Northeastern University

May Institute goes ONLINE!
Computation and statistics for mass spectrometry and proteomics

May 3 – May 14, 2021, Northeastern University, Boston MA
Organizer: Olga Vitek

May 4, 2021

11:00am-2:15pm **Hands-on, Brendan MacLean:** Part 1: Skyline processing of proteome-wide DIA of a controlled mixture

Starting soon

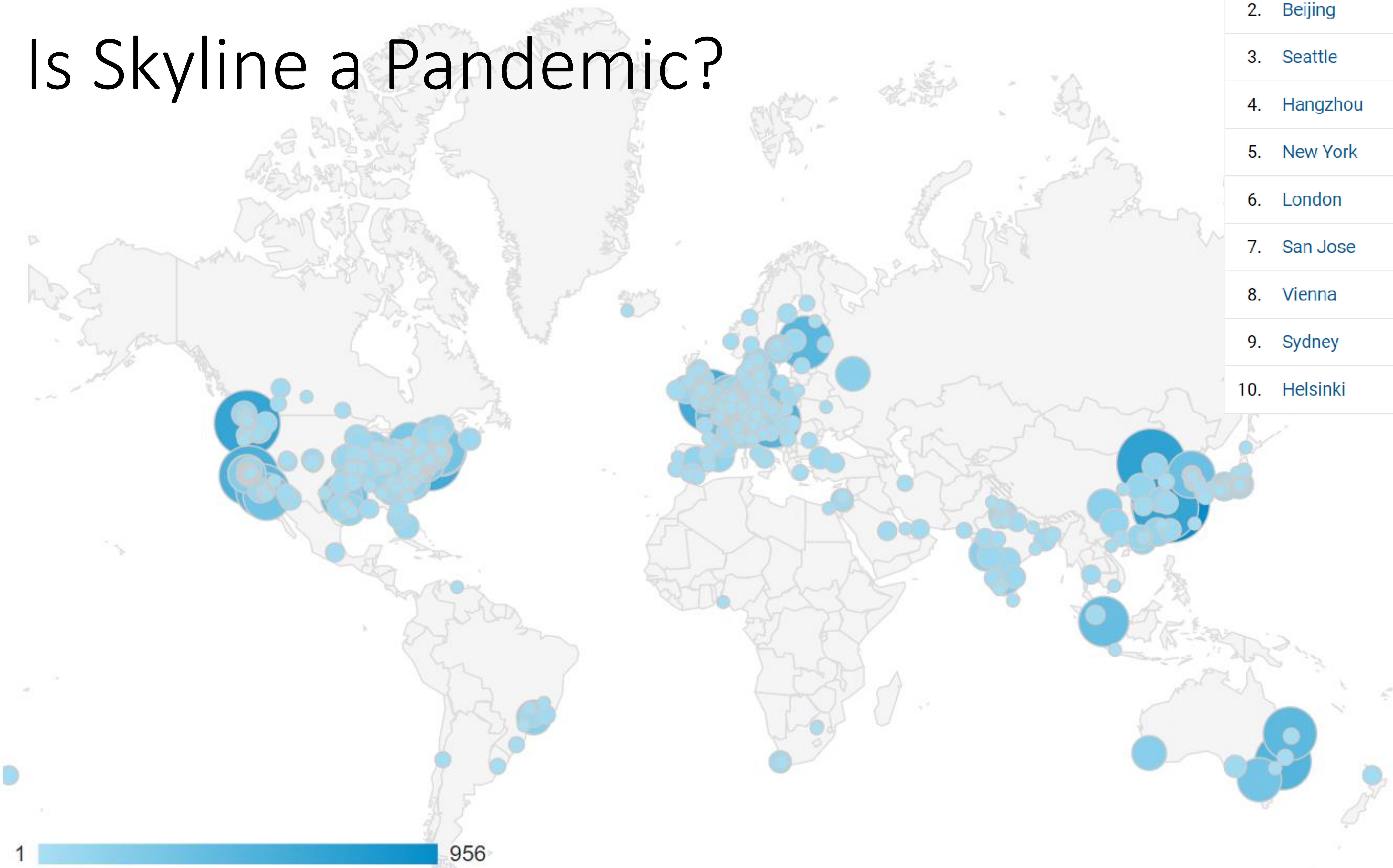
Materials at
https://computationalproteomics.khoury.northeastern.edu/

NIH National Institutes of Health

Chan Zuckerberg Initiative

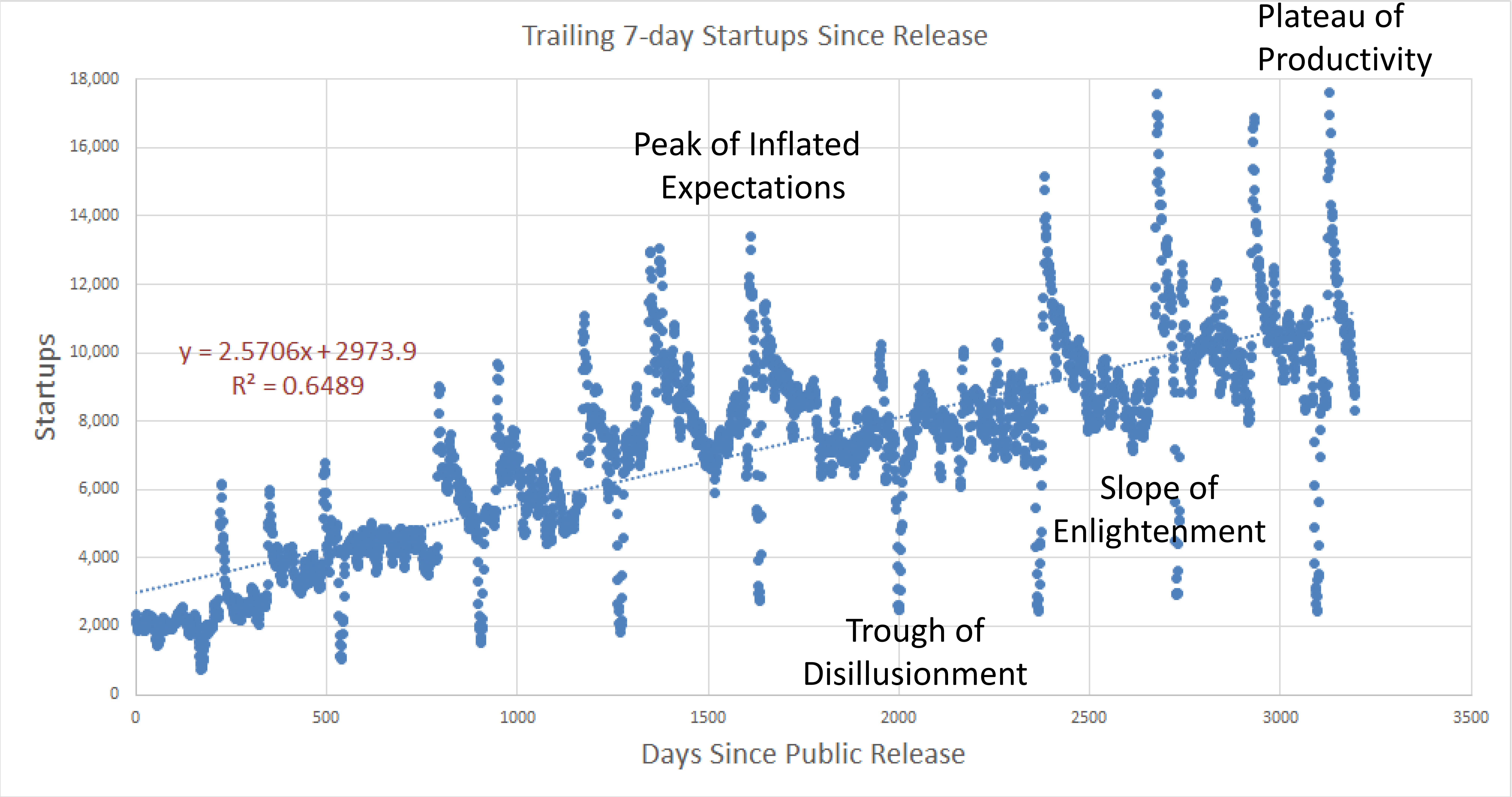
ASA Boston Chapter

Is Skyline a Pandemic?



1. Shanghai
2. Beijing
3. Seattle
4. Hangzhou
5. New York
6. London
7. San Jose
8. Vienna
9. Sydney
10. Helsinki

Viral Growth?



Plateau of Productivity

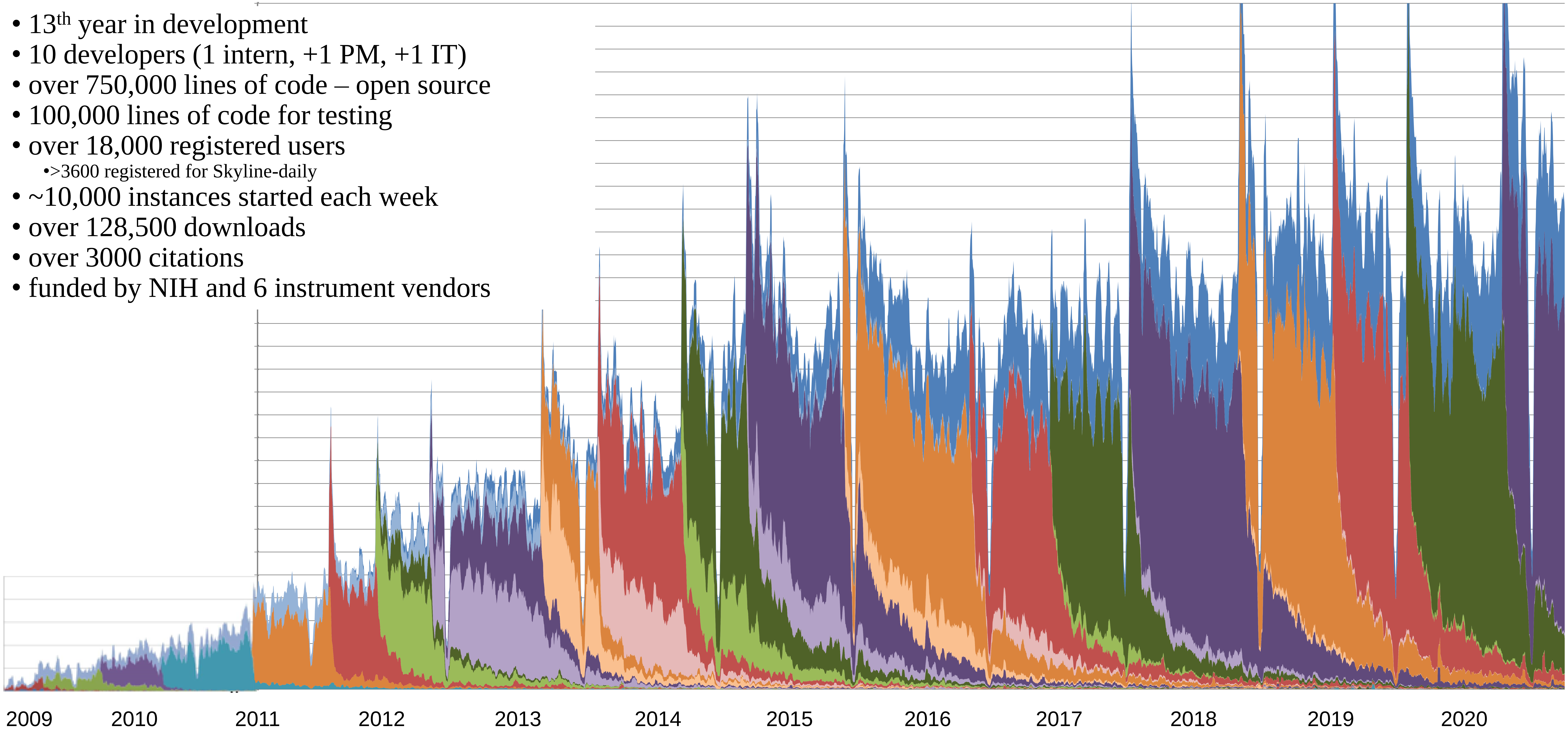
Peak of Inflated Expectations

Slope of Enlightenment

Trough of Disillusionment

Skyline Project Overview

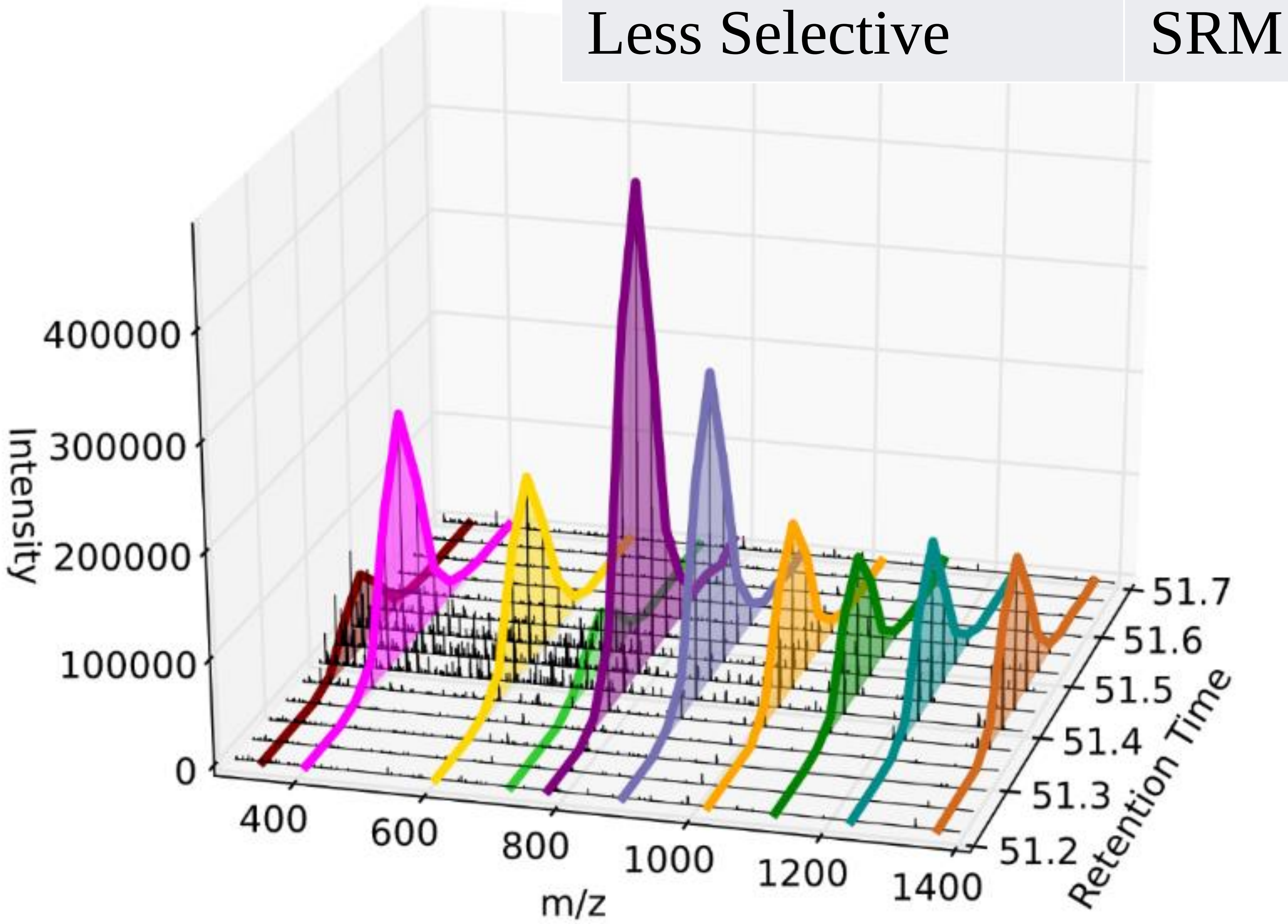
- 13th year in development
- 10 developers (1 intern, +1 PM, +1 IT)
- over 750,000 lines of code – open source
- 100,000 lines of code for testing
- over 18,000 registered users
  - >3600 registered for Skyline-daily
- ~10,000 instances started each week
- over 128,500 downloads
- over 3000 citations
- funded by NIH and 6 instrument vendors

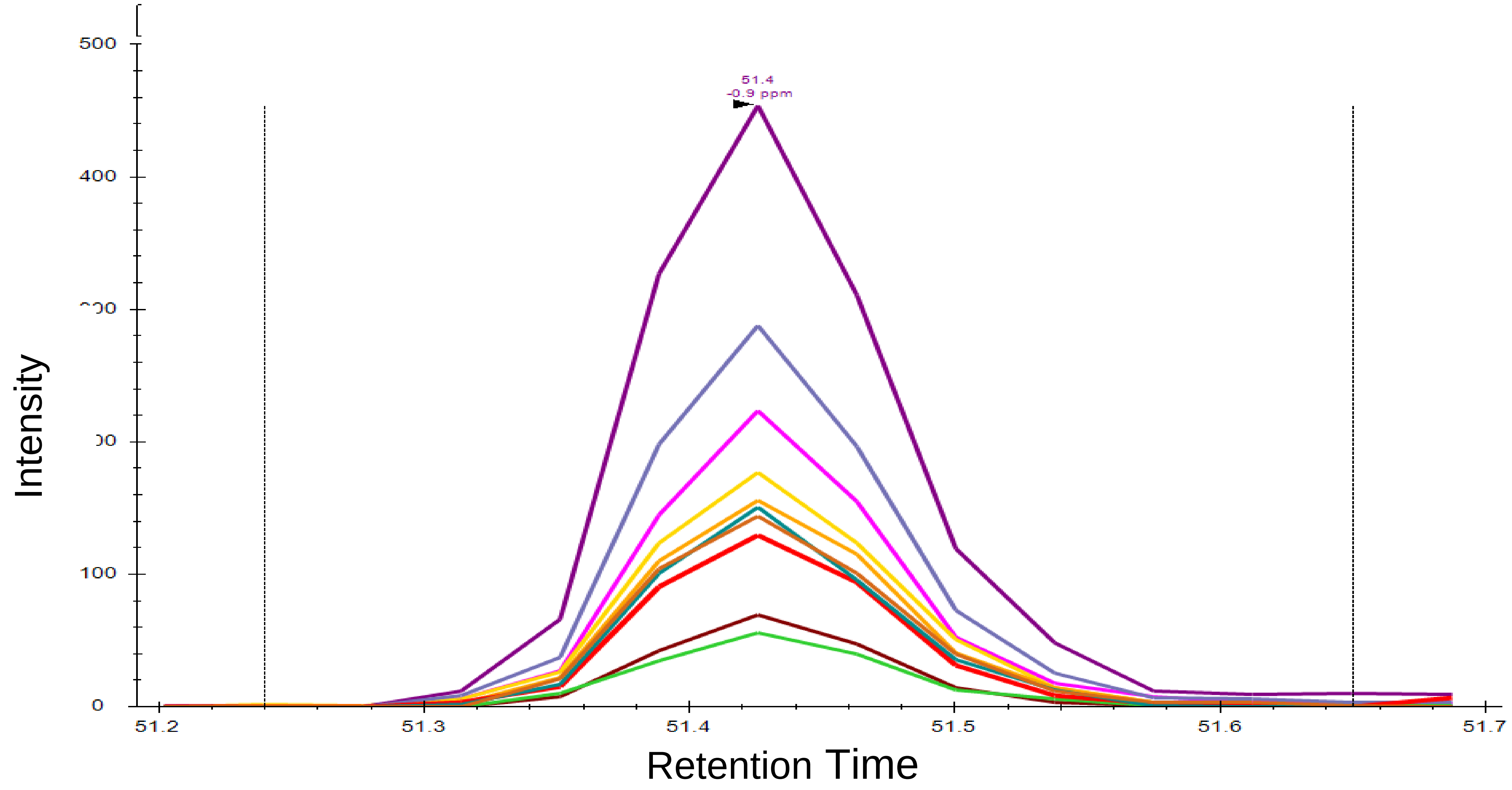


Chromatography-based Quantification

- SRM – Selected ion chromatograms
- PRM – Extracted ion chromatograms
- DIA/SWATH – Extracted ion chromatograms
- DDA – Extracted ion chromatograms from MS1-only

Acquisition	Targeted	Survey
More Selective	PRM	DIA
Less Selective	SRM	DDA





Multiple Instrument Vendors

Agilent

SHIMADZU
Excellence in Science

SCIEX

BRUKER

ThermoFisher
SCIENTIFIC

Waters
THE SCIENCE OF WHAT'S POSSIBLE.™

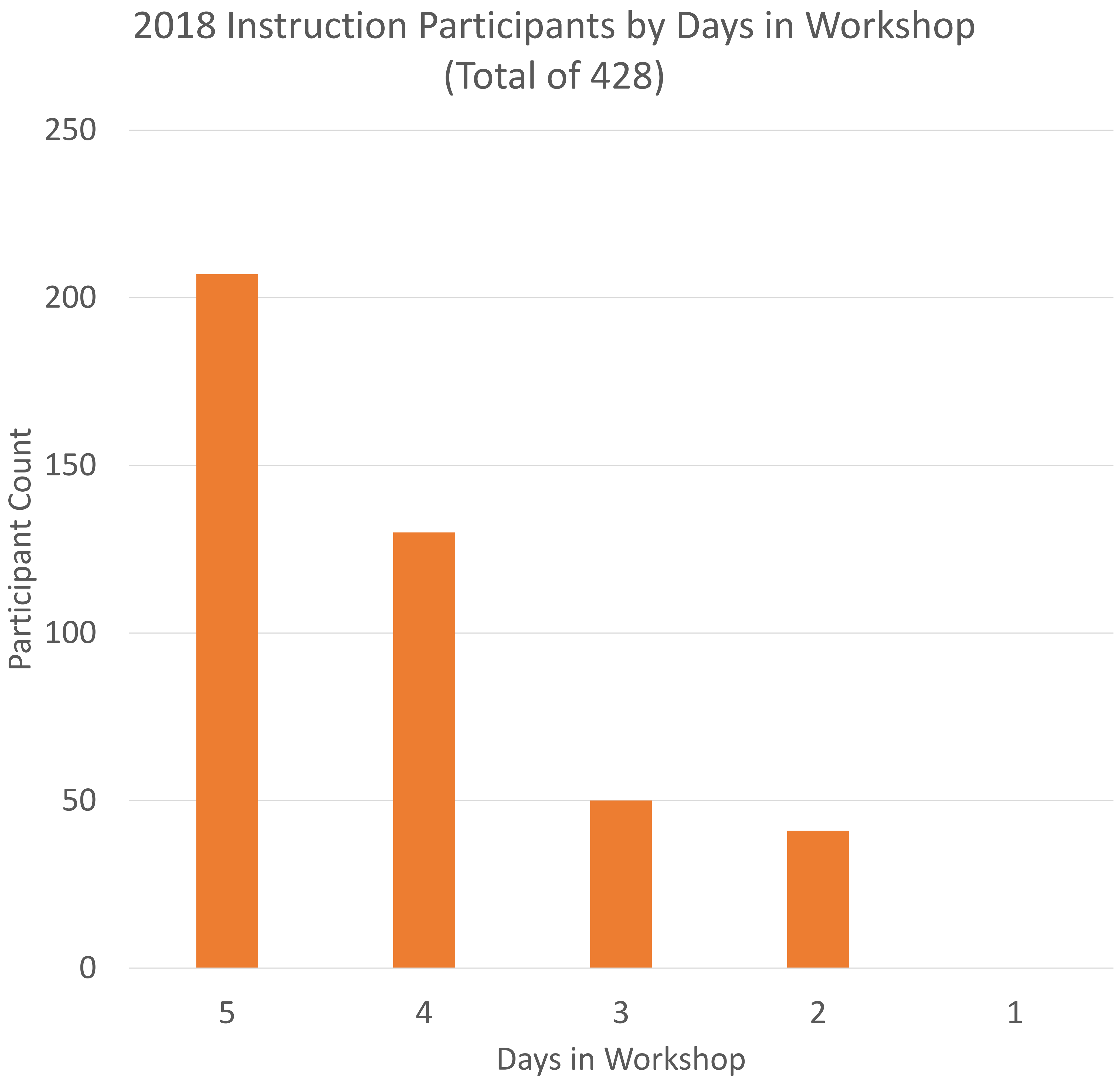
Teaching Targeted Proteomics in 2018

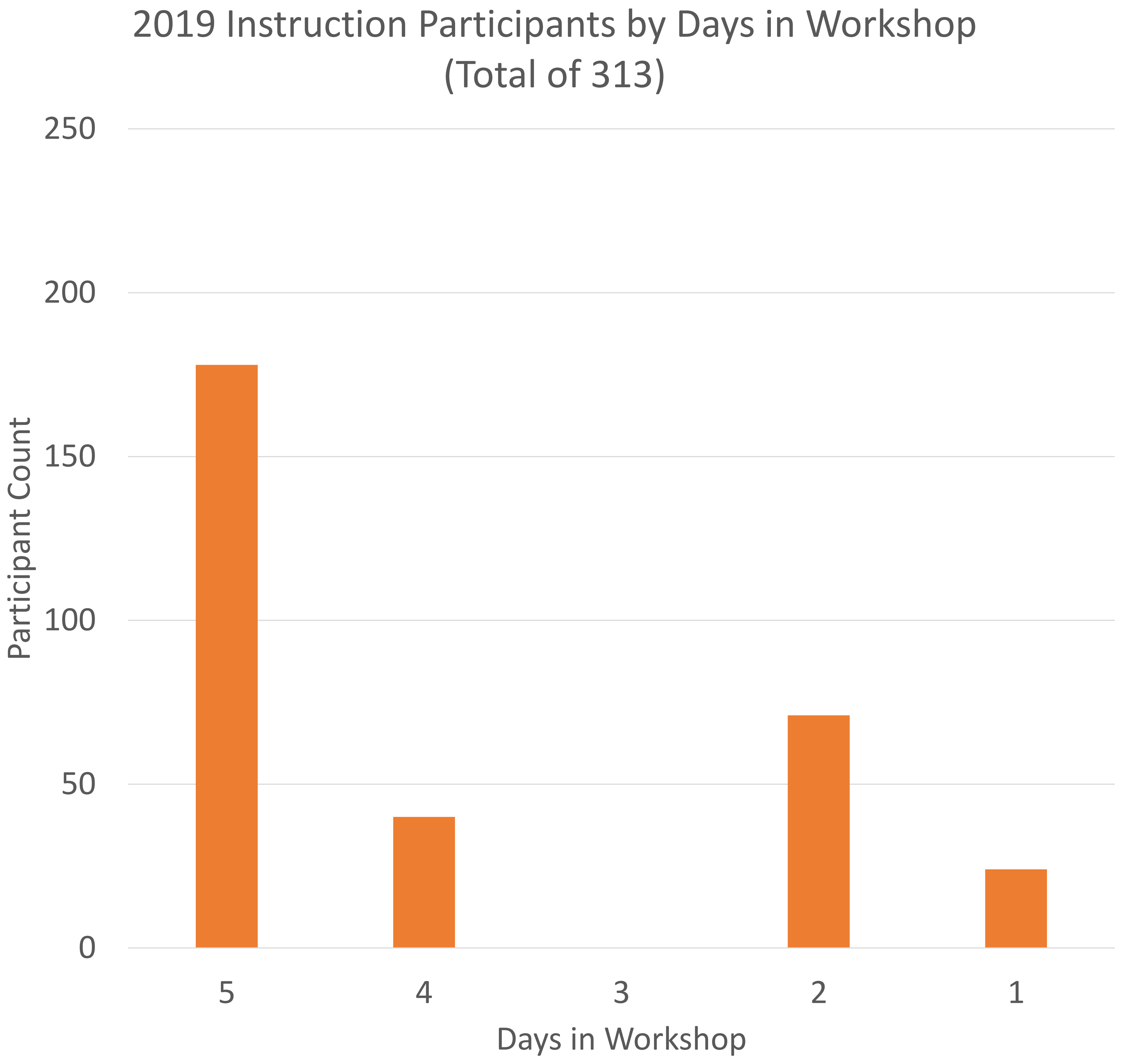
- 5-Buck Institute, Novato, CA – April 2-6 (32)
- 5-NEU, Boston + Statistics – April 30 – May 11(35)
- 5-ETH, Zurich – All DIA – July 2-6 (55)
- 5-U. of Wa., Seattle – July 30 – August 3 (28)
- 5-CRG, Barcelona – November 12-16 (25)
- 5-New! Duke, Durham, NC – December 10-14 (32)
- 4-New! Shanghai – October 22-26 (50)
- 4-IIT Bombay, Mumbai – February (40)
- 4-CNPEM, Campinas, Brazil – November 7-9 (50)
- 3-pre-Lorne, Melbourne – January 29-31 (50)
- 2-MSACL, Palm Springs – January 20&21 (8)
- 2-US HUPO, Minneapolis – March 10&11 (15)
- 2-ASMS, San Diego – June 2&3 (18)

Teaching Targeted Proteomics in 2019

- 5-Buck Institute, Novato, CA – March 18-22 (35)
- 5-NEU, Boston – April 29 - May 10 (50)
- 5-U. of Wa., Seattle – July 22-26 (28)
- 5-Brisbane, Australia – September 9-13 (40)
- 5-CRG, Barcelona – November 11-15 (25)
- 4-Cape Town, SA – November 18-22 (40)
- 2-US HUPO, Washington DC – March 2&3 (28)
- 2-MSACL, Palm Springs – March 31-April 2 (18)
- 2-ASMS, Atlanta – June 1&2 (25)
- 1-MRM Proteomics, Montreal – May 17 (12)
- 1-EUPSS, Brixen, Italy – July 29-31 (12)

Teaching Targeted Proteomics





So Much to Teach, So Much to Learn

“I realized I am only using 20% of what Skyline can do.”

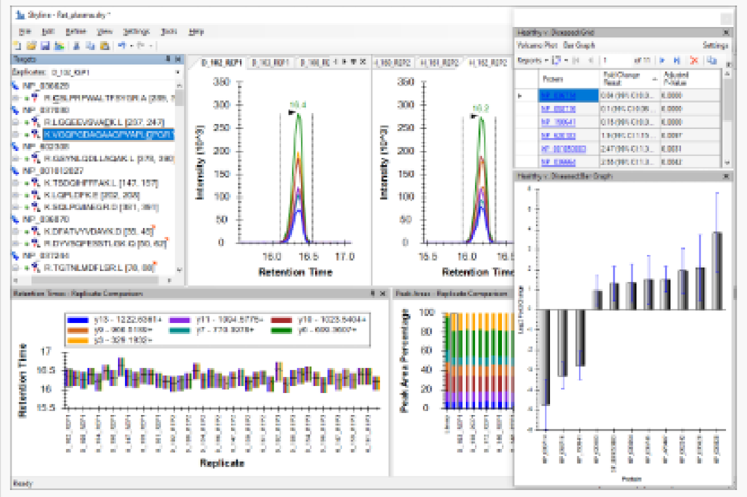
“We only taught you a fraction of what is possible.”

20 Tutorials – Working to Update All to 20.1+

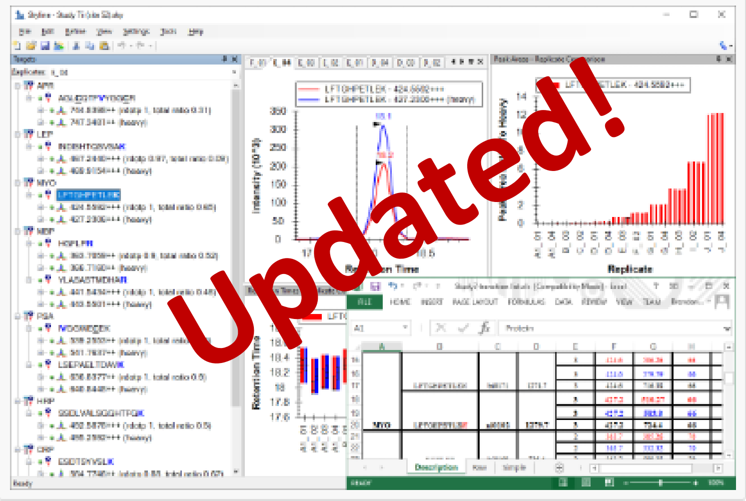
Targeted Method Editing (26 pages)

Targeted Method Refinement (28 pages)

Grouped Study Data Processing (70 pages)

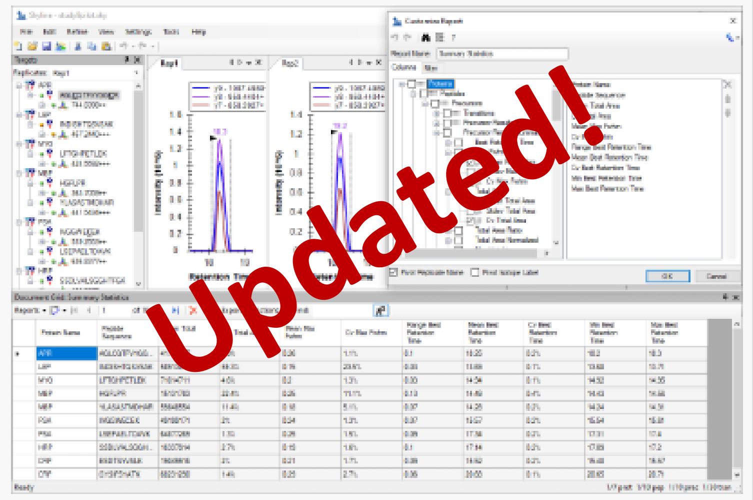


Existing & Quantitative Experiments (43 pages)

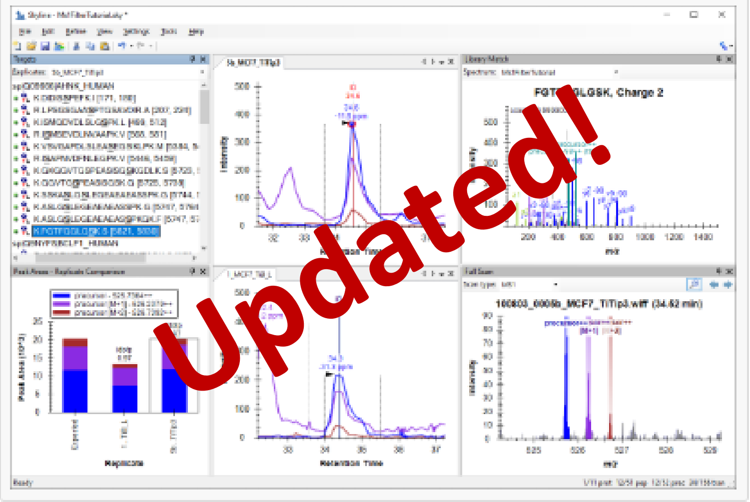


Absolute Quantification (19 pages)

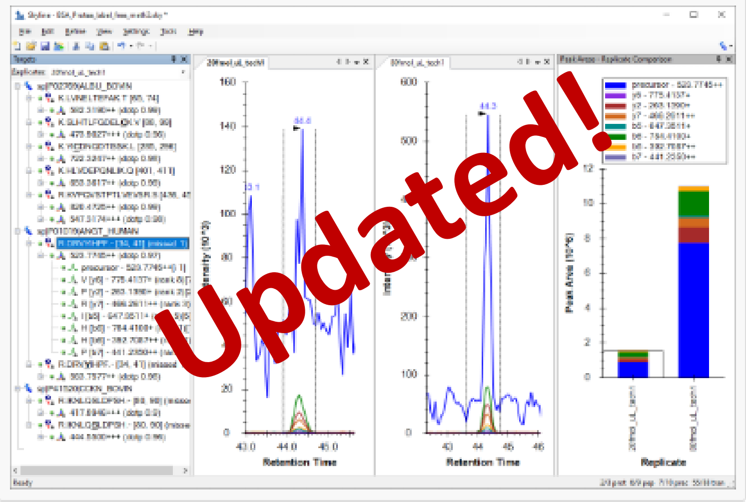
Custom & Live Reports (38 pages)



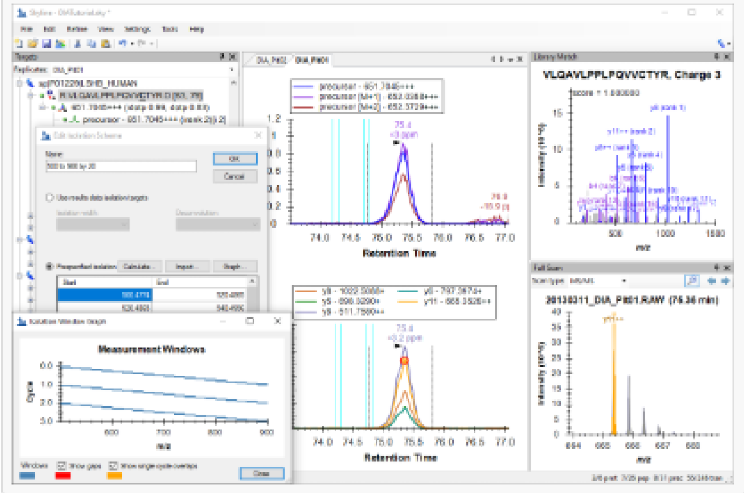
MS1 Full-Scan Filtering (38 pages)



Targeted MS/MS (PRM) (40 pages)



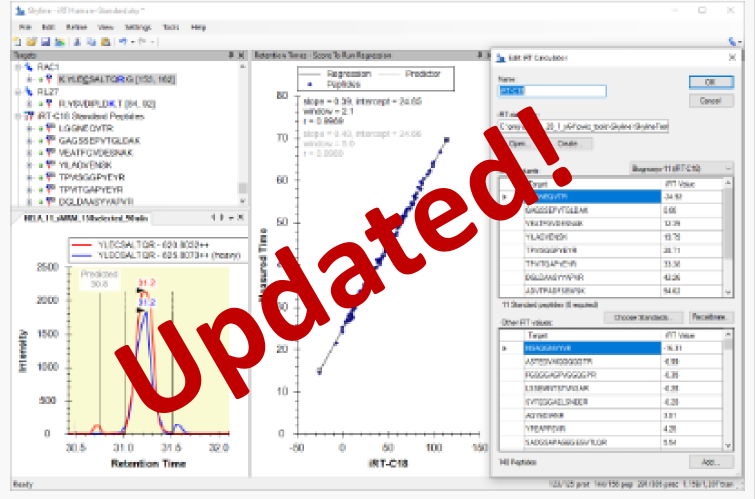
Basic Data Independent Acquisition (40 pages)



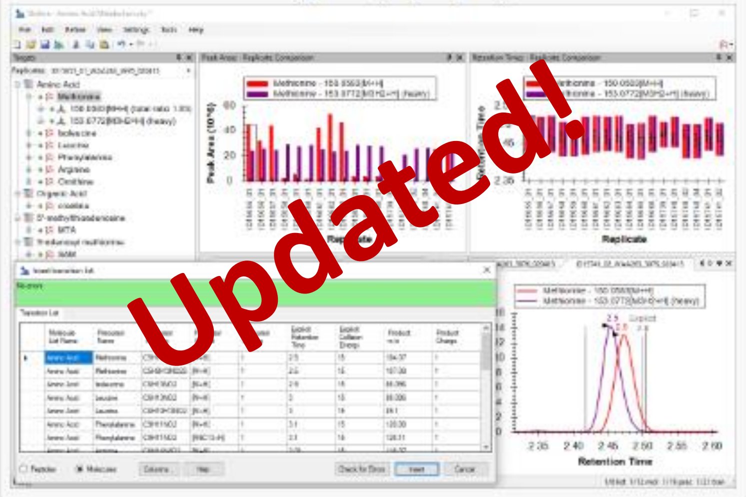
Analysis of DIA/SWATH Data (32 pages)

Advanced Peak Picking Models (28 pages)

iRT Retention Time Prediction (31 pages)



Small Molecule Targets (9 pages)

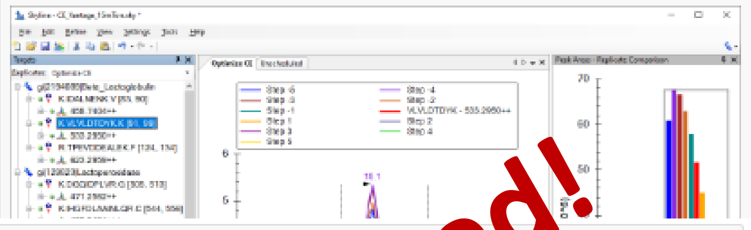


Small Molecule Method Development (28 pages)

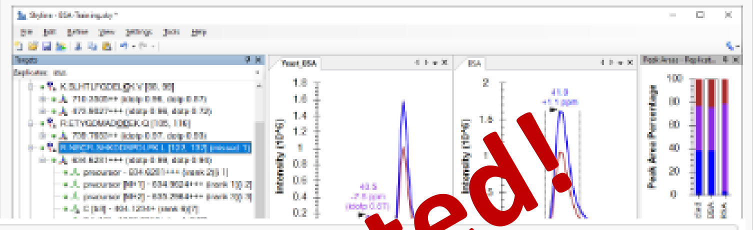
Small Molecule Quantification (23 pages)

Hi-Res Metabolomics (13 pages)

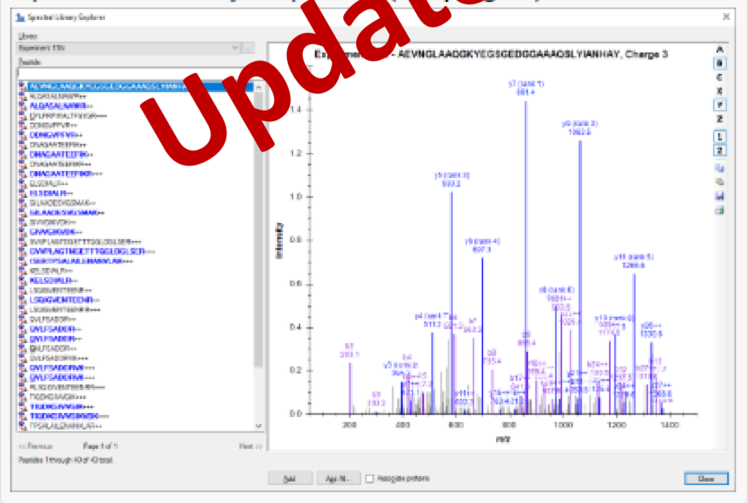
Collision Energy Optimization (12 pages)



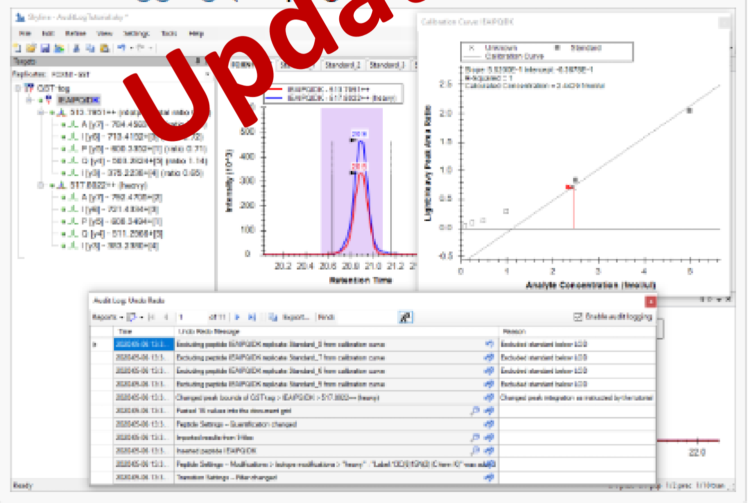
Ion Mobility Spectrum Filtering (26 pages)



Spectral Library Explorer ([illegible]2 pages)



Audit Logging (23 pages)



19 Tutorial Webinars – and Growing…

- Webinar #18:
  DIA/SWATH Data Analysis
  500+ attended

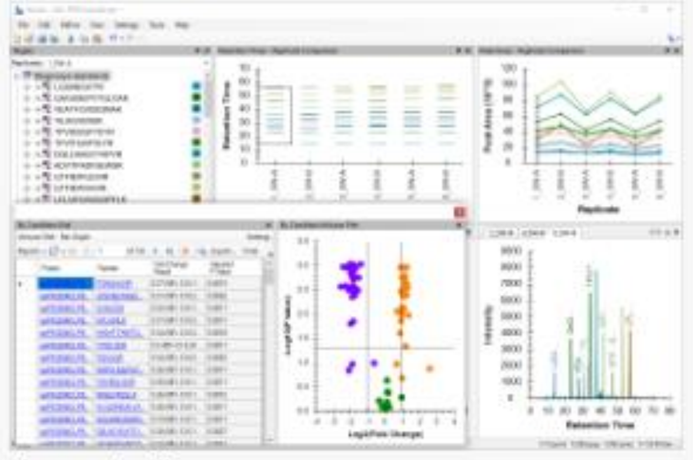

Brendan MacLean
(Principal Developer)

Presentation Slides


Ben Collins
(DIA/SWATH
researcher)

Presentation Slides

Analysis of DIA/SWATH Data
(32 pages)


[tutorial]

- Webinar #19:
  Ion Mobility Spectrum Filtering
  235 attended

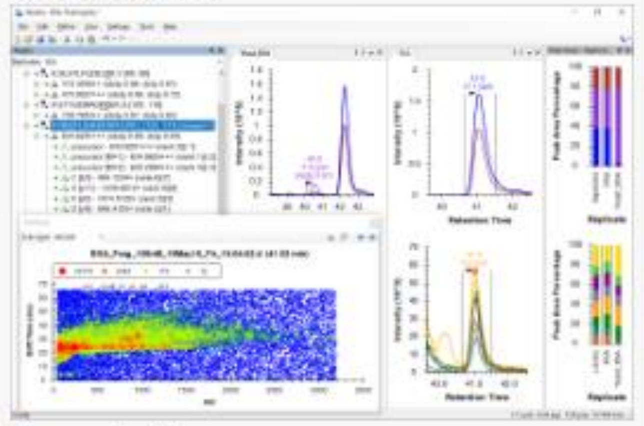

Brendan MacLean
(Principal Developer)

Presentation Slides


Erin Baker
(ion mobility researcher)

Presentation Slides

Ion Mobility Spectrum Filtering
(32 pages)


[tutorial]

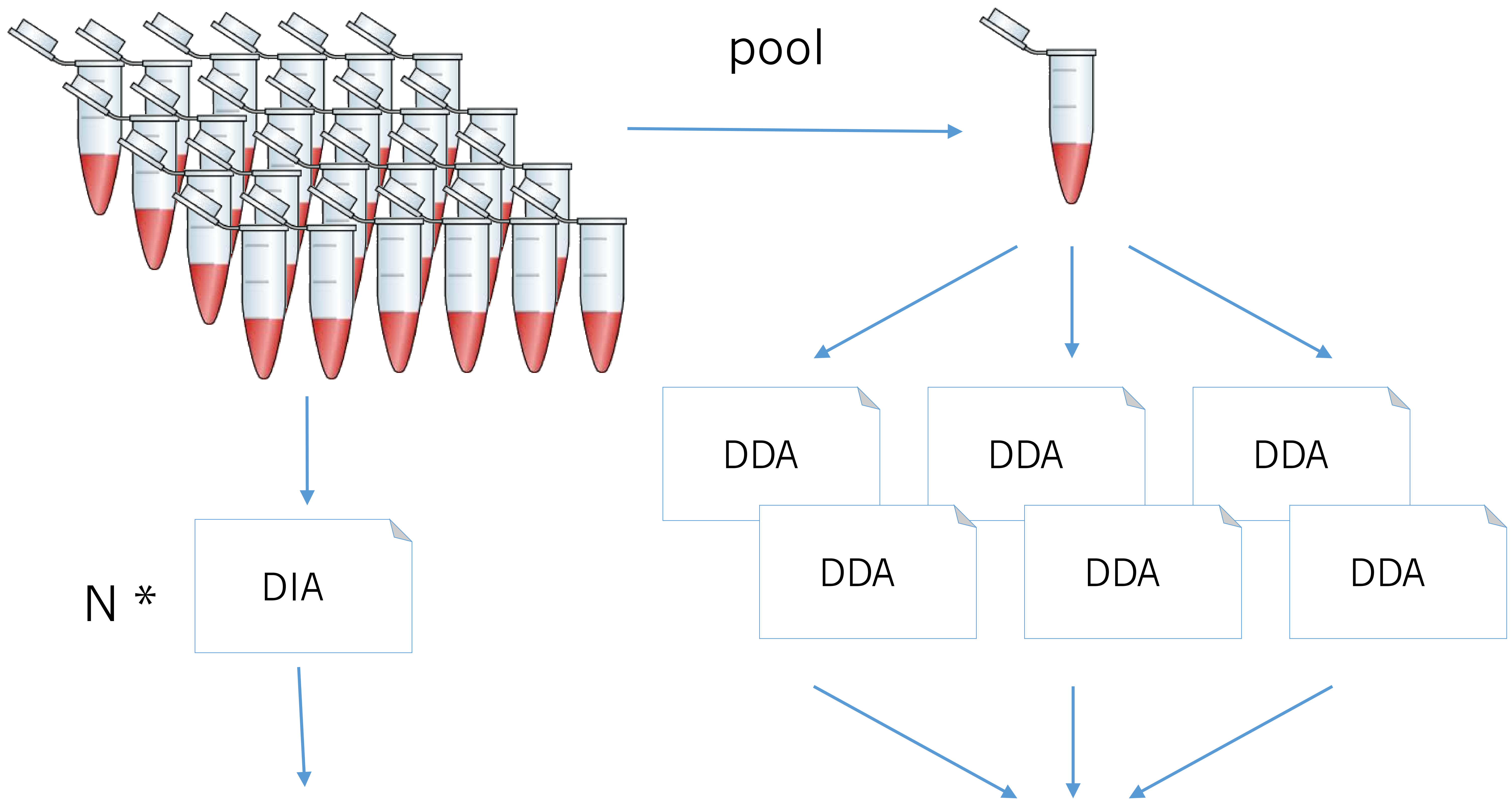
- Webinar #20:
  dia-PASEF Data Analysis

Coming Soon!!


Brendan MacLean
(Principal Developer)


Ben Collins
(DIA/SWATH
researcher)

Two Phase Experiment



N * Quantitative Samples

1 * Detection Samples

pool

N * DIA

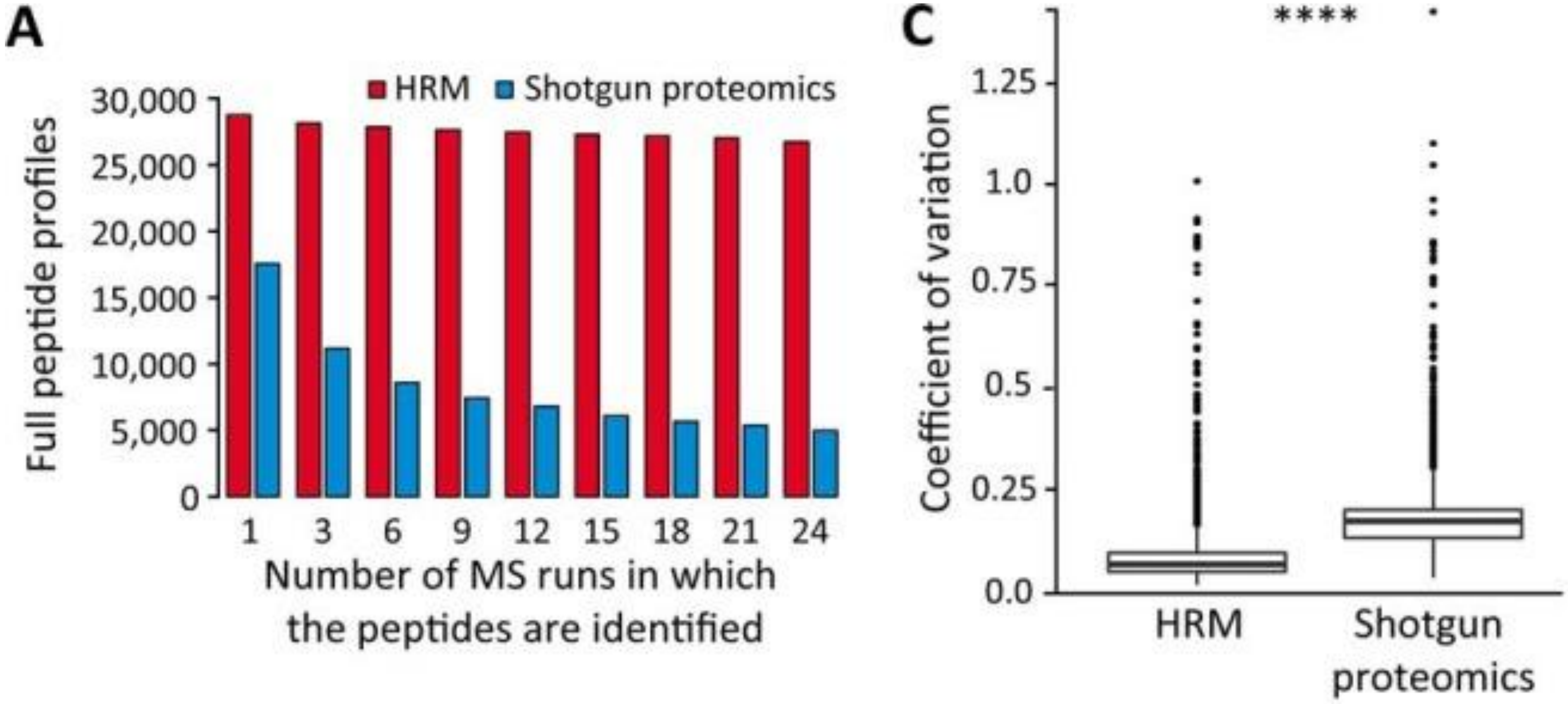
DDA DDA DDA
DDA DDA DDA

Quantify Peptides

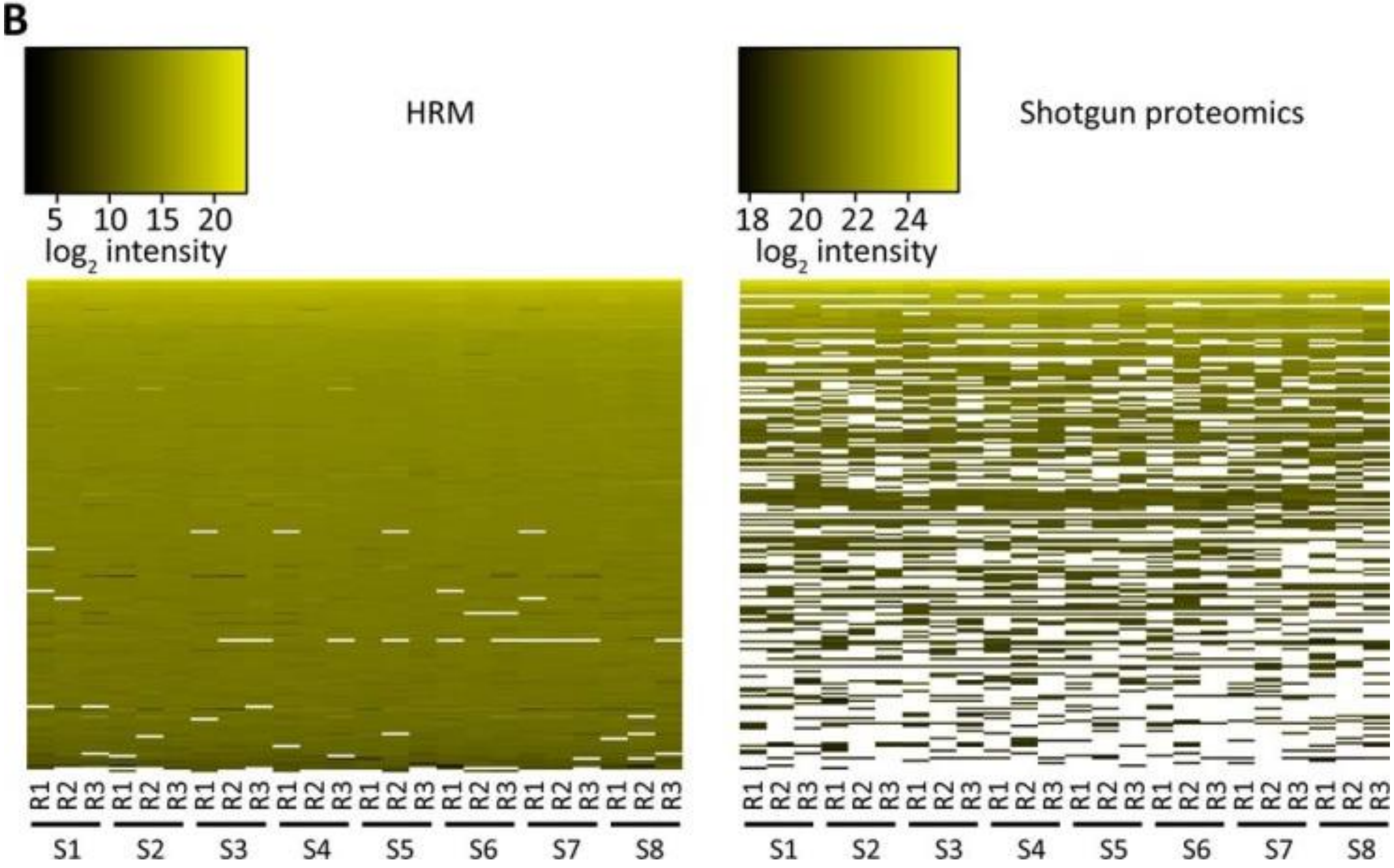
Detect Peptides
with Search Engine
Build Library

Bruderer, et al. & Reiter, *MCP, 2015*

DDA vs. DIA on IDs and CVs



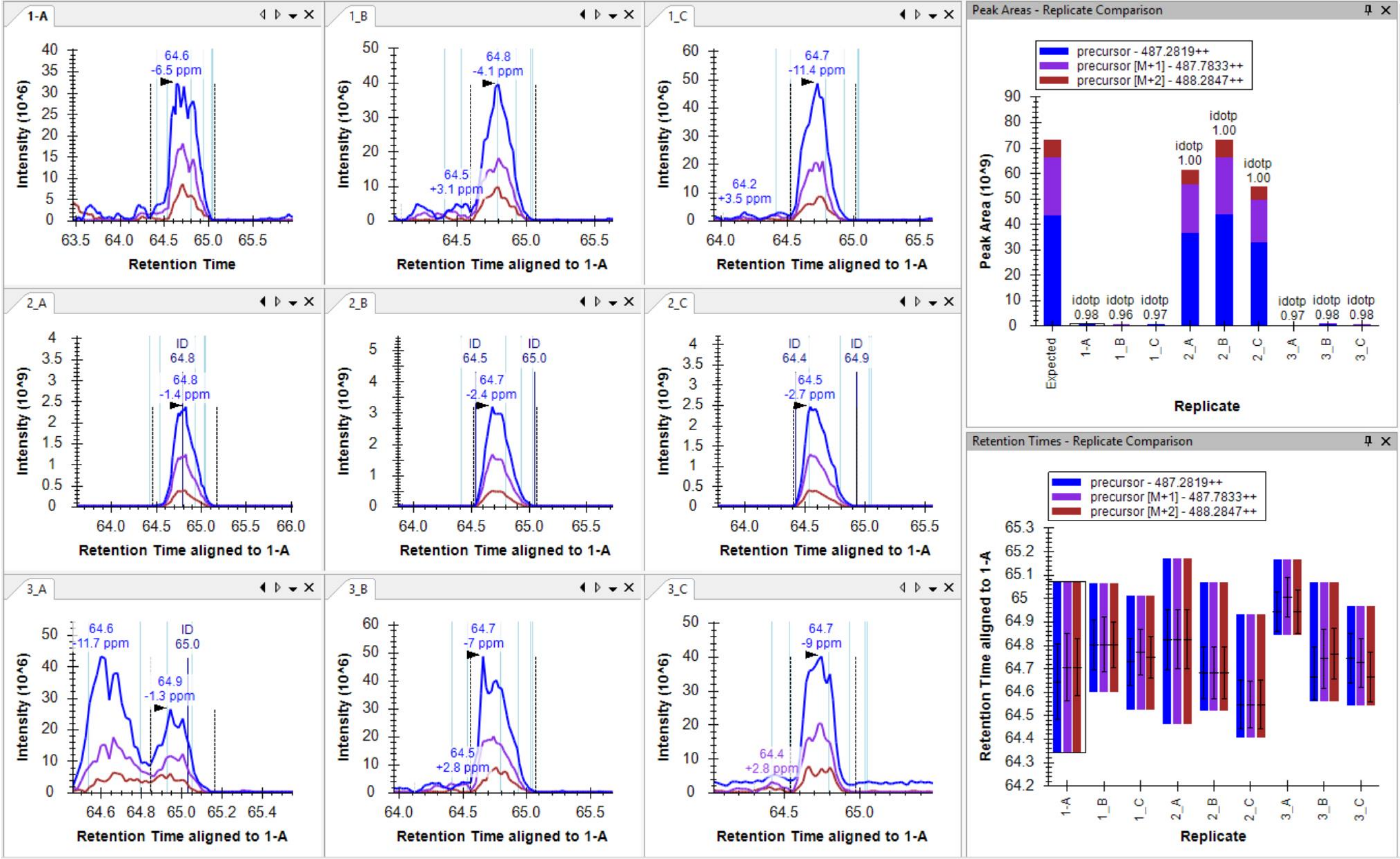
DDA vs. DIA on Coverage

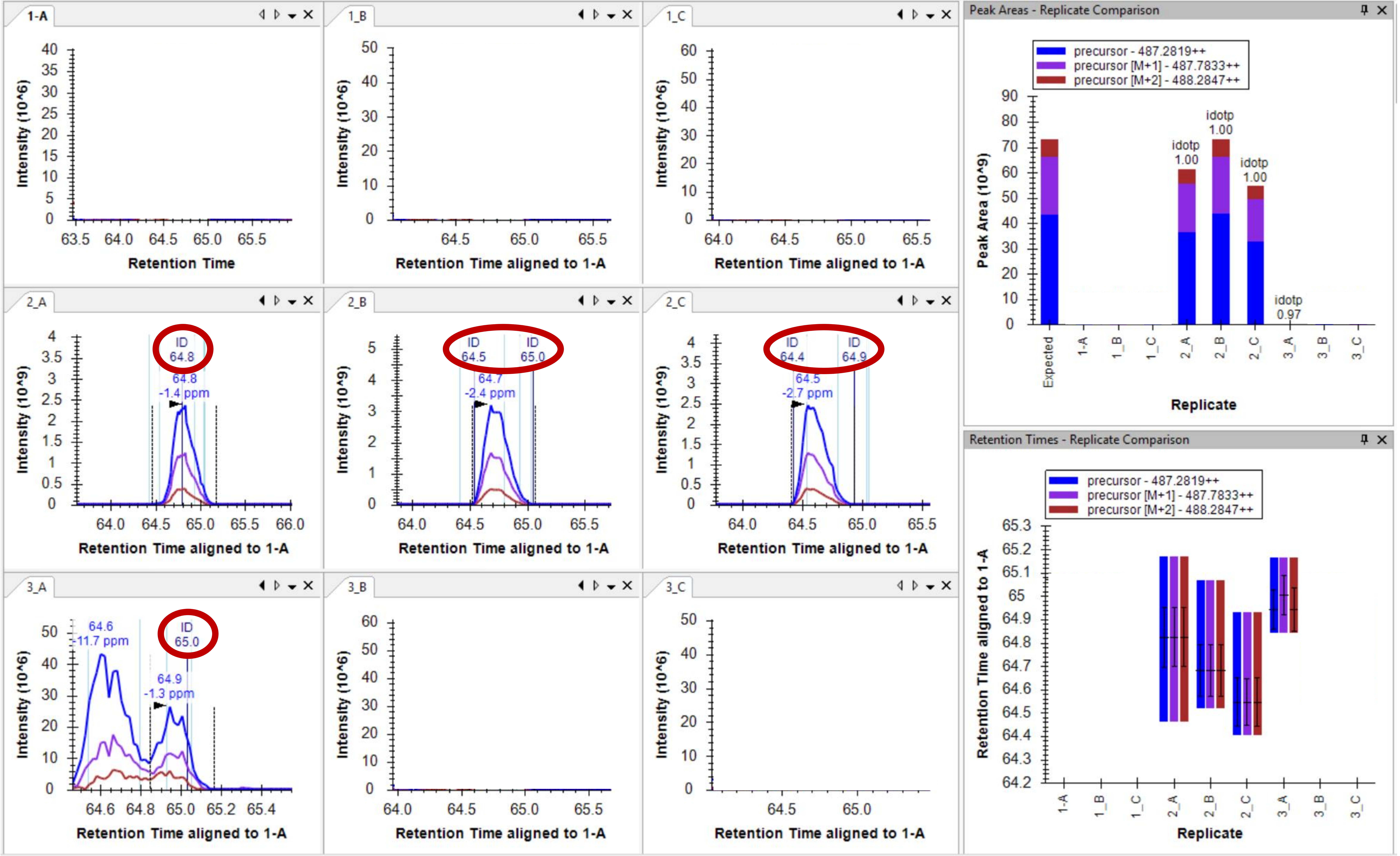


Note RT Alignment Restriction

“On average, 28,610 peptides were identified per measurement in HRM. The raw DDA spectra of the profiling standard sample set were analyzed with the software MaxQuant (29). *No feature alignment was used*. On average 17,547 peptides were identified per measurement in shotgun proteomics.”

Bruderer, MCP 2015


1-A
64.6
-6.5 ppm
Intensity (10^6)
Retention Time
1_B
64.8
-4.1 ppm
64.5
+3.1 ppm
Retention Time aligned to 1-A
1_C
64.7
-11.4 ppm
64.2
+3.5 ppm
2_A
ID
64.8
64.8
-1.4 ppm
Intensity (10^9)
2_B
ID
64.5
ID
65.0
64.7
-2.4 ppm
2_C
ID
64.4
ID
64.9
64.5
-2.7 ppm
3_A
64.6
-11.7 ppm
ID
65.0
64.9
-1.3 ppm
3_B
64.7
-7 ppm
64.5
+2.8 ppm
3_C
64.7
-9 ppm
64.4
+2.8 ppm
Peak Areas - Replicate Comparison
precursor - 487.2819++
precursor [M+1] - 487.7833++
precursor [M+2] - 488.2847++
Peak Area (10^9)
idotp 0.98
idotp 0.96
idotp 0.97
idotp 1.00
idotp 1.00
idotp 1.00
idotp 0.97
idotp 0.98
idotp 0.98
Expected
1-A
1_B
1_C
2_A
2_B
2_C
3_A
3_B
3_C
Replicate
Retention Times - Replicate Comparison
precursor - 487.2819++
precursor [M+1] - 487.7833++
precursor [M+2] - 488.2847++
Retention Time aligned to 1-A
Replicate

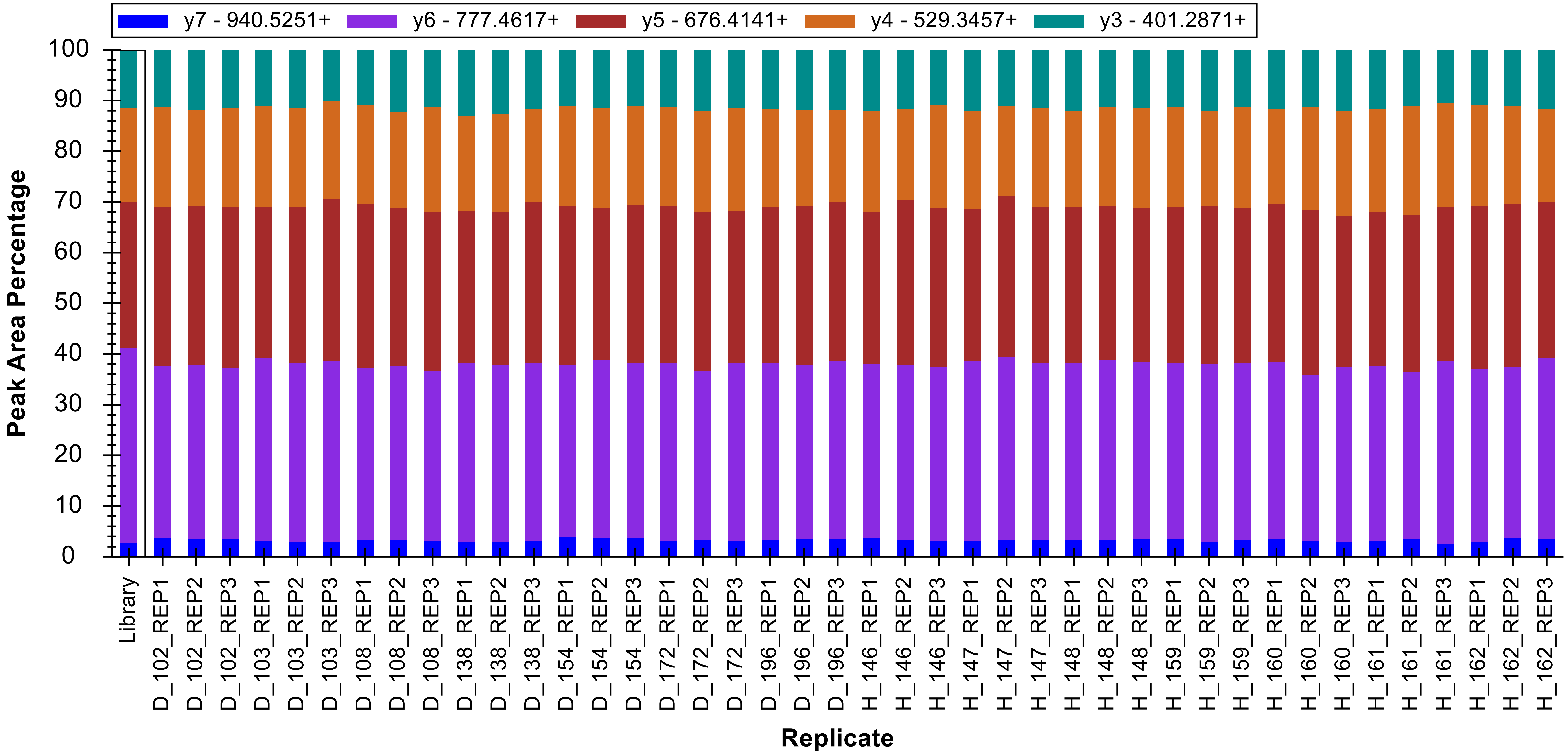

1-A
Intensity (10^6)
Retention Time
1_B
Retention Time aligned to 1-A
1_C
2_A
ID 64.8
64.8
-1.4 ppm
Intensity (10^9)
2_B
ID 64.5
ID 65.0
64.7
-2.4 ppm
2_C
ID 64.4
ID 64.9
64.5
-2.7 ppm
3_A
64.6
-11.7 ppm
ID 65.0
64.9
-1.3 ppm
3_B
3_C
Peak Areas - Replicate Comparison
precursor - 487.2819++
precursor [M+1] - 487.7833++
precursor [M+2] - 488.2847++
Peak Area (10^9)
idotp 1.00
idotp 1.00
idotp 1.00
idotp 0.97
Expected
1-A
1_B
1_C
2_A
2_B
2_C
3_A
3_B
3_C
Replicate
Retention Times - Replicate Comparison
precursor - 487.2819++
precursor [M+1] - 487.7833++
precursor [M+2] - 488.2847++
Retention Time aligned to 1-A
Replicate

Prior Knowledge and Consistency

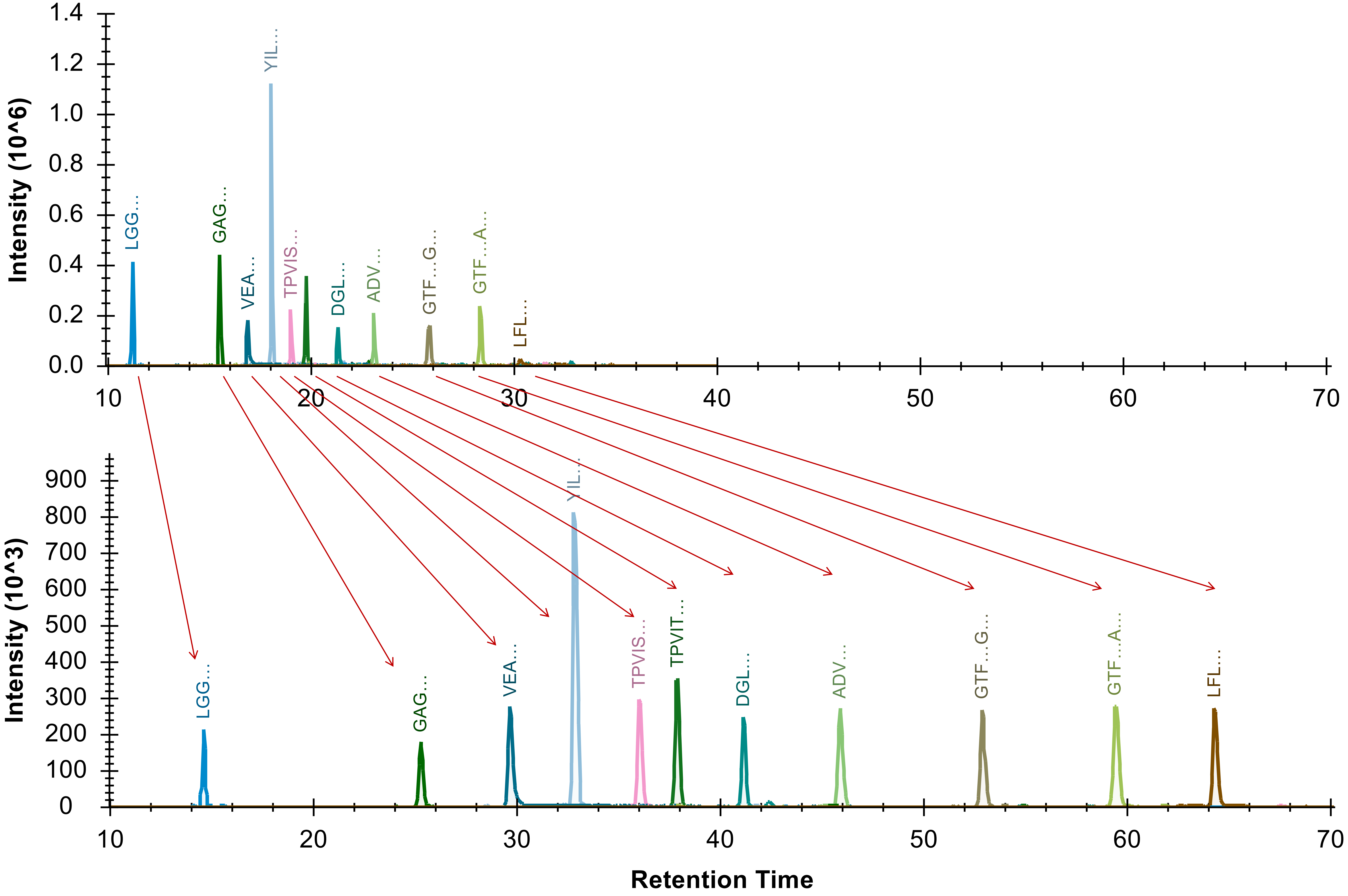
- Relative product ion abundance
  - Spectral libraries
- Retention time
  - iRT libraries

- Powerful enough to be used cross-lab / cross experiment
  - Measure, store, re-use
- More powerful run-to-run
  - Same instrument
  - Same chromatography

Relative Product Ion Abundance

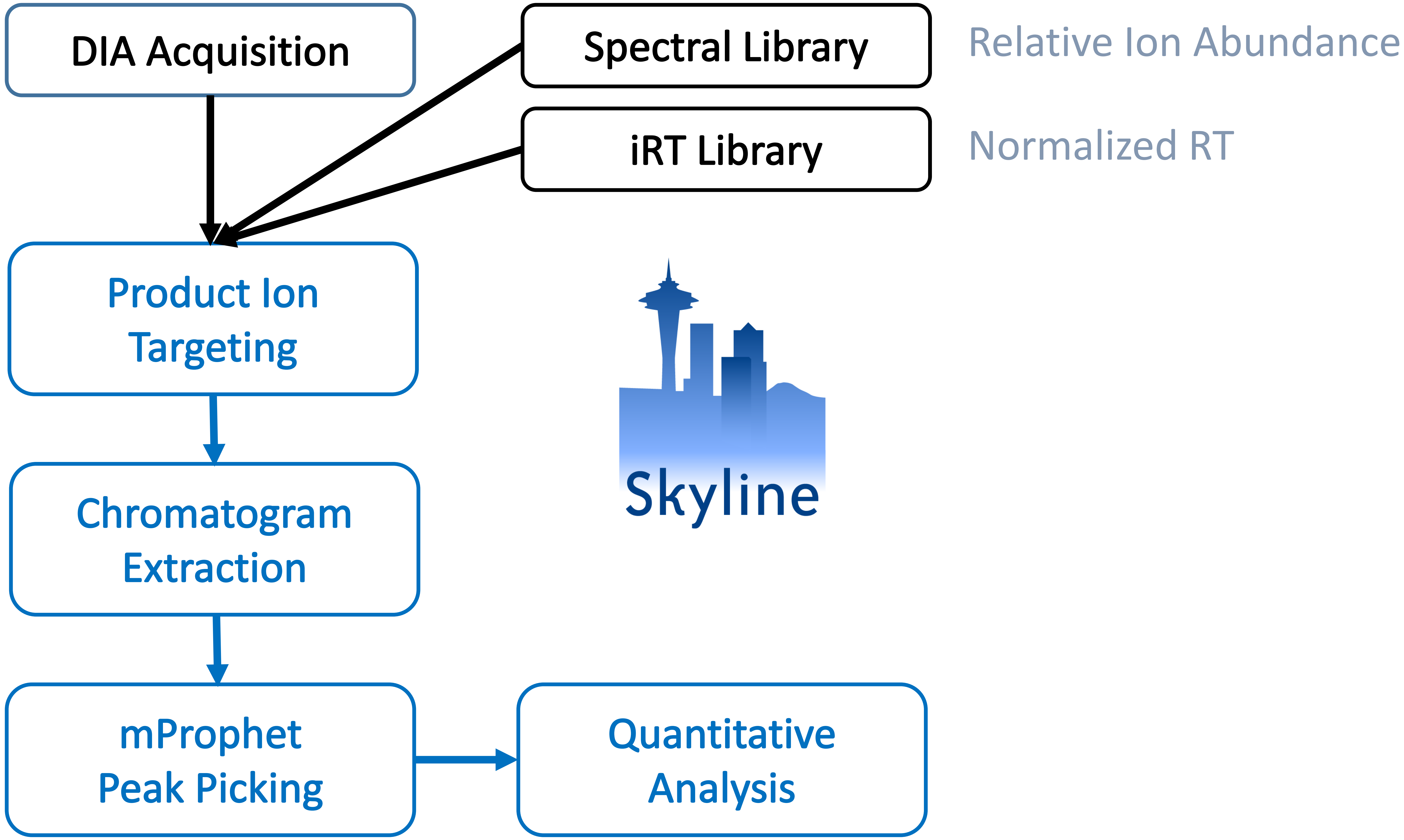


iRT time prediction



Escher, Reiter, et al. *Proteomics*, 2012

Prior Knowledge Workflow



DIA Acquisition

Spectral Library

Relative Ion Abundance

iRT Library

Normalized RT

Product Ion Targeting

Chromatogram Extraction

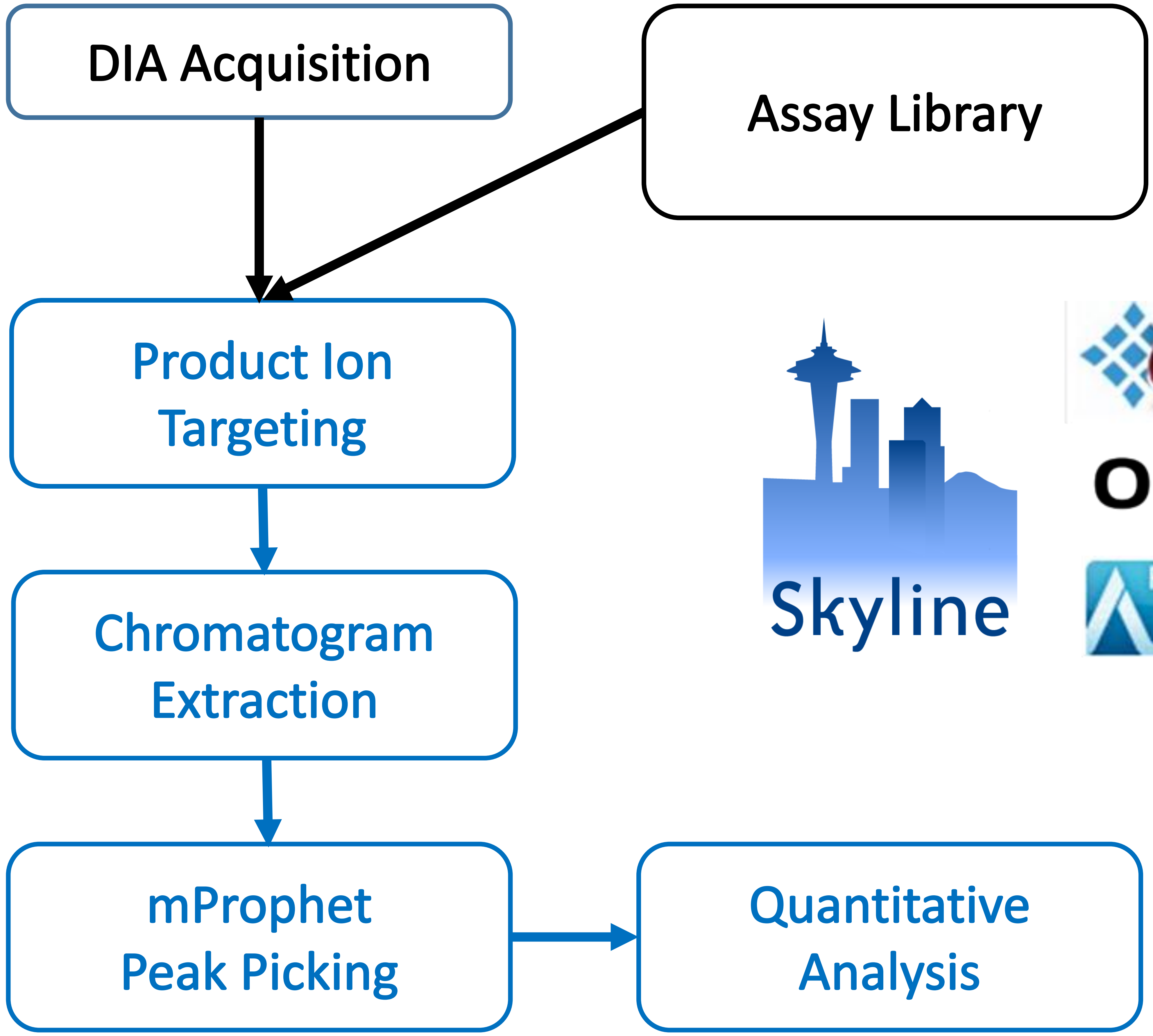
mProphet Peak Picking

Quantitative Analysis

Skyline

Reiter, Rinner, et al.
Nature Methods, 2011

Prior Knowledge Workflow



DIA Acquisition

Assay Library

Product Ion Targeting

Chromatogram Extraction

mProphet Peak Picking

Quantitative Analysis

Relative Ion Abundance

Normalized RT

Skyline

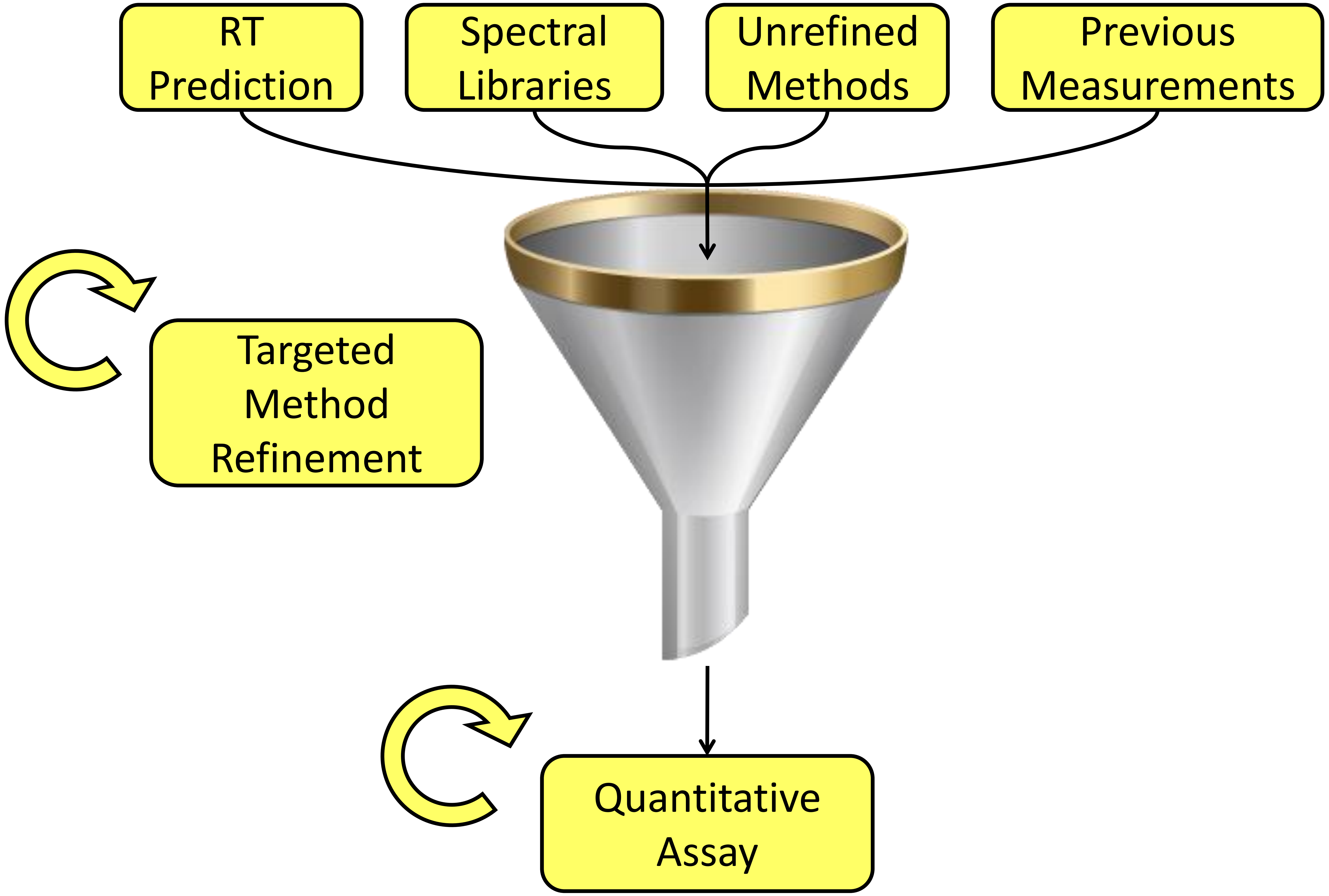
Spectronaut

OpenSWATH by ETH zürich

PeakView

Reiter, Rinner, et al.
Nature Methods, 2011

Supporting a Broader Hypothesis



RT Prediction

Spectral Libraries

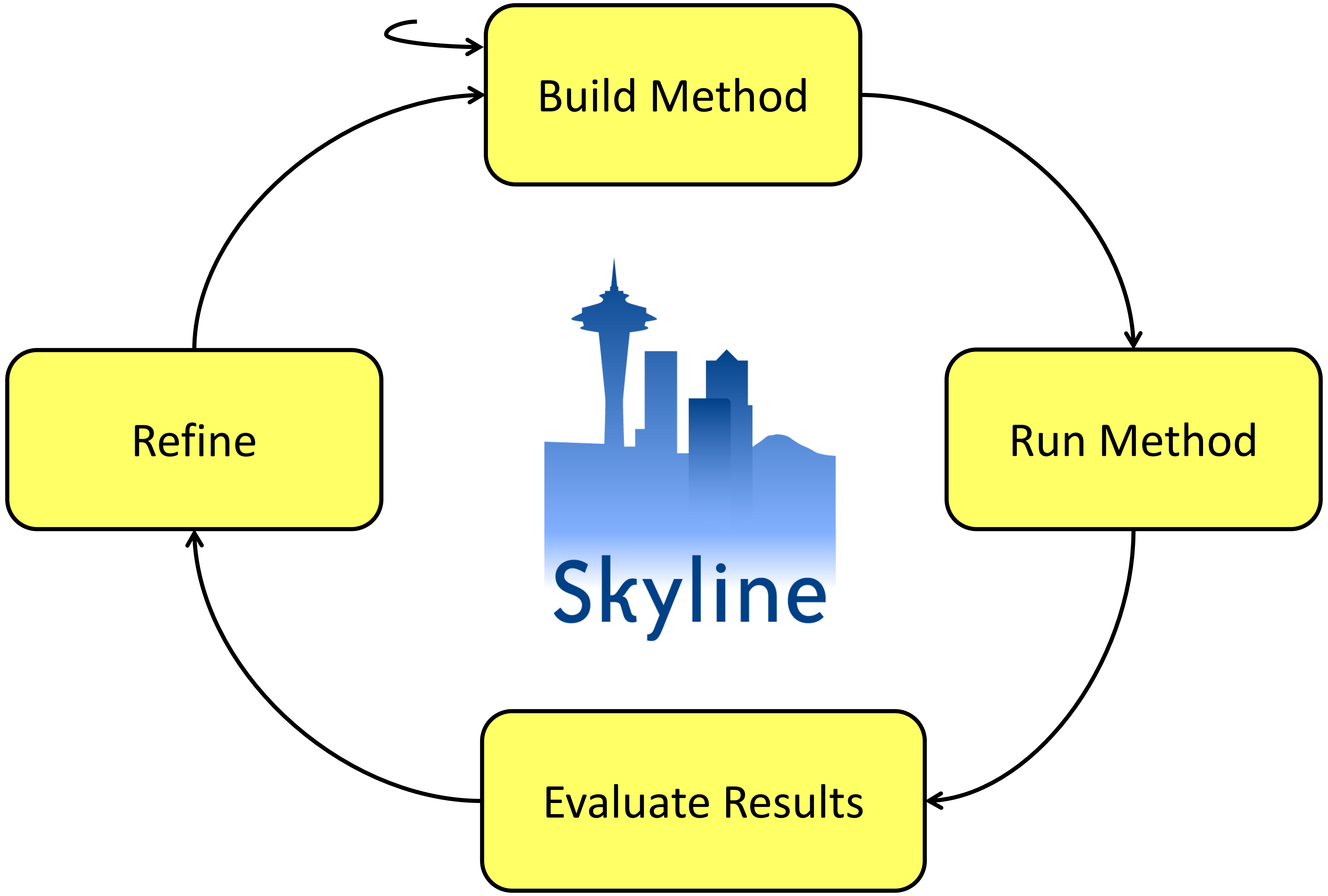
Unrefined Methods

Previous Measurements

Targeted Method Refinement

Quantitative Assay

Targeted Method Refinement

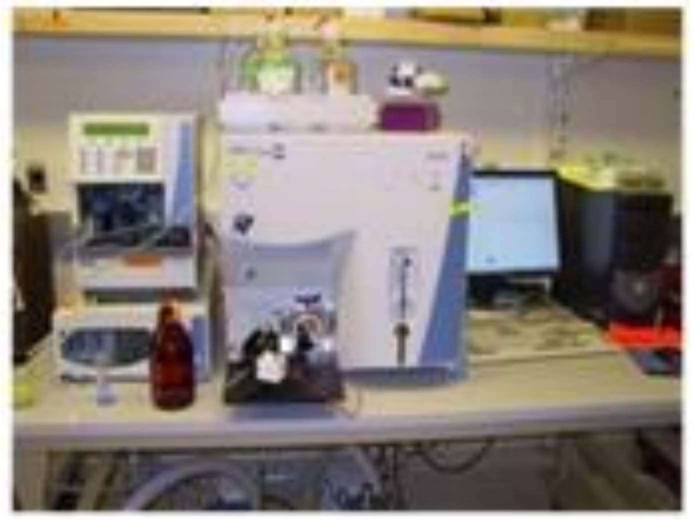


Build Method

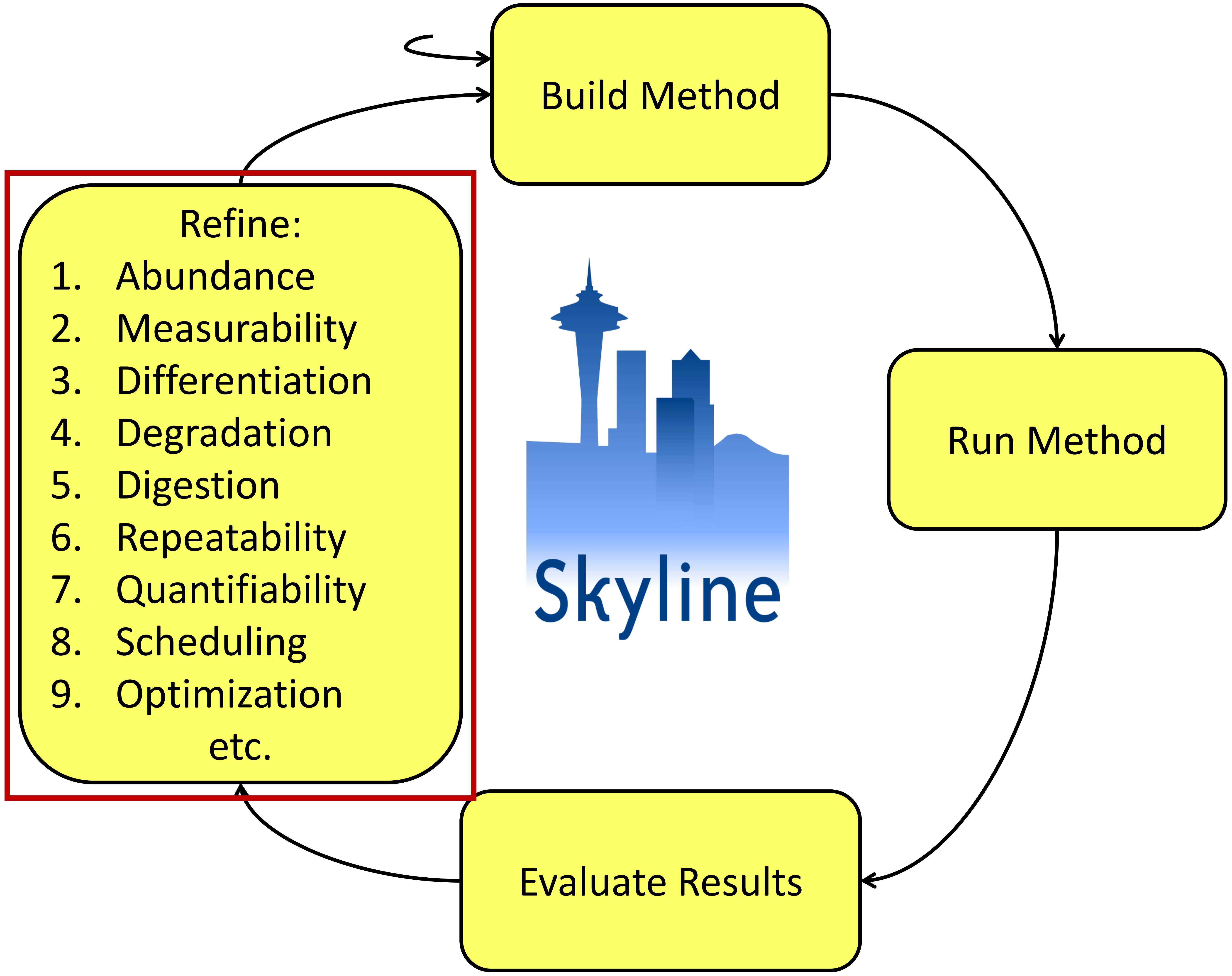
Run Method

Evaluate Results

Refine



Targeted Method Refinement



Build Method

Run Method

Evaluate Results

Refine:

1. Abundance
2. Measurability
3. Differentiation
4. Degradation
5. Digestion
6. Repeatability
7. Quantifiability
8. Scheduling
9. Optimization

etc.

What you see is what you believe

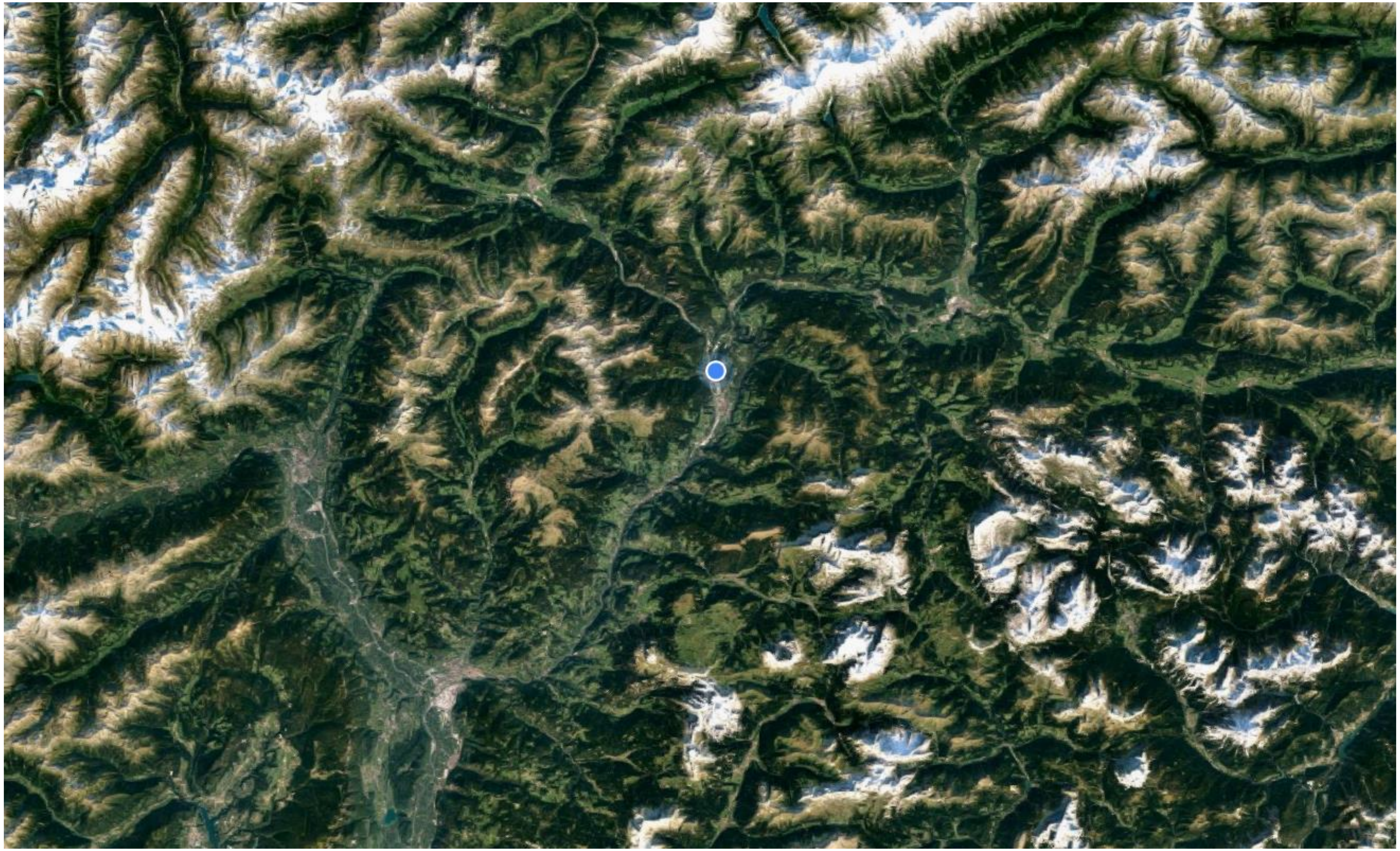
See MORE!!

with

Skyline

“WISIATI” – Daniel Kahneman 2011

The Satellite View



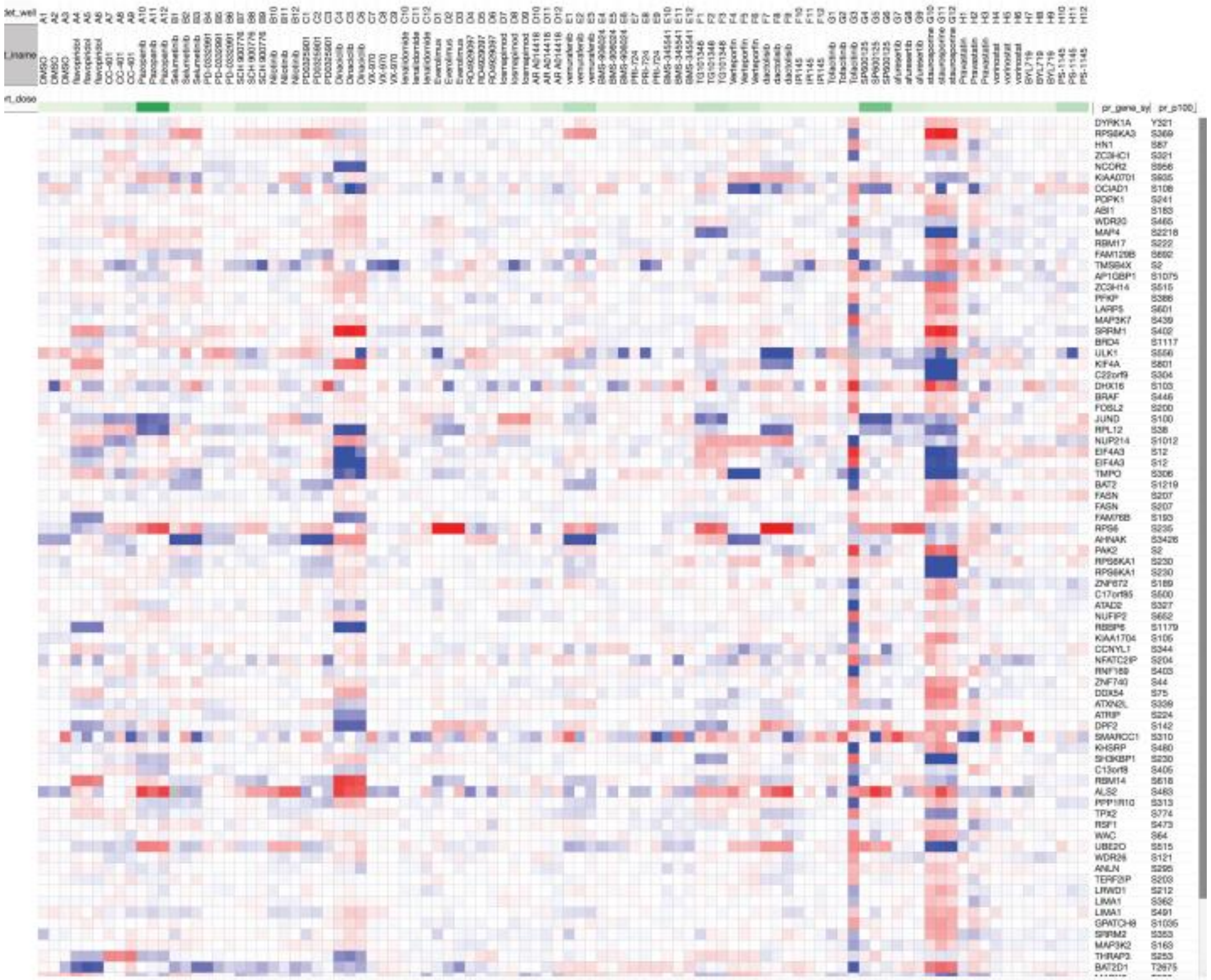
The Satellite View

The Satellite View

Proteomics Satellite View

Sample

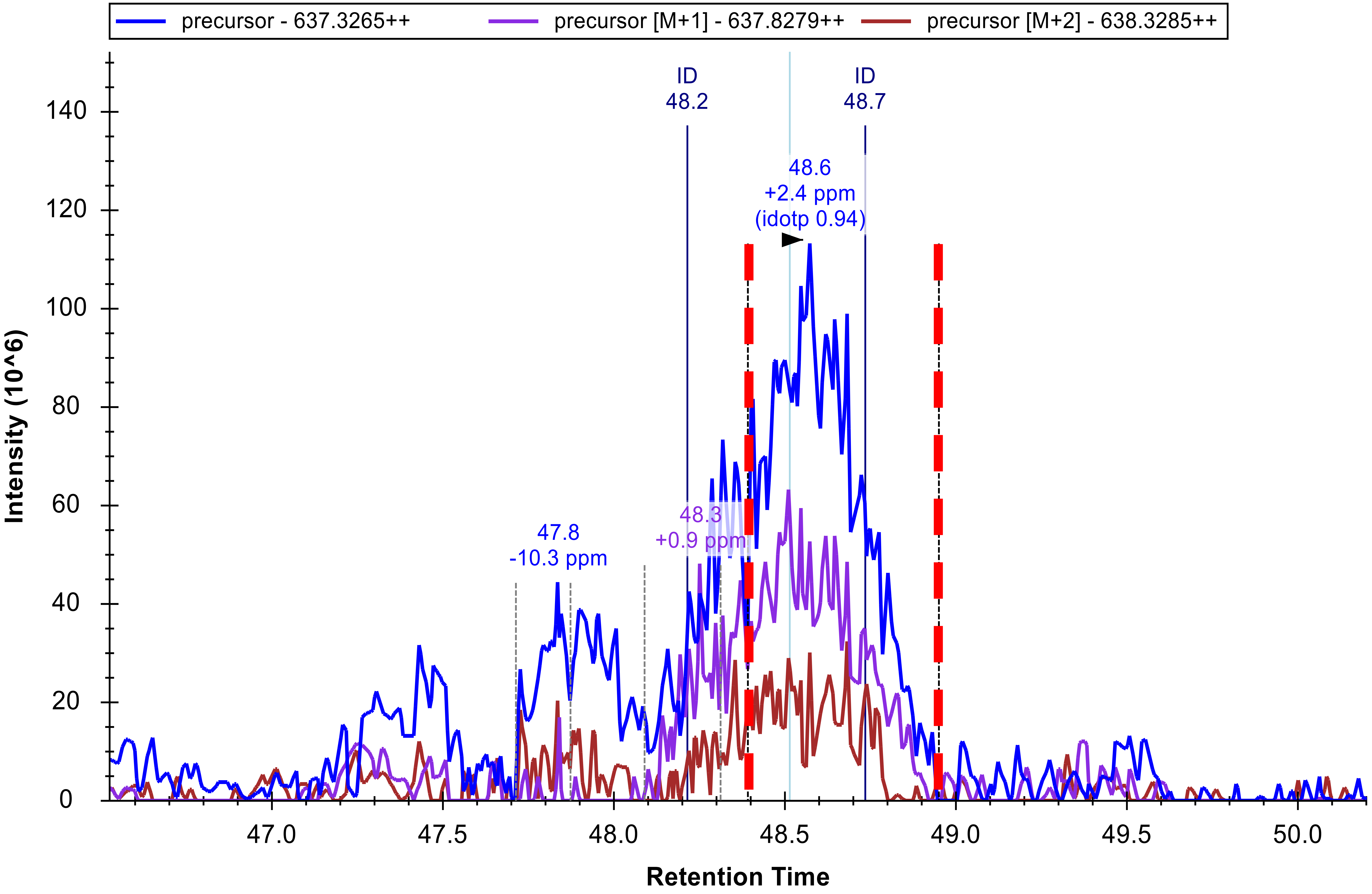
Query



What’s underneath these values?

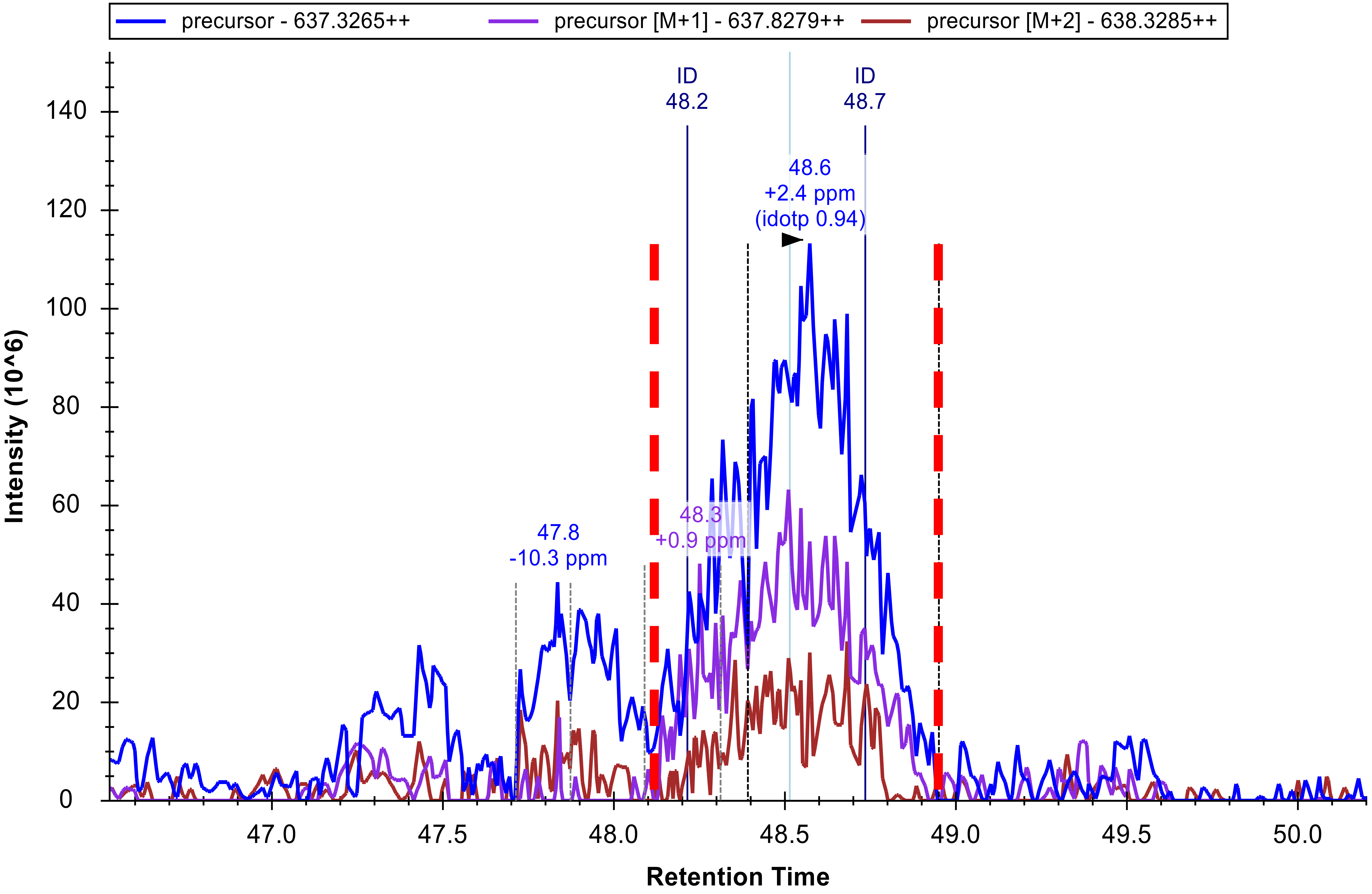
Sources of Error

- Incorrect Integration Boundaries



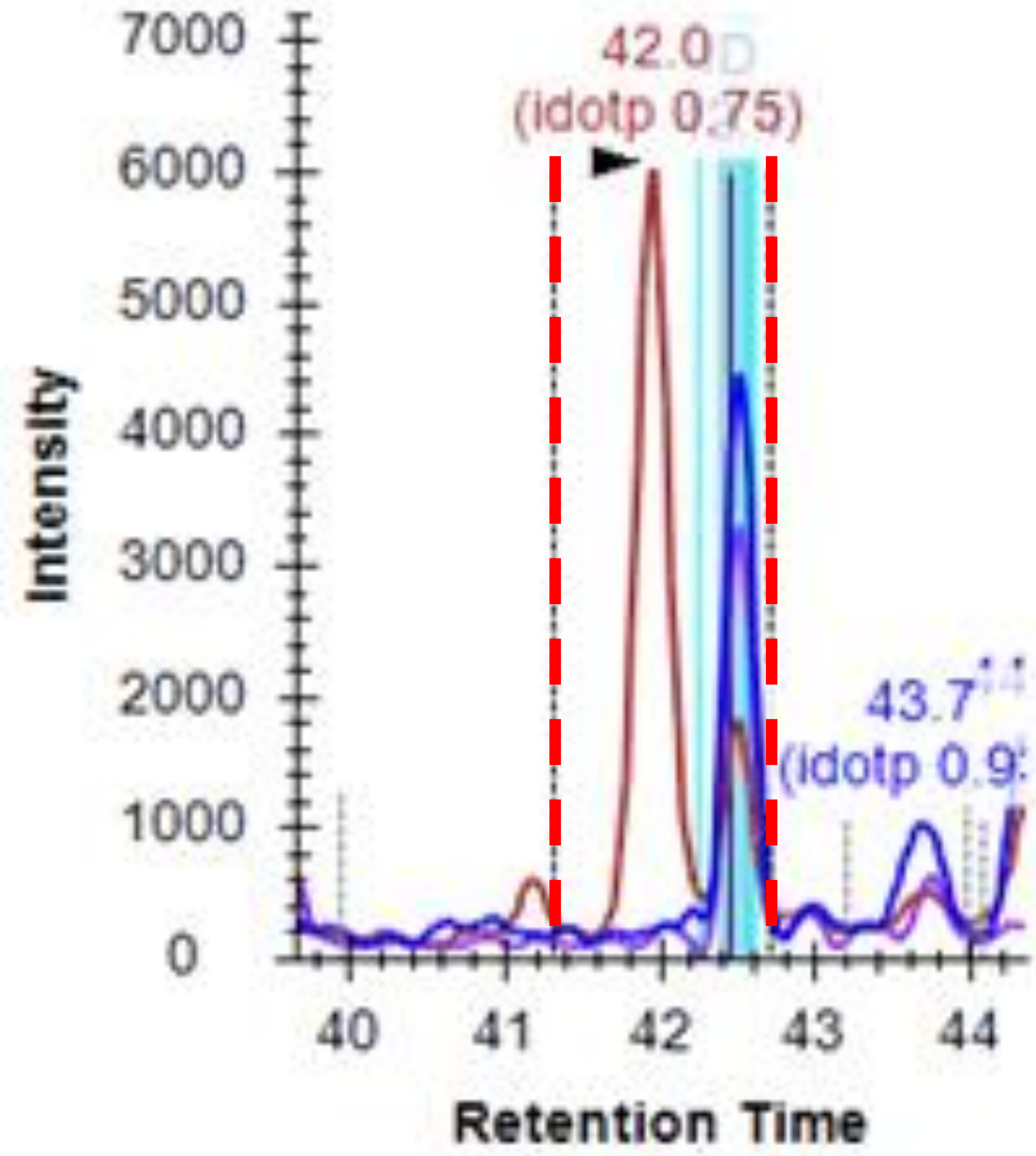
Sources of Error

- Incorrect Integration Boundaries



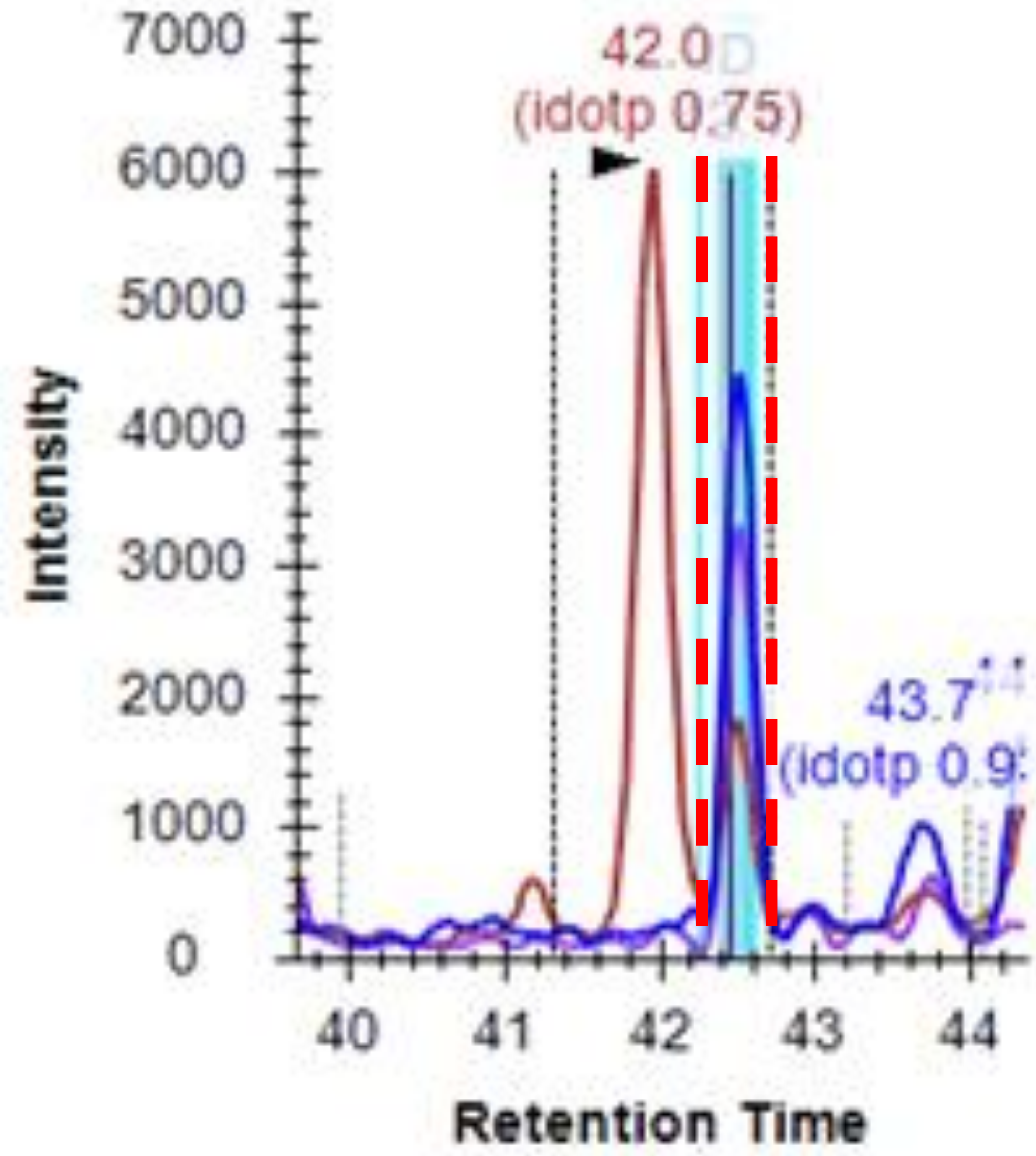
Sources of Error

- Interference
- Incorrect integration boundaries



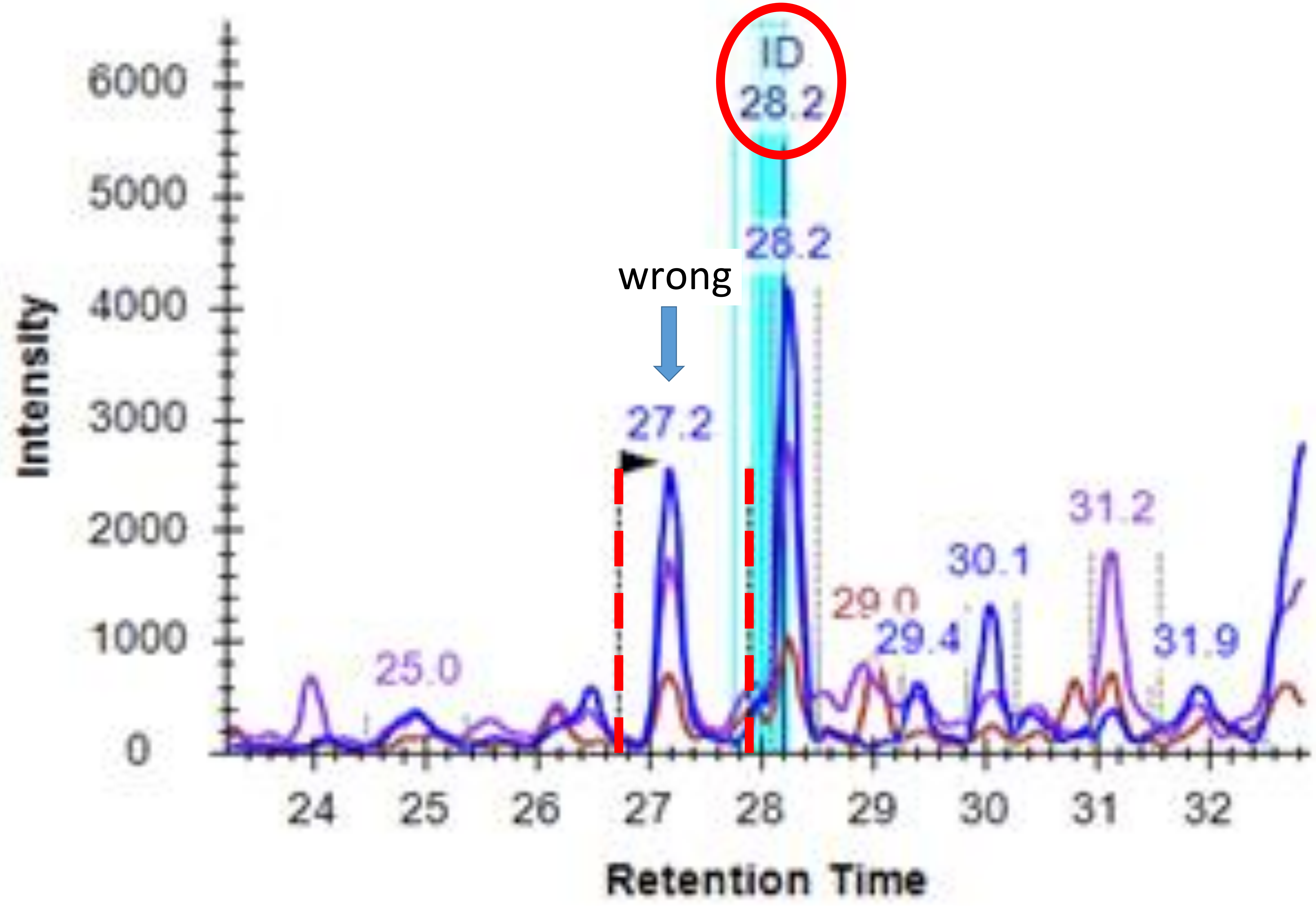
Sources of Error

- Interference
- Incorrect integration boundaries

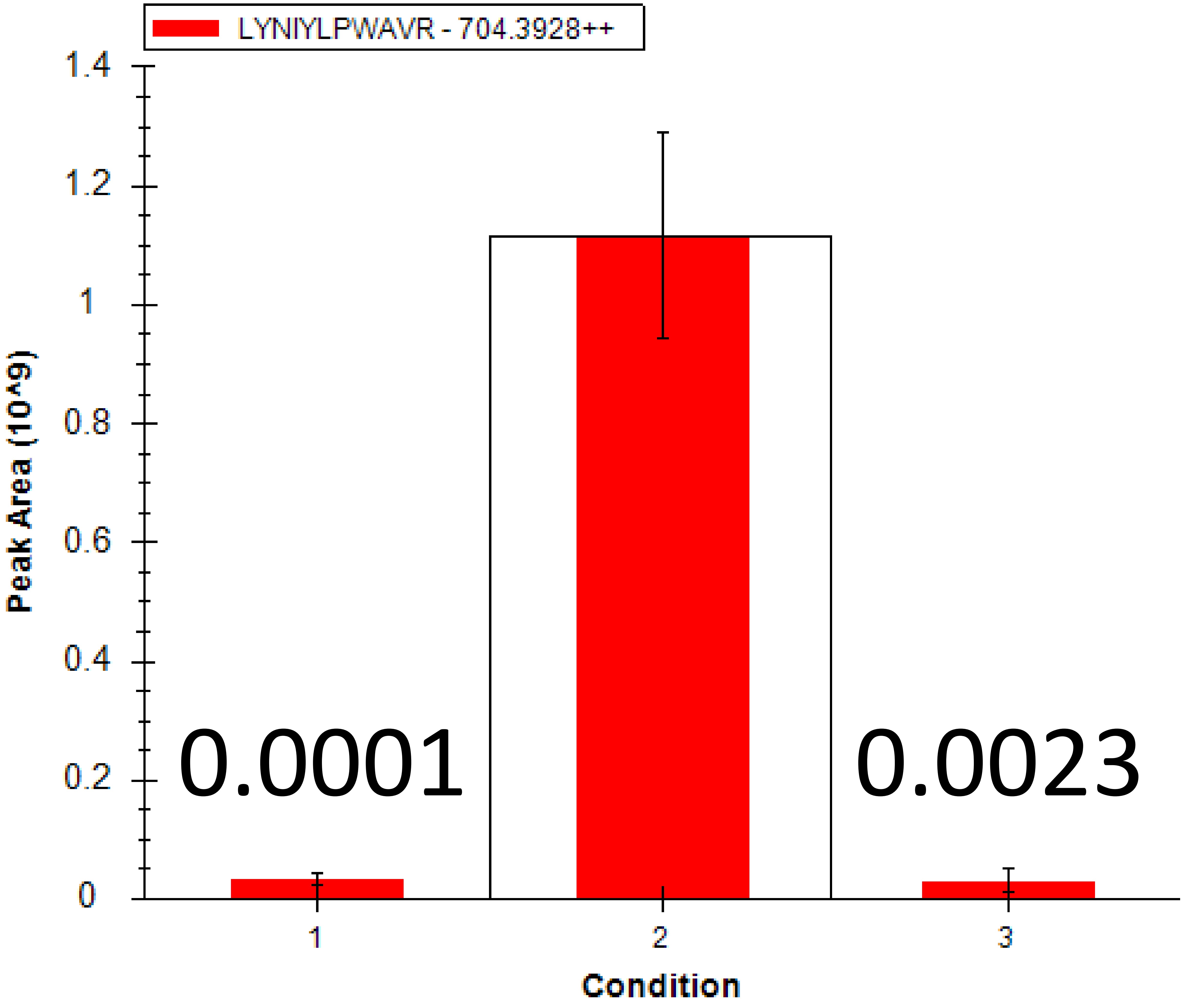


Sources of Error

- Incorrectly picked peaks



Statistically Significant Differences



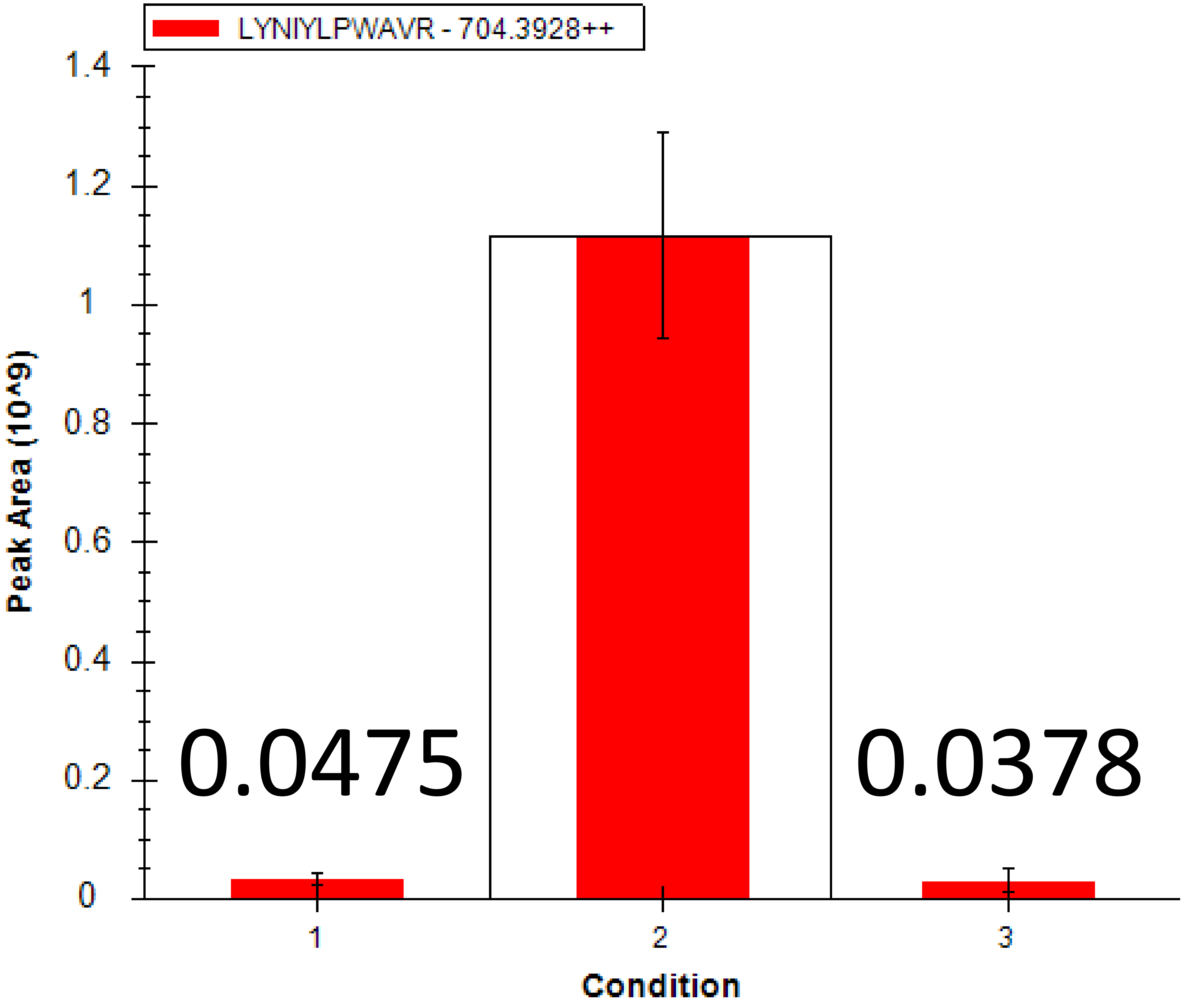
p values

0.0001

0.0023

Statistically Significant Differences

Adjusted
p values

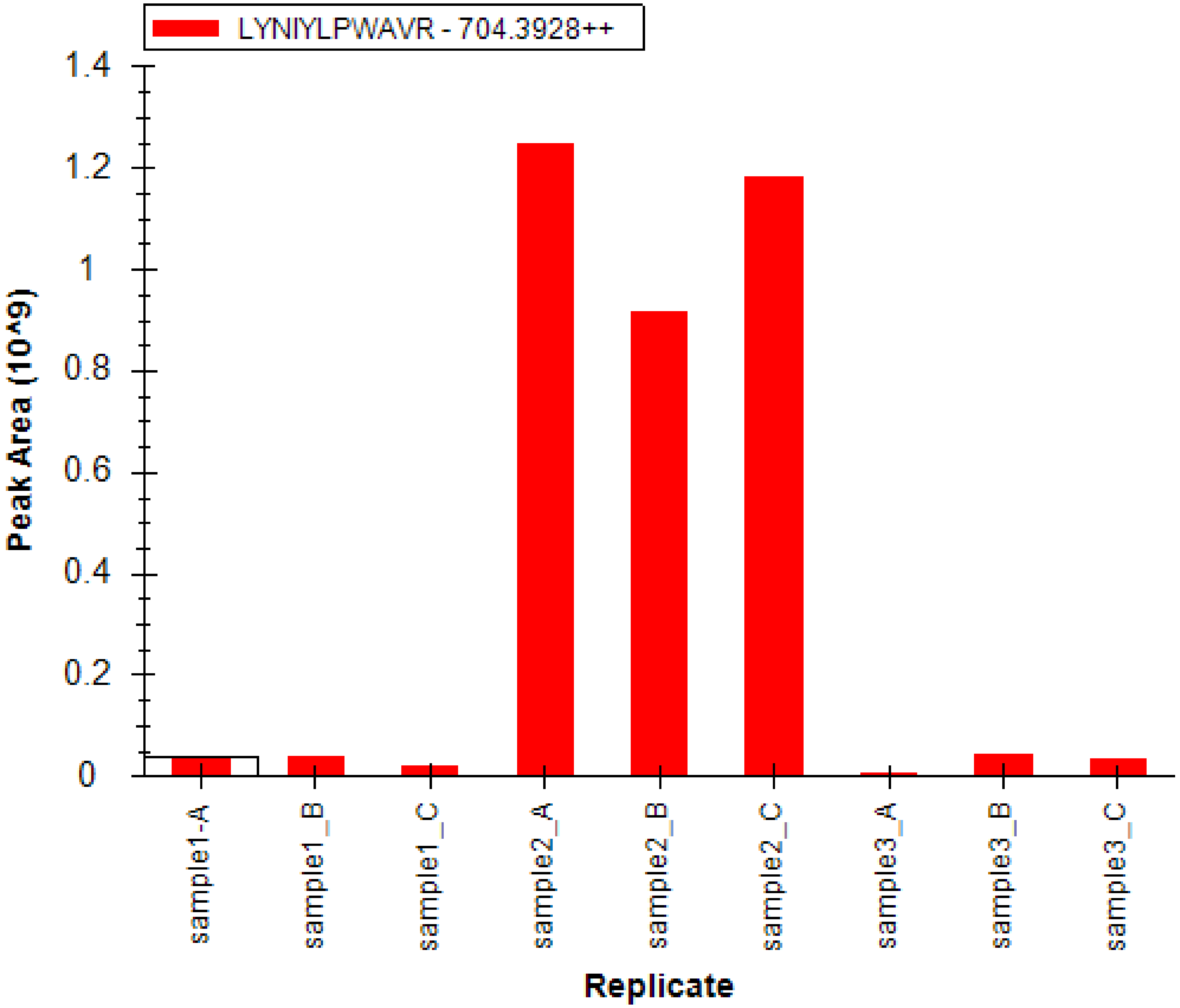


0.0475

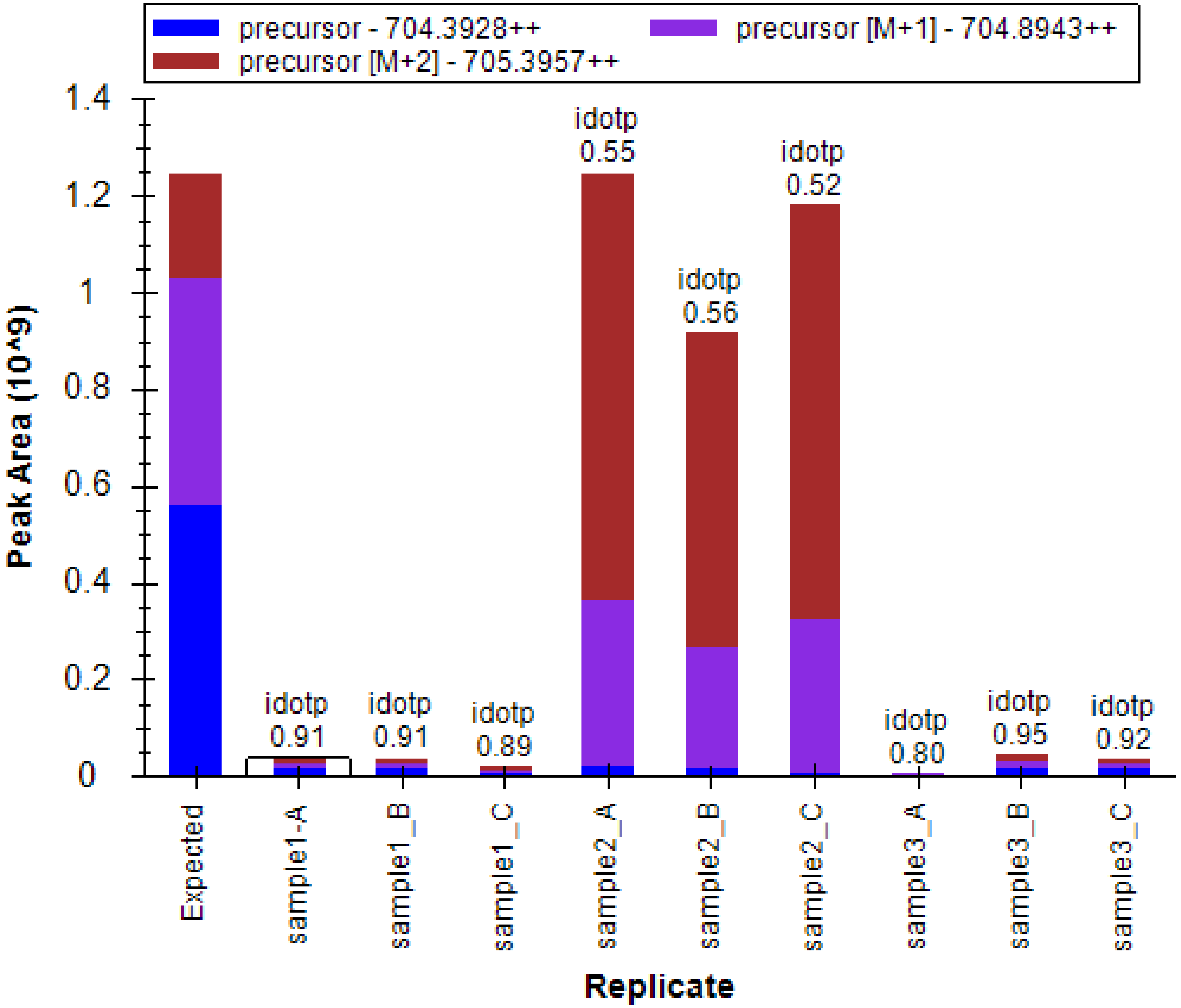
0.0378

26,500
peptides

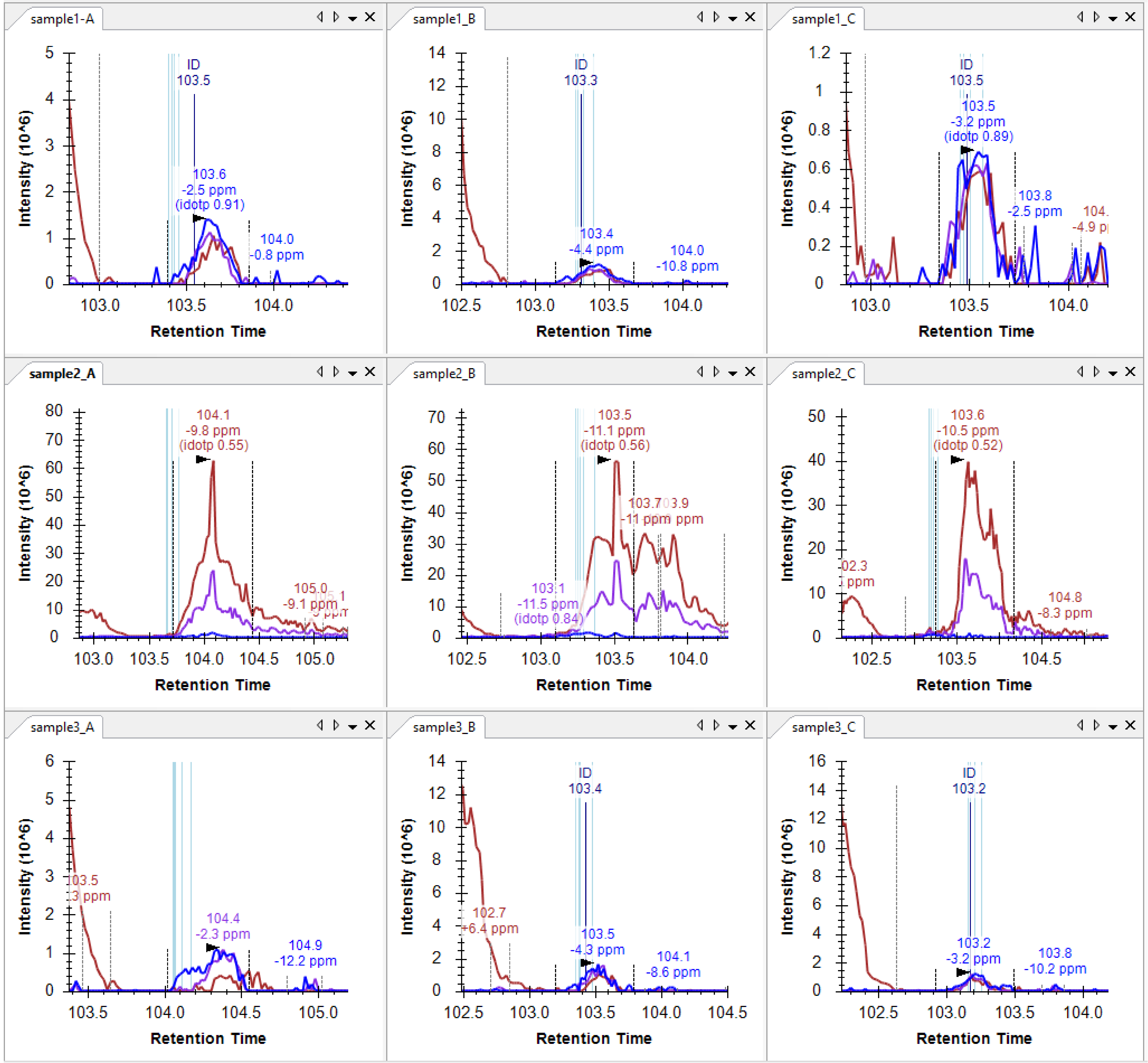
Statistically Significant Differences



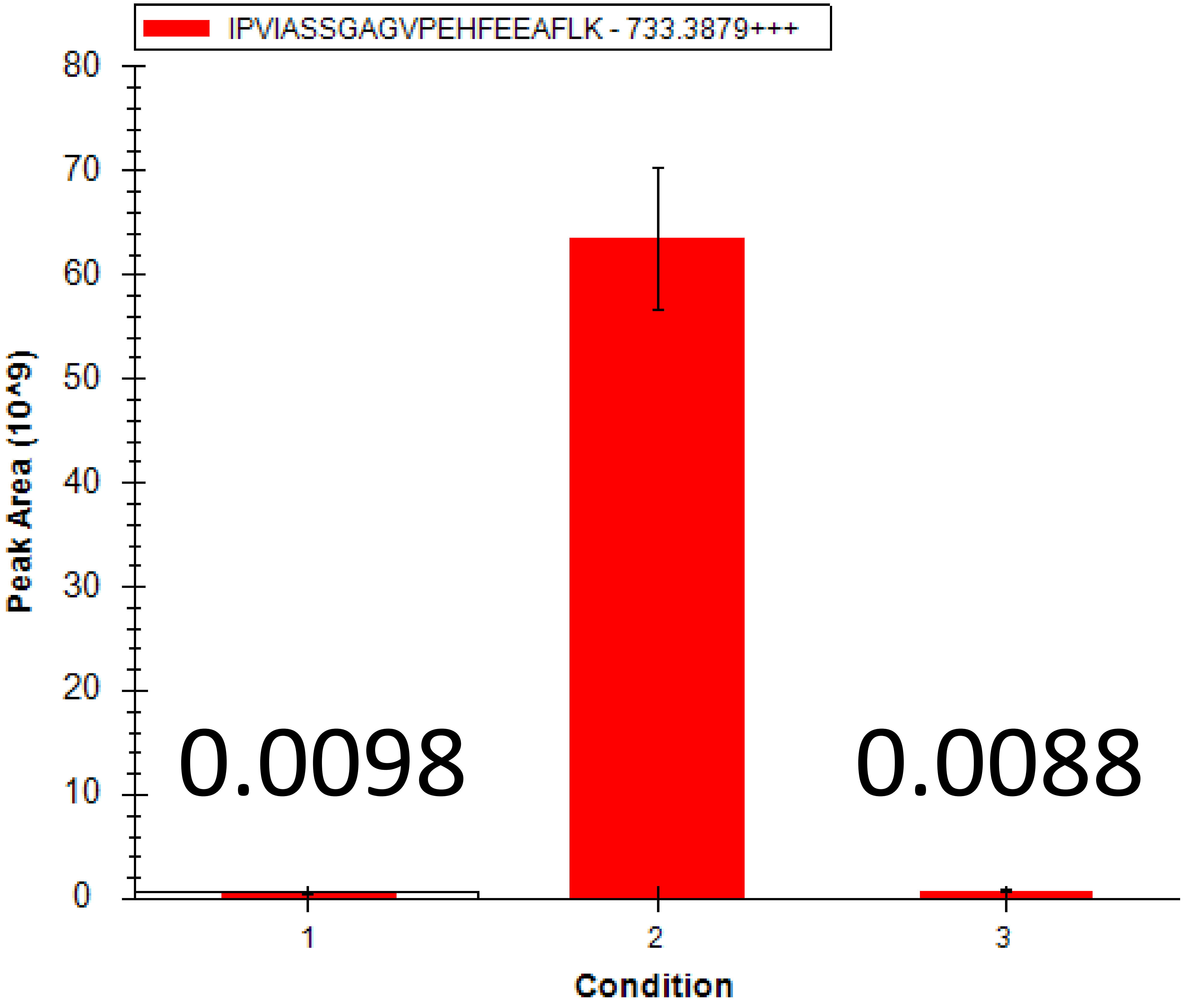
Statistically Significant Differences



The Peak Underneath



Statistically Significant Differences



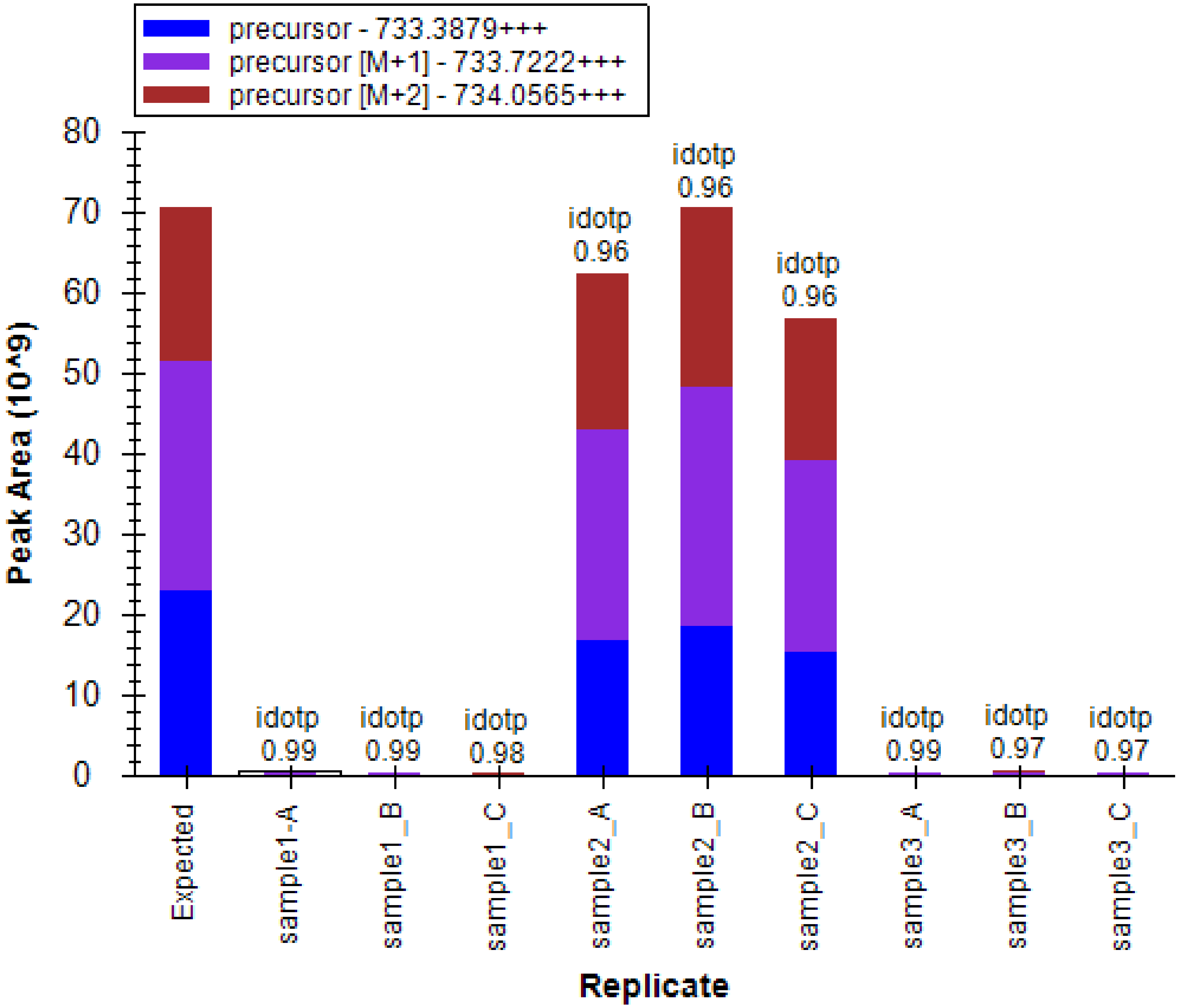
Adjusted
p values

0.0098

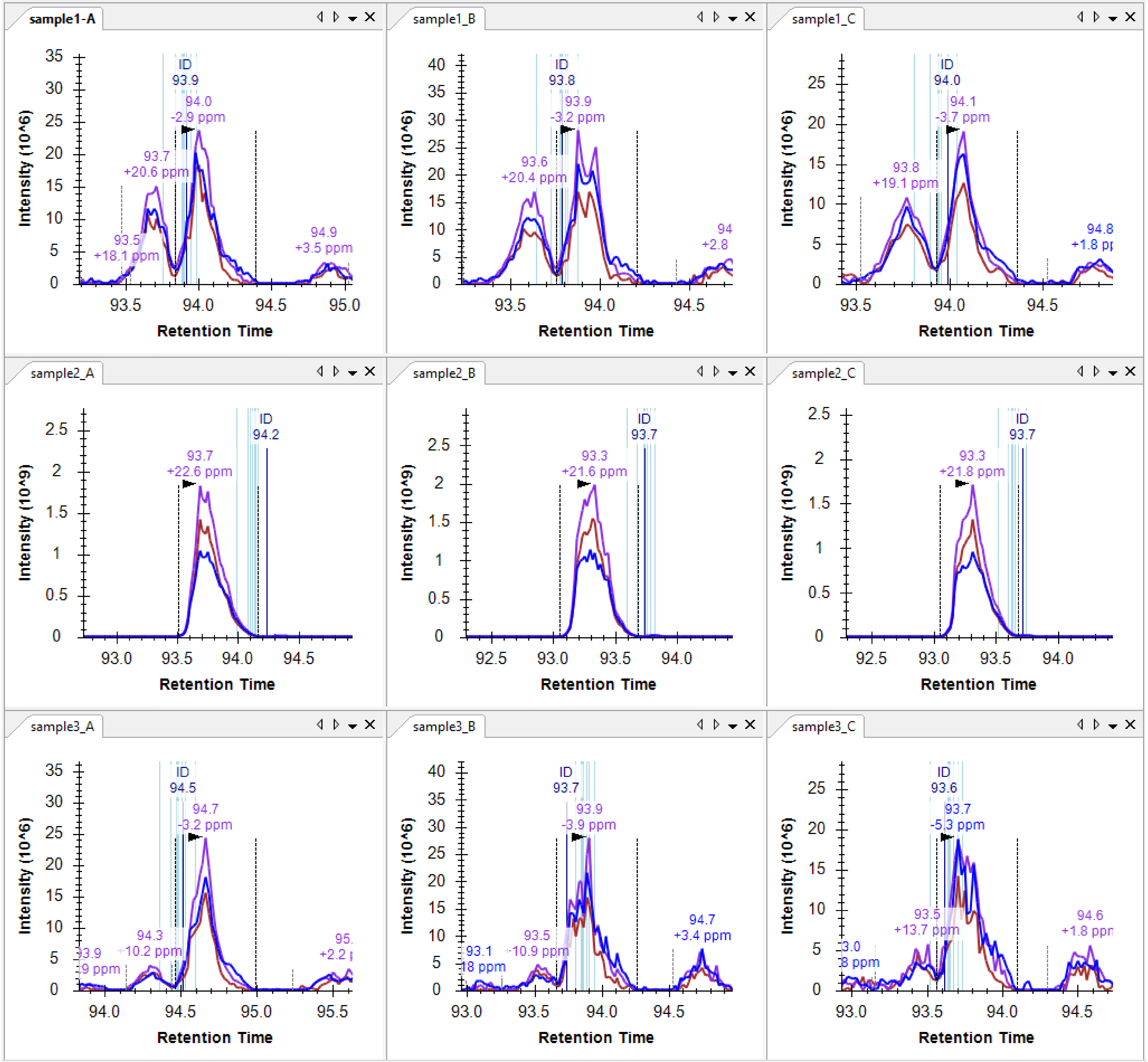
0.0088

26,500
peptides

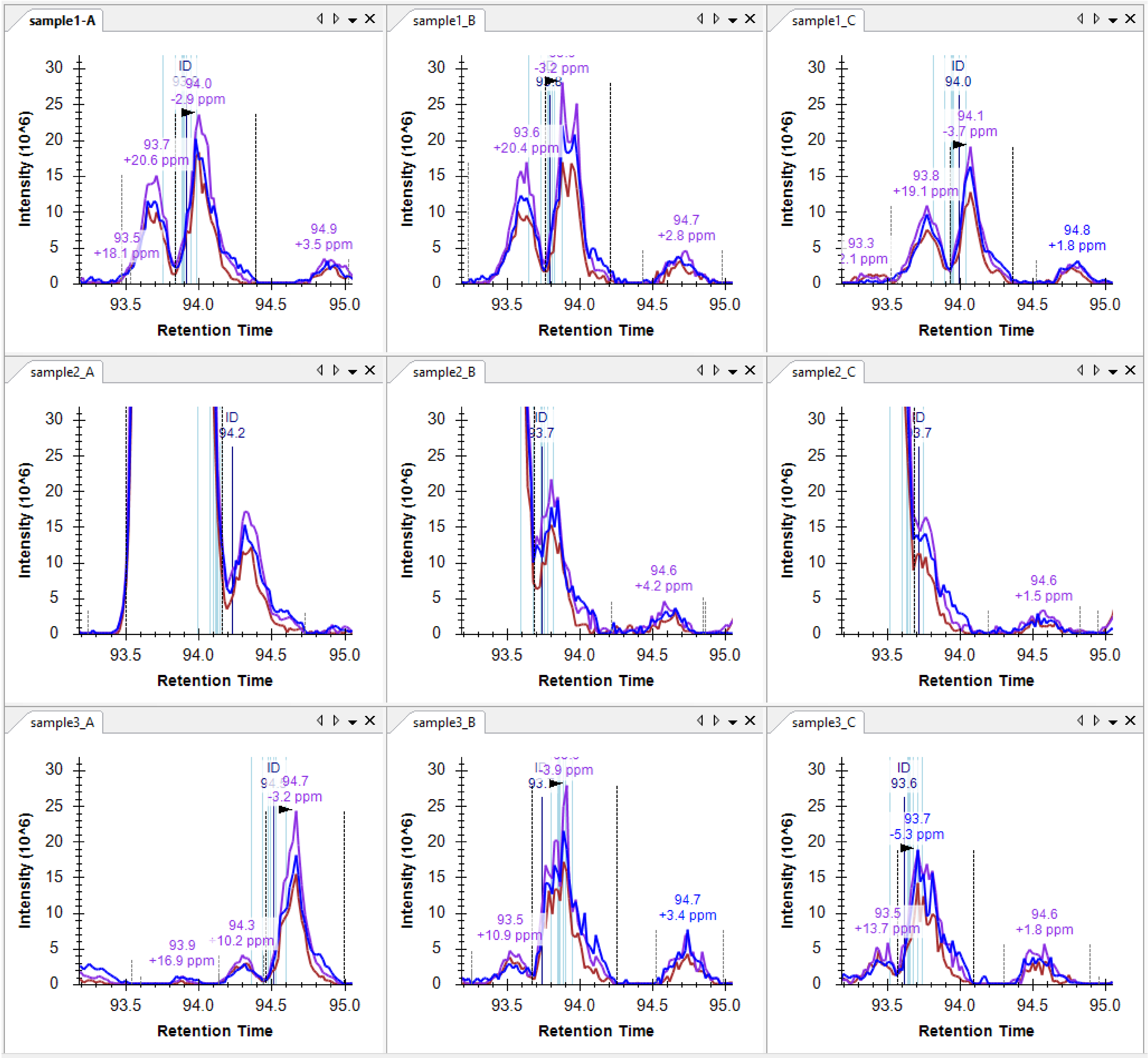
Statistically Significant Differences



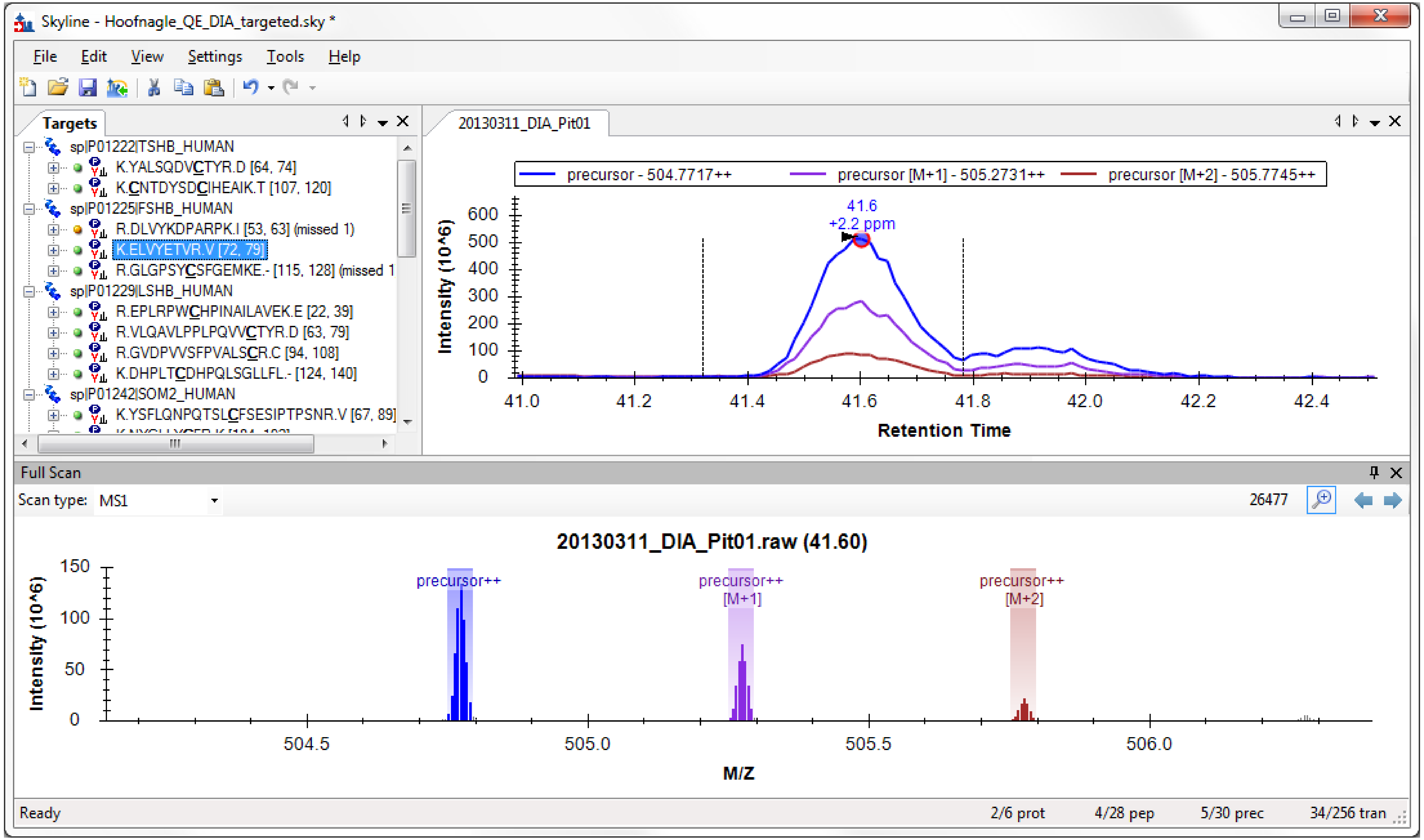
The Peak Underneath



The Peak Underneath



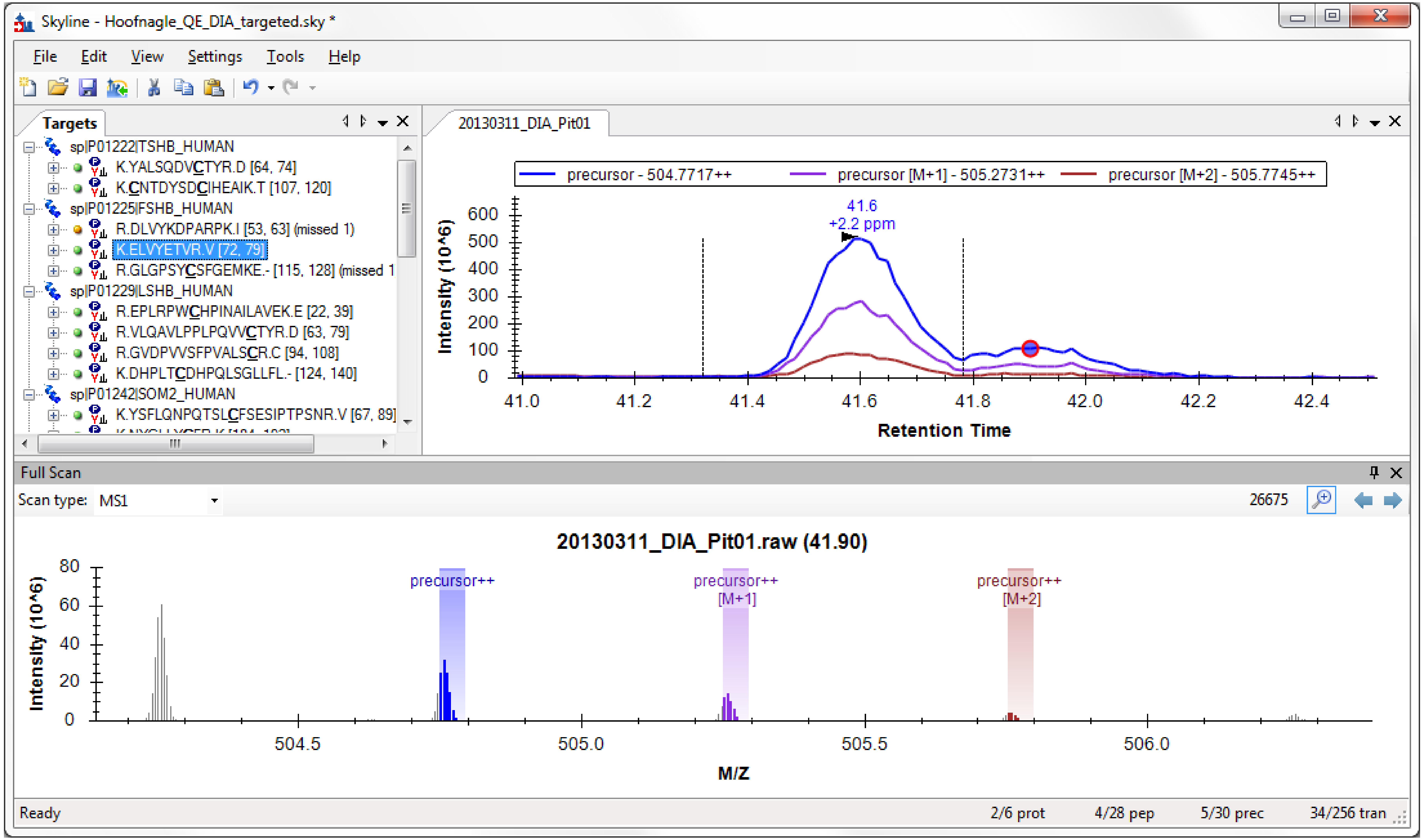
The Spectra Underneath the Peaks



Scan type: MS1

Ready 2/6 prot 4/28 pep 5/30 prec 34/256 tran

“Like pulling the blinders off”



- sp|P01222|TSHB_HUMAN
  - K.YALSQDVCTYR.D [64, 74]
  - K.CNTDYSDCIHEAIK.T [107, 120]
- sp|P01225|FSHB_HUMAN
  - R.DLVYKDPARPK.I [53, 63] (missed 1)
  - K.ELVYETVR.V [72, 79]
  - R.GLGPSYCSFGEMKE.- [115, 128] (missed 1
- sp|P01229|LSHB_HUMAN
  - R.EPLRPWCHPINAILAVEK.E [22, 39]
  - R.VLQAVLPPLPQVVCTYR.D [63, 79]
  - R.GVDPVVSFPVALSCR.C [94, 108]
  - K.DHPLTCDHPQLSGLLFL.- [124, 140]
- sp|P01242|SOM2_HUMAN
  - K.YSFLQNPQTSLCFSESIPTPSNR.V [67, 89]

Scan type: MS1

Skyline Team


Brendan MacLean


Nick Shulman


Brian Pratt


Vagisha Sharma


Rita Chupalov


Kaipo Tamura


Matt Chambers


Nat Brace


Brian Connolly


Ali Marsh