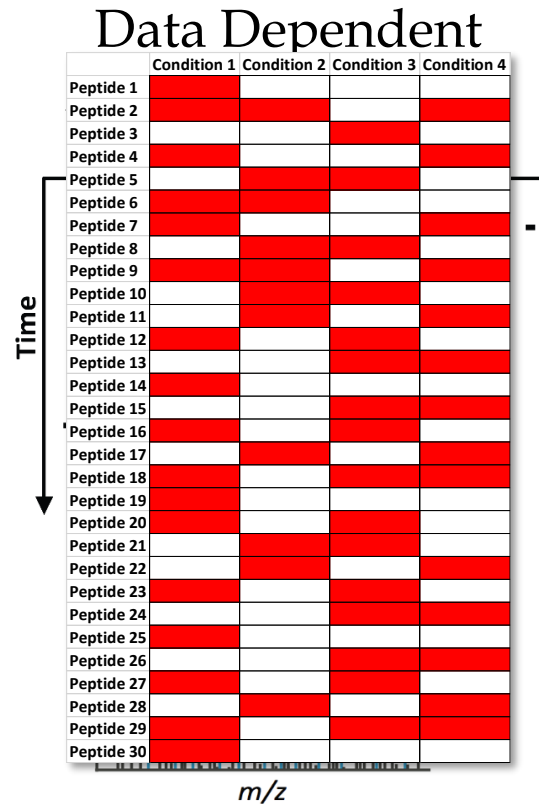# DIA Acquisition and Data Analysis on Ion Trapping Mass Spectrometers:
How we perform quantitative analyses using DIA in our Lab

MacCoss Lab

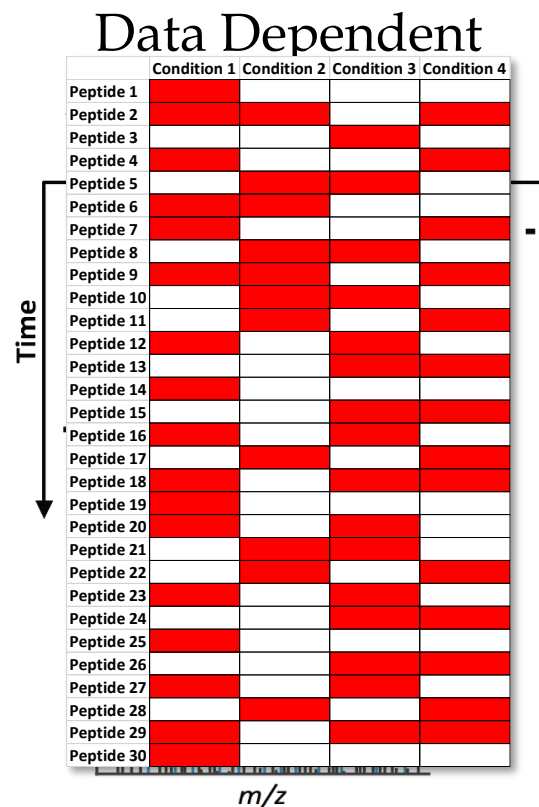# Mass Spectrometry Data Acquisition Strategies Used in Proteomics
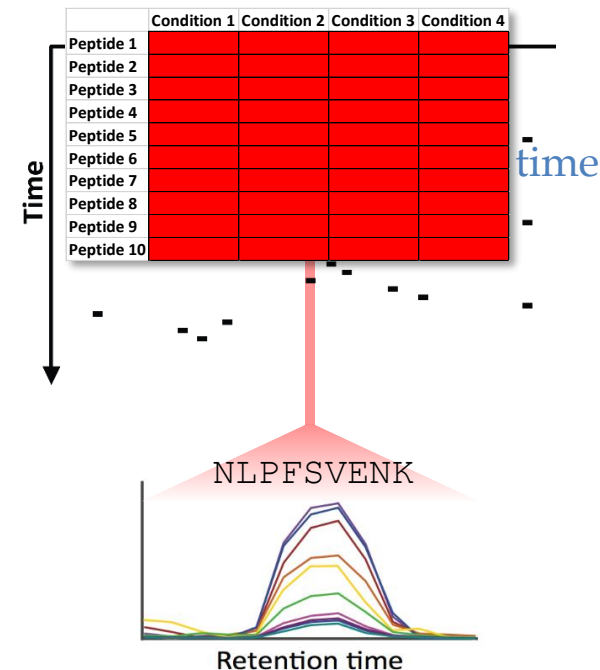
Mass Spectrometry Data Acquisition Strategies Used in Proteomics
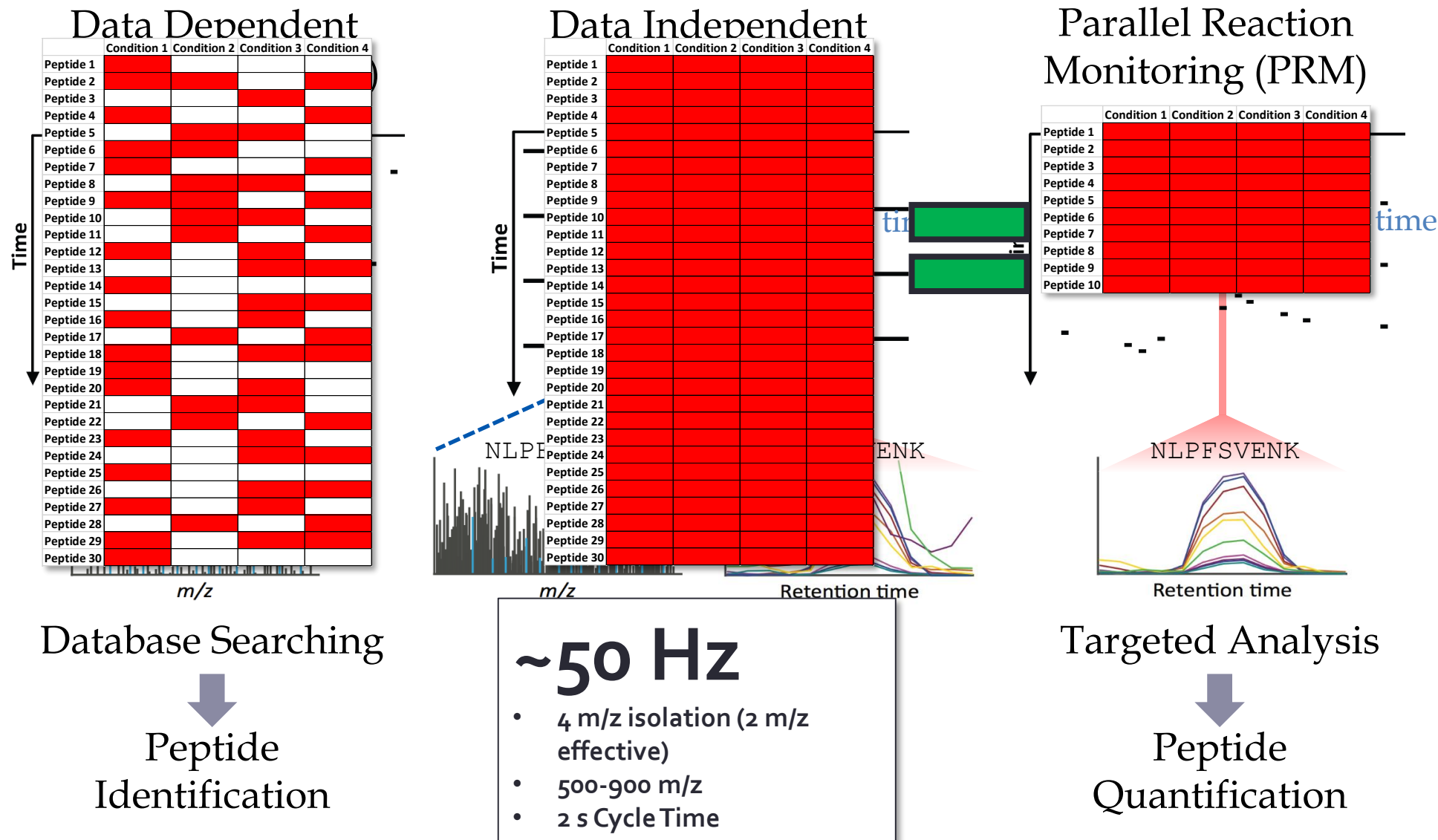
Targeted Chromatogram Extraction
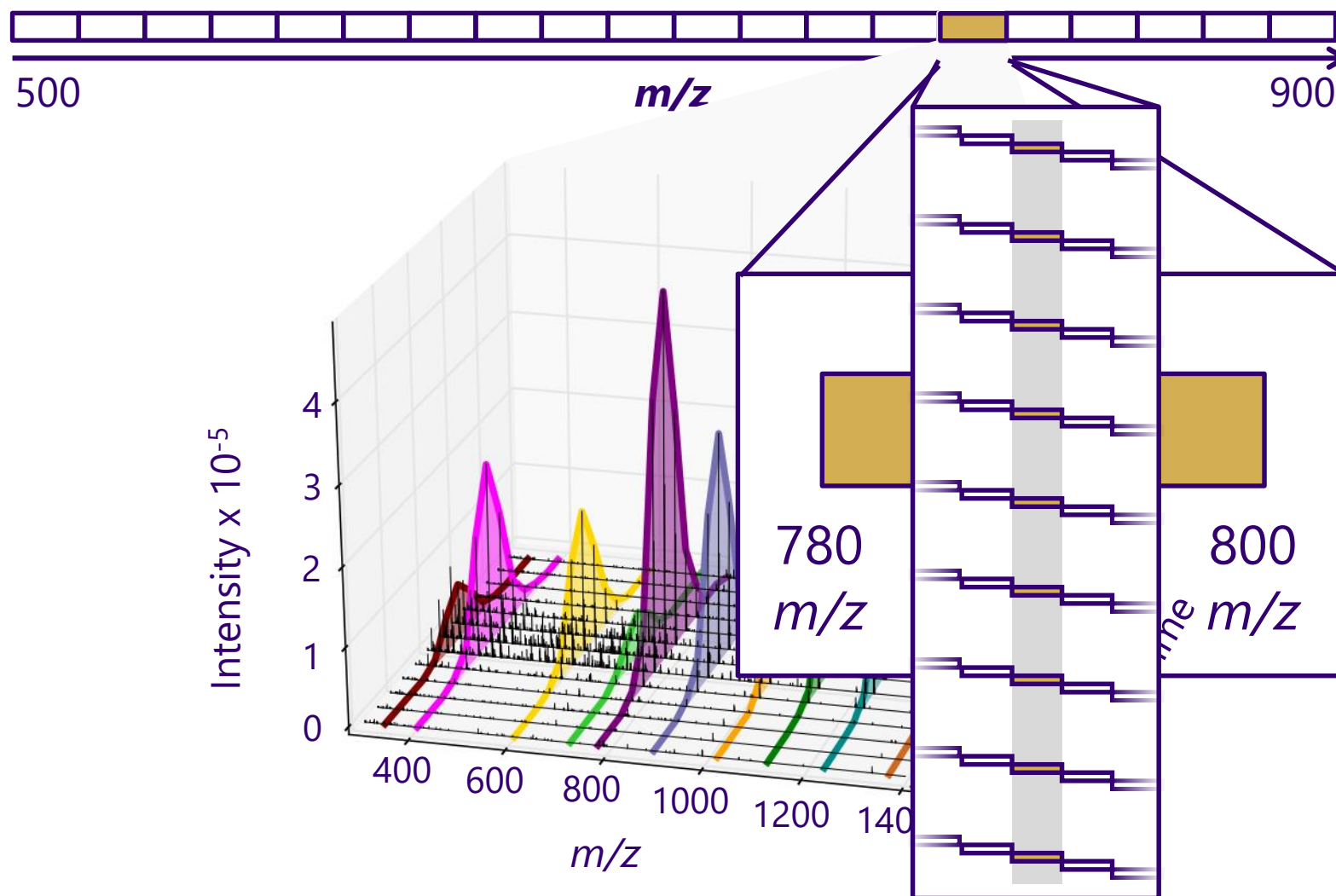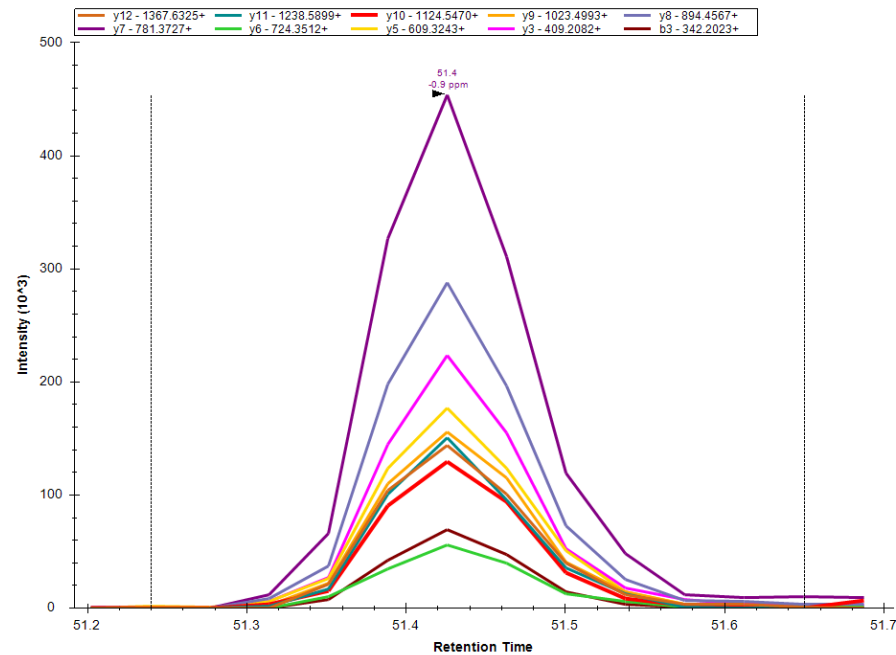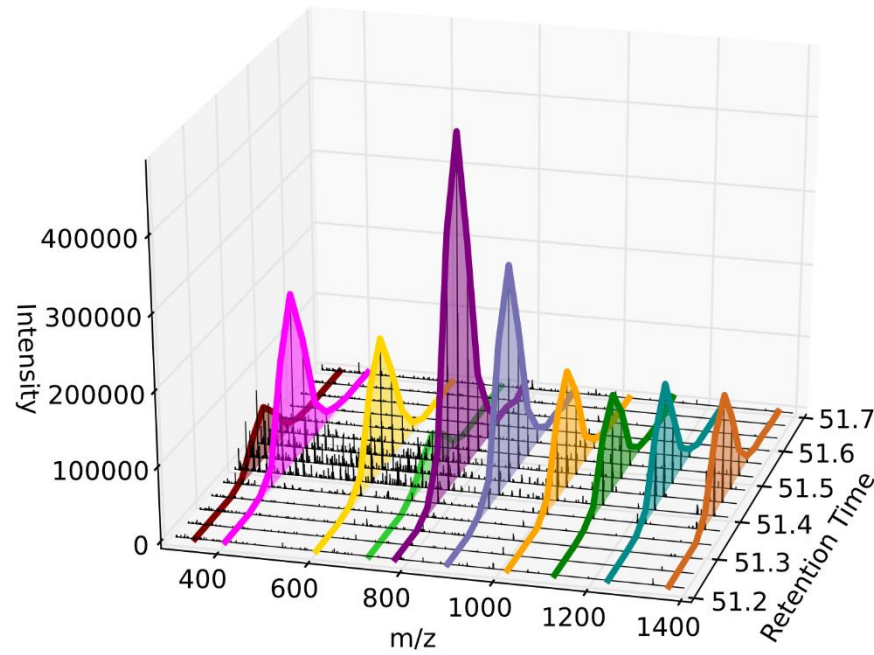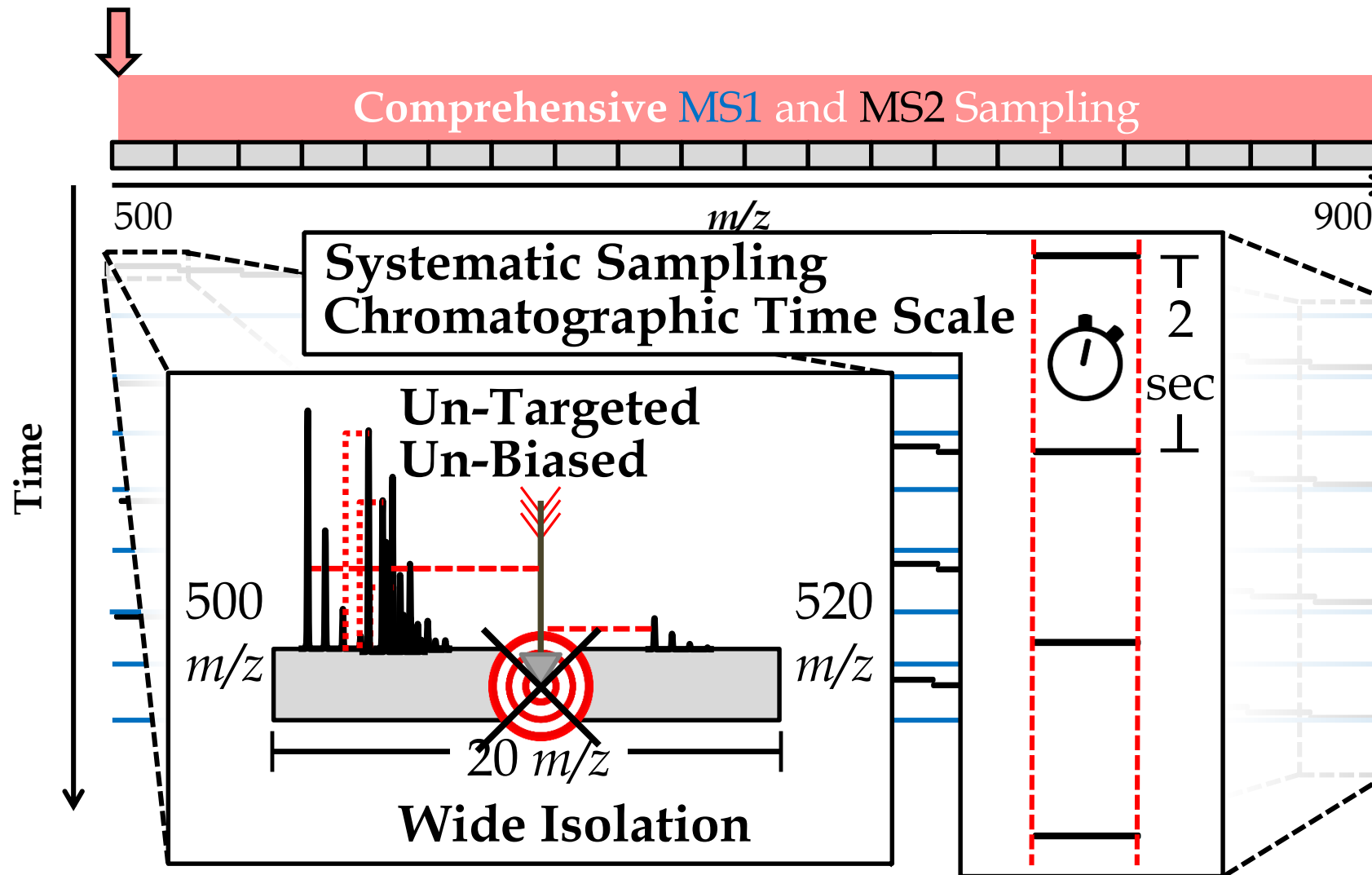## VLENTFEIGSDSIFDK++ (790.4 *m/z*)

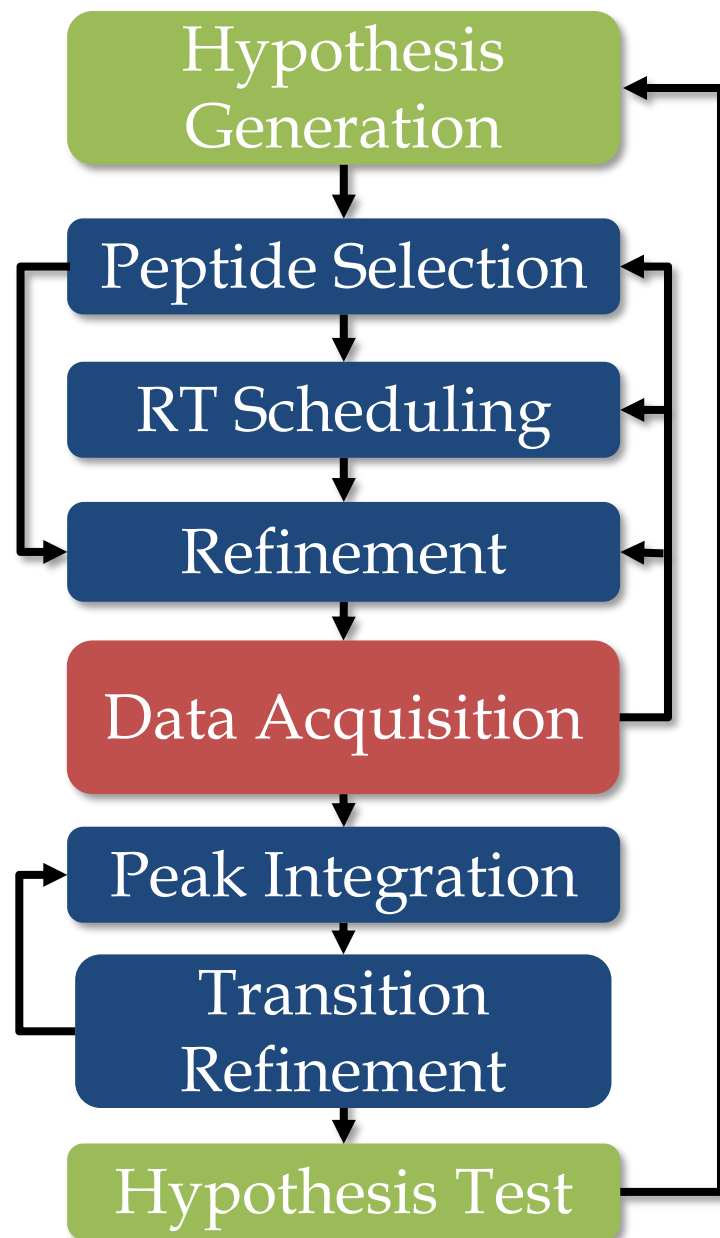# Extraction of Fragment Ions from DIA Data

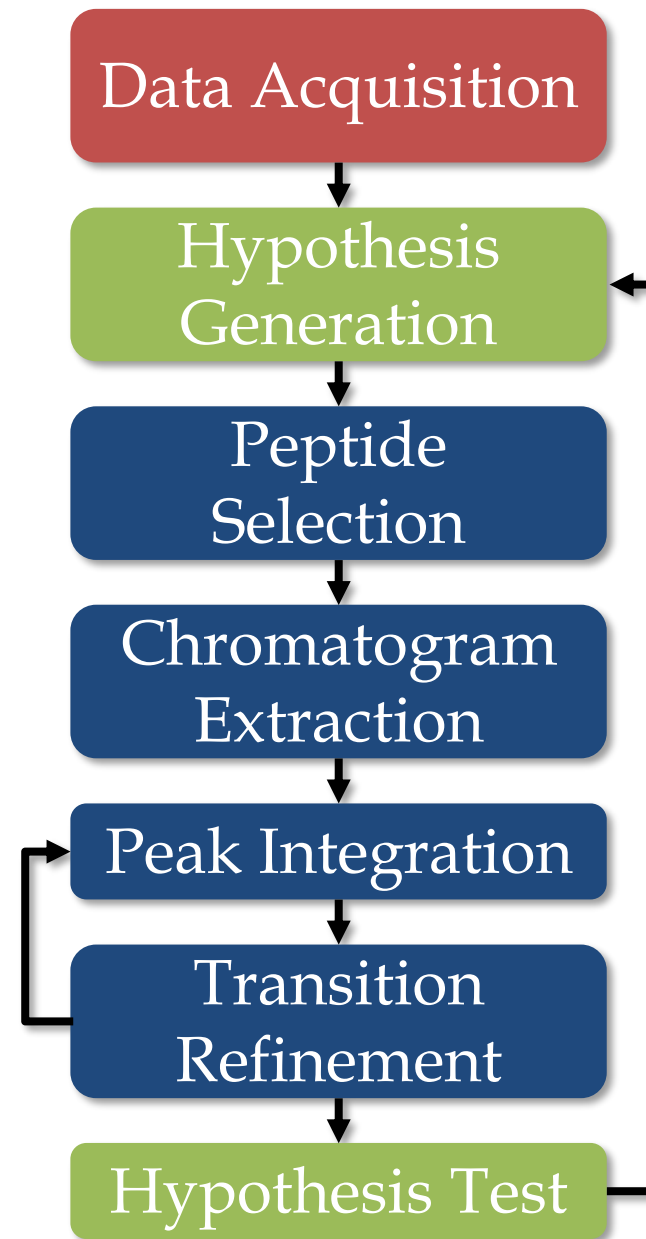

Venable et al Nat Methods 2004
Dong et al Science 2007

DIA Data Acquisition

Quantification via MS2 is much more accurate than MS1
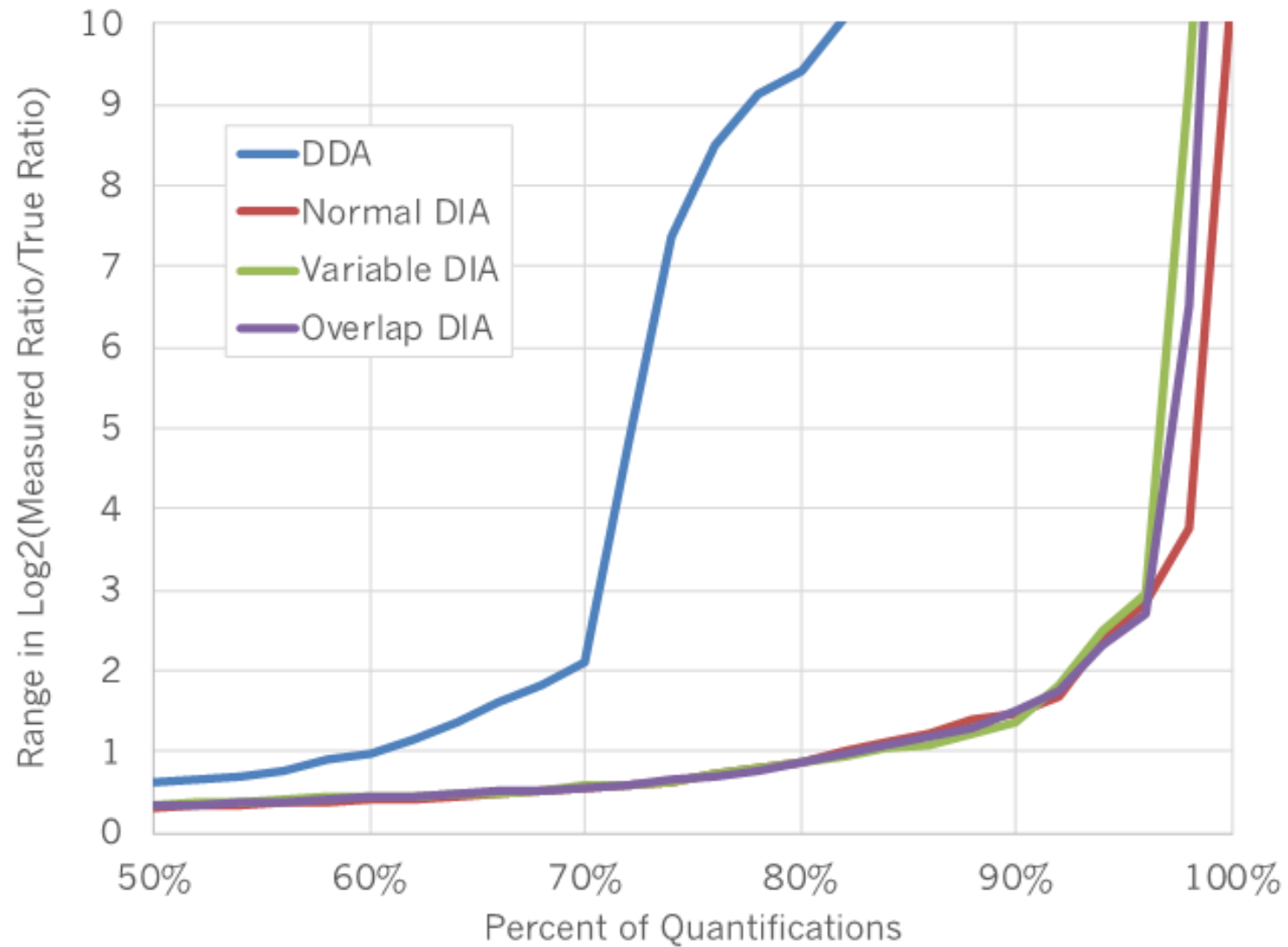
The Promise of DIA

- Acquire a "molecular image" or "digital archive" of the sample
  - Mine it over and over
  - No scheduling
- Direct queries (and p-values) for peptides of interest
- Better quantitation than DDA
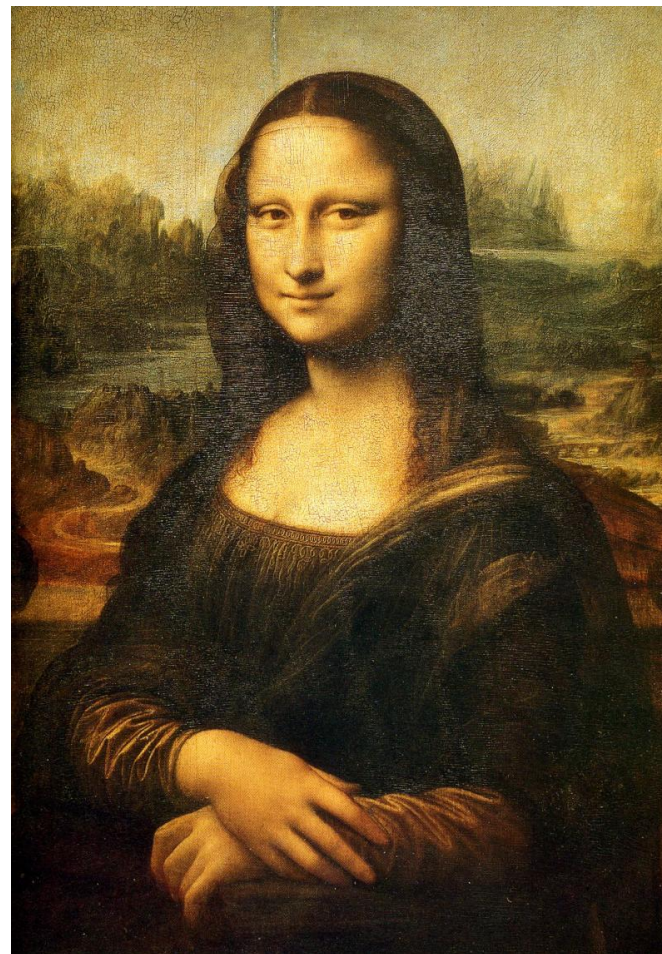
Donald Trump   Barack Obama   Hillary Clinton   Melania Trump   Michelle Obama   SCOTUS   J. Chief of Staff   Jimmy Carter   Bill Clinton   George W. Bush   Dan Quayle

Tweet

Needs for Quantitative Proteomics

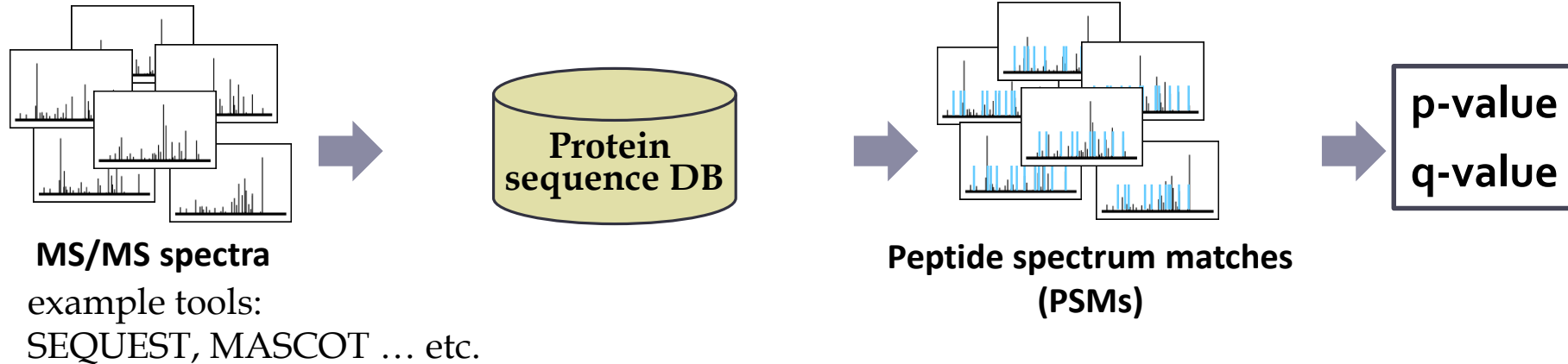- Are we really _identifying_ peptides? Focus on detectable peptides and quantifiable peptides

- Good selectivity

- We want as good or better peptide detections than DDA

- No missing data across many runs
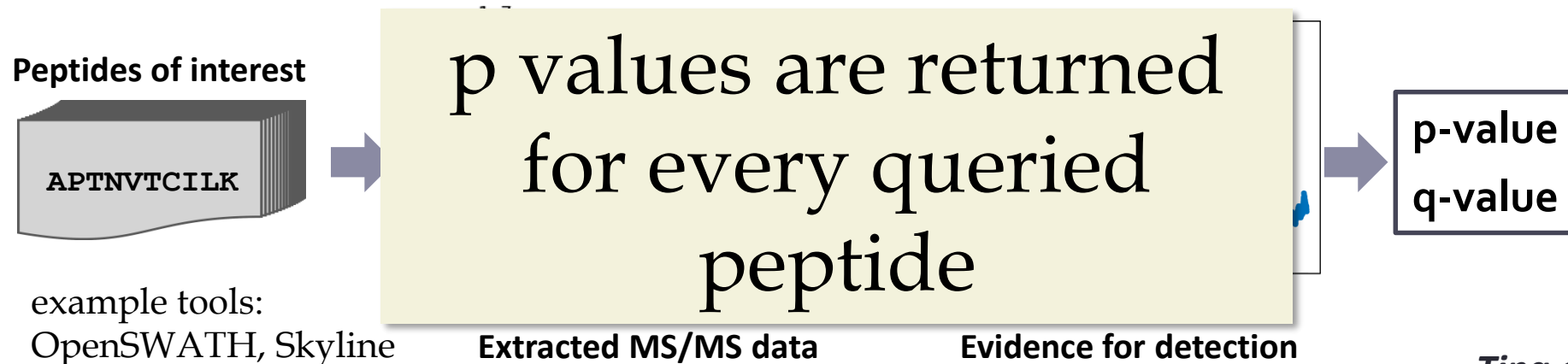
Detecting Peptides from DIA Data
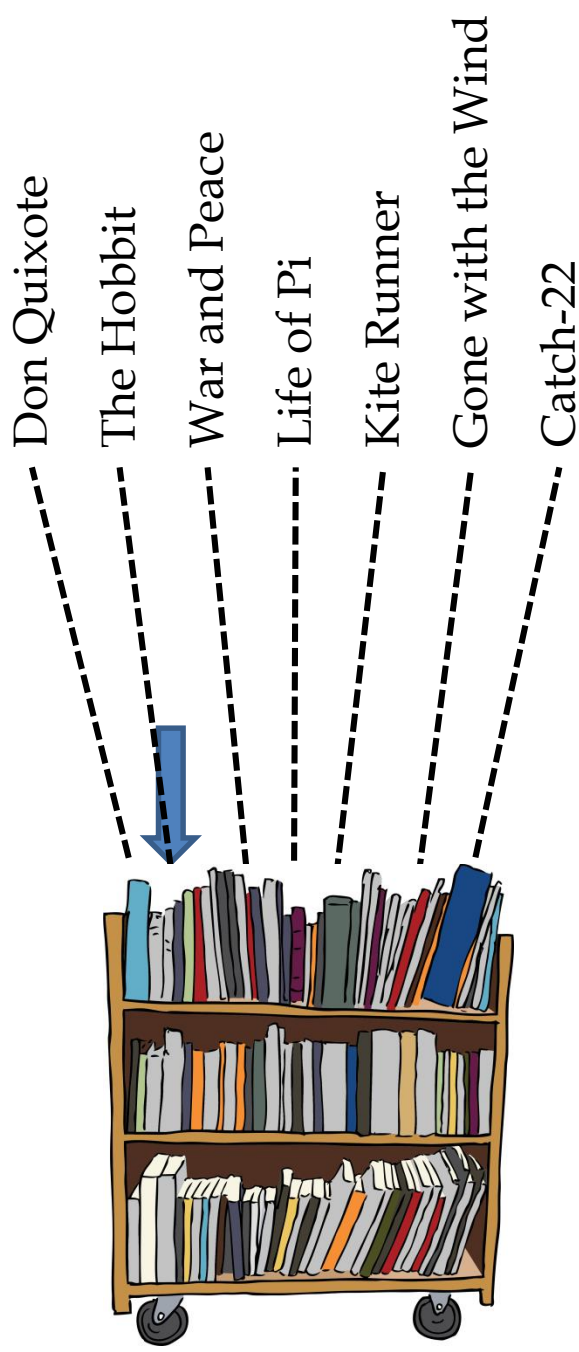
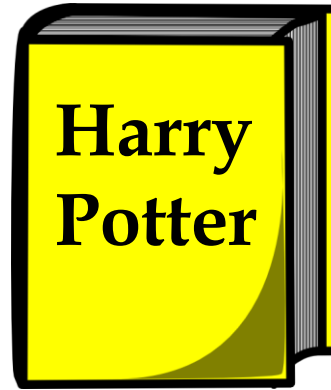**Spectrum-centric analysis** **What peptides best explain the data?**

MS/MS spectra

example tools:
SEQUEST, MASCOT ... etc.

Peptide spectrum matches (PSMs)

p-value
q-value

Peptide-centric analysis **Which peptides are detected in our data?**

Peptides of interest

APTNVTCILK

p values are returned for every queried peptide

example tools:
OpenSWATH, Skyline

Extracted MS/MS data   **Evidence for detection**

p-value
q-value

Ting et al, MCP, 2015

Don Quixote

The Hobbit

War and Peace

Life of Pi

Kite Runner

Gone with the Wind

Catch-22

Harry Potter

Conclusion: Harry Potter was not identified, so the library does not have it.

Harry Potter

Conclusion: Harry Potter was detected, the library has it

PECAN: Peptide Detection Directly from DIA Data

Peptides of interest

APTNVTCILK

DIA

Centroid .mzML

PECAN

1. Peptide processing / data extraction
2. Background estimation
3. Peak picking / feature g
4. Combine results / statis

Protein sequence DB

Skyline

2017 17
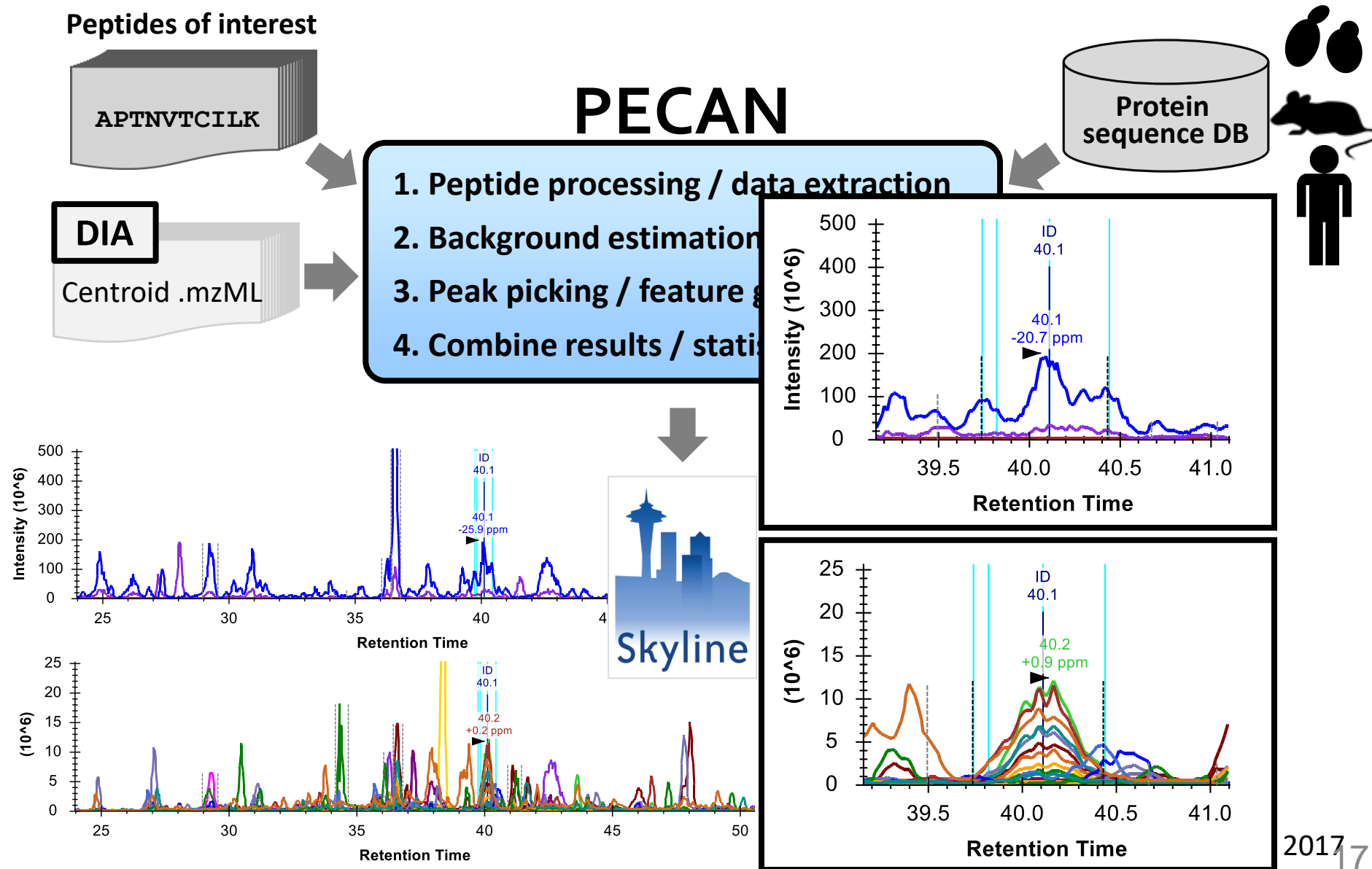
DIA is all about balance (and sacrifice)

500   **Precursor *m/z***   900

20 x **20** *m/z*-wide windows

40 x **10** *m/z*-wide windows

80 x **5** *m/z*-wide windows

2 injections of
40 x **5** *m/z*-wide windows
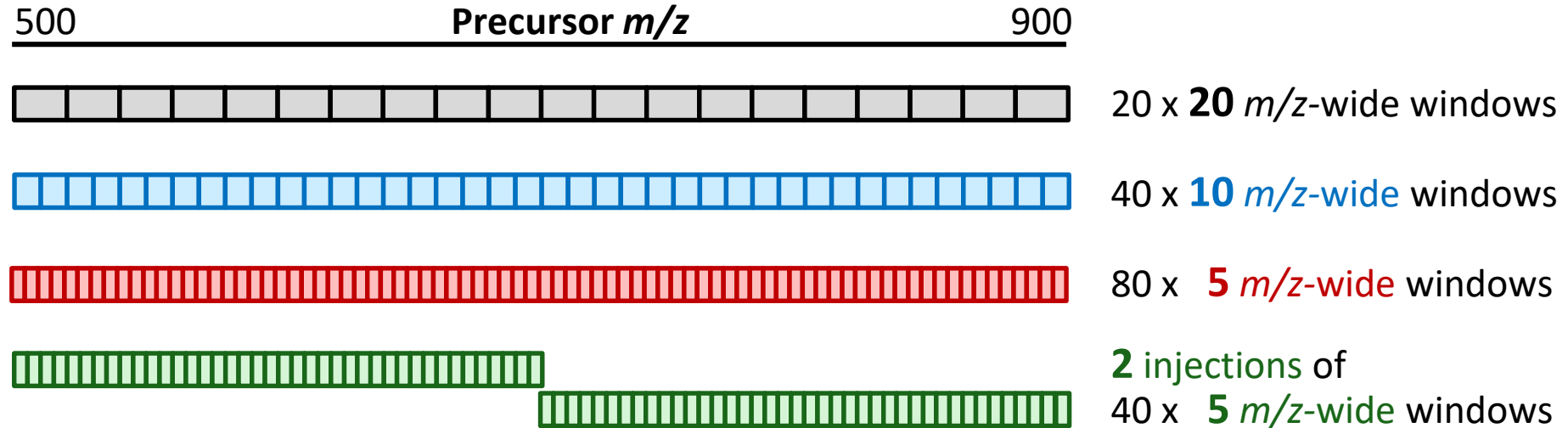
Instrument MS/MS scan speed: 10 Hz

30 seconds retention time   30 seconds retention time   30 seconds retention time   30 seconds retention time
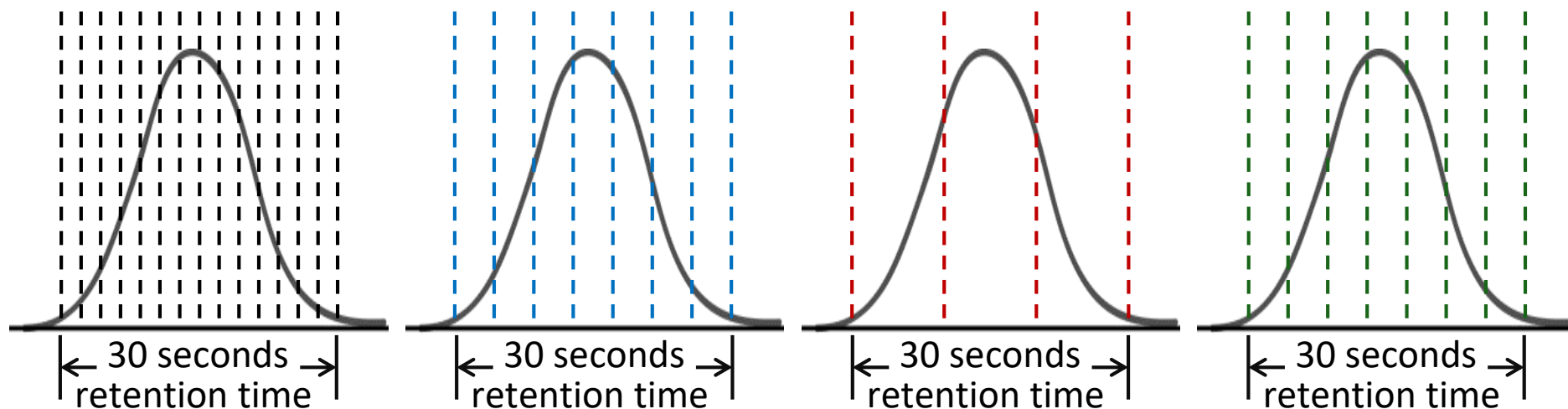
Archive Quality

20 x **20** *m/z*-wide windows

30 x **10** *m/z*-wide windows

2 injections of
40 x **5** *m/z*-wide windows

Problem: Detection of Peptides in DIA Data is Inversely Proportional to Isolation Width
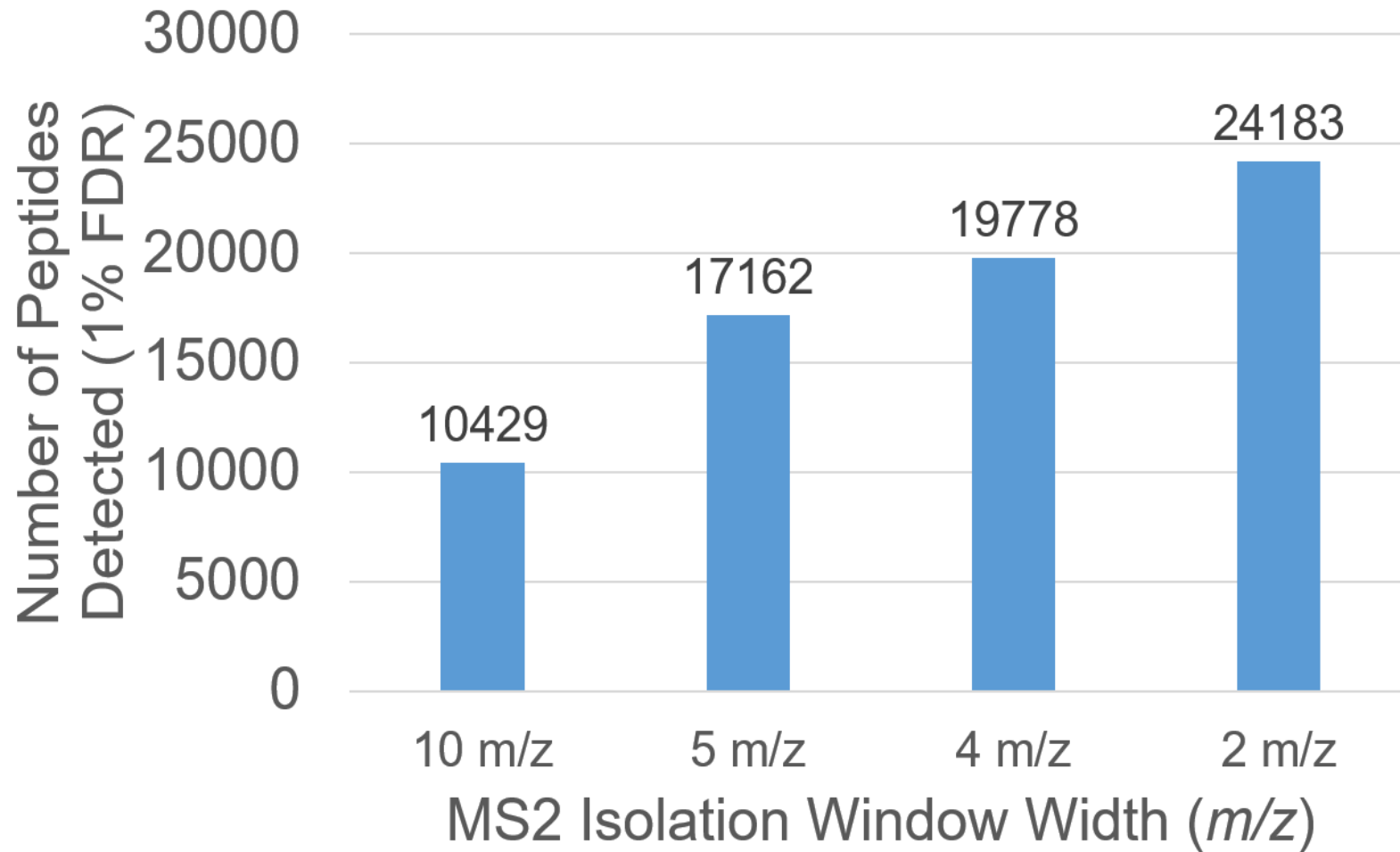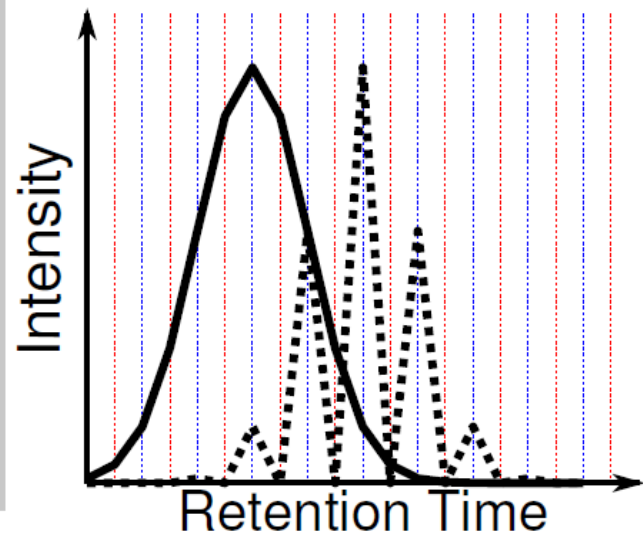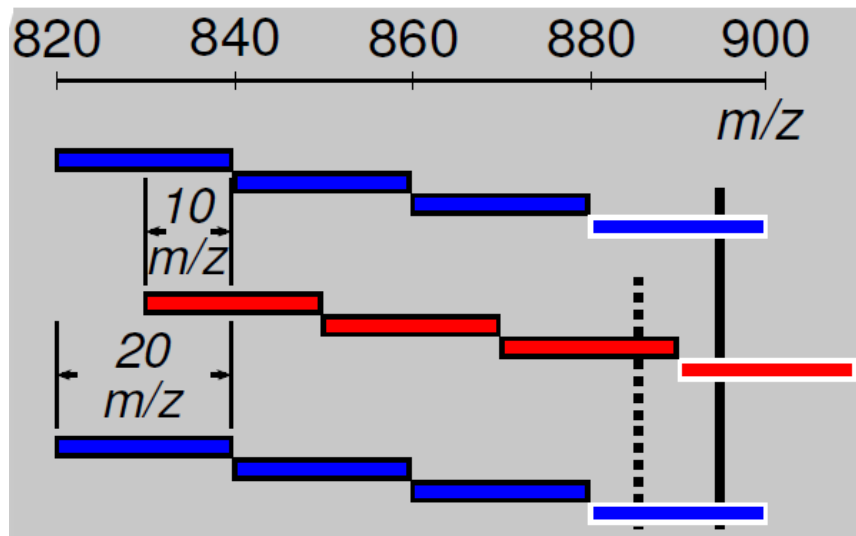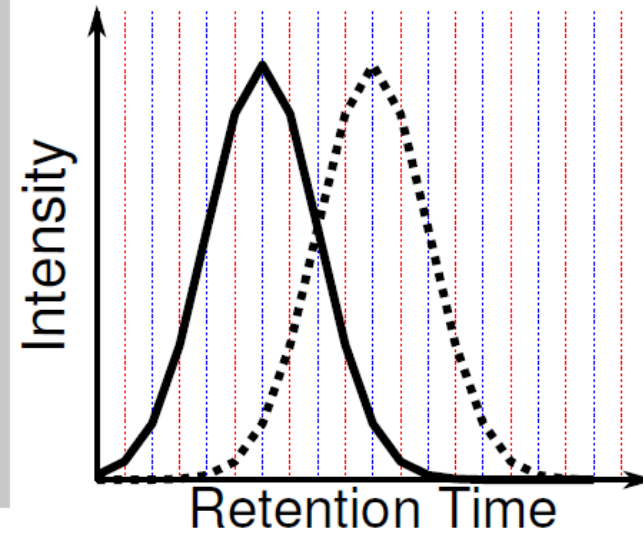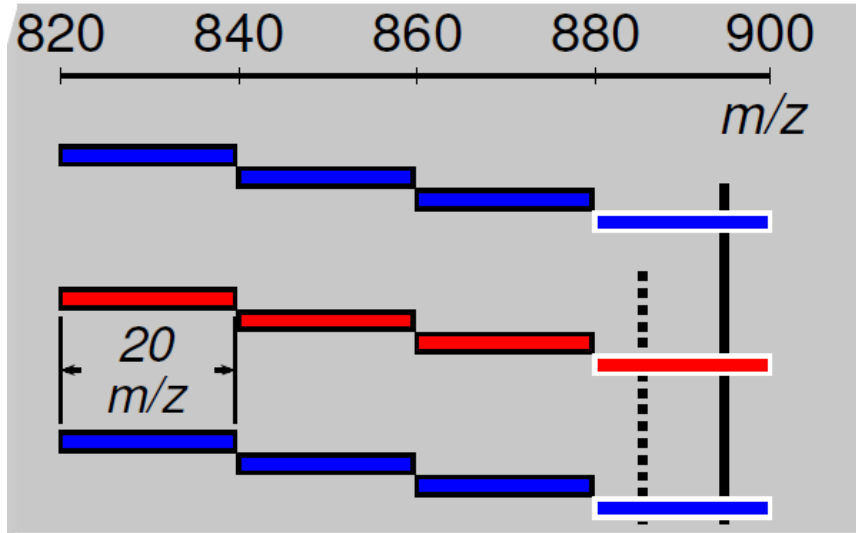
Needs for Quantitative Proteomics

- **Are we really identifying peptides? Focus on detectable peptides and quantifiable peptides**

- **Good selectivity**

- **We want as good or better peptide detections than DDA**
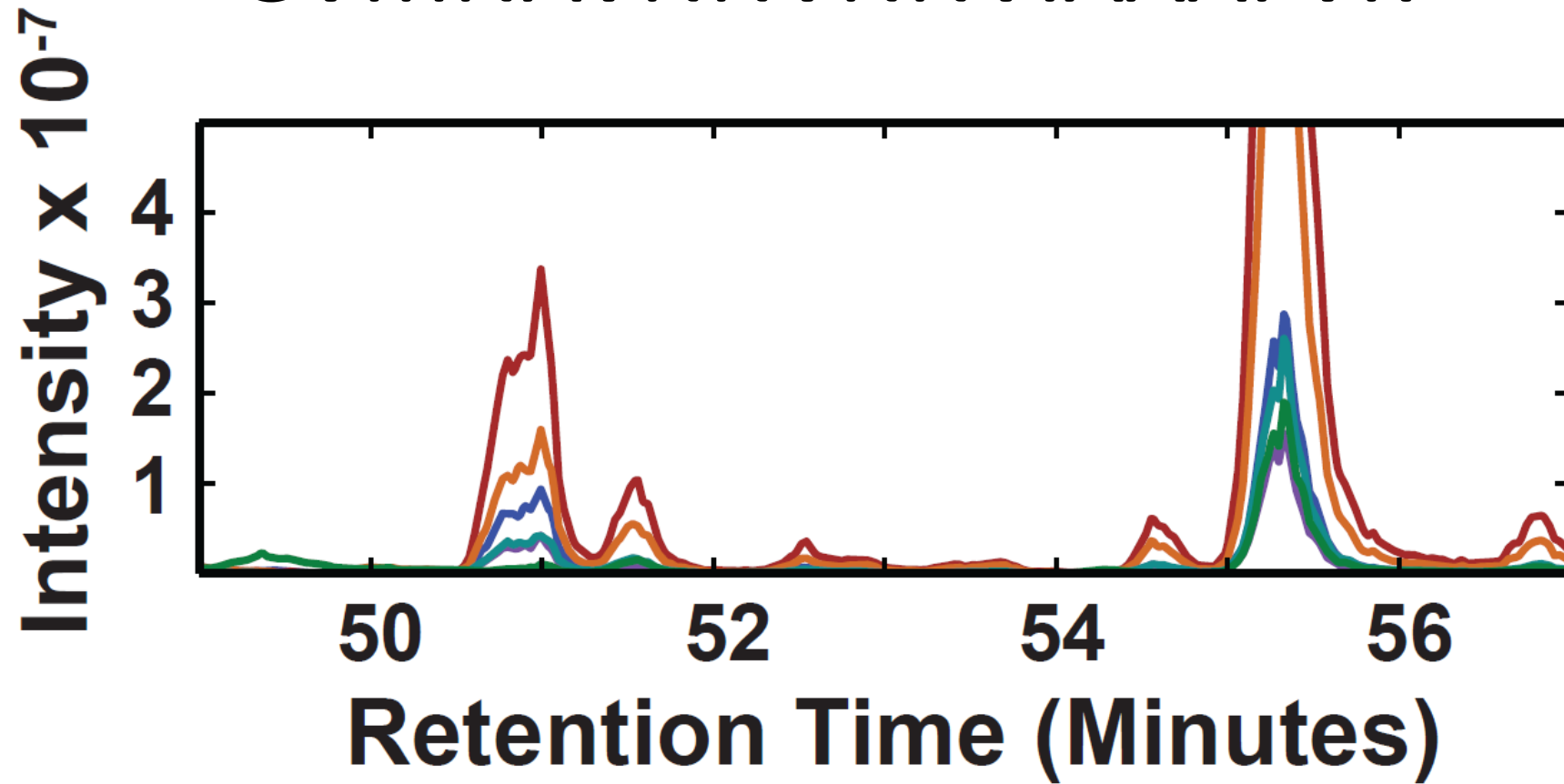
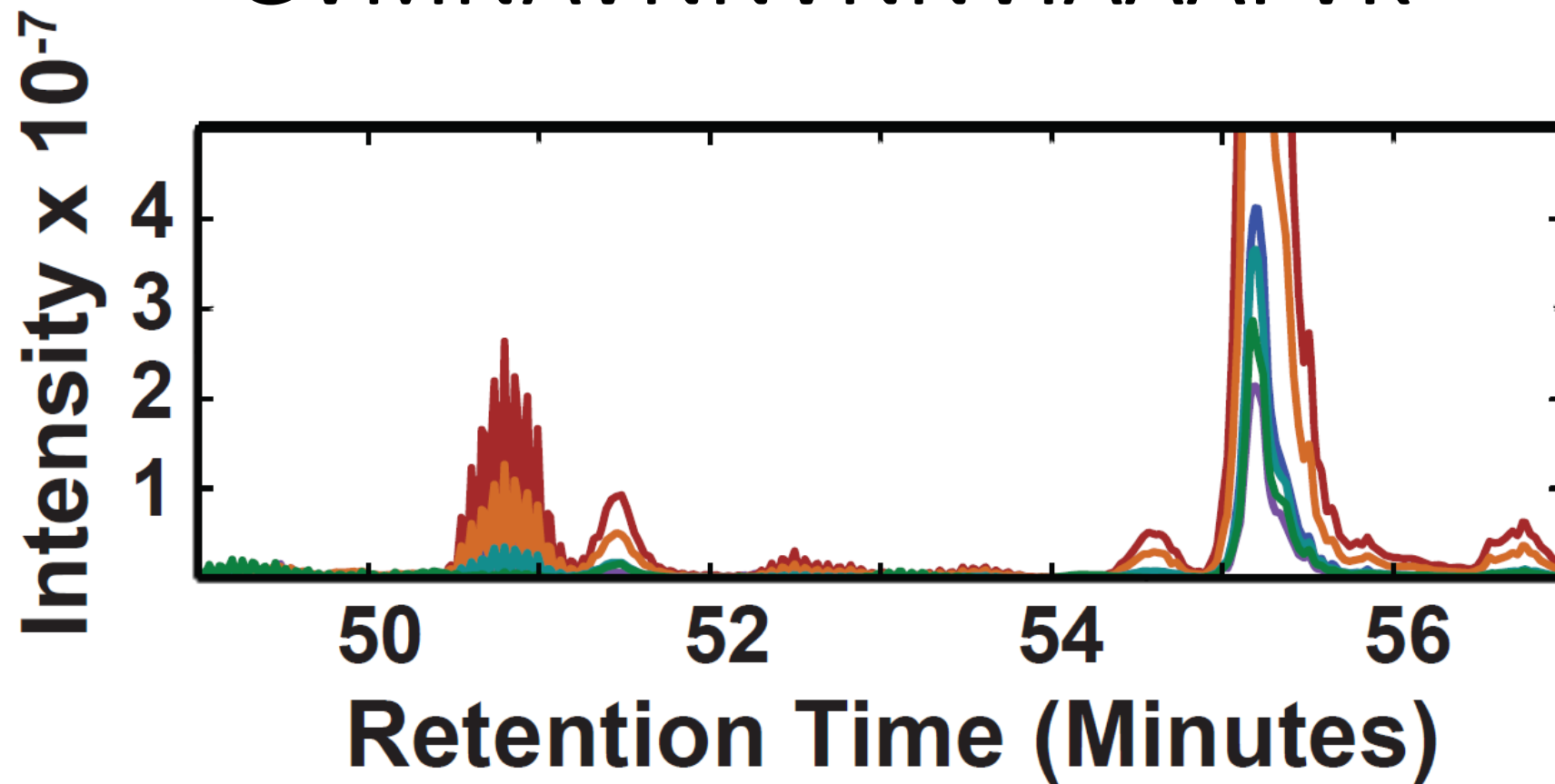- **No missing data across many runs**
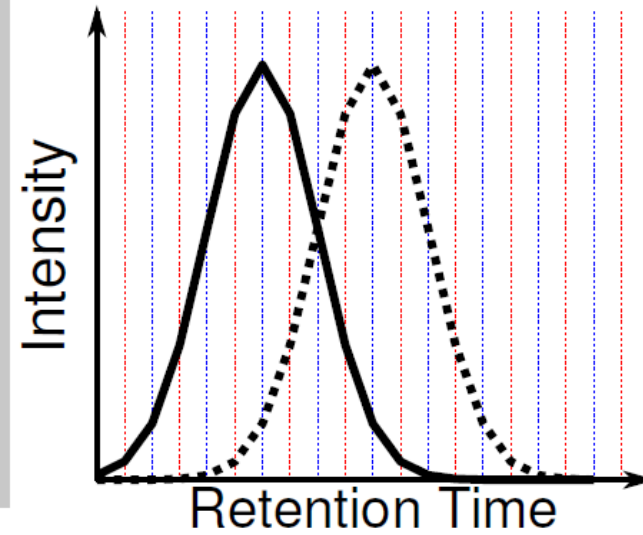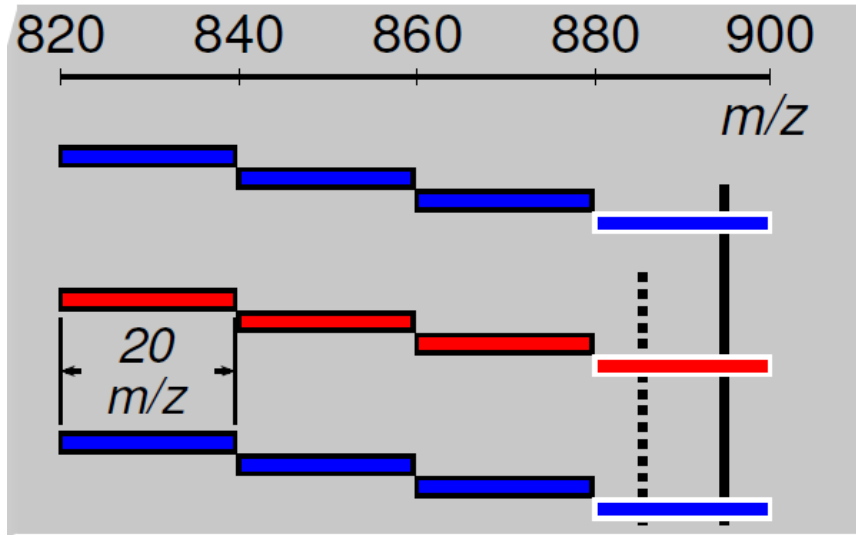
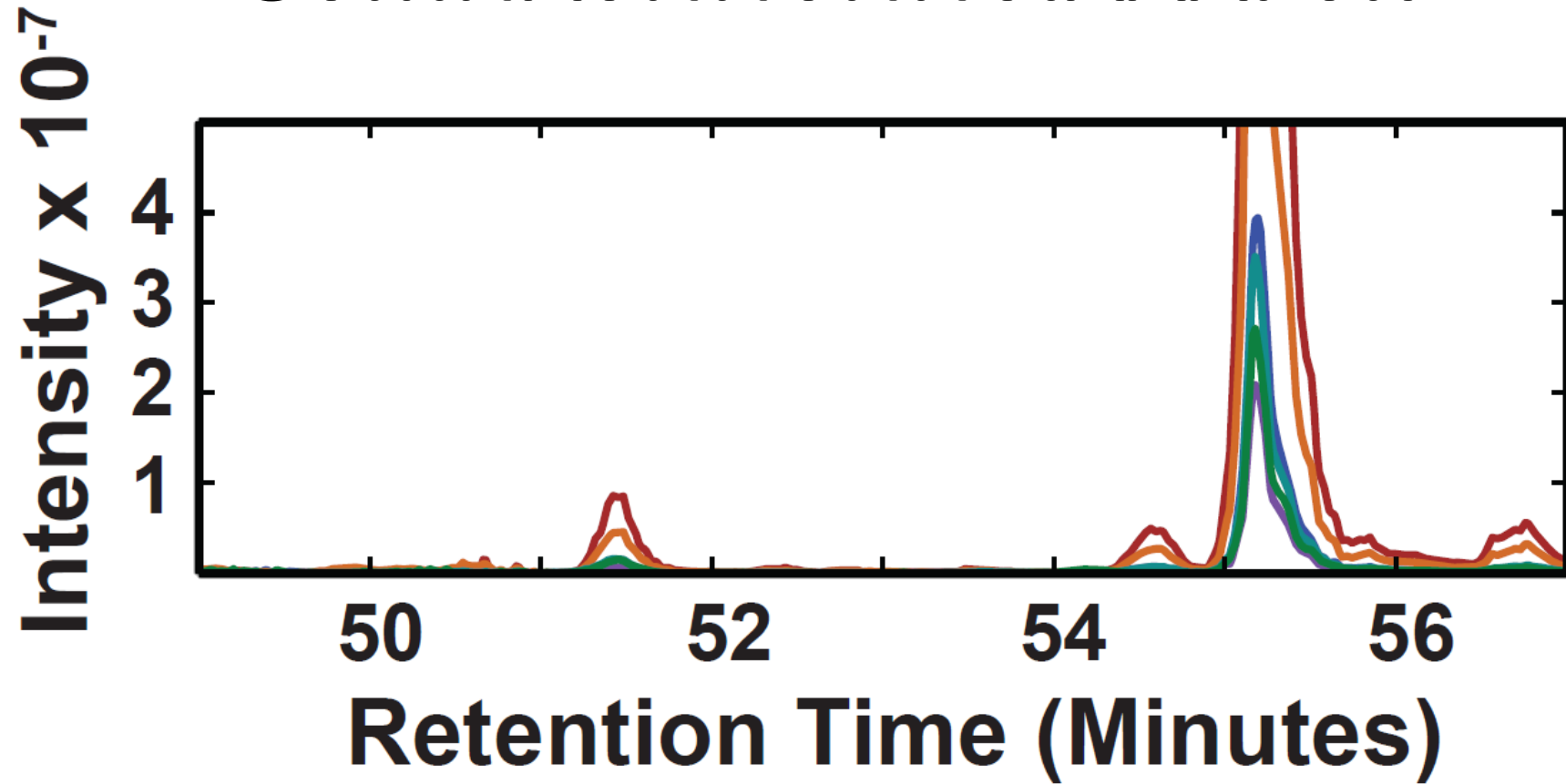# 20 *m/z* DIA

GVMNAVNNVNNVIAAAFVK

20 *m/z* DIA + Overlap

GVMNAVNNVNNVIAAAFVK

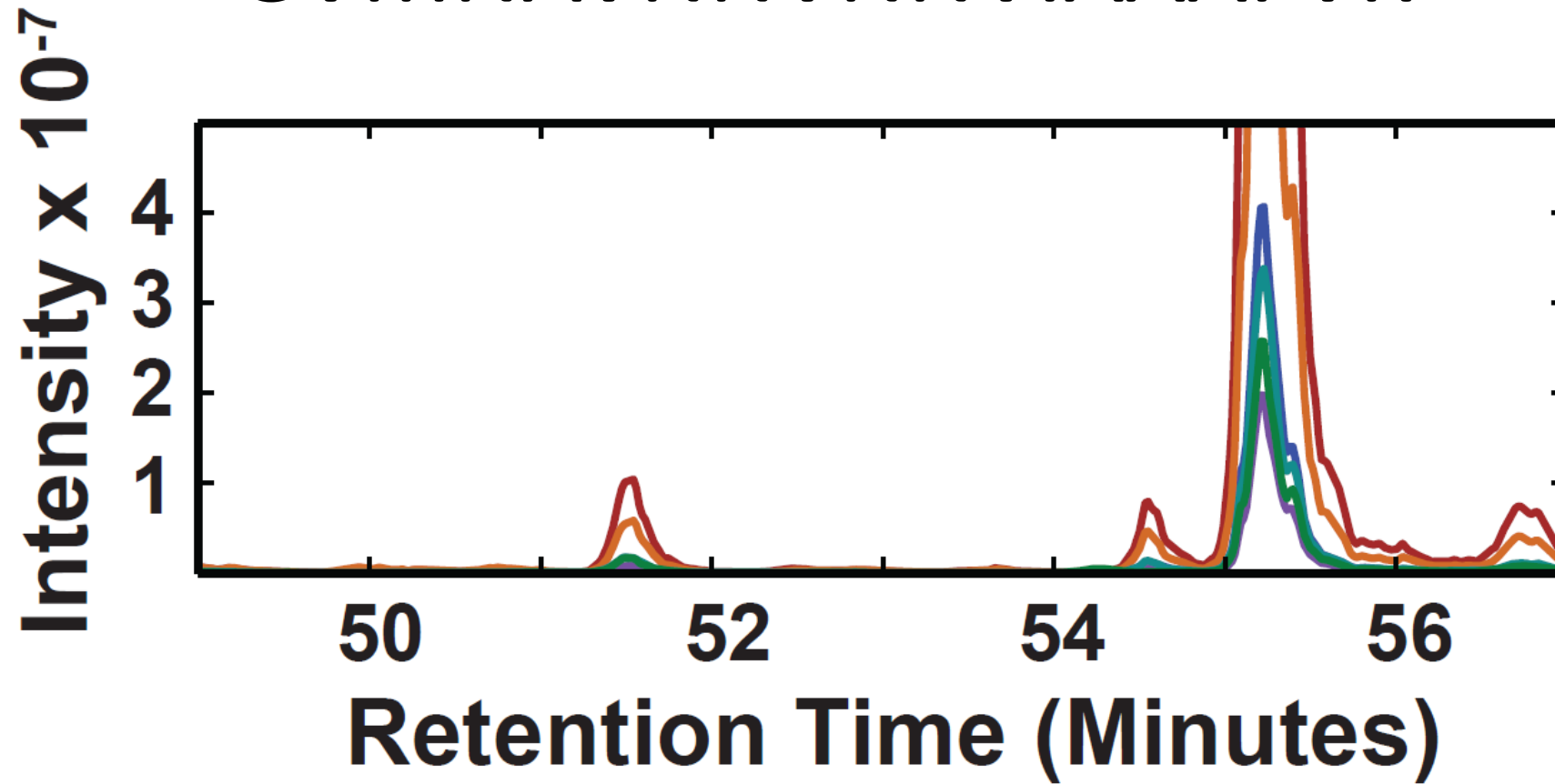# 20 *m/z* DIA + Overlap + Demultiplexing

GVMNAVNNVNNVIAAAFVK

10 *m/z* DIA

GVMNAVNNVNNVIAAAFVK

No reason not to do overlap windows

http://proteowizard.sourceforge.net/
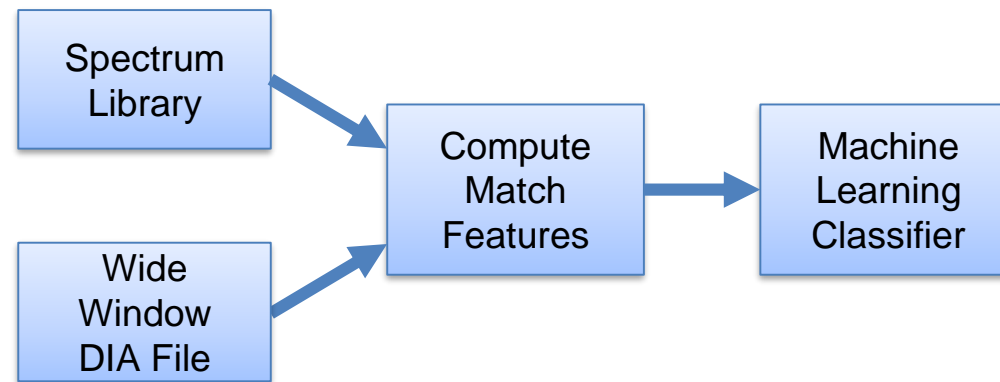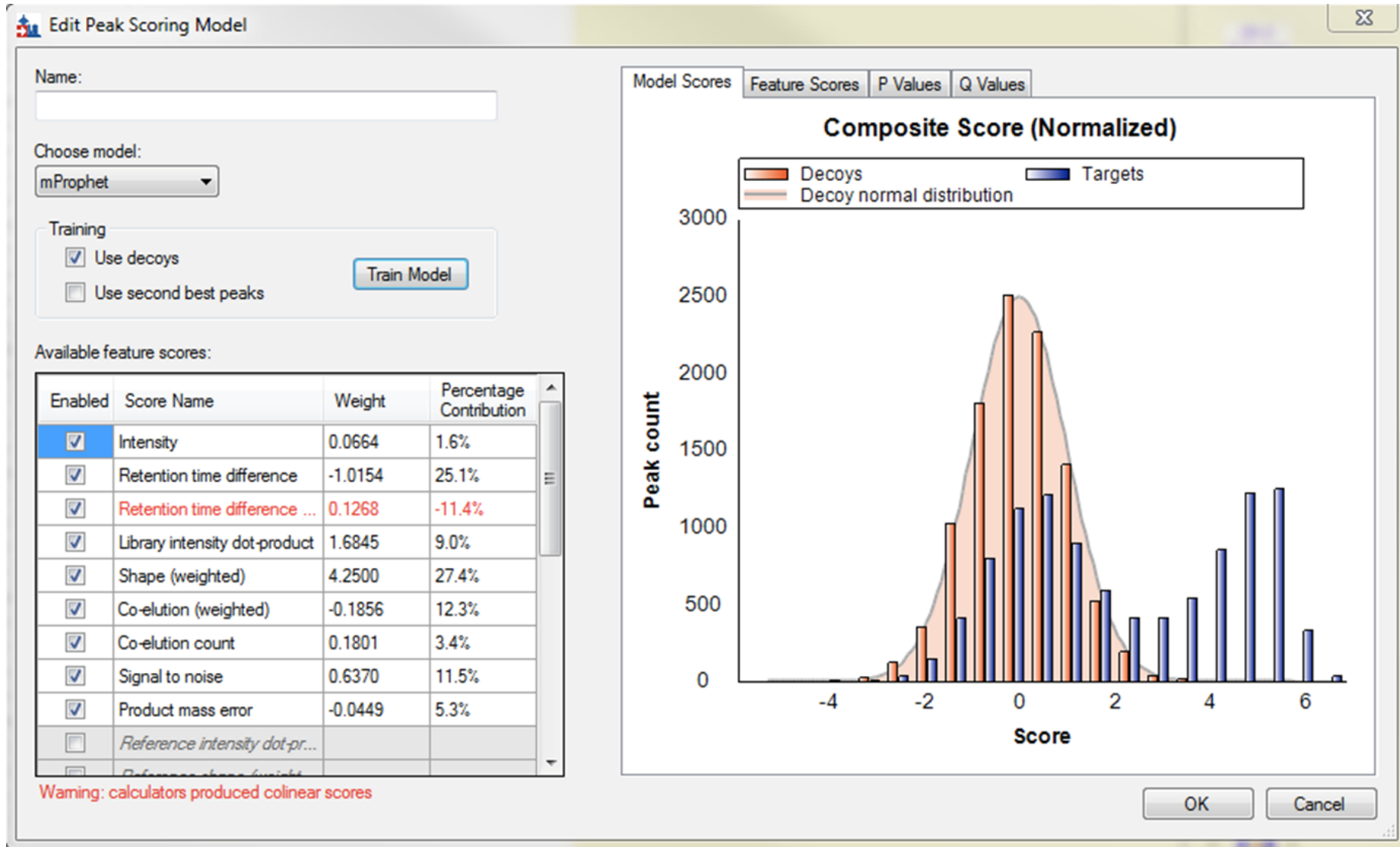
Needs for Quantitative Proteomics

- **Are we really identifying peptides? Focus on detectable peptides and quantifiable peptides**

- **Good selectivity**

- **We want as good or better peptide detections than DDA**

- **No missing data across many runs**

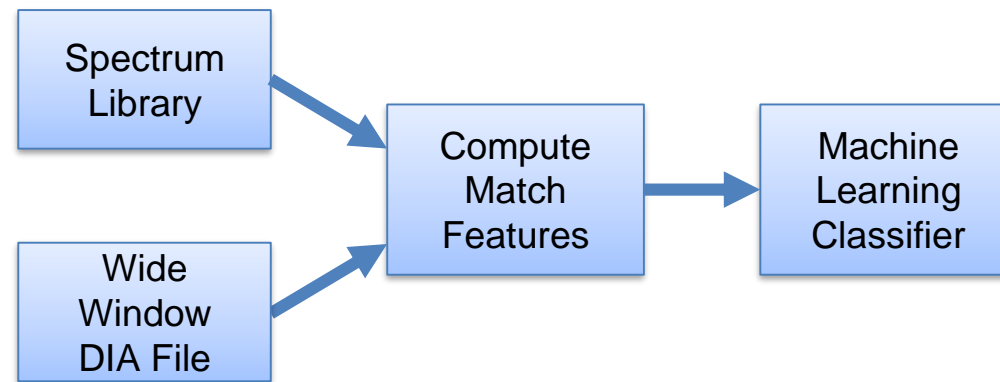# Peptide Detections: Typical DIA library search workflow

Spectrum Library → Compute Match Features

Wide Window DIA File → Compute Match Features → Machine Learning Classifier

Typical DIA library search

Peptide Detections: EnclyclopeDIA workflow

Percolator from Käll L et al, Nat Methods. 2007 Nov;4(11):923-5.
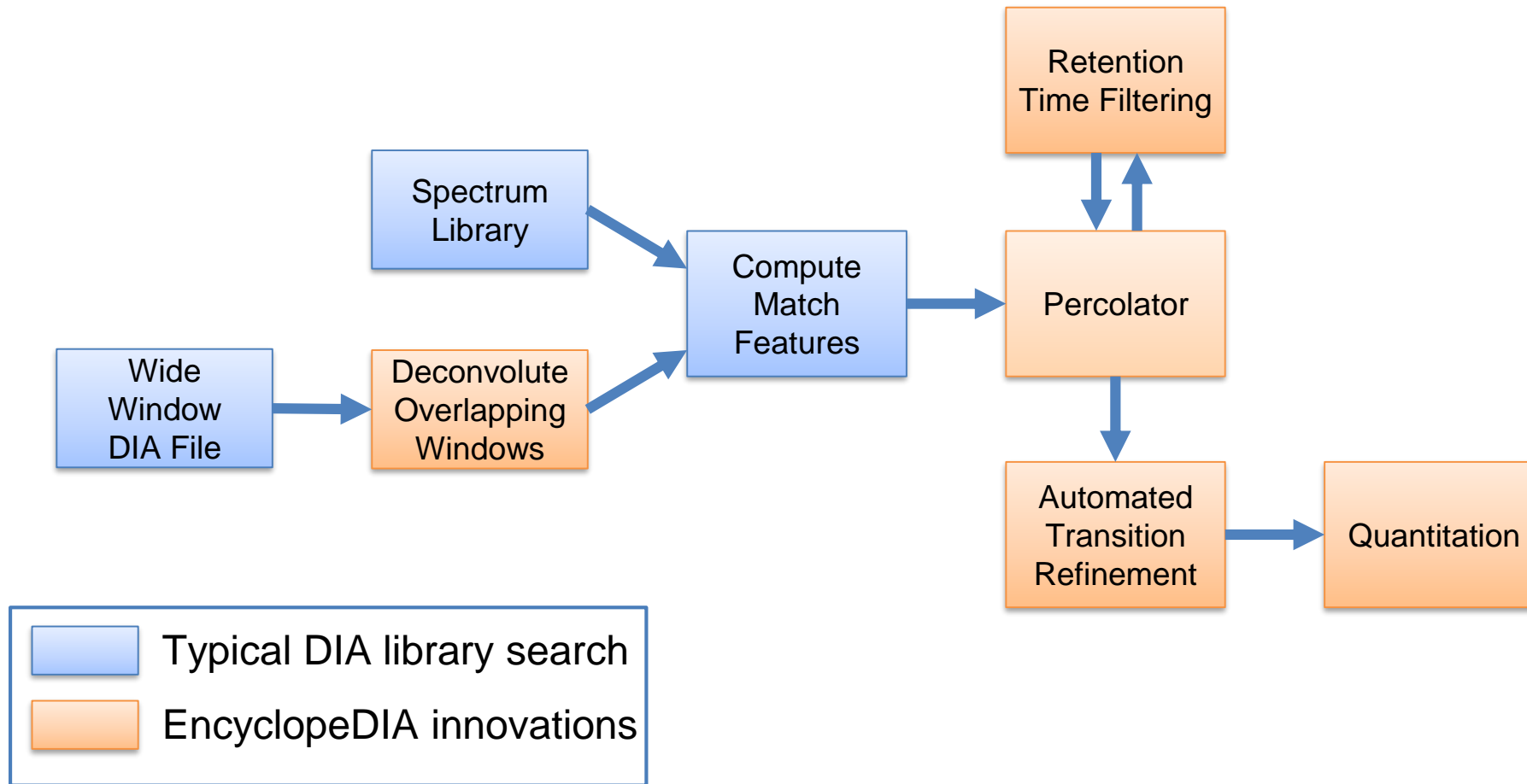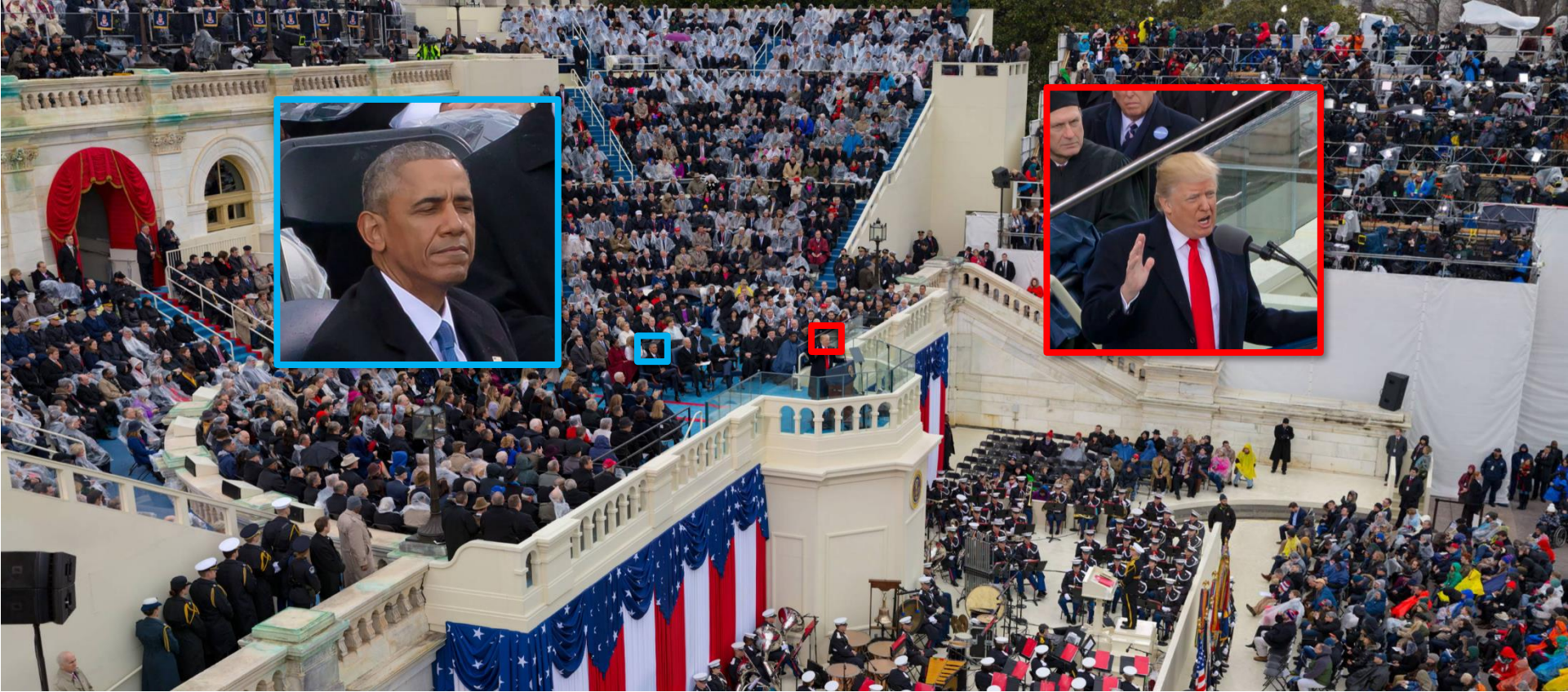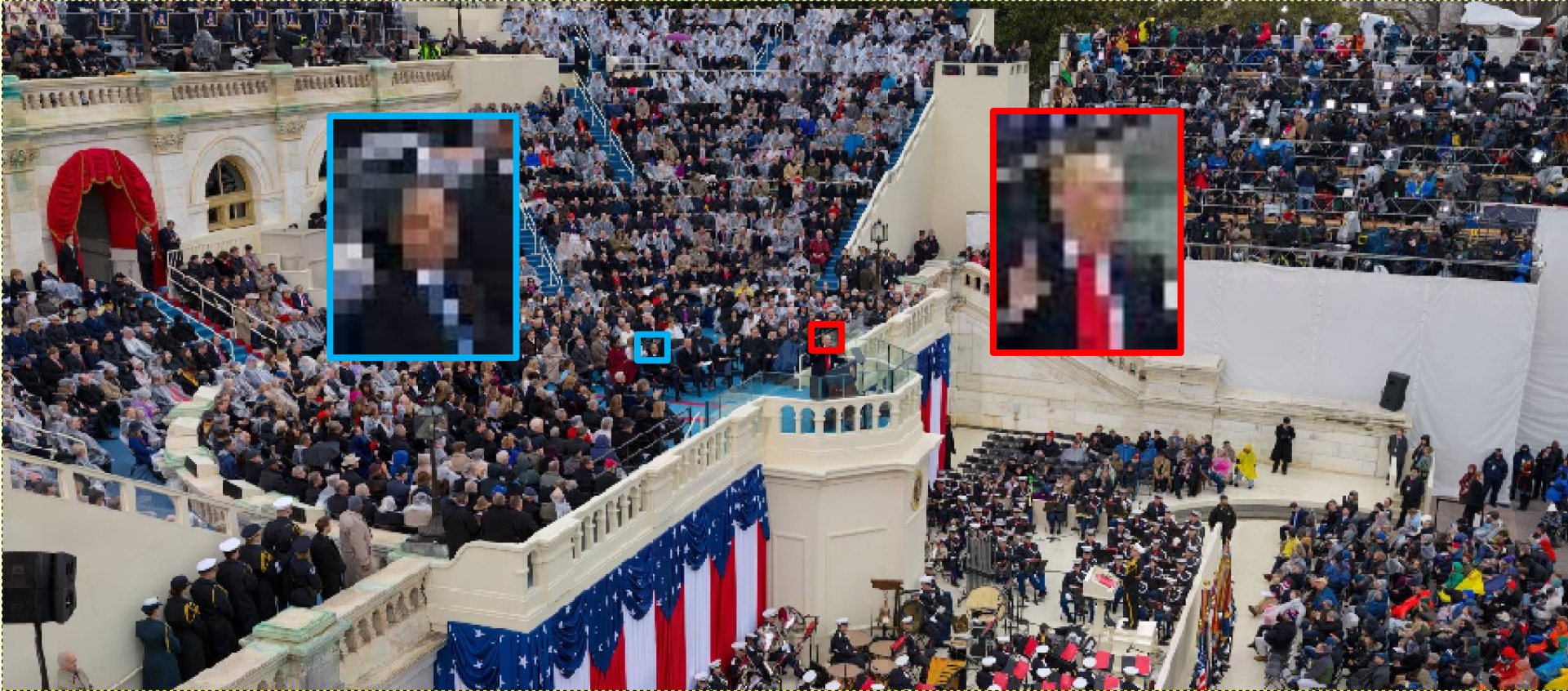
Spectrum Library Approach
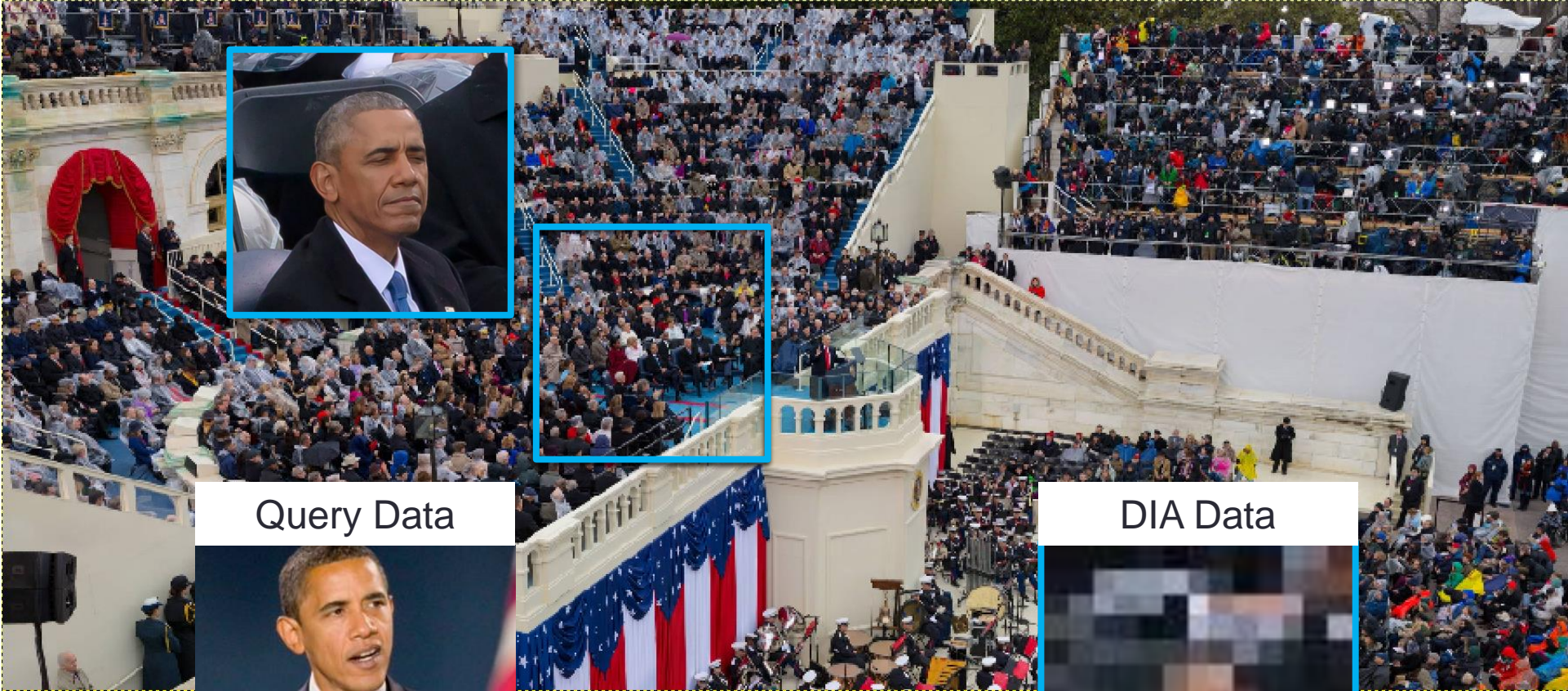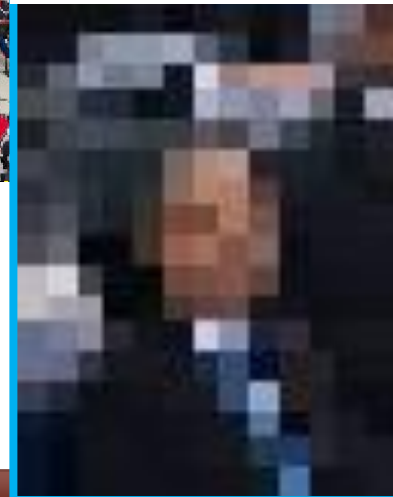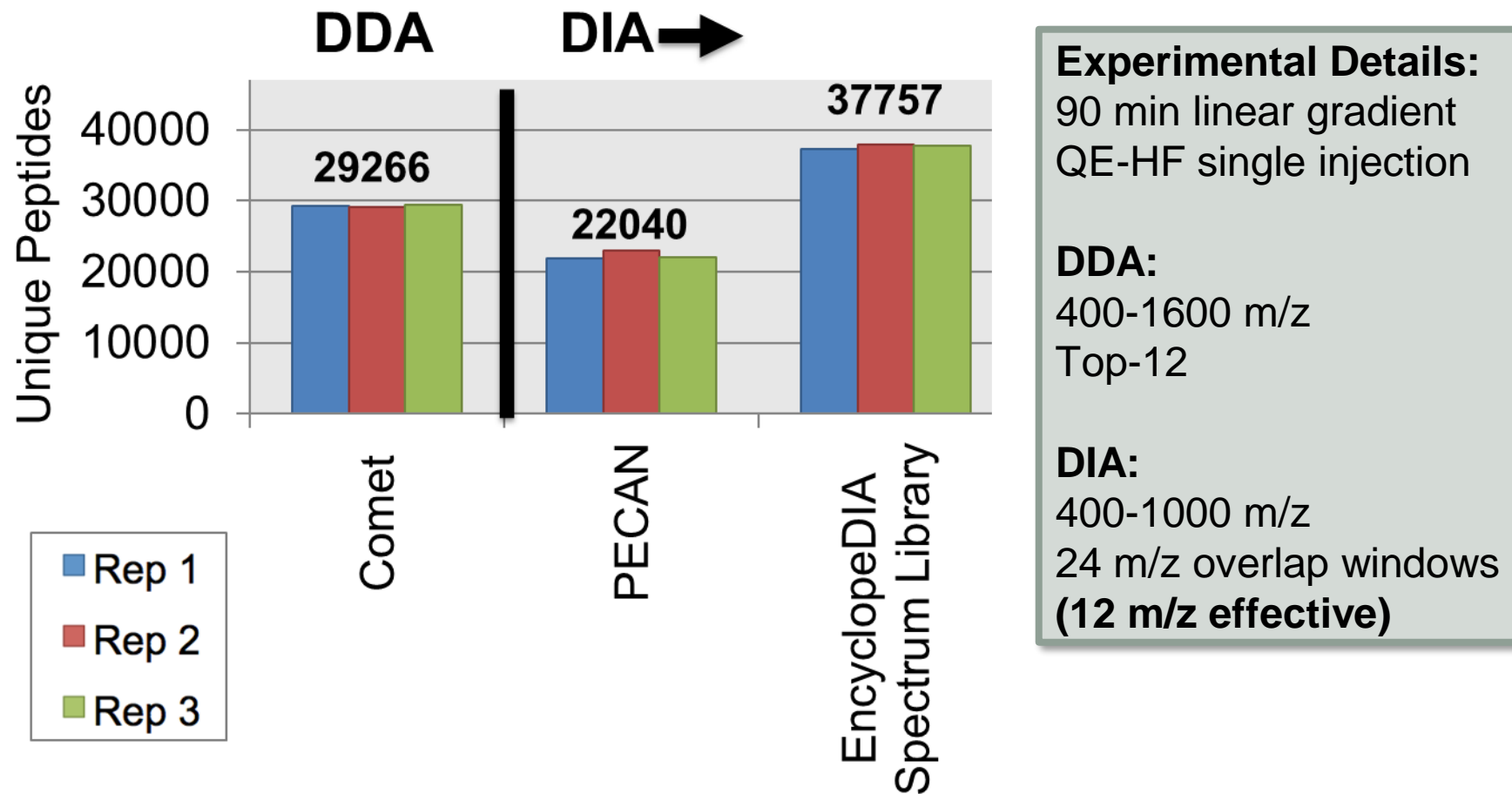

Query Data

DIA Data

DIA is a powerful tool for detecting peptides in HeLa

Experimental Details:
90 min linear gradient
QE-HF single injection

DDA:
400-1600 m/z
Top-12

DIA:
400-1000 m/z
24 m/z overlap windows
(12 m/z effective)

Retention Time Fit · Data Used In Fit · Data Removed From Fit
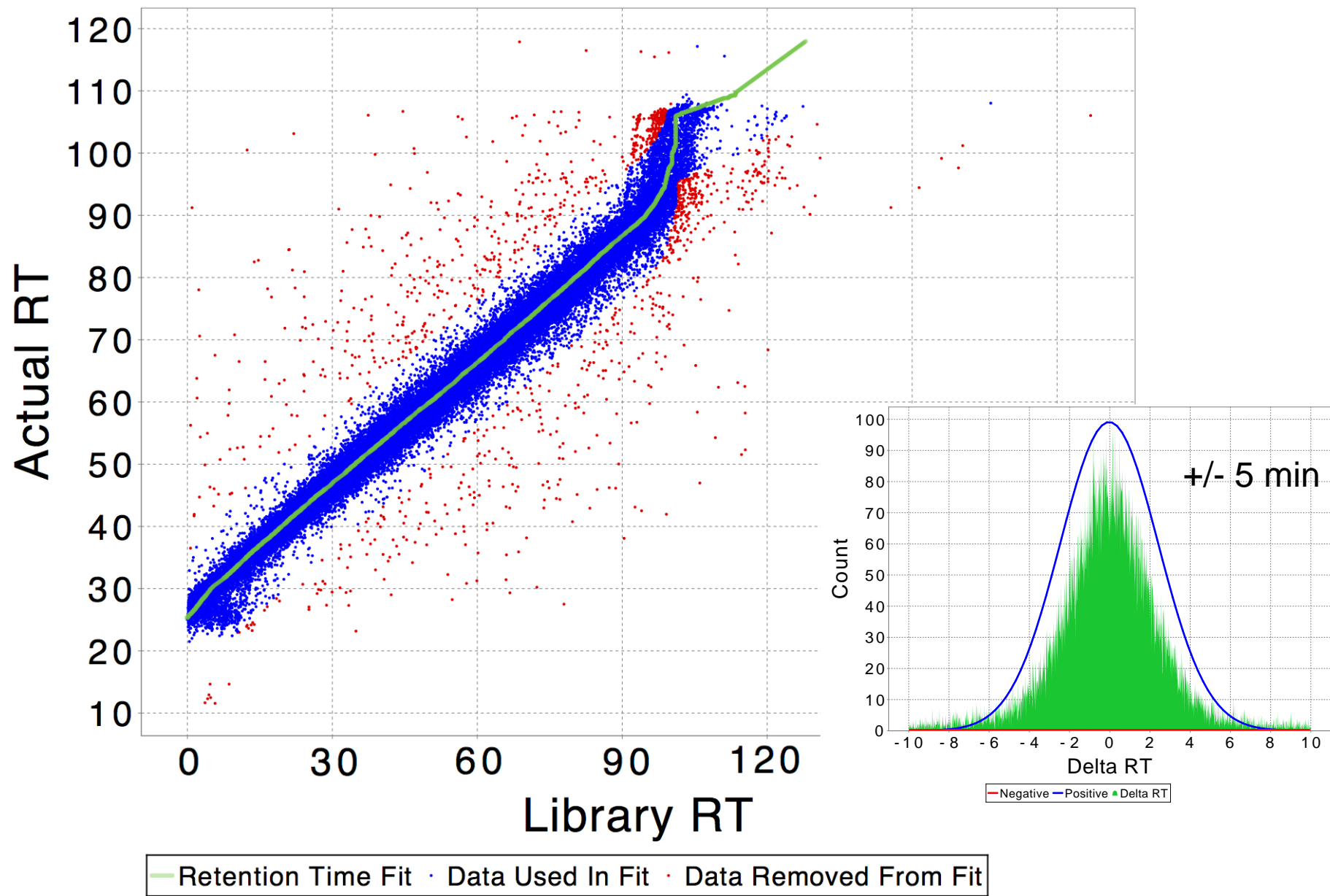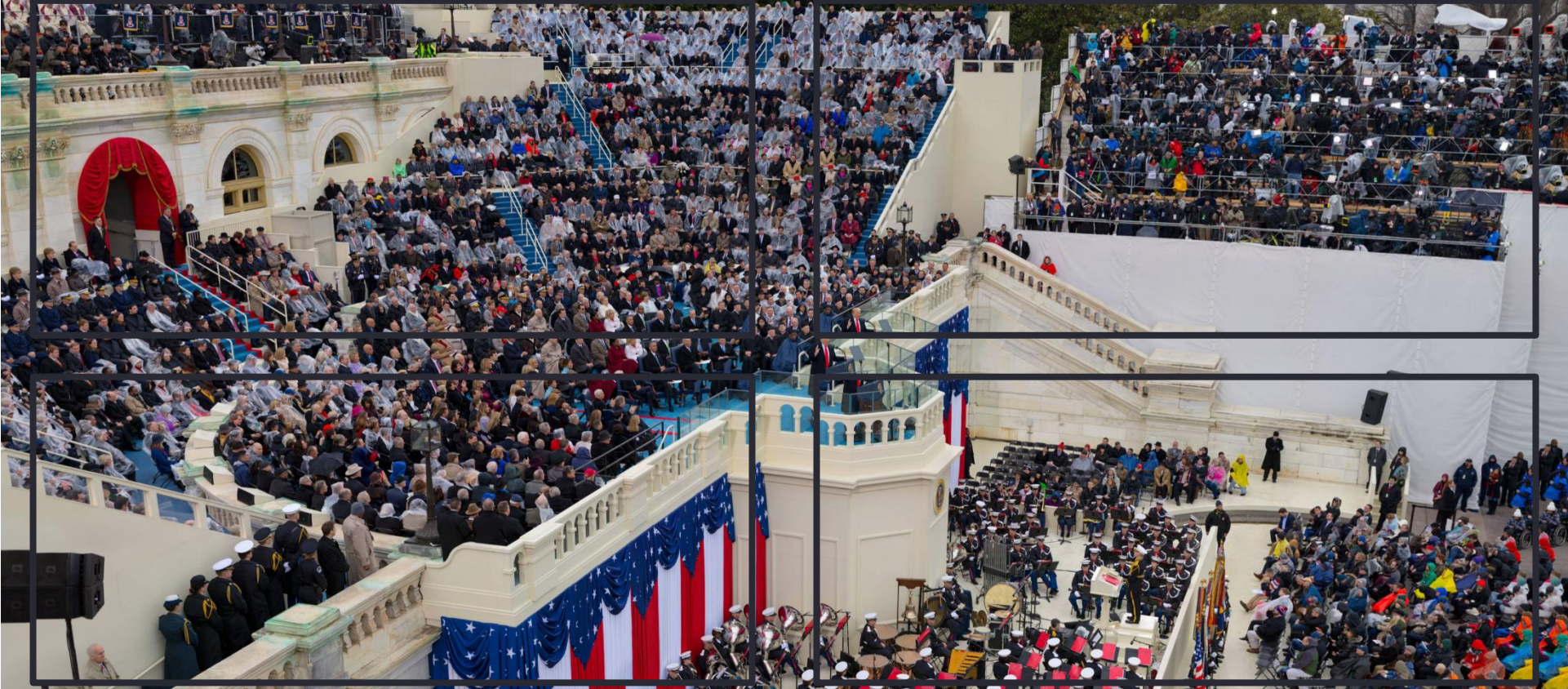
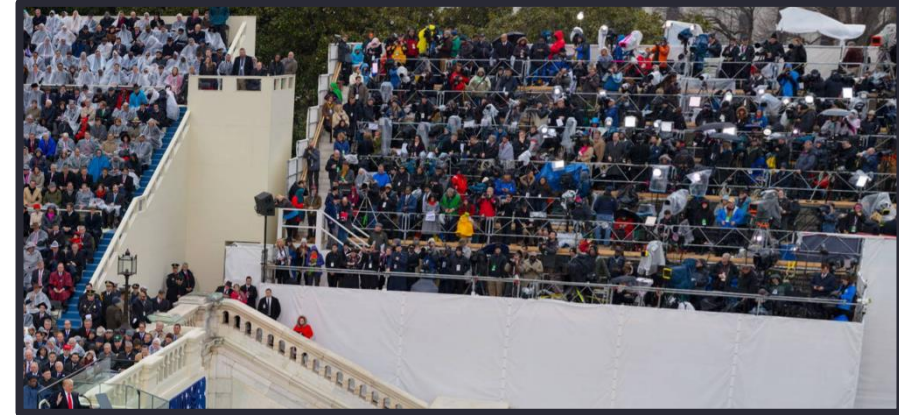# Separating Detection from Quantitation
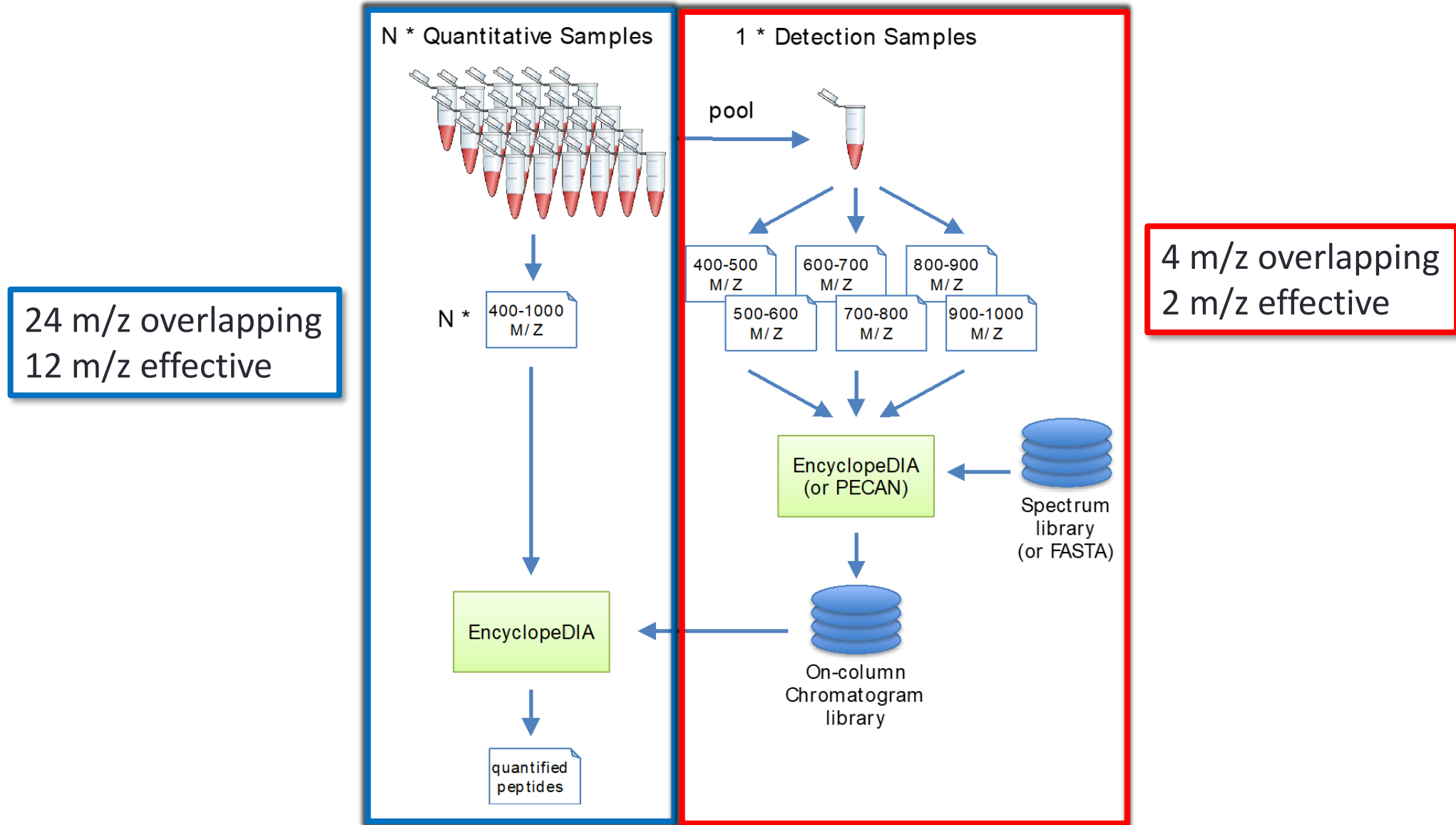
DIA Chromatogram Library Approach
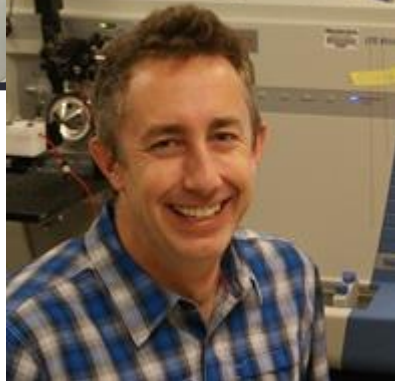

Query Data

DIA Data

Chromatogram Library
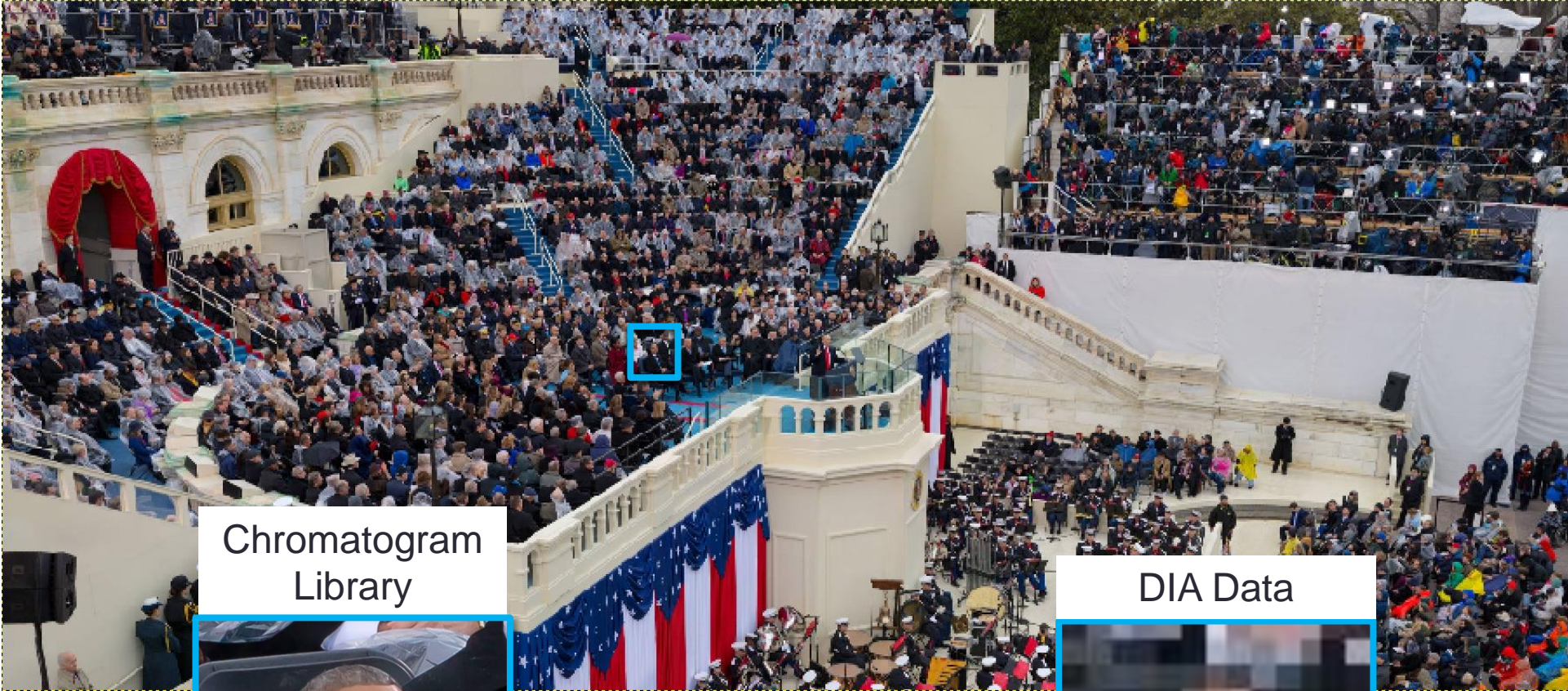
DIA Chromatogram Library Approach

Query Spectrum

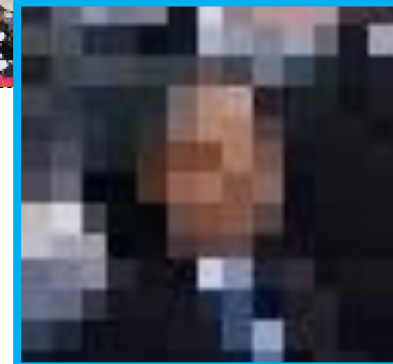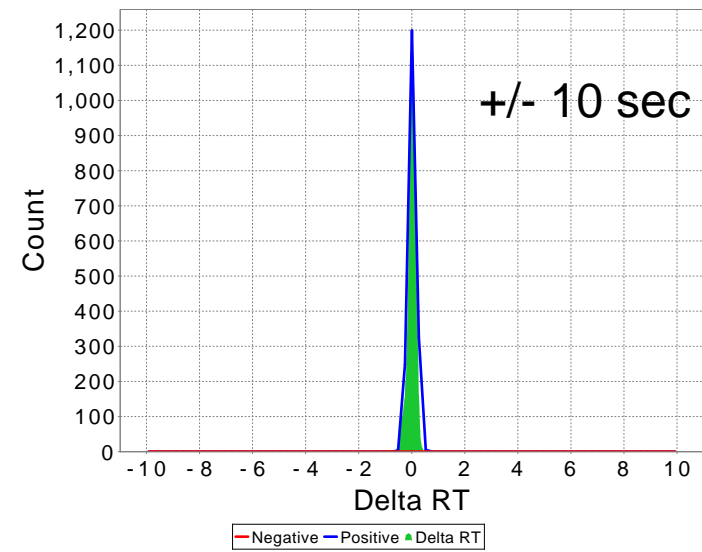= Not Found → Chromatogram Library
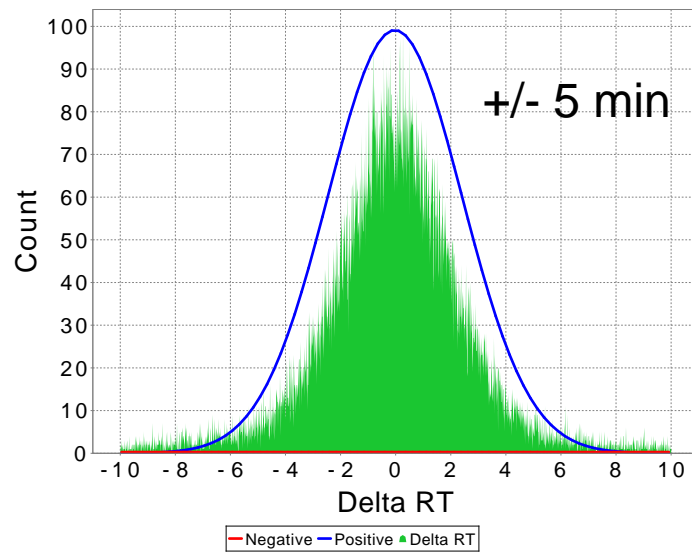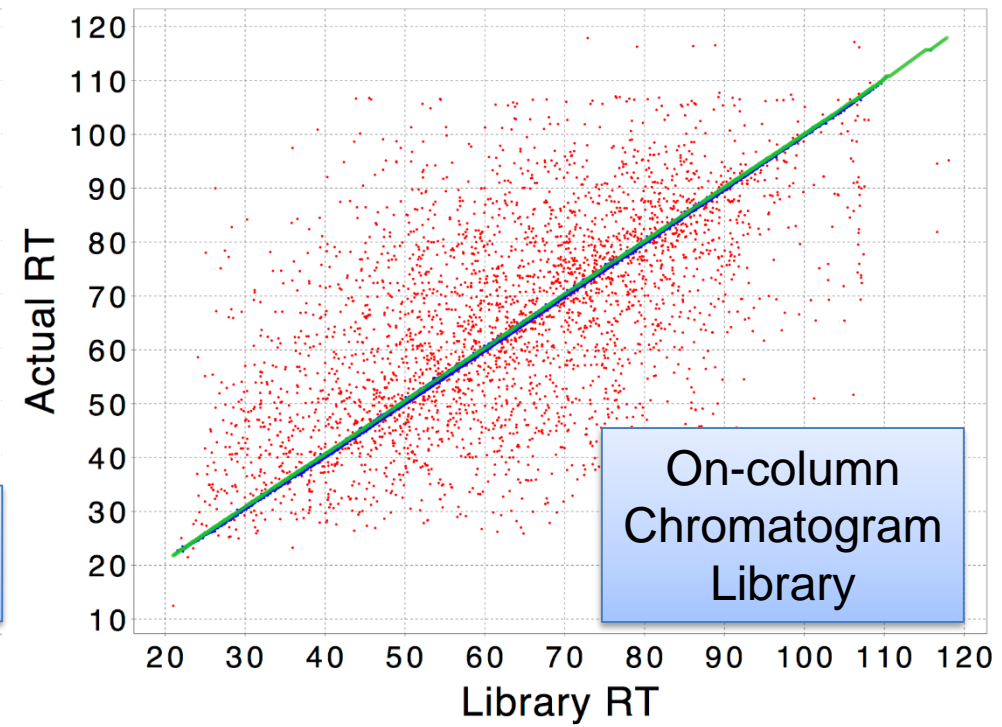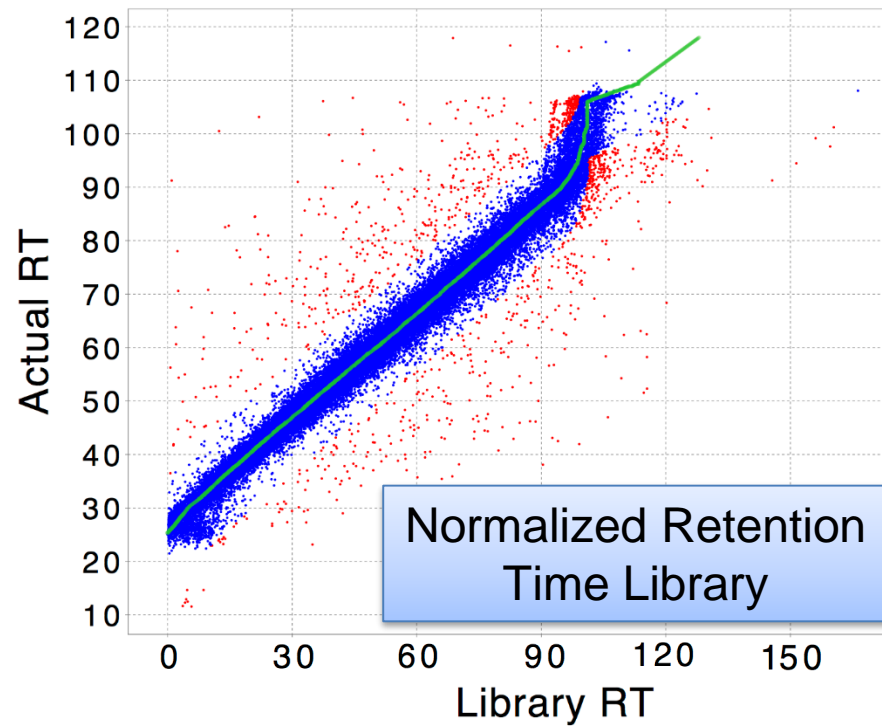
DIA Chromatogram Library Approach
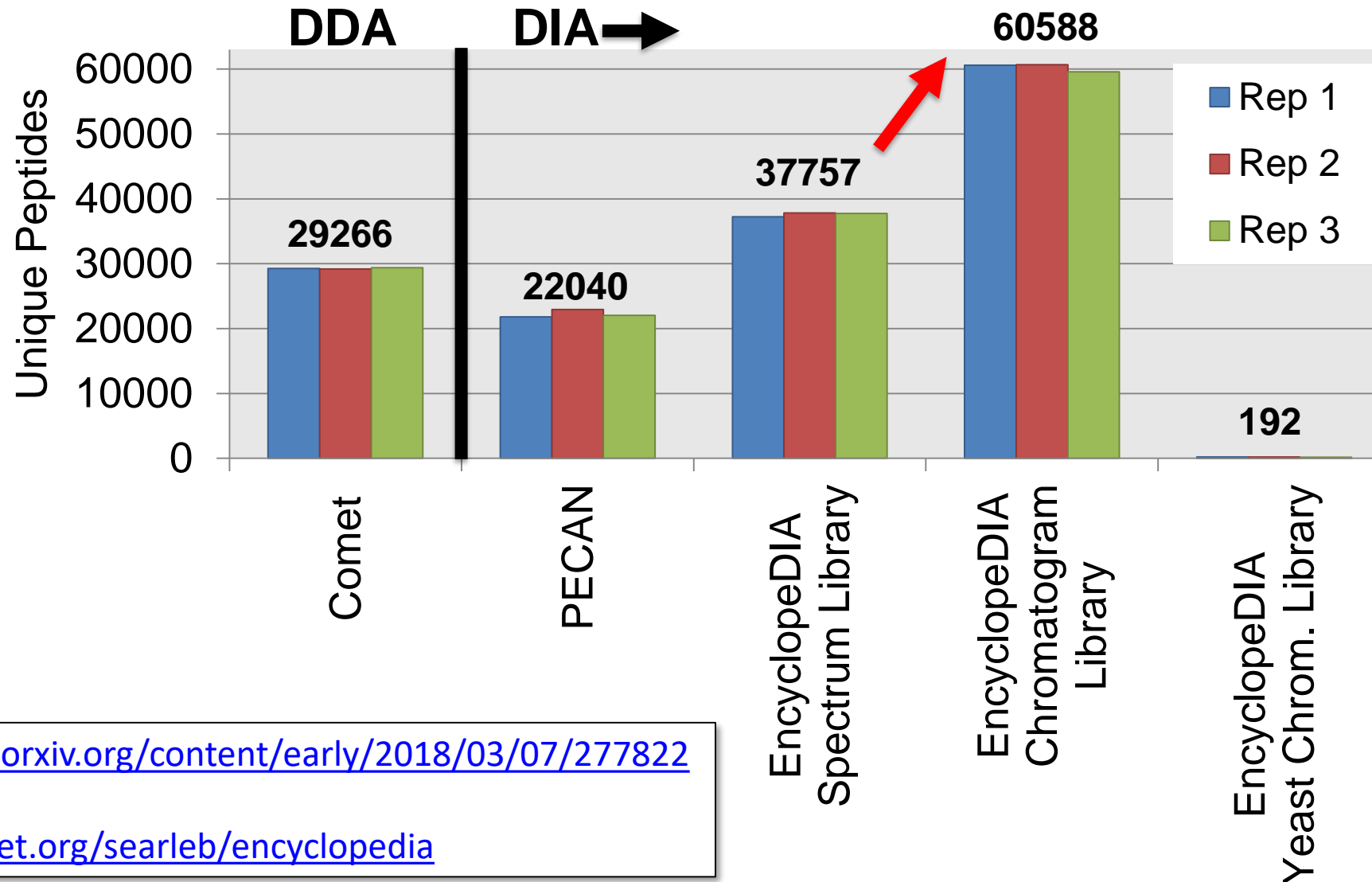

Chromatogram Library

DIA Data

Chromatogram libraries are significantly more powerful than spectrum libraries

Advantages of On-Column Chromatogram Library

- If you can't detect a peptide with a narrow precursor window then you will never detect it with a wide-window.

- On-column RT calibration overcomes selectivity limitations of poor precursor isolation in DIA

- Simplifies workflow. No need to perform extensive fractionation to generate sample specific library data. Just 4-6 runs with narrow windows

- Reduces multiple testing. Only look for peptides you know are in the sample in the wide window data.

Don't Query for Lots Peptides Not in Your Sample!

0.1% FDR

Incorrect PSMs

Correct PSMs

Discriminant score
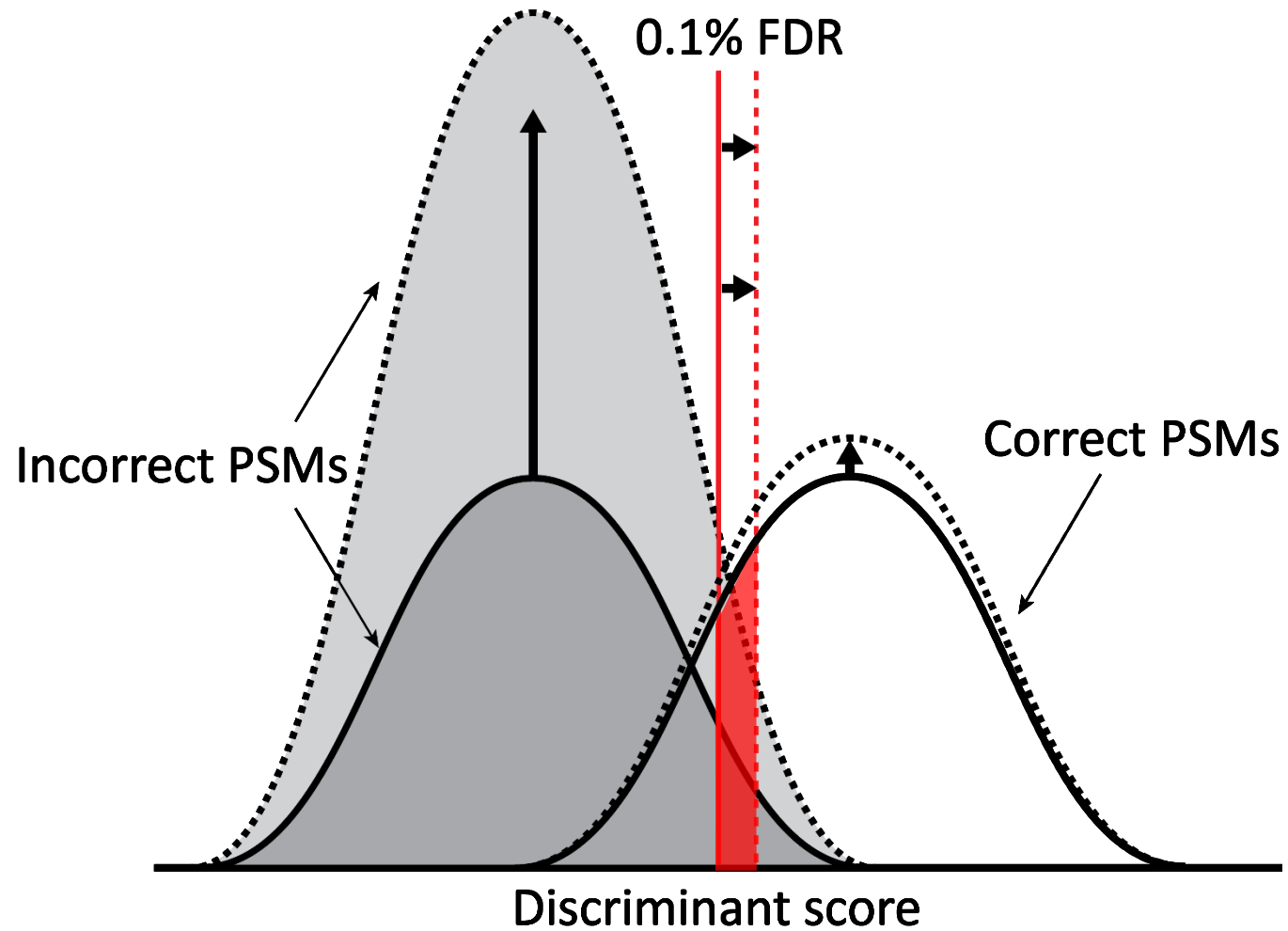
Conclusions

- We have the ability to perform DIA with selectivity of 50% the precursor isolation window size across the entire m/z range.

- Dynamic range and sensitivity that approximates PRM but is comprehensive

- Tools that can handle peptide detection without libraries

- Using gas phase fractionation to build on-column DIA chromatogram libraries.

- We focus on quantitative figures of merit instead of binary presence and absence.

- We need to be putting all of our assays on the same scale so that our quantitative data are commutable between labs, platforms, etc…