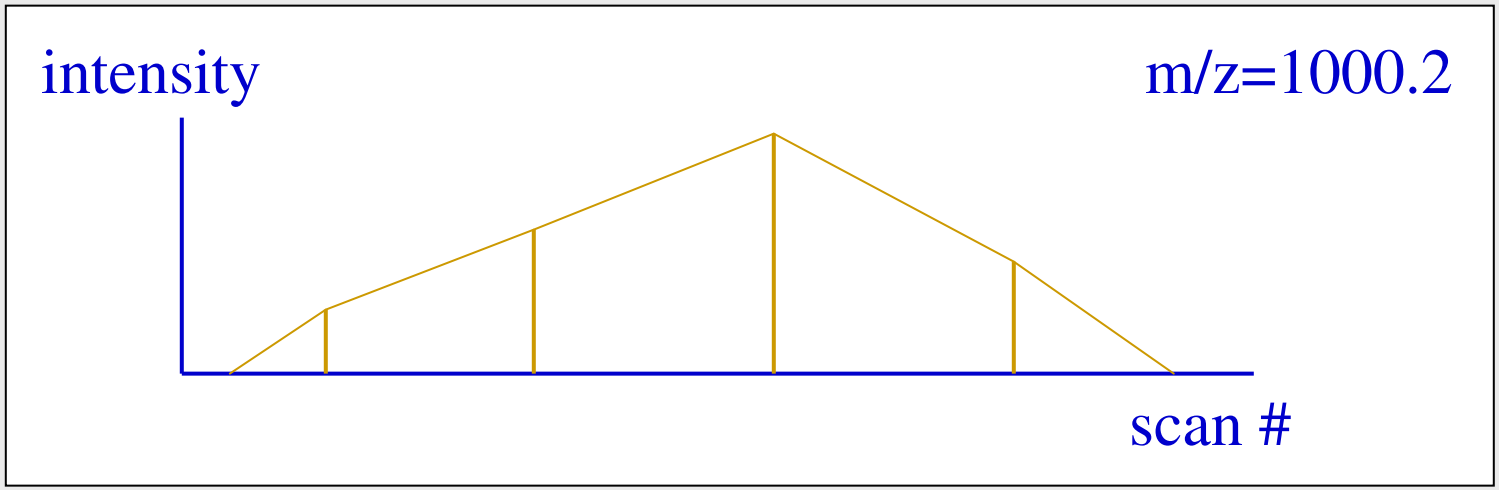
2D view: m/z, intensity




3D view: m/z, intensity, time




Single Ion Current (SIC) Trace



# Peptide Quantification

- Area of SIC is proportional to peptide abundance
- Ionization efficiency of each peptide is different
  - Depends on the peptide molecular properties (e.g. number of basic residues)
- MS Technology is NOT quantitative

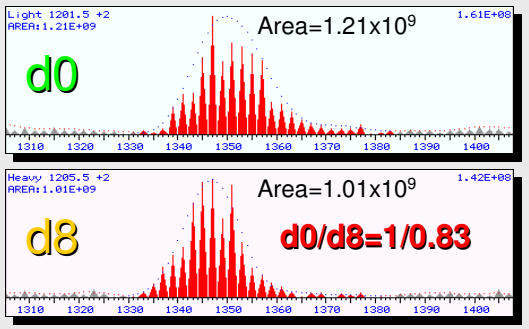
# Outline

- Principles of quantitative proteomics
- Labeling vs. non-labeling approaches for quantitative proteomics
- SpecArray: software tool for quantitative proteomics without isotopic labeling
- Corra: Framework to generate candidate biomarkers using non labeling methods

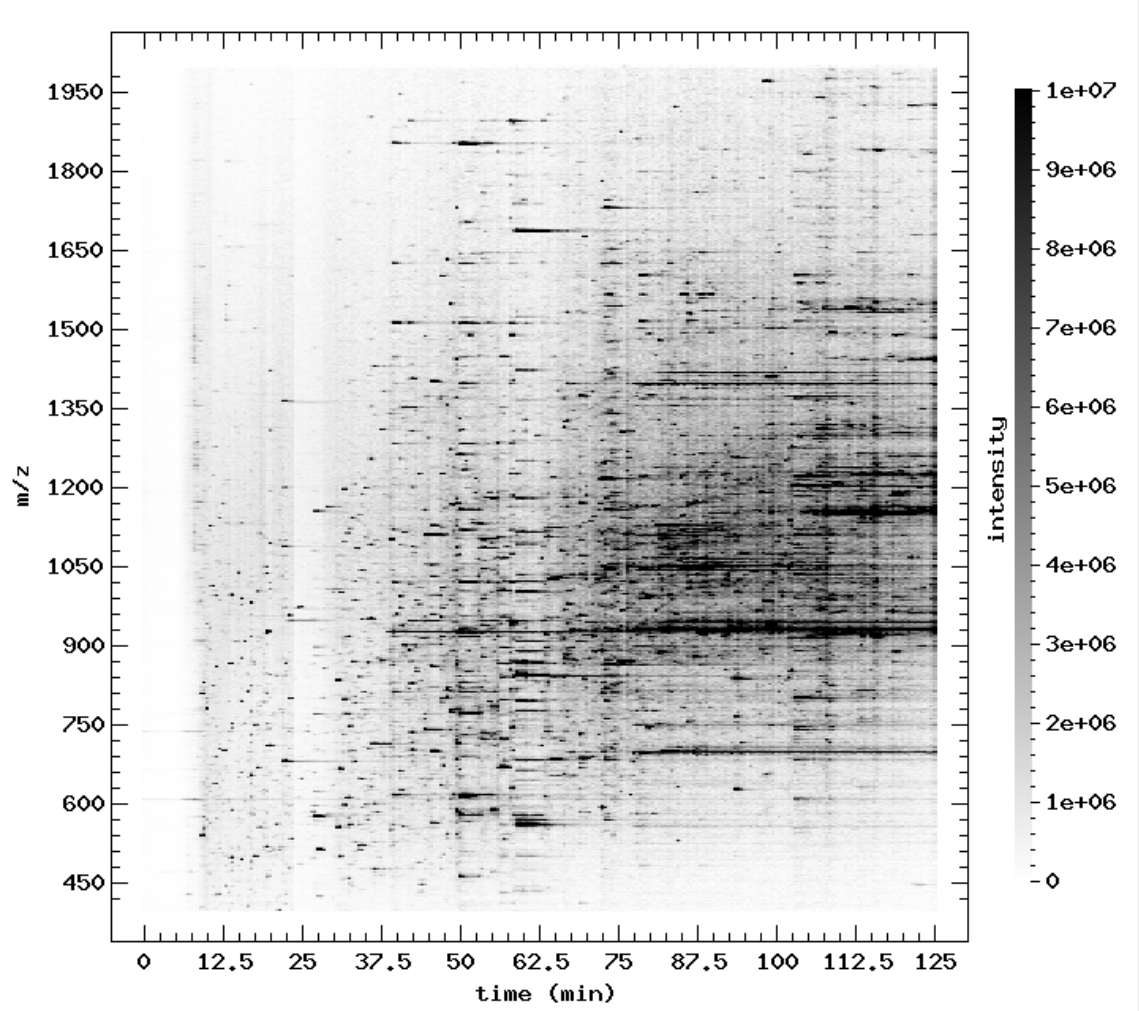
# Stable Isotopic Labeling and Quantitative Proteomics

- Samples labeled with different stable isotopes
- Chemically identical
- Distinguishable by MS in mass shift
- Peptide abundance ratio measured by ratio of SIC areas
- Peptides are identified before quantification

ASAPRatio

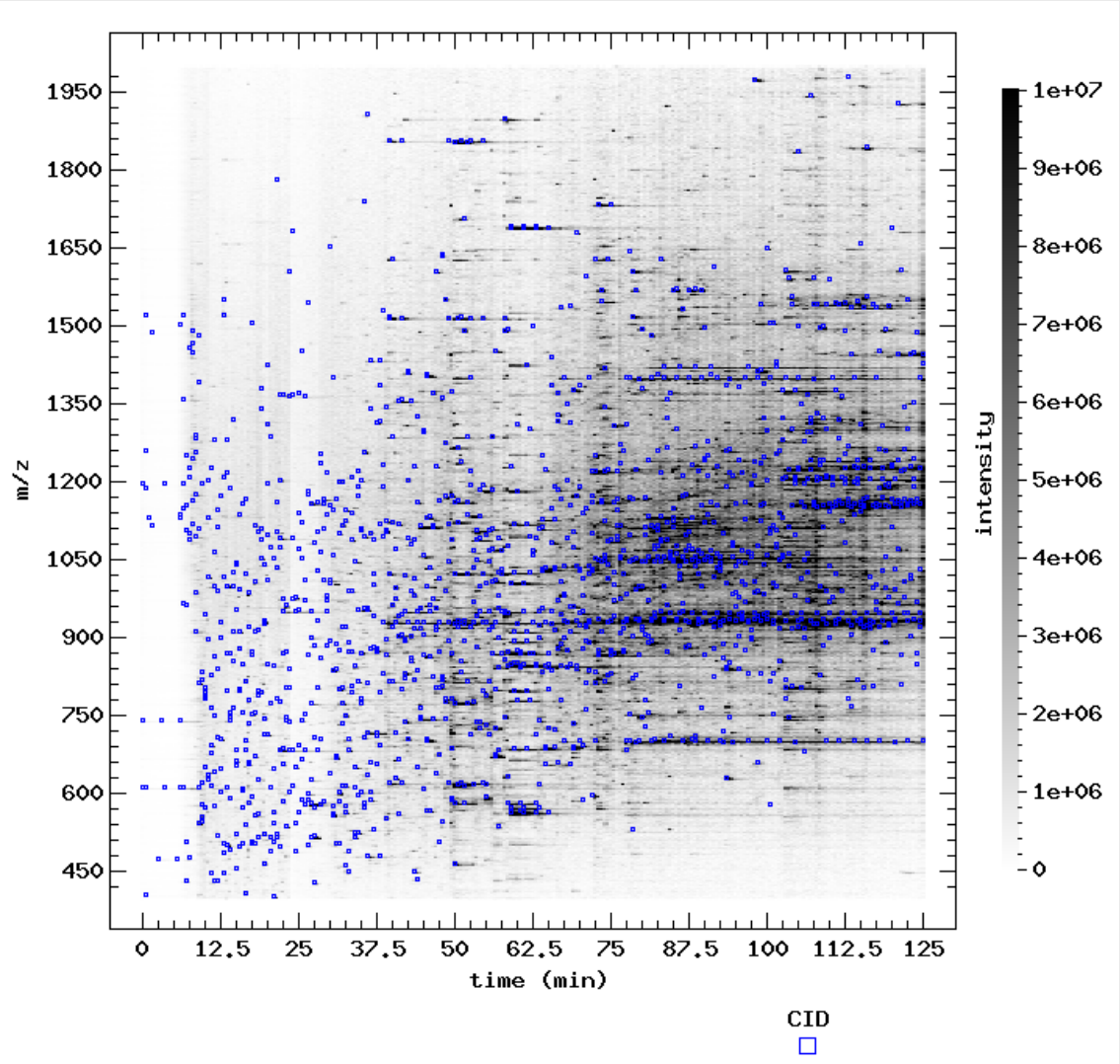


# Typical LC-MS/MS Analysis




Features: 2720

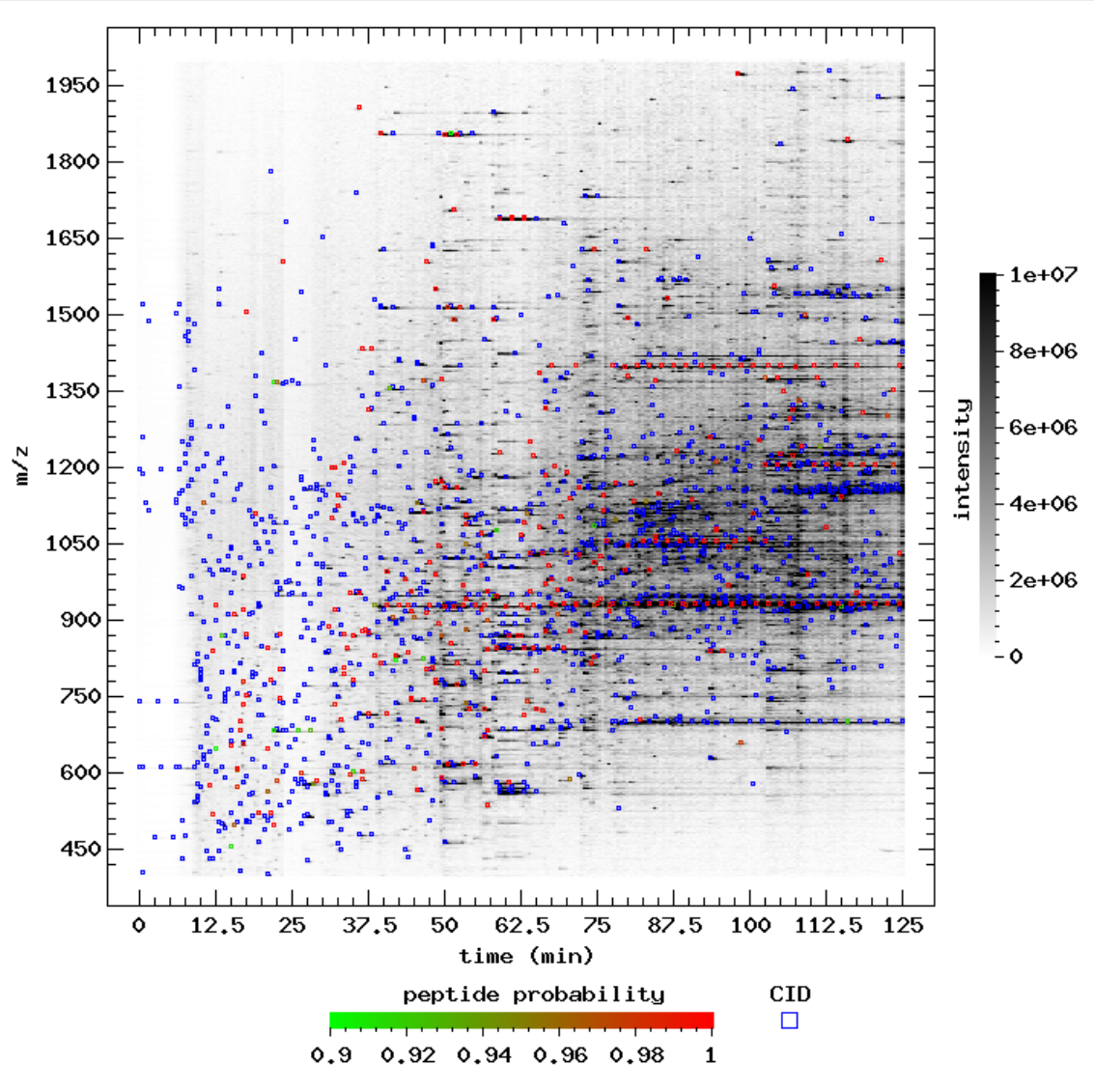
# Typical LC-MS/MS Analysis




Features: 2720

CIDs: 1633

# Typical LC-MS/MS Analysis




Features: 2720

CIDs: 1633

IDs: 363

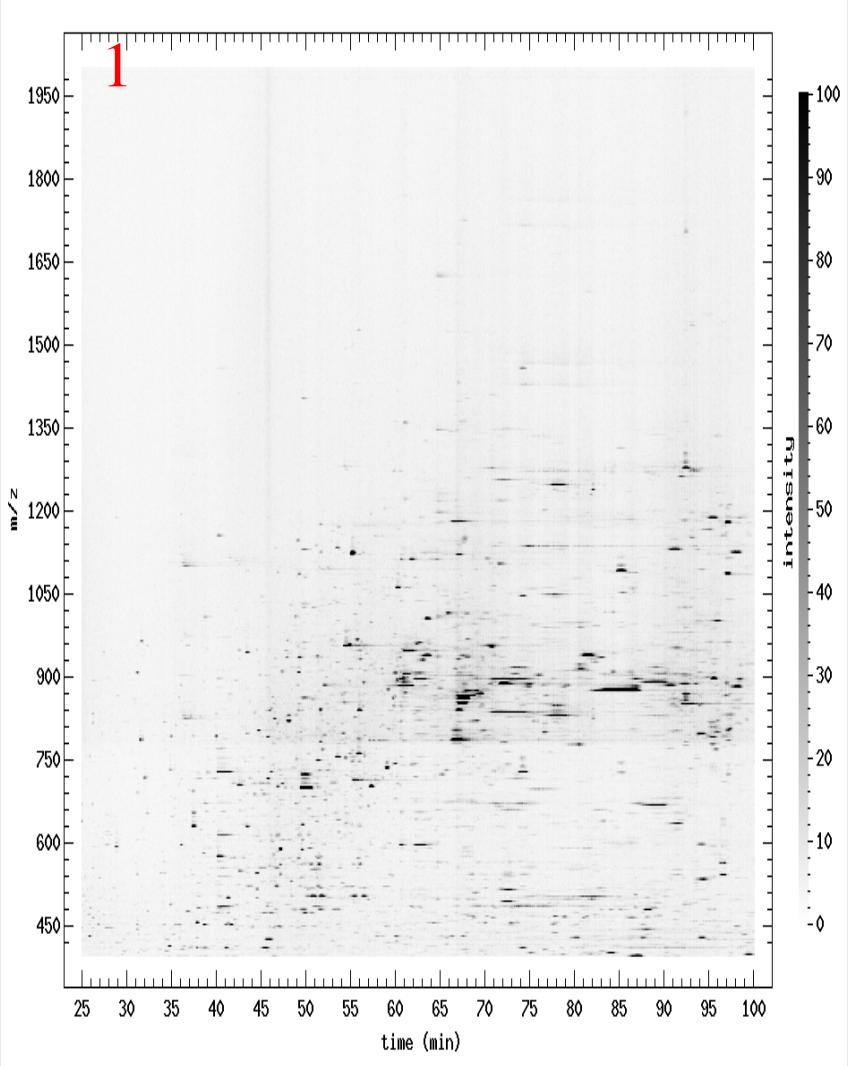
ID/CID: 22%

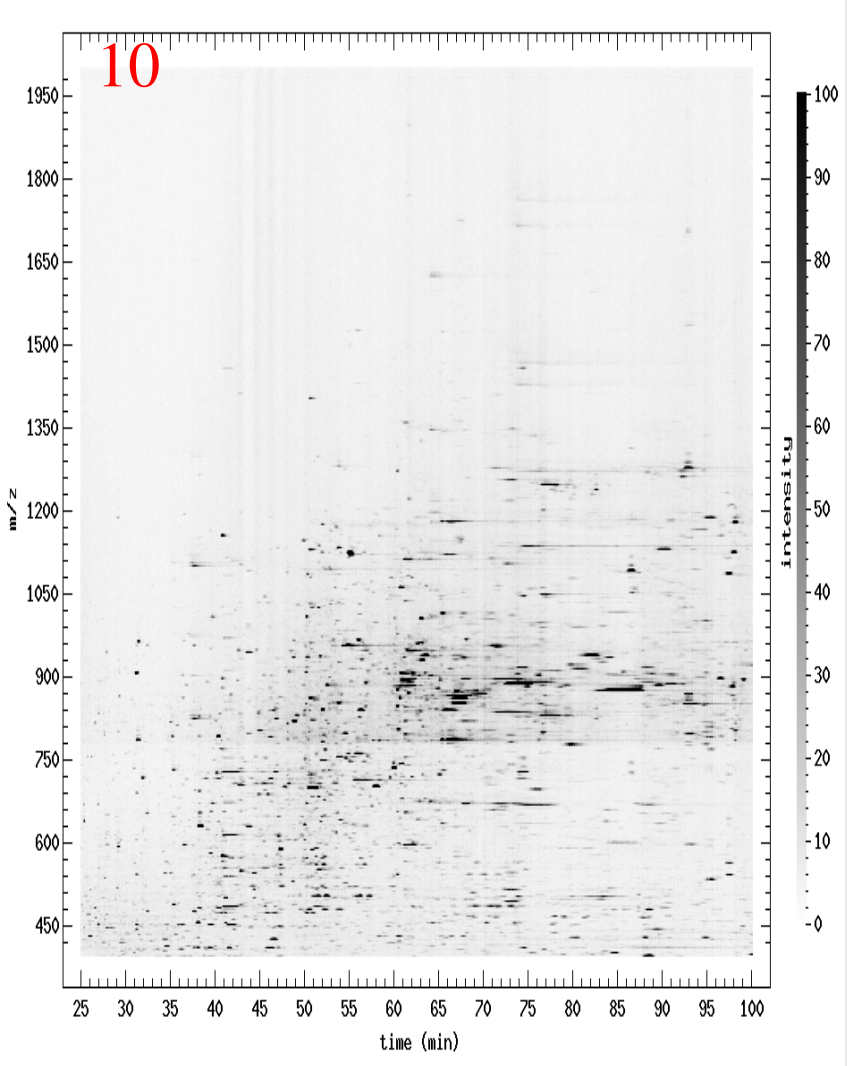
ID/feature: 13%

# Limitations of LC-MS/MS Approach to Large-Scale Protein Profiling

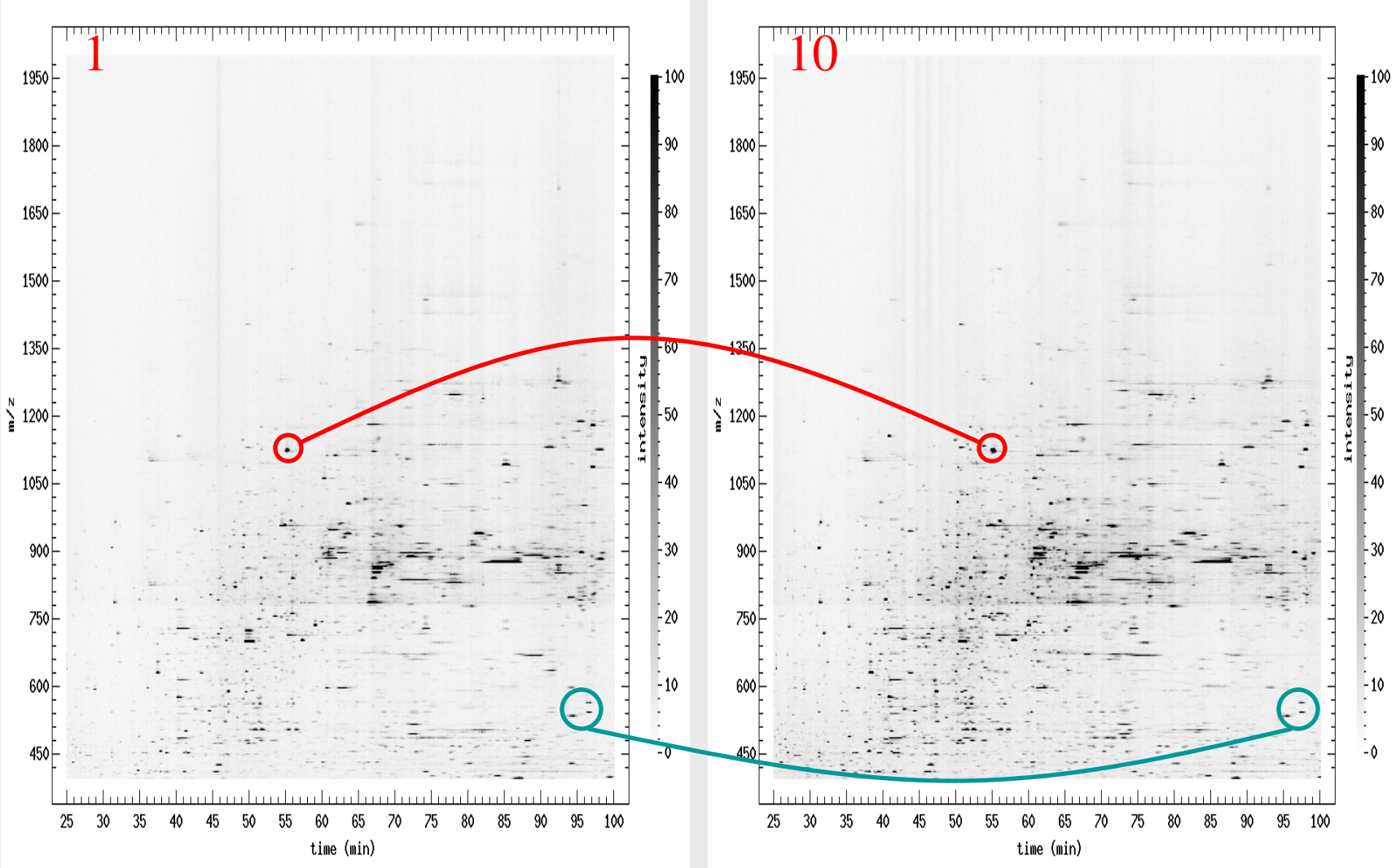
- Sample size limited:
  - ICAT (2), iTRAQ (4), …
- Difficult to trace protein abundance across a large number of samples
- Most peptides cannot be identified
- Difficult to identify & quantify low-abundance proteins

# LC-MS Approach



# LC-MS Approach



# LC-MS Platform for Large-Scale Protein Profiling

- Samples are NOT labeled
- Samples are analyzed under identical settings
- Peptide abundance is evaluated by MS signal intensity in different runs
- Reproducibility in LC-MS analysis critical
- Peptide alignment crucial
- Followed by target LC-MS/MS

# Challenges in LC-MS Platform

- Highly reproducible LC-MS analysis
  - Retention time shift
  - Fluctuations in MS signal intensity
  - Peptide identification in separated MS/MS
- Complex samples
  - Overlapping signals
  - Misaligned peptides
- Large sample size
  - Column degradation

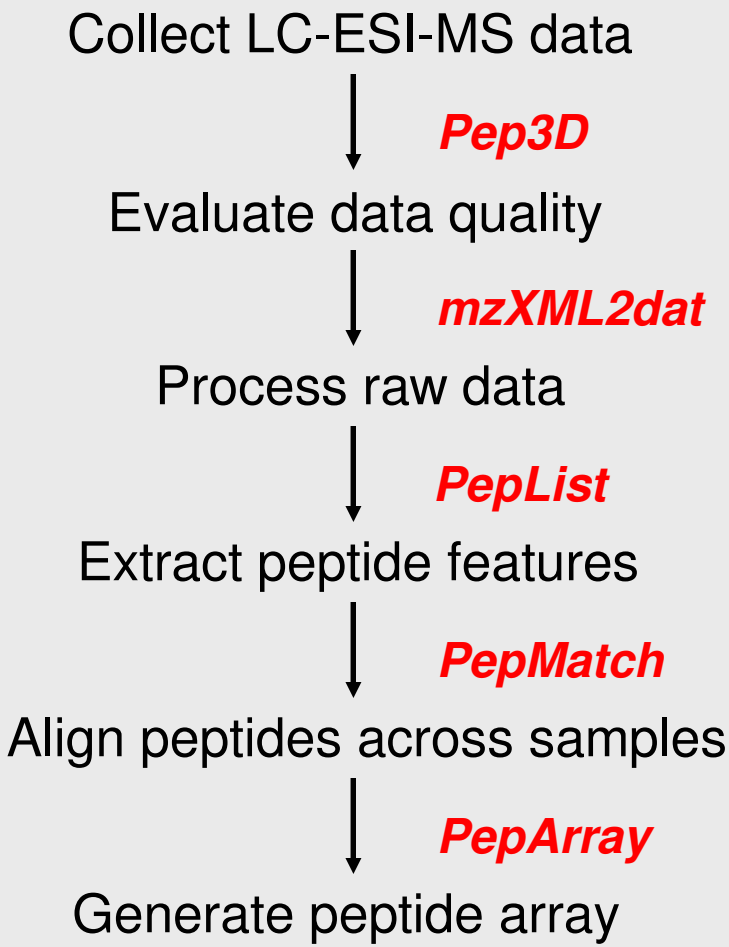
# Outline

- Principles of quantitative proteomics
- Labeling vs. non-labeling approaches for quantitative proteomics
- SpecArray: software tool for quantitative proteomics without isotopic labeling
- Corra: Framework to generate candidate biomarkers using non labeling methods

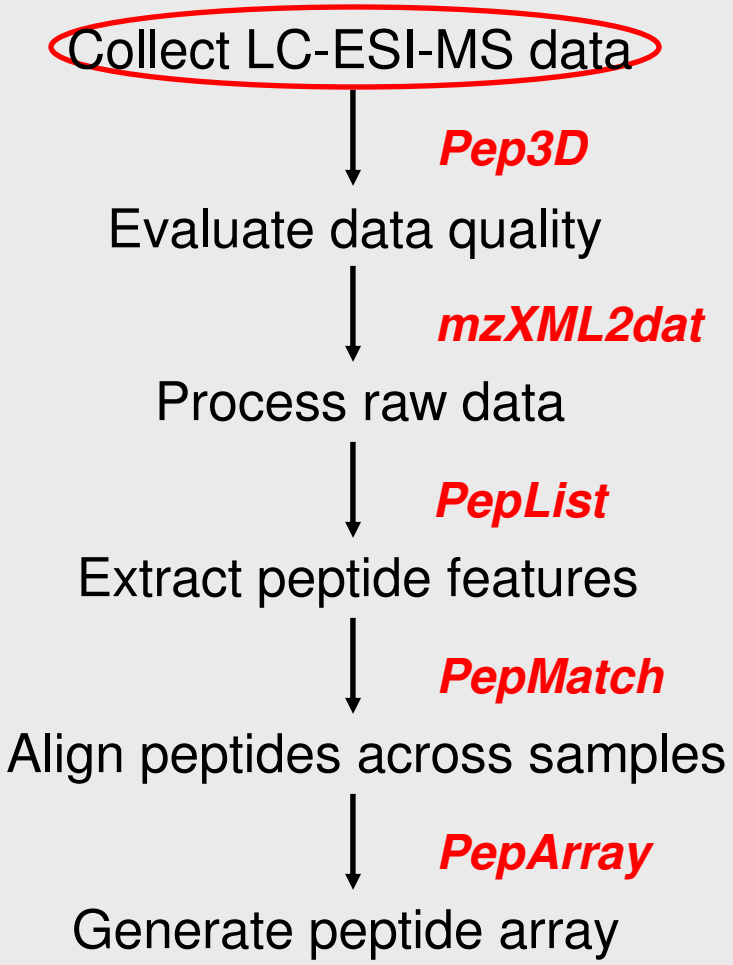
# SpecArray

- Software Suite for the Generation and Comparison of Peptide Arrays from Sets of Data Collected by Liquid Chromatography-Mass Spectrometry
- Xiao-Jun Li *et. al.* Molecular & Cellular Proteomics 4.9, 2005
- SpecArray v1.2.0 is available in http://sourceforge.net/projects/sashimi or http://tools.proteomecenter.org/software.php

# SpecArray Software Suite



# Collect LC-MS Data




**Data collection:**

- ✓ **LC-MS in profiling mode**
- ✓ **No or minimal MS/MS**
- ✓ **Peptide abundance evaluated by MS signal intensity (after proper normalization)**
- ✓ **Converted into mzXML file format**

# Evaluate Data Quality

Collect LC-ESI-MS data

↓ ***Pep3D***

Evaluate data quality

↓ ***mzXML2dat***

Process raw data

↓ ***PepList***

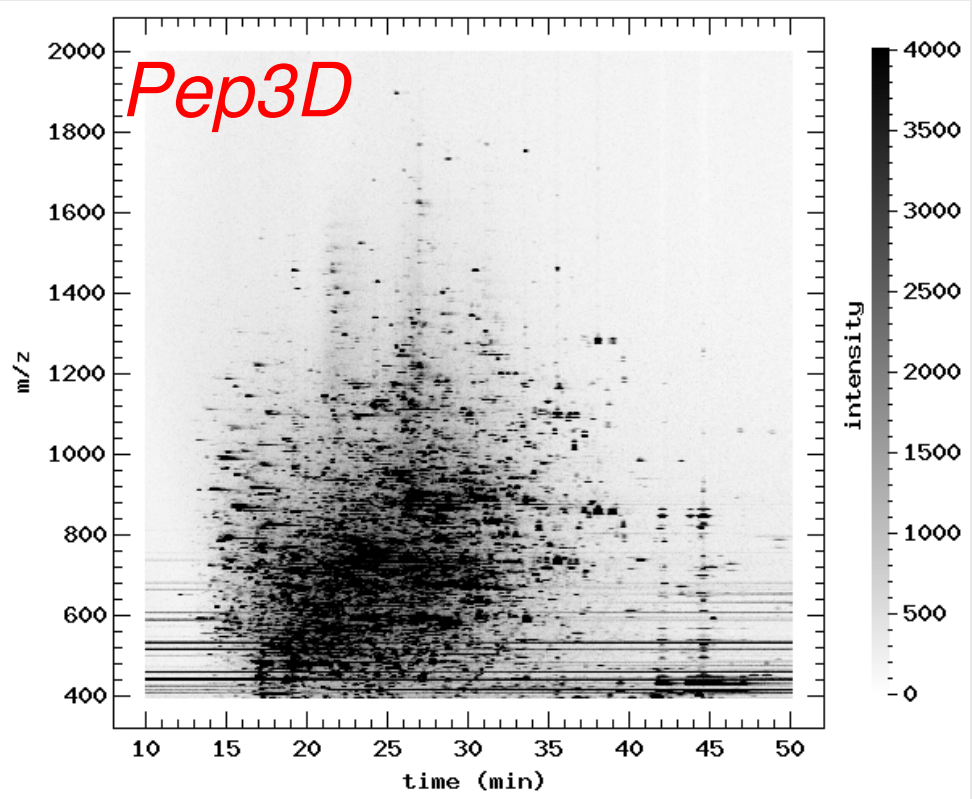
Extract peptide features

↓ ***PepMatch***

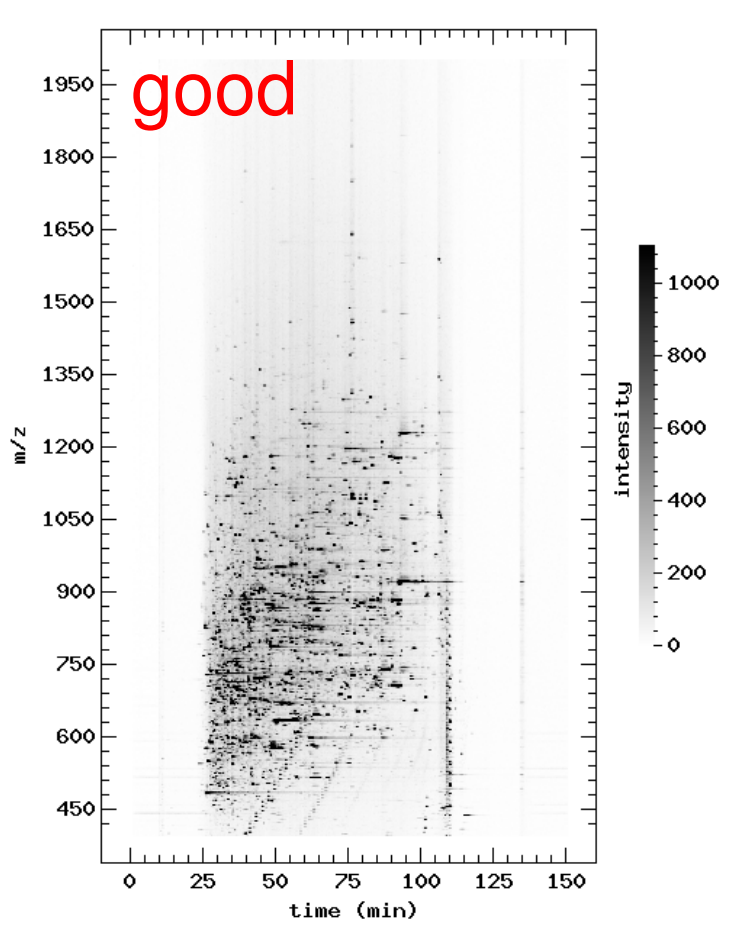
Align peptides across samples

↓ ***PepArray***

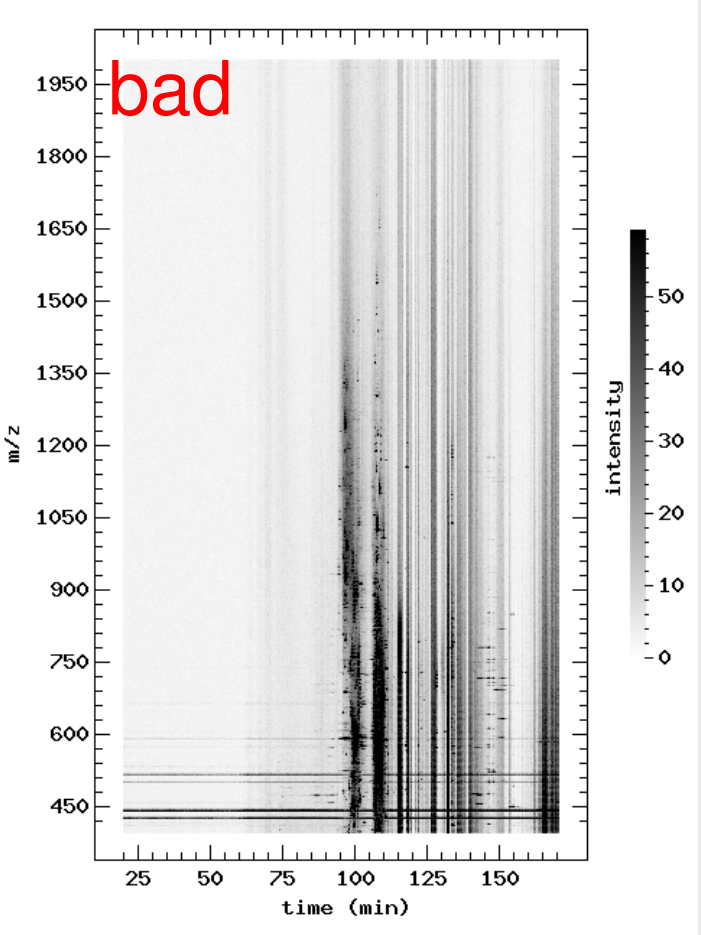
Generate peptide array



# LC-MS Data of Same Sample




**MicroTOF**




**QTOF**

# Process Raw Data

Collect LC-ESI-MS data

↓ ***Pep3D***

Evaluate data quality

↓ ***mzXML2dat***

Process raw data

↓ ***PepList***

Extract peptide features

↓ ***PepMatch***

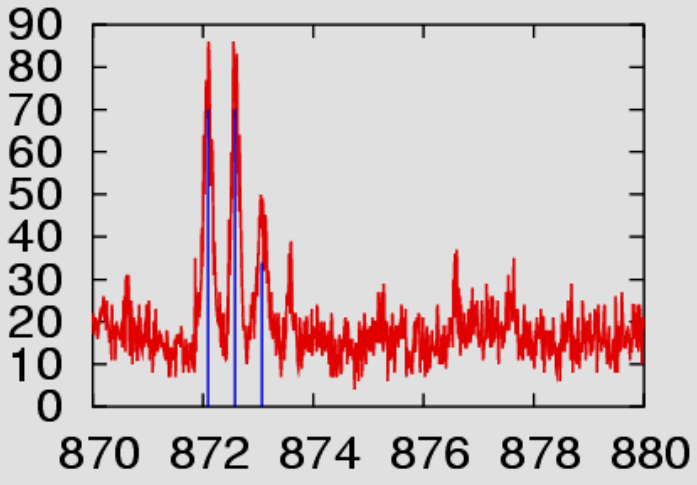
Align peptides across samples

↓ ***PepArray***

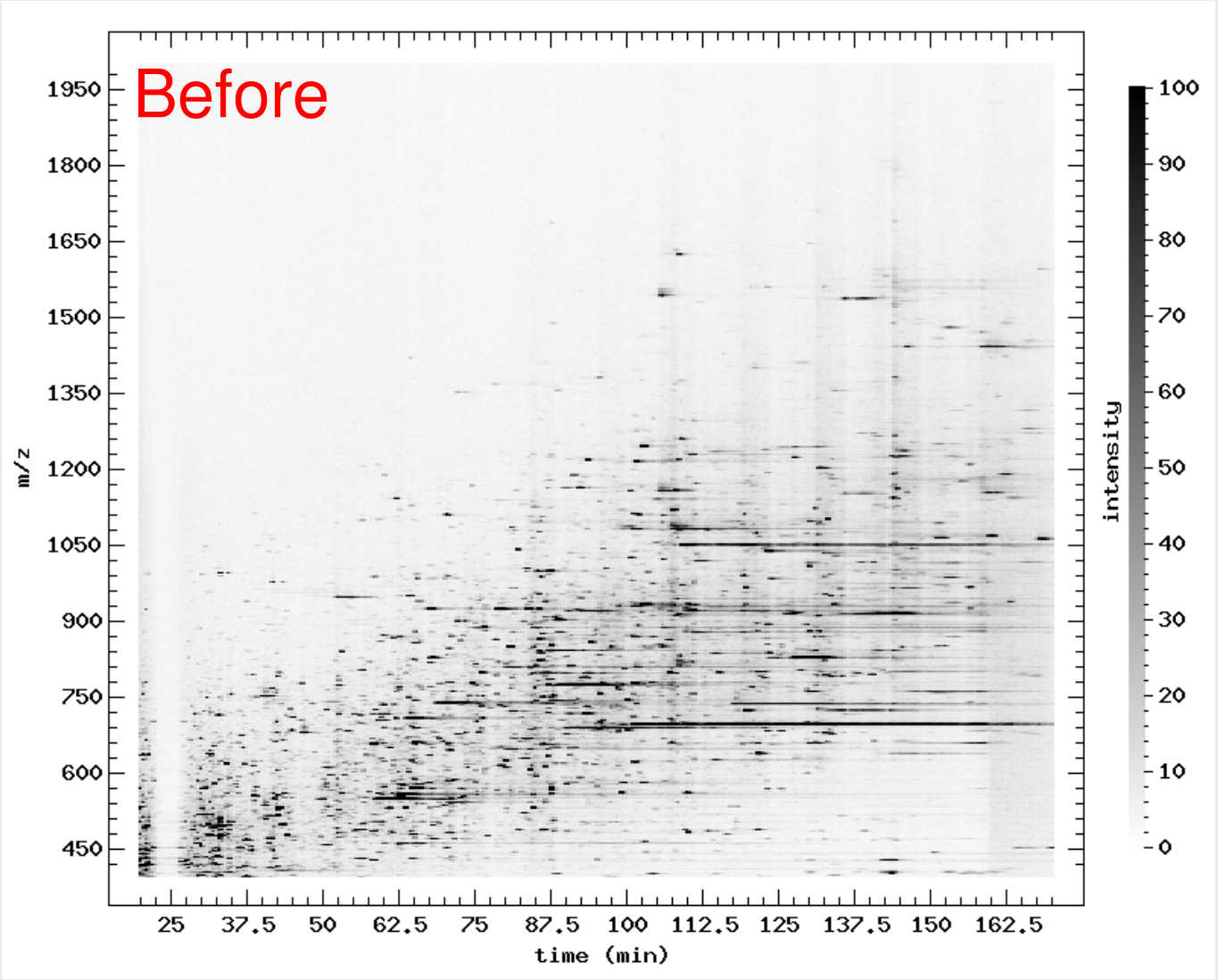
Generate peptide array

***mzXML2dat*:**

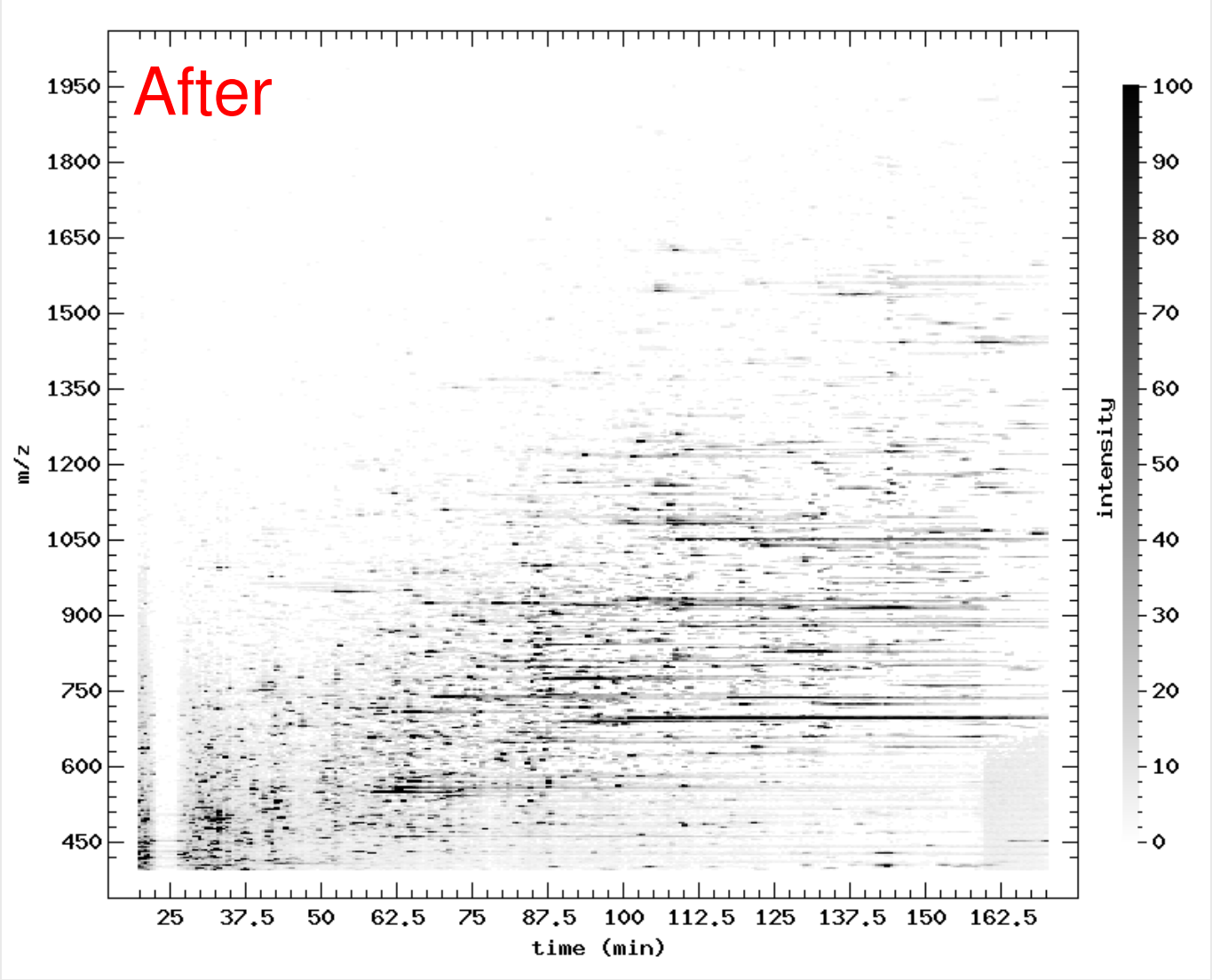
- ✓ **Smoothing, denoising, centroiding, subtracting background, estimating S/N**
- ✓ **~2GB → ~5MB**



# Pep3D Images



# Pep3D Images



# Extract Peptide Features

Collect LC-ESI-MS data

↓ ***Pep3D***

Evaluate data quality

↓ ***mzXML2dat***

Process raw data

↓ ***PepList***

Extract peptide features

↓ ***PepMatch***

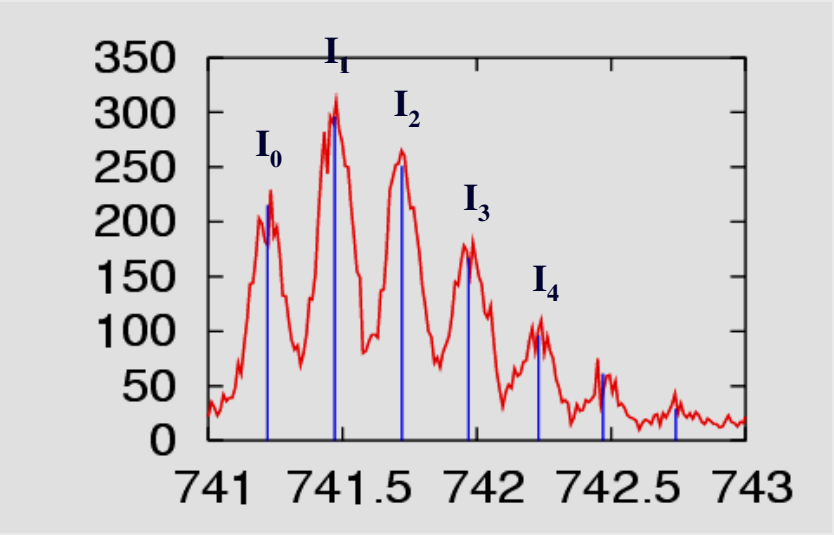
Align peptides across samples

↓ ***PepArray***

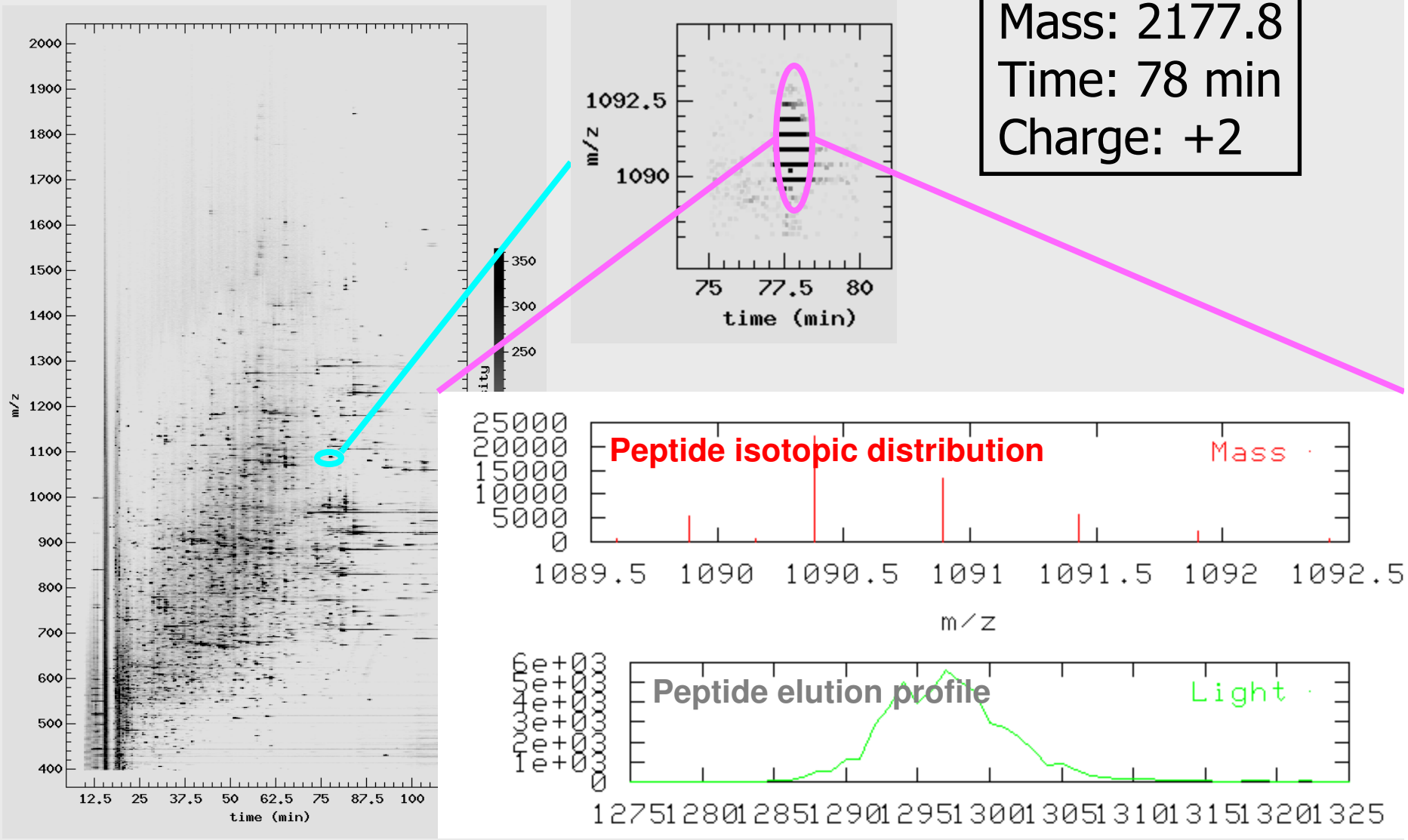
Generate peptide array

***PepList*:**

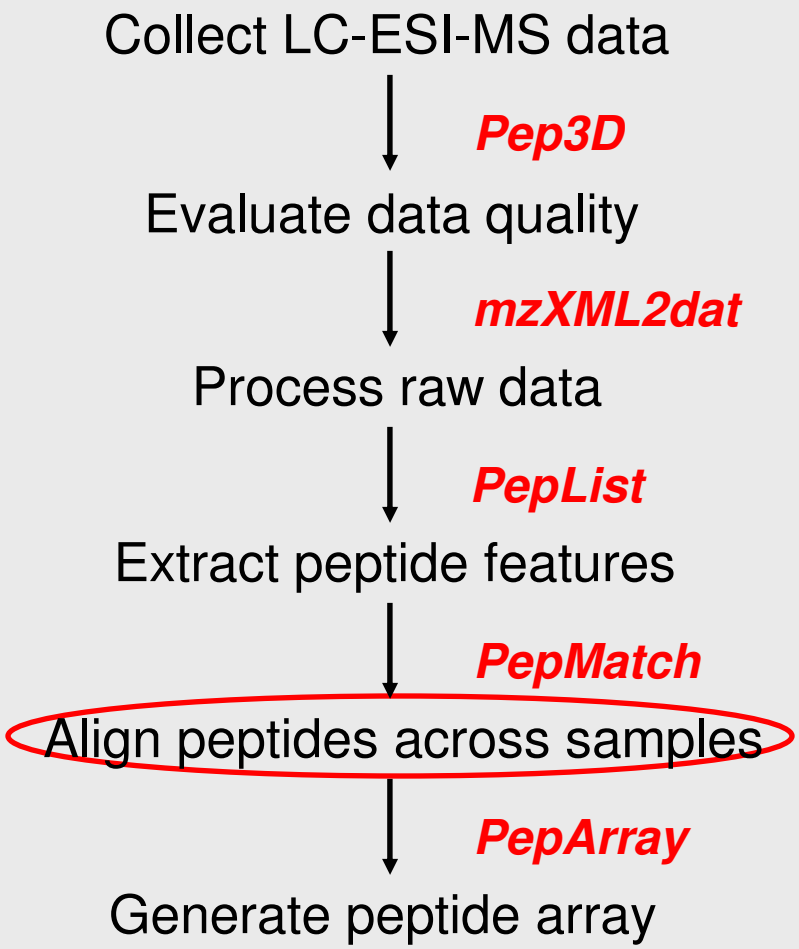
- ✓ **Peptide isotopic distribution**
- ✓ **Peptide elution profile**
- ✓ **Data range: m/z, time**



# Peptide Feature Detection



# Align Peptides Across Samples




***PepMatch*:**

- ✓ **Align peptides by mass, charge, retention time**
- ✓ **Reproducibility**

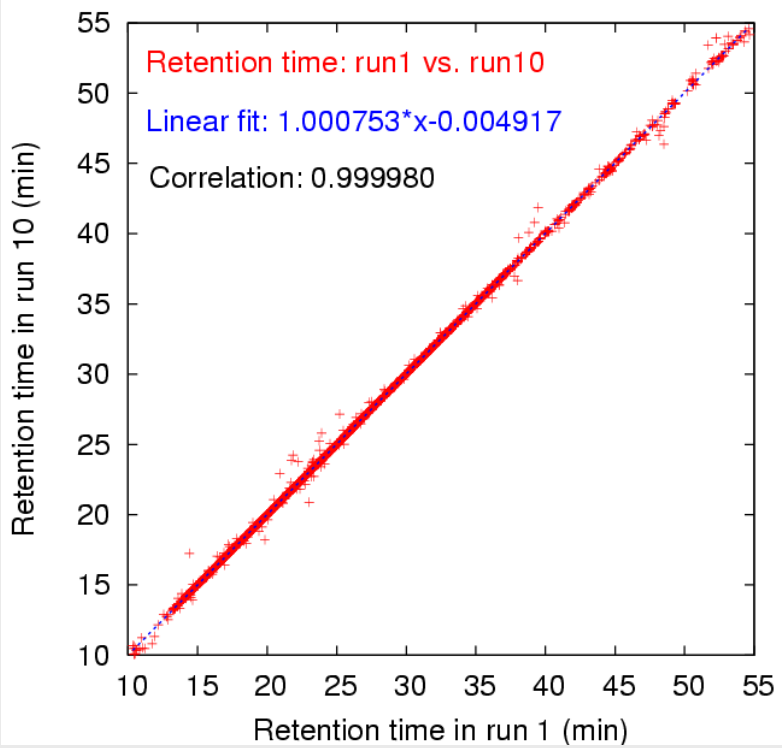
  m/z: **Highly reproducible**

  retention time: **Relative order reproducible**

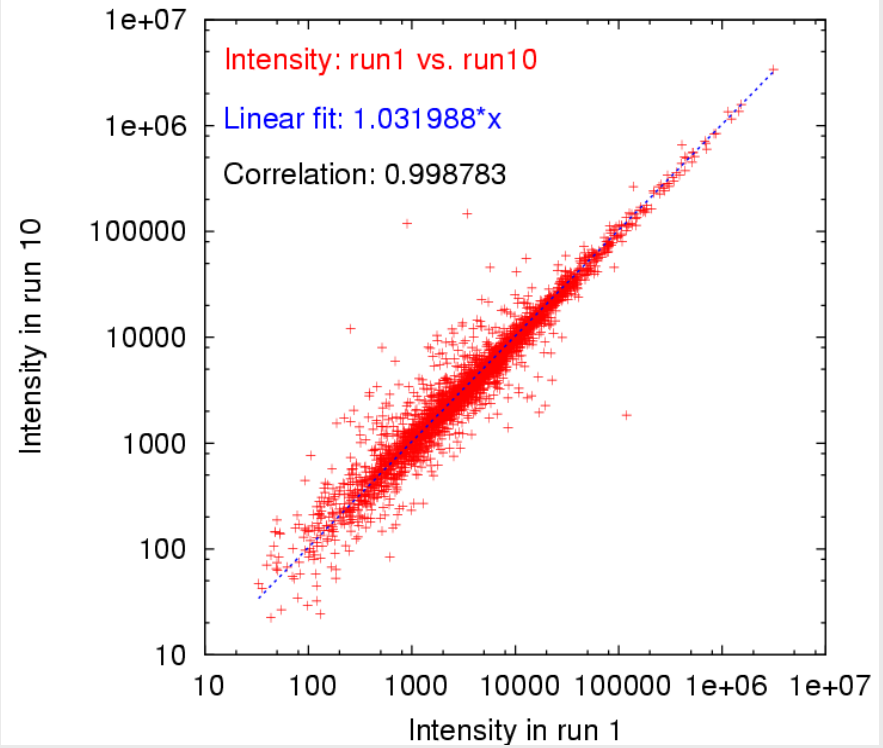
  signal intensity: **Relative intensity reproducible**

- ✓ **Align multiple samples**
- ✓ **Combine aligned peptides**

# Peptide Alignment Between Two Samples

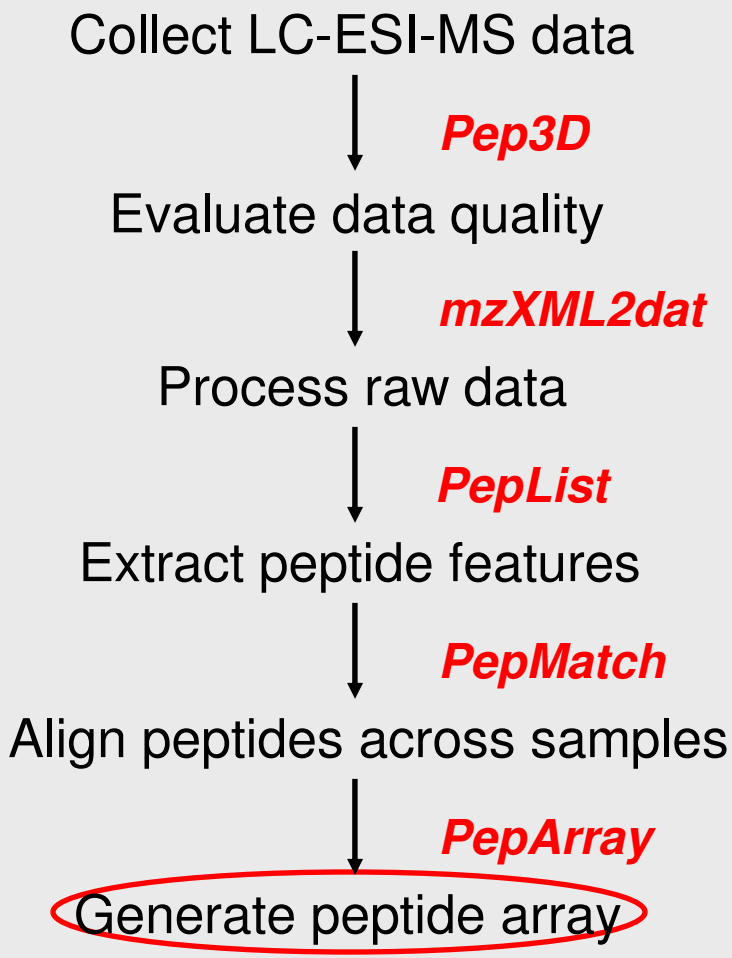



retention time




intensity

# Generating Peptide Expression Array




***PepArray*:**

- ✓ **Sample-dependent normalization**
- ✓ **Search missed features**
- ✓ **User specified information**

**Minimal sample size of each group**

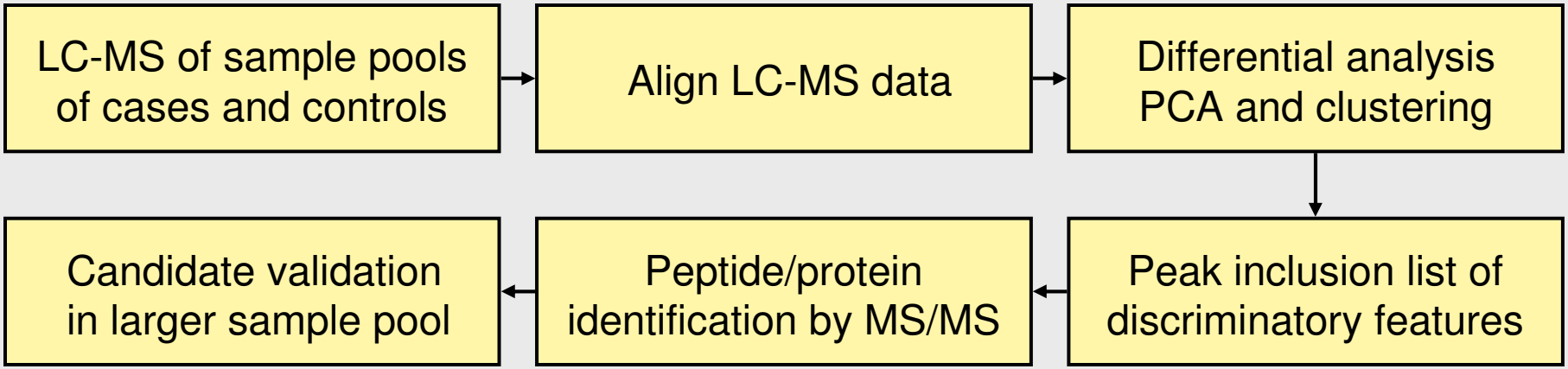
# Outline

- Principles of quantitative proteomics
- Labeling vs. non-labeling approaches for quantitative proteomics
- Introduction to ASAPRatio: software tool for quantitative proteomics using isotopic labeling
- Introduction to SpecArray: software tool for quantitative proteomics without isotopic labeling
- Corra: Framework to generate candidate biomarkers.

# What is Corra?

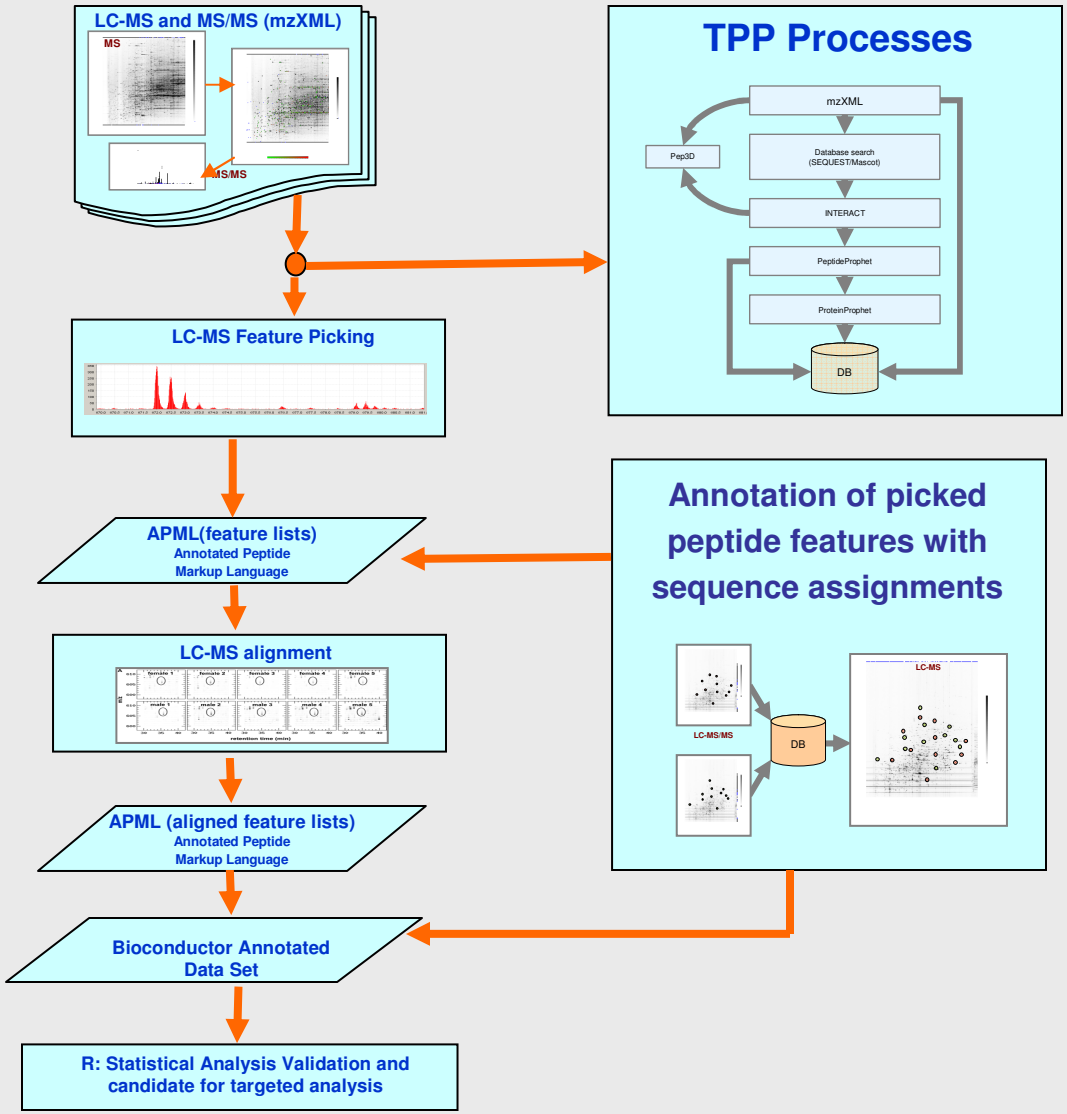
- Corra is a Scottish Goddess of prophecy.
- Corra is a framework to generate candidate biomarkers in high throughput MS data processing environment.
- Corra's goal is to detect differentially expressed features by maximizing the number of such features while controlling the false discovery rate.
- Corra uses LC-MS (and LC-MS/MS) and experiment design information
- Work in progress
  - Validation of work flow using Latin Square data
  - Validation of reproducibility

# Corra: MS1 Data Analysis (A Discovery-Based Approach)



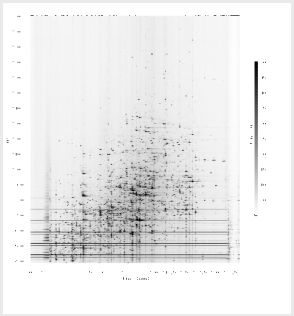

Discriminatory peptide features initially determined just from MS1 data analysis
Follow-up MS/MS subsequently determines peptide identity

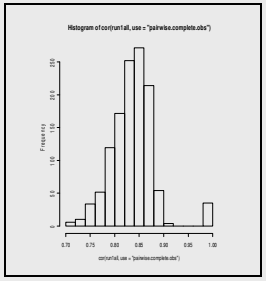
# Corra Framework Overflow

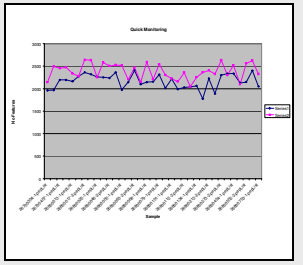


# Corra Framework Overview

Inputs:
- mzXML
- Search Result Files
- Sample Information

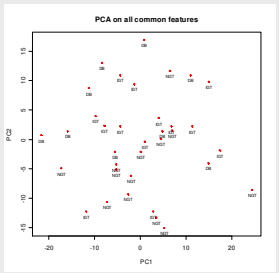


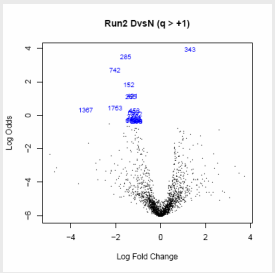



Corra
- Quick Quality Control on the fly
- LC-MS or LC-MS/MS Process
- Annotation of Samples and aligned Features
- Simple and Basic Statistical Analysis (PCA, Volcanoplot)
- Simple Cross-validation

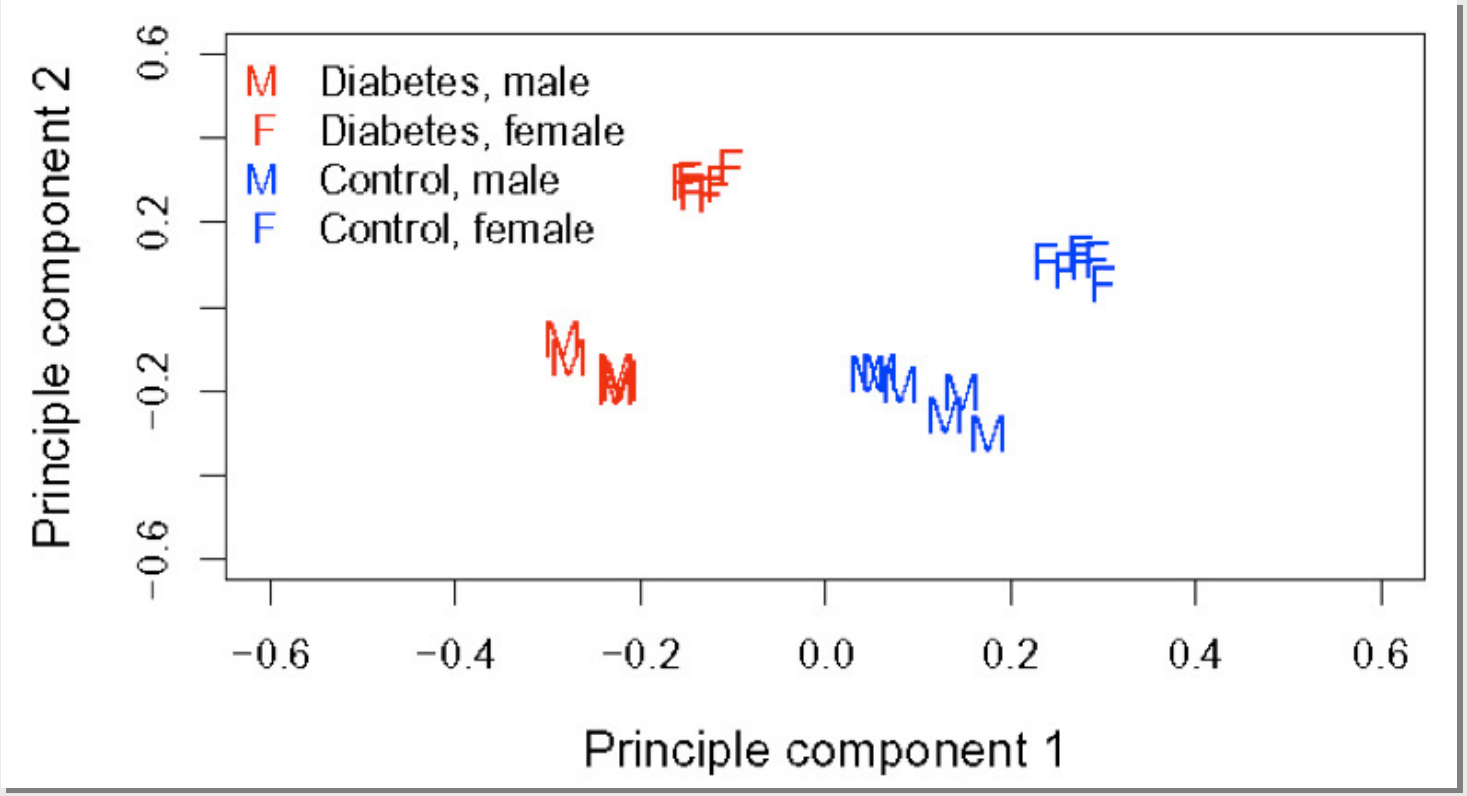
Outputs:
- APML for Feature Lists
- APML for Aligned Feature Lists
- Candidate Target Analysis Lists (feature or peptide or protein)
- Weighted Panels of Proteins which has a certain statistical predictability



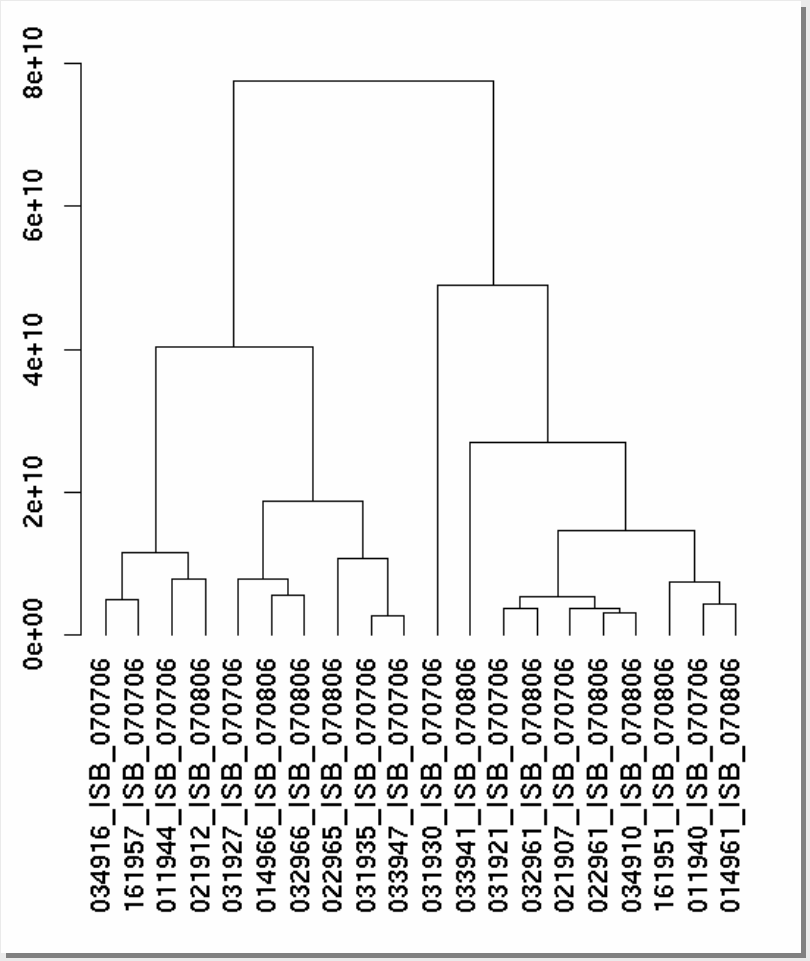

| | m.z | rt | charge | M | t | P.Value | adj.P.Val | B |
|---|---|---|---|---|---|---|---|---|
| 343 | 439.468 | 35.666 | 4 | 1.309641 | 5.747842 | 5.92E-06 | 0.010247 | 3.929592 |
| 285 | 1164.557 | 121.915 | 4 | -1.55887 | -5.54884 | 9.78E-06 | 0.010247 | 3.480909 |
| 742 | 734.352 | 65.715 | 3 | -2.03654 | -5.30905 | 2.34E-05 | 0.016372 | 2.687205 |
| 152 | 974.18 | 80.755 | 4 | -1.41841 | -4.82714 | 6.15E-05 | 0.032238 | 1.825438 |
| 421 | 992.613 | 16.686 | 4 | -1.26019 | -4.53599 | 0.00013 | 0.040325 | 1.150351 |
| 255 | 981.175 | 45.484 | 4 | -1.32148 | -4.52185 | 0.000135 | 0.040325 | 1.11754 |
| 1753 | 669.88 | 90.033 | 2 | -2.03223 | -4.39905 | 0.000293 | 0.071341 | 0.442843 |
| 1367 | 861.883 | 47.297 | 2 | -3.3339 | -6.34749 | 0.000107 | 0.040325 | 0.299392 |
| 453 | 1104.57 | 112.804 | 4 | -1.17287 | -4.15356 | 0.000346 | 0.071341 | 0.264205 |
| 680 | 758.056 | 67.722 | 3 | -1.21069 | -4.10977 | 0.000387 | 0.071341 | 0.163092 |
| 332 | 739.094 | 92.16 | 4 | -1.07081 | -4.06259 | 0.000437 | 0.071341 | 0.05428 |
| 326 | 987.174 | 45.883 | 4 | -1.25356 | -3.99535 | 0.000518 | 0.071341 | -0.10052 |
| 183 | 726.855 | 82.543 | 4 | -1.12979 | -3.94417 | 0.00059 | 0.071341 | -0.2181 |
| 1252 | 905.114 | 81.87 | 3 | -1.18671 | -3.93147 | 0.00061 | 0.071341 | -0.24723 |
| 324 | 923.705 | 131.279 | 4 | -1.10673 | -3.9304 | 0.000611 | 0.071341 | -0.24969 |
| 668 | 889.396 | 67.162 | 3 | -1.31687 | -3.90338 | 0.000655 | 0.071341 | -0.31164 |
| 1179 | 448.137 | 25.178 | 3 | -1.25939 | -3.8901 | 0.000677 | 0.071341 | -0.34206 |
| 234 | 531.773 | 102.71 | 4 | -1.05553 | -3.88587 | 0.000684 | 0.071341 | -0.35175 |
| 370 | 448.18 | 24.596 | 4 | -1.05587 | -3.88262 | 0.00069 | 0.071341 | -0.35918 |
| 644 | 892.412 | 71.75 | 3 | -1.13239 | -3.91765 | 0.000715 | 0.071341 | -0.36621 |

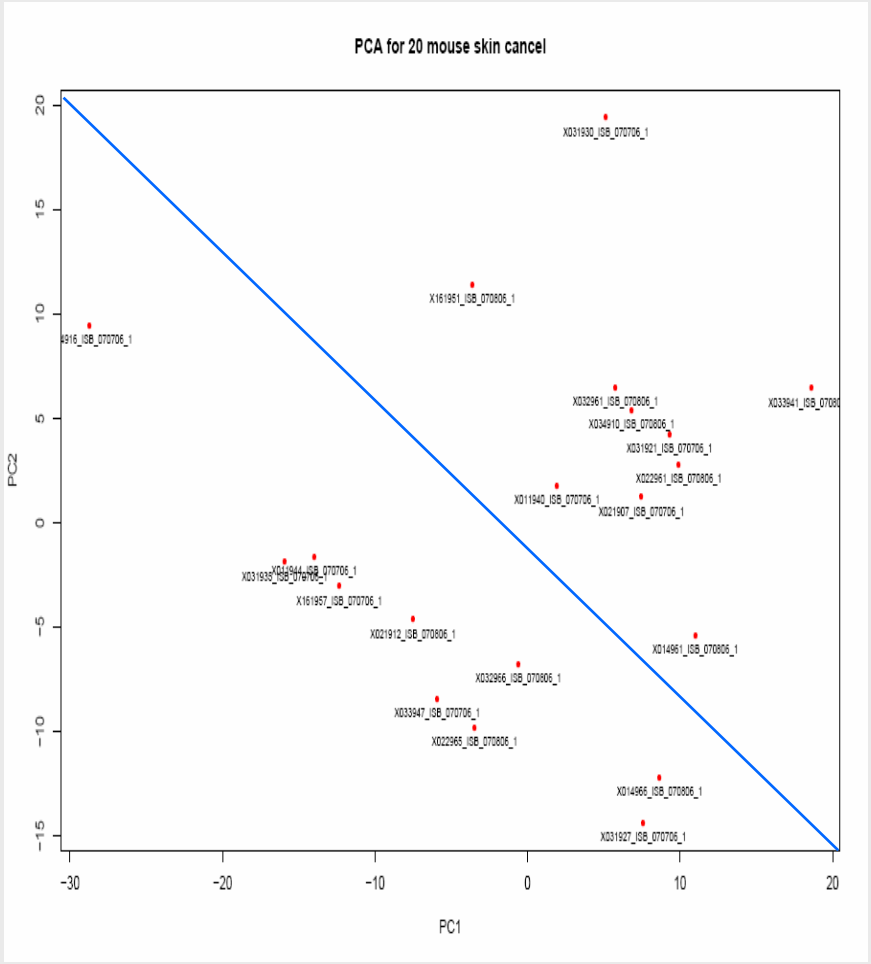
# Corra Application Examples (LC-MS: Diabetes Pilot Study)




24 runs = 8 individuals x 3 replicate runs

# Corra Application Examples (LC-MS: NCI Mouse Skin Cancer)



# Summary

- Label free quantitative LC-MS and LC-MS/MS is a powerful method for identifying and quantifying proteins in complex samples
- Corra framework is very promising in large-scale protein profiling
- Software suites have been developed for both LC-MS and LC-MS/MS platforms