Meeting the Design, Development and Implementation Challenges of >100-Plex Quantitative Assays for Proteins in Plasma: A Large-Scale, NCI-CPTAC Interlaboratory Study

Susan E. Abbatiello, Birgit Schilling, Lisa Zimmerman, Corbin Whitwell, Brendan MacLean, Daniela Tomazela, Pawel Sadowski, Angela Jackson, Mousumi Ghosh, Hasmik Keshishian, Terri A. Addona, Jeffrey R. Whiteaker, Simon Allen, Michael Burgess, Xingdong Feng, Nell Sedransk, D.R. Mani, Steven C. Hall, Steven A. Carr, CPTAC Network

BROAD INSTITUTE

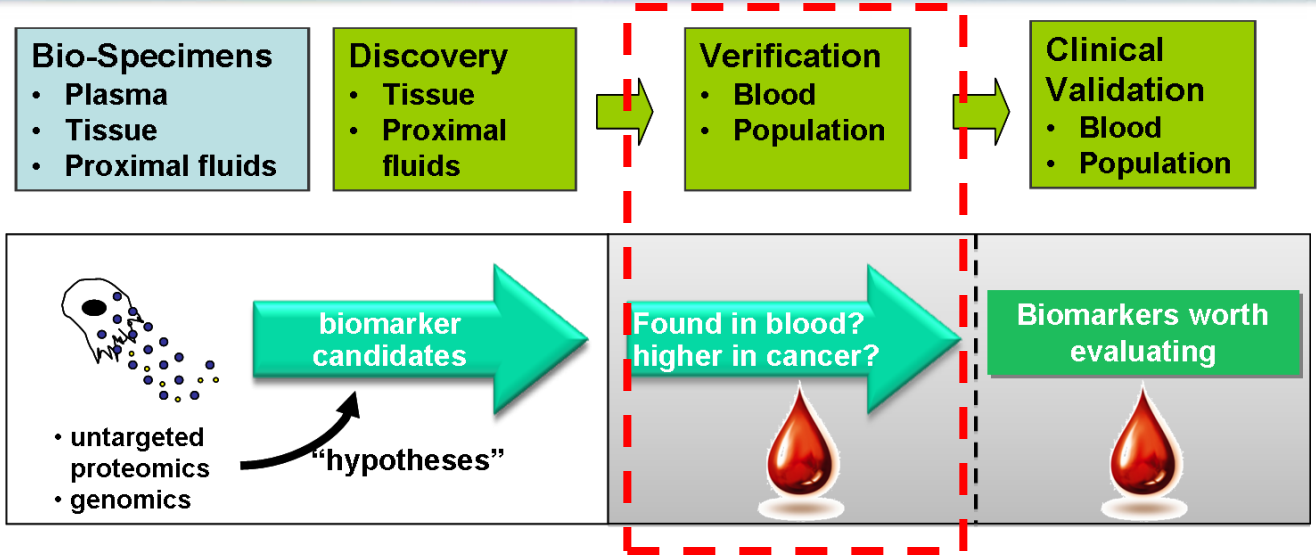
CLINICAL PROTEOMIC TECHNOLOGIES FOR CANCER

DEPARTMENT OF HEALTH & HUMAN SERVICES · USA

NATIONAL INSTITUTES OF HEALTH

NATIONAL CANCER INSTITUTE

CPTAC – Clinical Proteomic Technologies Assessment for Cancer



NCI established CPTC October 2006 to Support Biomarker Development

- **Evaluate and standardize proteomic validation platforms for analysis of cancer-relevant proteomic changes in human clinical specimens.**

Is SID-MRM-MS Technology Reproducible, Transferrable, and Sensitive? Yes!

Candidate Protein Biomarkers

↓

Define "Signature peptides" for candidate biomarkers

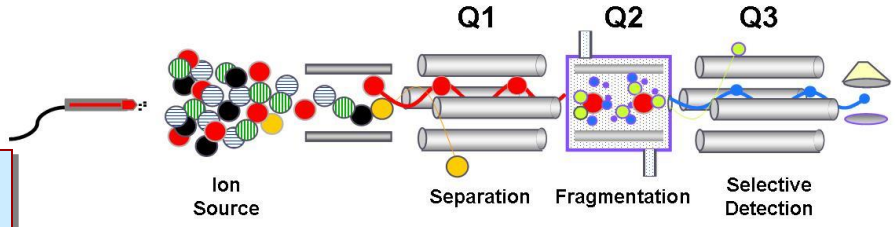
↓

Synthesize $^{13}C_6$-labeled versions of signature peptides for use as internal standards

→ **Spike heavy ($^{13}C_6$)-labeled peptides**

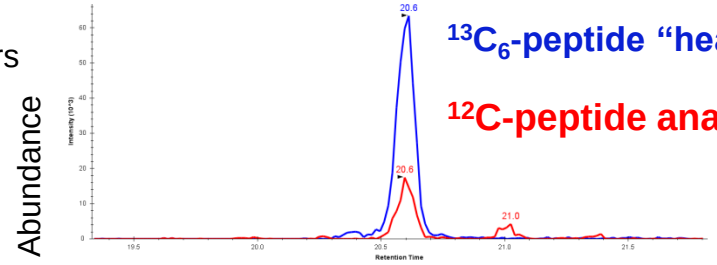
Endogenous ^{12}C signature peptides

↓ **MRM-MS**



- Observed ratio gives precise, relative quantitation across samples
- 10's to 100's peptides can be simultaneously quantified

↓ **Ratio ^{13}C-peptide to ^{12}C-peptide by SID-MRM-MS**



$^{13}C_6$-peptide "heavy standard"

^{12}C-peptide analyte

Whiteaker, et al, JPR 2007……Breast cancer
Keshishian et al, MCP, 2007 and 2009…Cardiovascular markers
Hoofnagle et al, Clin. Chem. 2008……Thyroglobulin
Addona et al, Nat. Biotech. 2009……..……Interlab study
Kuhn et al, Clin Chem 2009……IL-33, Troponin I
Williams et al, JPR 2009……C-Reactive Protein
Ossola et al, Methods Mol. Bio., 2011…..Glycated peptides
Selevsek et al, Proteomics, 2011……..Urine proteins

Pushing the Envelope in SID-MRM-MS Technology

Reproducibility:

Steady retention times and peak areas

Transferability:

Easy method transfer between laboratories

Throughput:

Rapid analysis time or larger number of targets per assay

Sensitivity:

Low ng/mL quantitation limits in plasma

Accuracy:

Getting as close to absolute quantification as possible

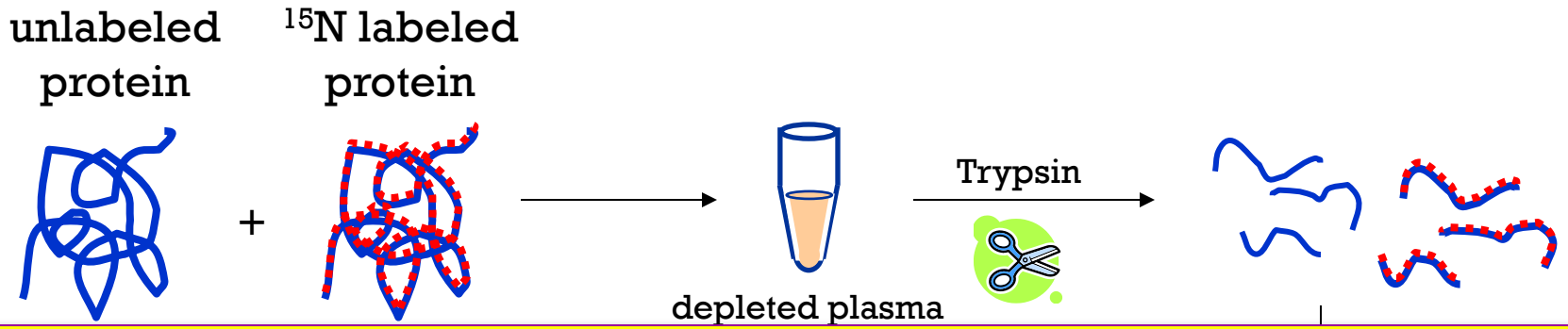
Minimizing Subjectivity and Manual Intervention:

Automated data quality filtering and interference detection

Statistical Pipeline:

Streamlined data processing/analysis for LOD/LOQ, variability, figures of merit

CPTAC VWG Study 9 – Targeting 35 Proteins in Depleted Plasma, 123 Peptide Targets



Goals:

- Demonstrate cancer relevancy
- Prove feasibility of > 100-plex (34 proteins) assays in plasma
- Improve LOD and LOQ by depleting abundant proteins
- Demonstrate true quantitative accuracy and evaluate depletion/digestion recovery using heavy labeled proteins
- Conduct blinded verification study to assess accuracy, precision and reproducibility across multiple sites and instrument platforms
- Evaluate system suitability test in context of this large-scale inter-lab study

35 proteins, 10 participating sites, 15 instruments, 4 Vendors

From 10's of Peptides to 100's of Peptides: It's a Different Game

- ^{15}N Protein Characterization
- Automated transition selection
- Autointerference
- Retention time shift
- Digestion
- Calculating LOD/LOQ – how many points in curve?
- Automated data processing

Protein Targets

(5) Aldolase C
(5) Annexin A1
(5) Annexin A4
(5) Annexin A7
(5) Calreticulin
(5) Chloride Intracellular Channel 1
(5) Ezrin
(5) Fascin Homolog 1
(2) Ferritin Light Chain
(5) Flap structure-specific endonuclease I
(5) Galectin
(5) Glutathione S-transferase pi
(5) Glyoxalase I
(5) Growth Factor Receptor Bound Protein 2
(5) Heat Shock 27 kDa Protein 1
(5) Interleukin 18
(5) PDZ and LIM/ELFIN/CLIM1/CLP36
(4) Peroxiredoxin 2
(4) Peroxiredoxin 4
(1) Protein S100-A1
(2) Protein S100-A2
(2) Protein S100-B
(3) RAD23 Homolog B
(4) Synuclein, gamma
(4) Tropomyosin 1
(5) Ubiquitin conjugating enzyme E2C
(4) Ubiquitin conjugating enzyme E2I

Control Proteins

(1) Aprotinin
(2) C-reactive protein
(1) Horseradish peroxidase
(1) Leptin
(2) Myelin basic protein
(1) Myoglobin

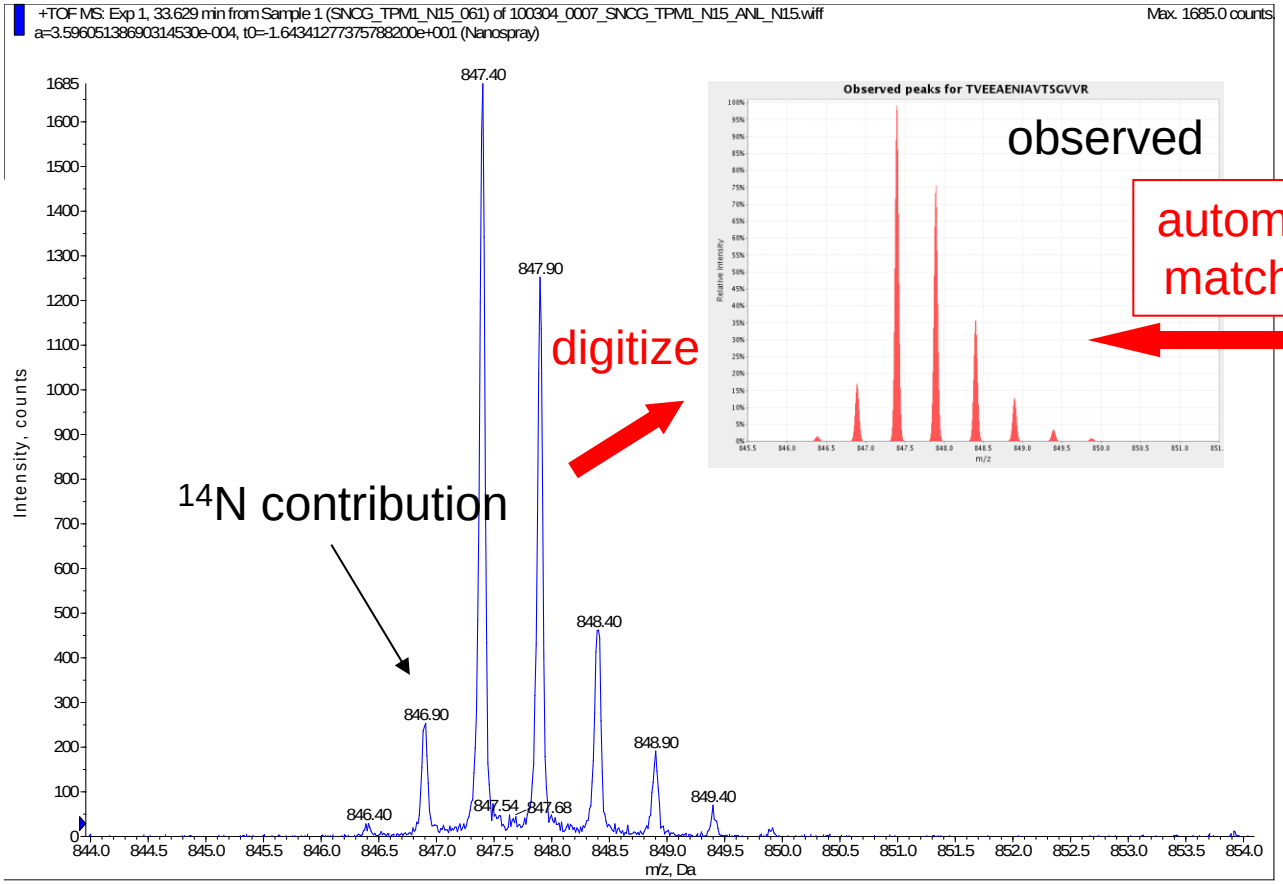
Protein Standard Characterization

Data-dependent MS experiments searched against a database

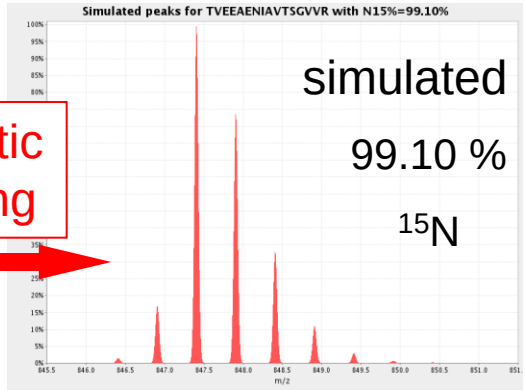
15N Distinct Peptides (#)	14N Distinct Peptides (#)	Protein Name
30	29	Annexin A1 OS=Homo sapiens GN=ANXA1 PE=1 SV=2
36	29	Annexin A4 OS=Homo sapiens GN=ANXA4 PE=1 SV=4
18	17	Chloride intracellular channel protein 1 OS=Homo sapiens GN=CLIC1 PE=1 SV=4
	33	Ezrin OS=Homo sapiens GN=EZR PE=1 SV=4
25	32	Fascin OS=Homo sapiens GN=FSCN1 PE=1 SV=3
28	27	Fructose-bisphosphate aldolase C OS=Homo sapiens GN=ALDOC PE=1 SV=2
11	9	Galectin-1 OS=Homo sapiens GN=LGALS1 PE=1 SV=2
16	14	Gamma-synuclein OS=Homo sapiens GN=SNCG PE=1 SV=2
18	18	Growth factor receptor-bound protein 2 OS=Homo sapiens GN=GRB2 PE=1 SV=1
14	17	Heat shock protein beta-1 OS=Homo sapiens GN=HSPB1 PE=1 SV=2
16	11	Interleukin-18 OS=Homo sapiens GN=IL18 PE=1 SV=1
18	15	Peroxiredoxin-2 OS=Homo sapiens GN=PRDX2 PE=1 SV=5
15	16	Peroxiredoxin-4 OS=Homo sapiens GN=PRDX4 PE=1 SV=1
7	9	SUMO-conjugating enzyme UBC9 OS=Homo sapiens GN=UBE2I PE=1 SV=1
28	25	Tropomyosin alpha-1 chain OS=Homo sapiens GN=TPM1 PE=1 SV=2
11	12	Ubiquitin-conjugating enzyme E2 C OS=Homo sapiens GN=UBE2C PE=1 SV=1
16	15	UV excision repair protein RAD23 homolog B OS=Homo sapiens GN=RAD23B PE=1 SV=1

^{15}N Incorporation in Labeled Proteins

Determined ^{15}N incorporation is 99.10 % for synuclein gamma peptide TVEEAENIAVTSGVVR



automatic matching



Data from Qstar Elite, Buck Institute

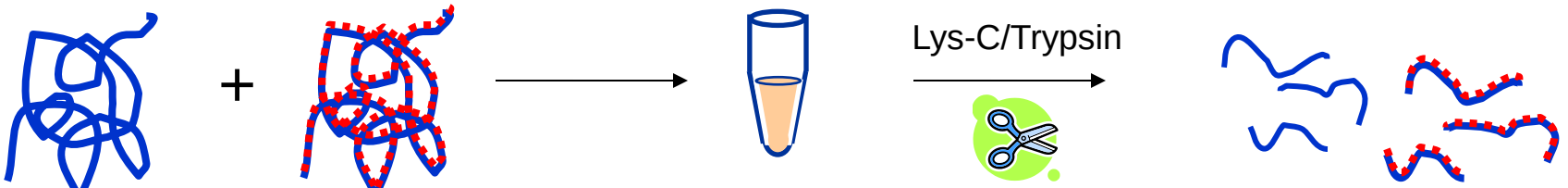
M.P. Cusack

Selected ^{15}N Stable Isotope Incorporation Efficiency

Gamma Synuclein	99.10 % ^{15}N
Tropomyosin 1 (alpha)	99.13 % ^{15}N
Ubiquitin-conjugating enzyme E2I (UBC9)	99.28 % ^{15}N
Ubiquitin-conjugating enzyme E2C	99.03 % ^{15}N
Growth factor receptor-bound protein 2	99.53 % ^{15}N
Heat shock 27kDa protein 1	99.18 % ^{15}N
Chloride intracellular channel 1	99.13 % ^{15}N
Annexin A1	99.55 % ^{15}N
Annexin A4	99.18 % ^{15}N
RAD23 homolog B	98.95 % ^{15}N
Interleukin 18 (interferon-gamma-inducing factor)	99.20 % ^{15}N
Lectin, galactoside-binding, soluble, 1 (galectin 1)	98.98 % ^{15}N
Fascin homolog 1, actin-bundling protein	99.10 % ^{15}N
Peroxiredoxin 4	98.95 % ^{15}N
Peroxiredoxin 2	99.10 % ^{15}N
Aldolase C	99.13 % ^{15}N

2 tryptic peptides for each ^{15}N labeled protein were analyzed for isotopic purity (QSTAR Elite, Buck Institute)

Peptide and Transition Selection is Streamlined using Skyline



DDA LC-MS/MS, Database search

Skyline

Spectral Library

Selection of 123 target peptides

Skyline

Top 5 Product Ions

MRM-MS Data Acquisition

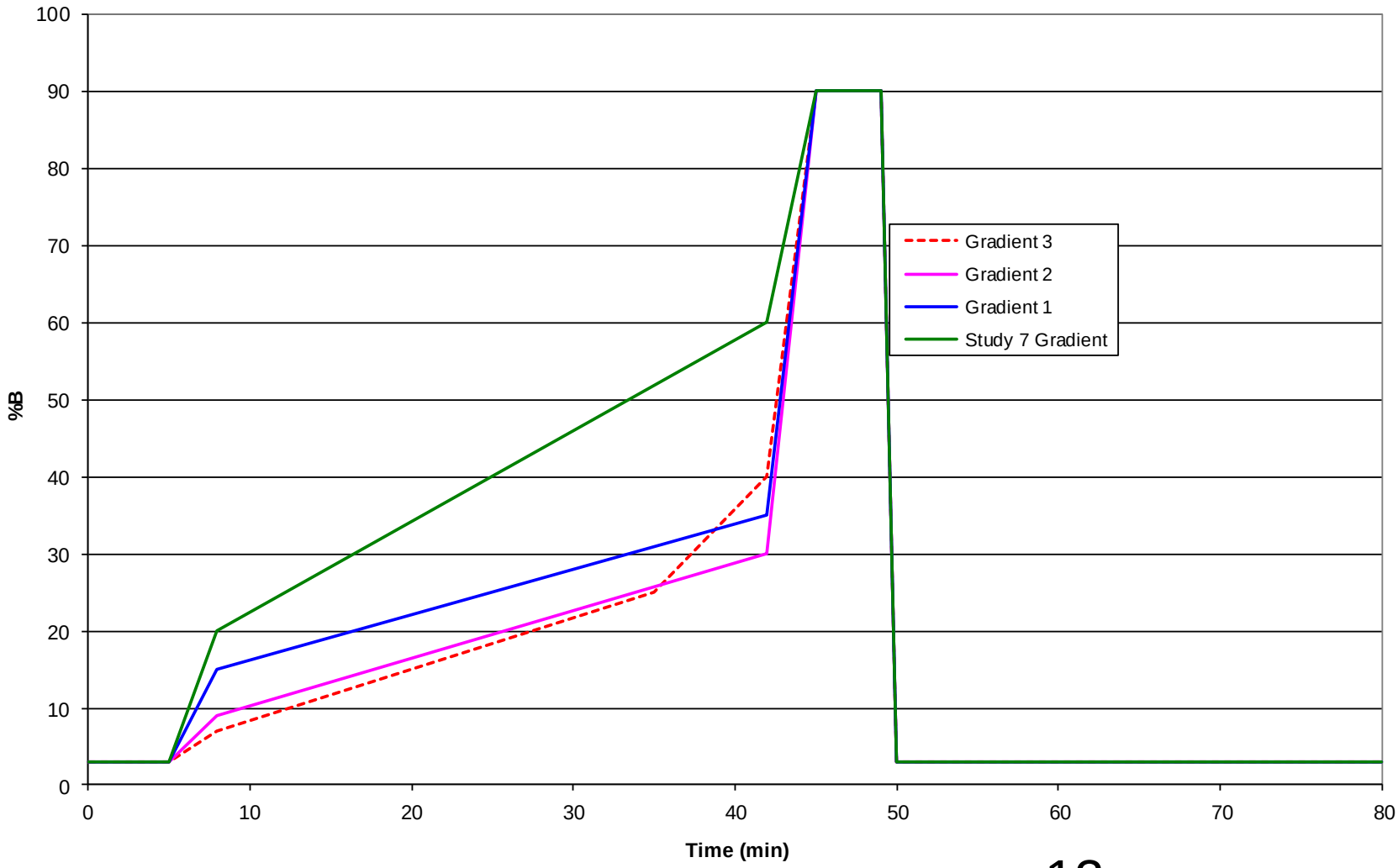
Skyline

- Selection of best 3 ions
- CE Calculation

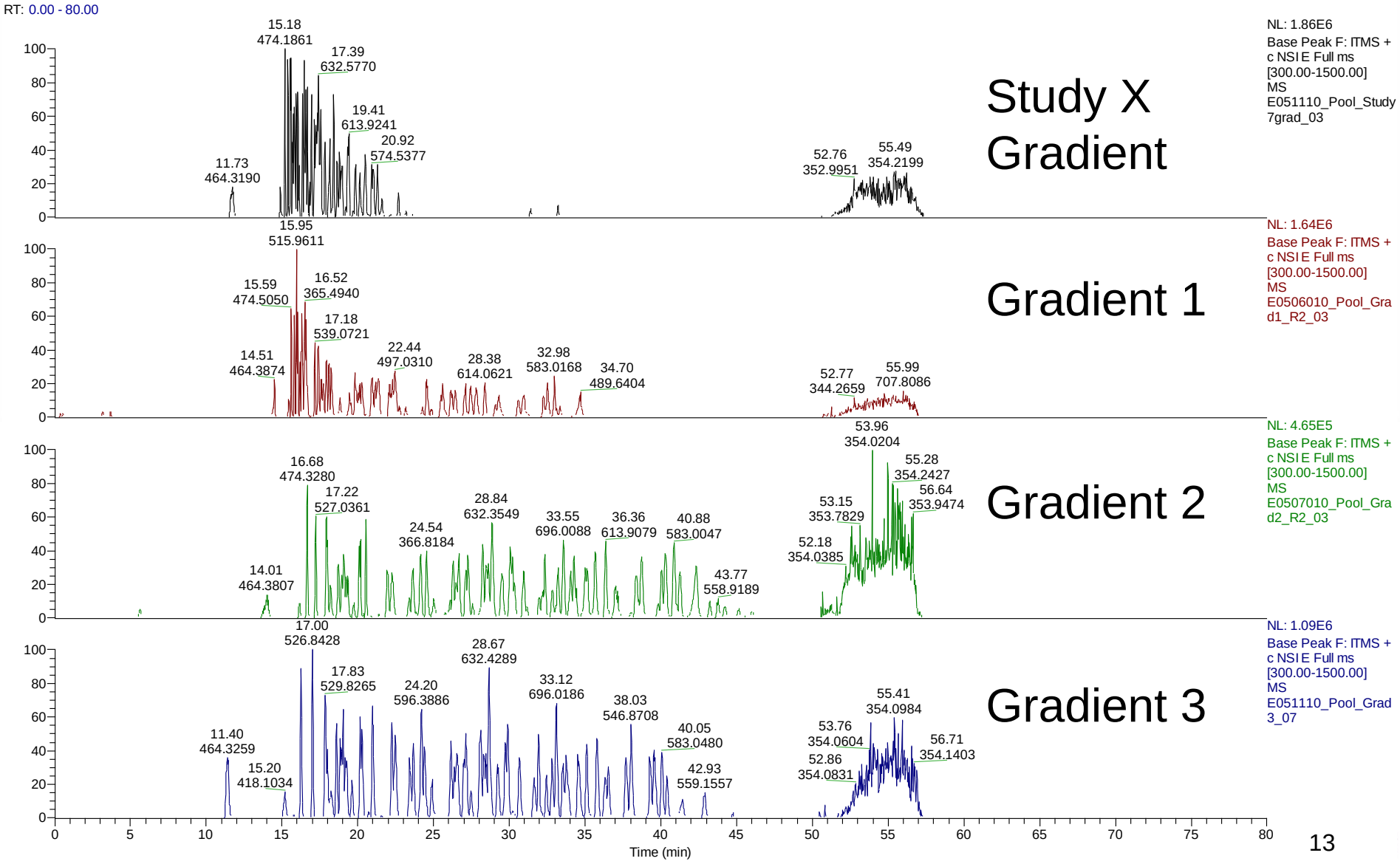
Final Transition List
L/H: 738 Transitions
L/H/15N: 1107 Transitions

CE publication: B. MacLean et al, 2010, Anal Chem

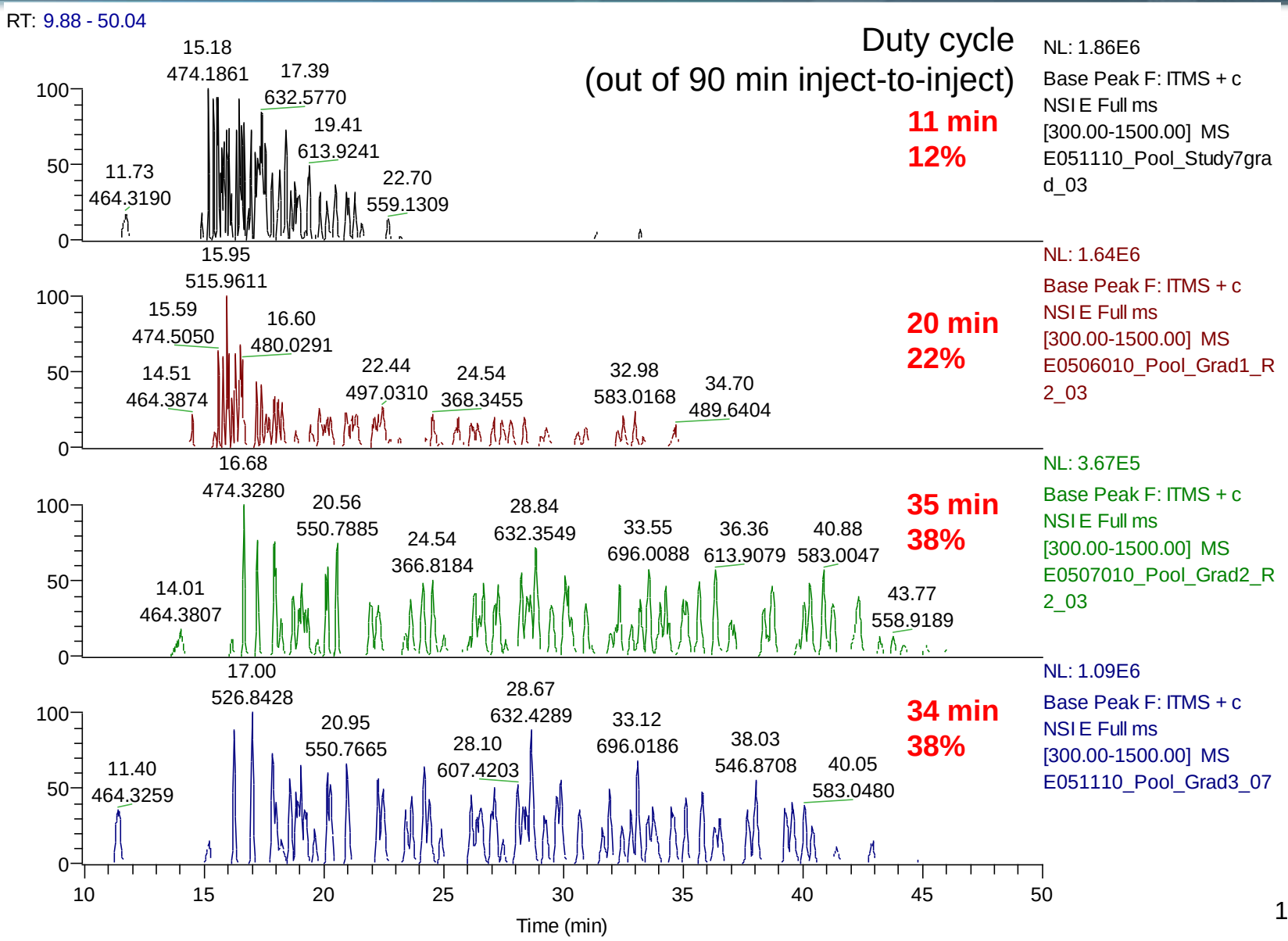
Gradient Development



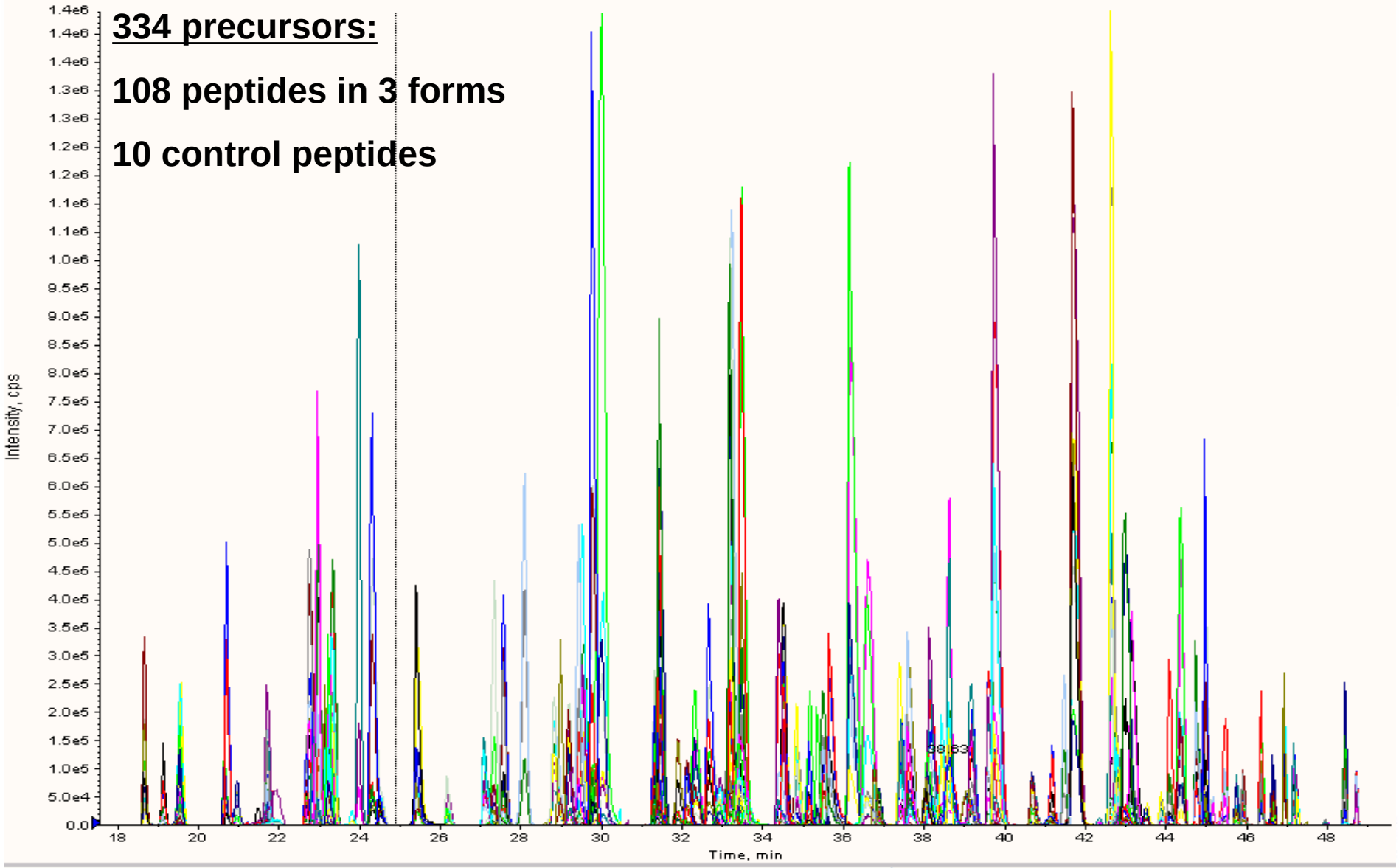
Gradient Optimization – Separate those Peptides!



Zoomed in View of Gradients and Separation Time

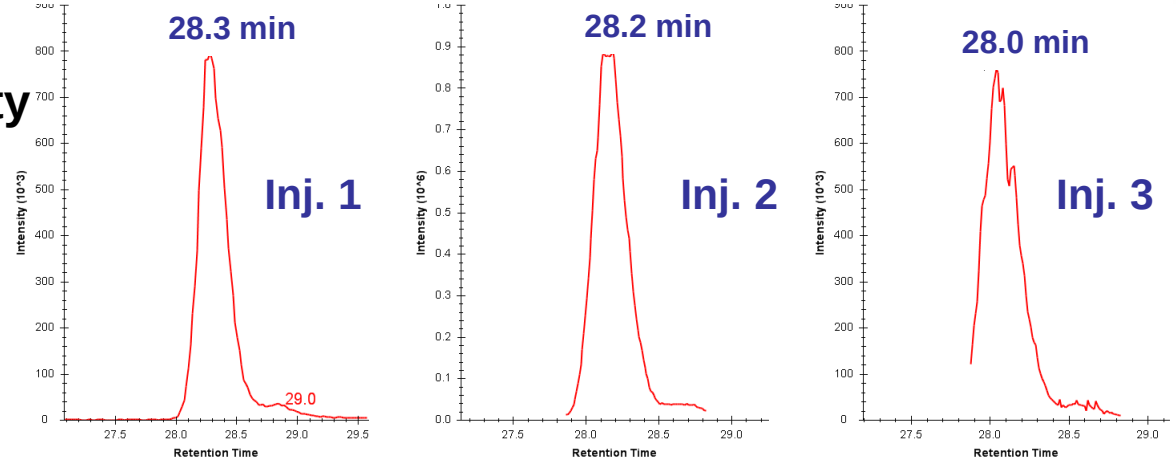


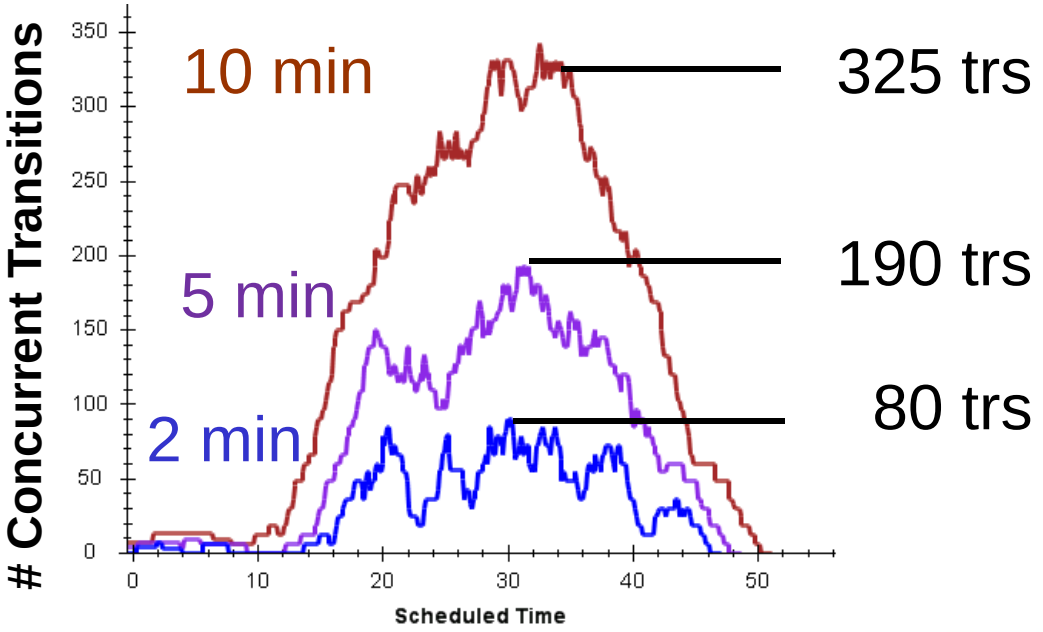
1000 Q1/Q3 Pairs – AB Sciex 4000 QTRAP



Retention Time Scheduling: The best way to acquire data?

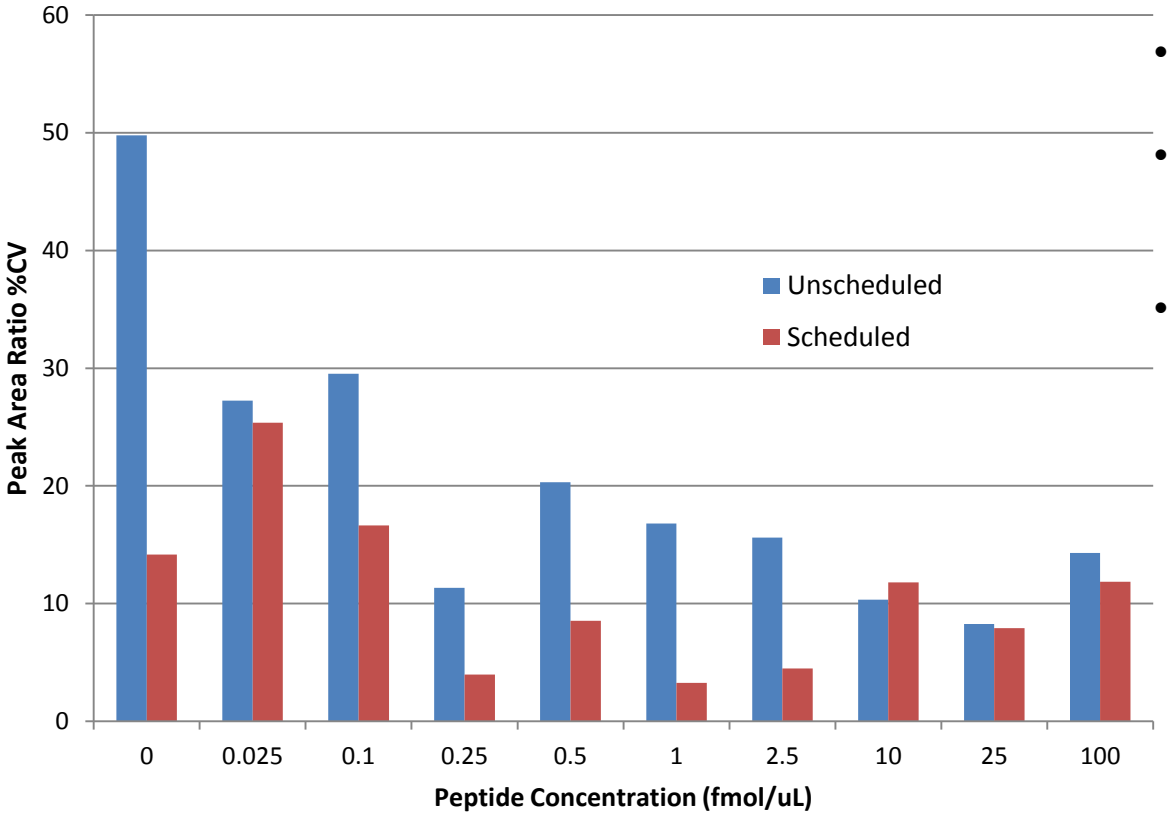
- Scheduling puts rigorous demands on RT reproducibility
- Peak width and RT drift are often limiting factors
- Different peptides shift to various degrees.





- Large numbers of transitions require narrow RT windows or longer cycle times
- Cycle times may be governed by chromatographic peak width
- Scheduled runs with narrow RT windows require BABYSITTING

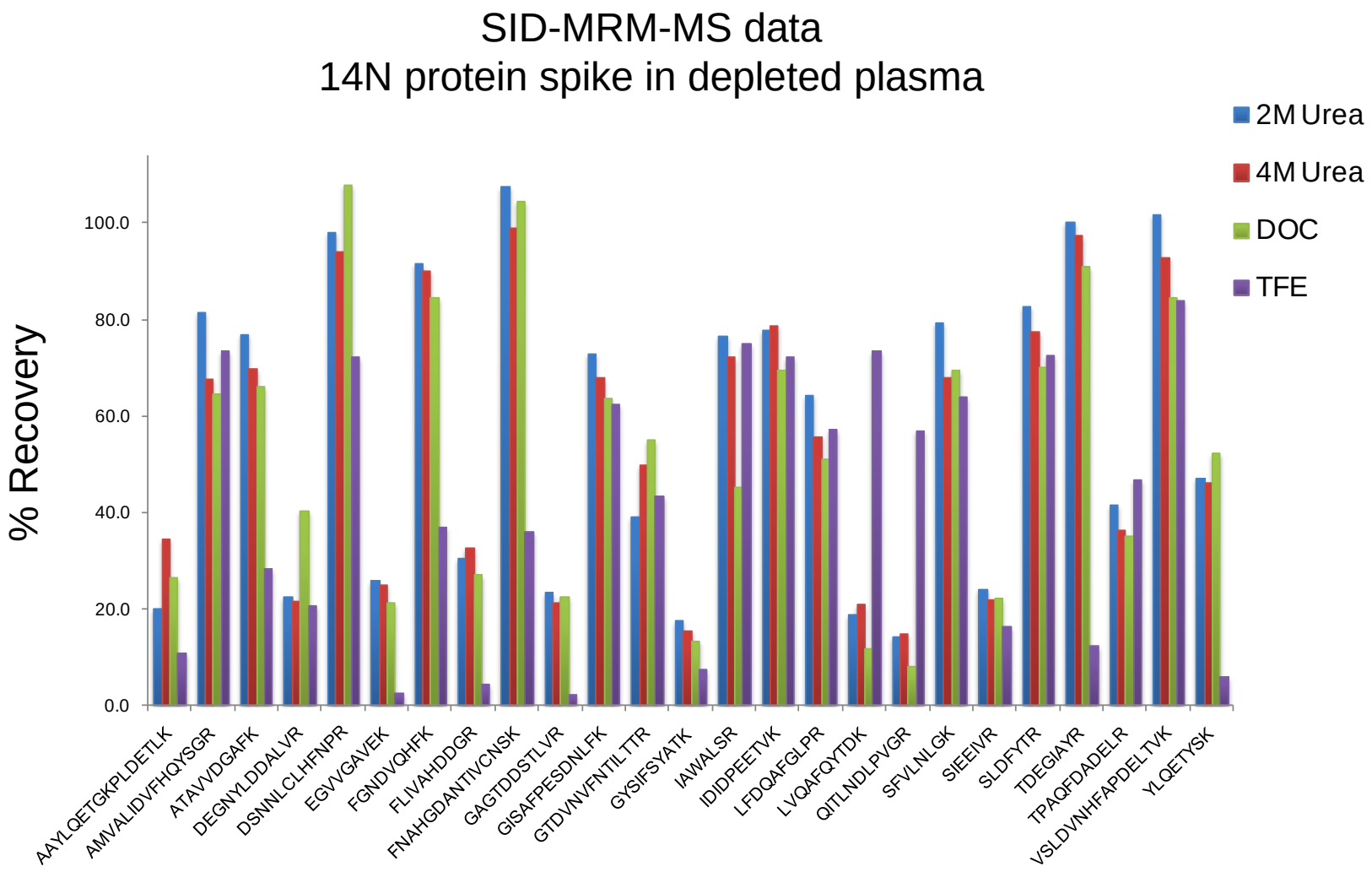
The positive effect of scheduling!



- Same Cycle Time = 1.8 sec
- Unscheduled Dwell Time
  - ~2 msec
- Scheduled Dwell Time
  - 19-70 msec

✓ Longer Dwell times allow better signal measurement and better reproducibility!

Recovery and Reproducibility of Four Digestion Methods Evaluated are Similar



Poster: ASMS 2011, J. Markell

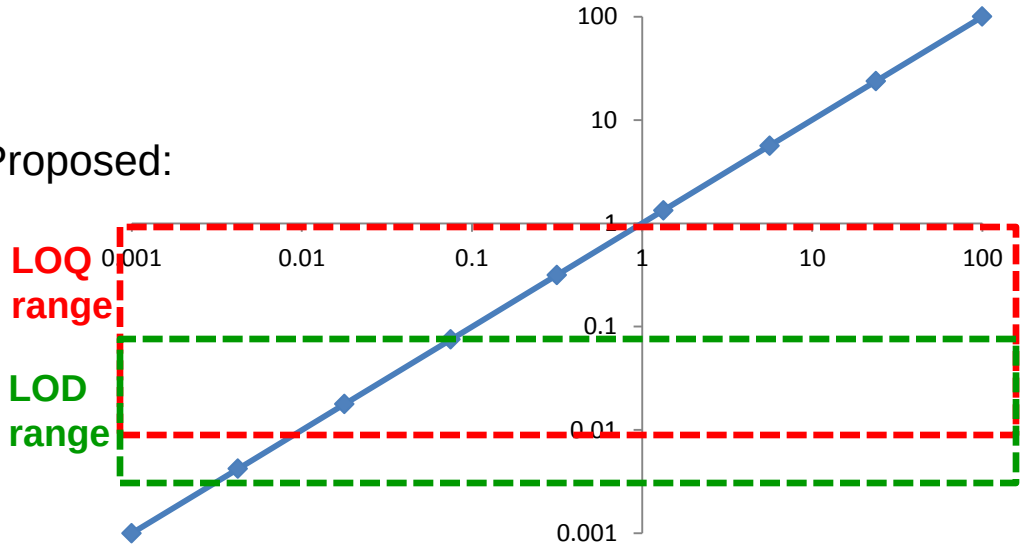
LOD/LOQ Calculations: How Many Points in the Curve are Needed?

What is the ideal concentration range?

$$LOD = \bar{s}_{blank} + t_{0.95} \times (\sigma_{blank} + \sigma_{low})/\sqrt{n}$$

(fmol/µL)
250
113
51
23
10
4.6
2.0
0.9
0.42
0.19
0.09
0.04
0.017
0.008
0.004
0.002

Proposed:

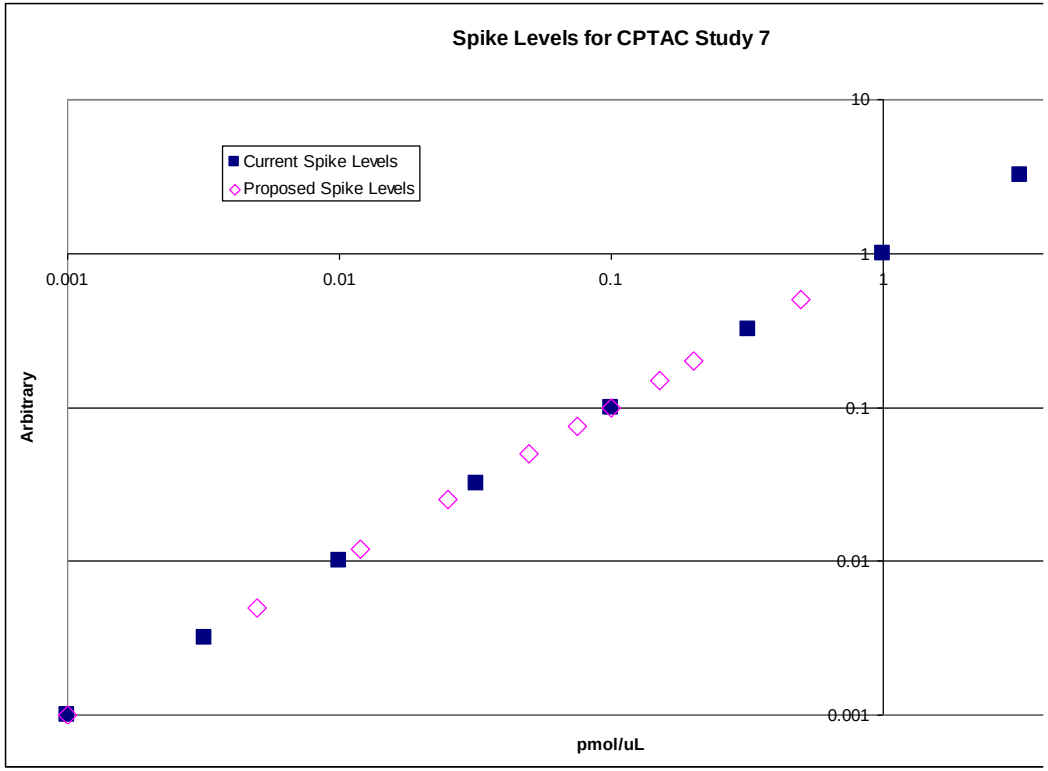


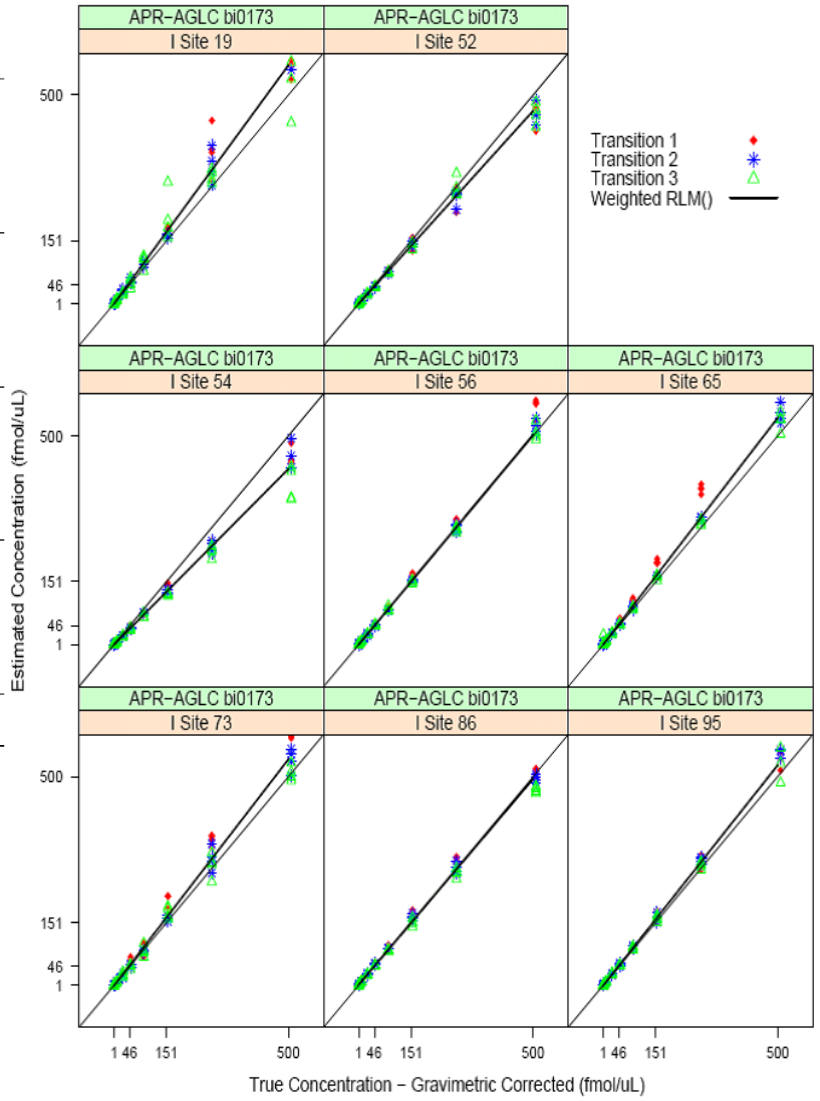
- Generate preliminary curves (16 pts)
- Pick a range and number of points to cover most peptides

Linnet & Kondratovich, (2004) Clin Chem

Keshishian et al, (2009) MCP

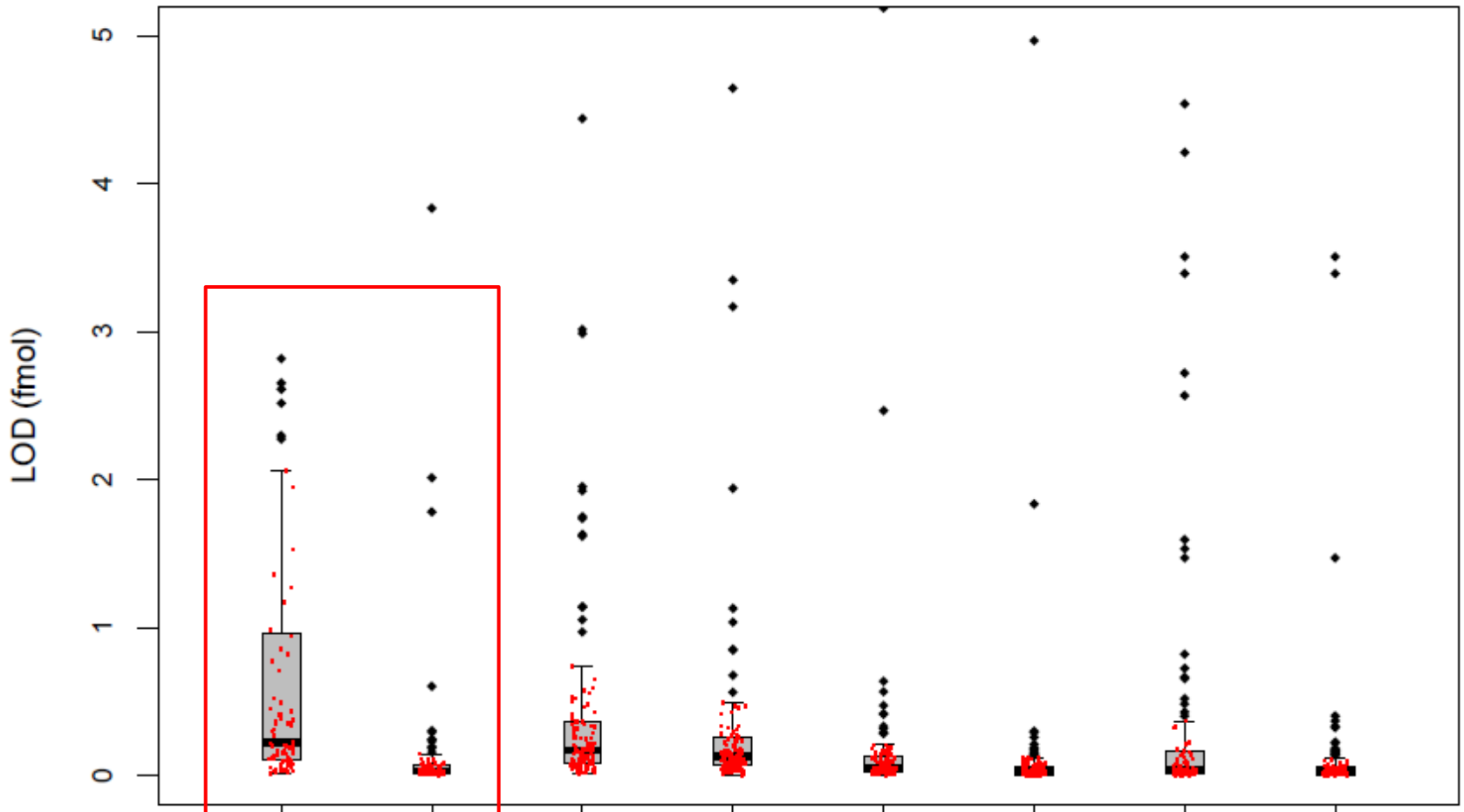
Mistakes from Study 9 (Addona *et al*, 2009 Nat. Biotech)





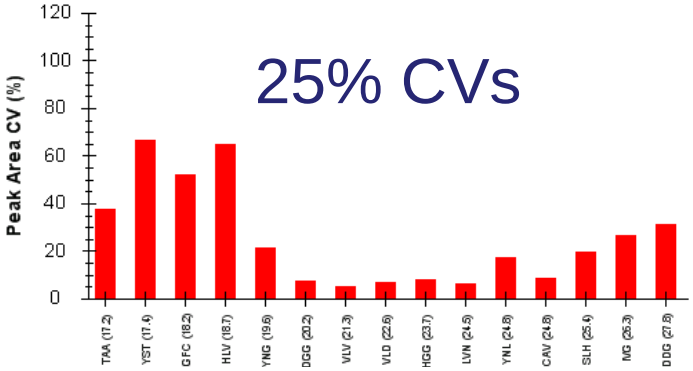
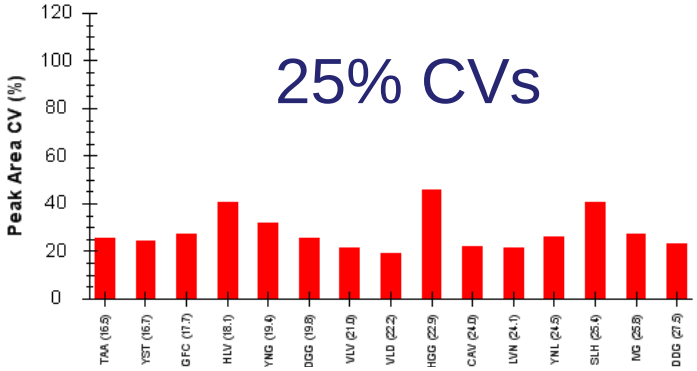
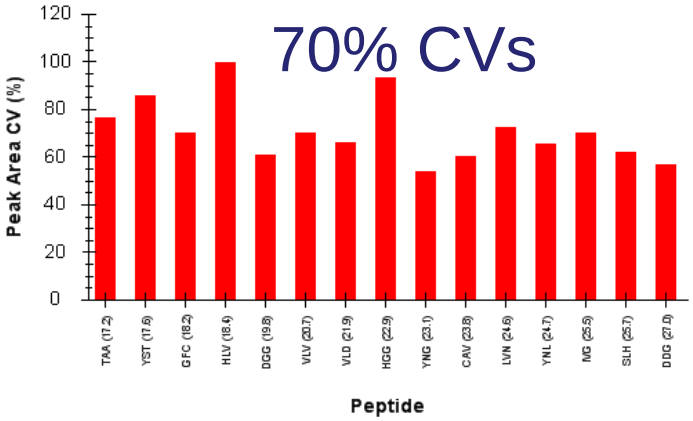
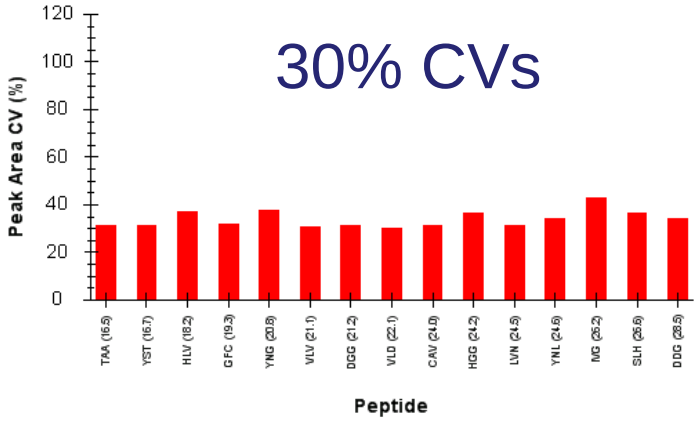
Spike levels didn't go low enough to detect LOD empirically

16 Point Curve at Selected CPTAC Sites Shows Good Reproducibility and Sensitivity



	AB Sciex 5500	AB Sciex 4000	Agilent 6410	Agilent 6460	Thermo Vantage	Waters Xevo	Waters Xevo TQS	Waters Xevo TQS
Median LOD (fmol/uL)	0.22	0.032	0.17	0.13	0.055	0.027	0.044	0.023
Outliers	15	16	15	15	15	12	23	19

LOD is Highly Dependent Upon System Performance: Chromatography and Ionization

	QTRAP 5500 Median LOD: 220 amol	4000 QTRAP Median LOD: 32 amol
Pre-Assay System Suitability Runs (5)		
Throughout Assay System Suitability Runs (24)		

Unstable ESI was a major factor in poor detection and reproducibility

System Suitability assessment detects poor system performance

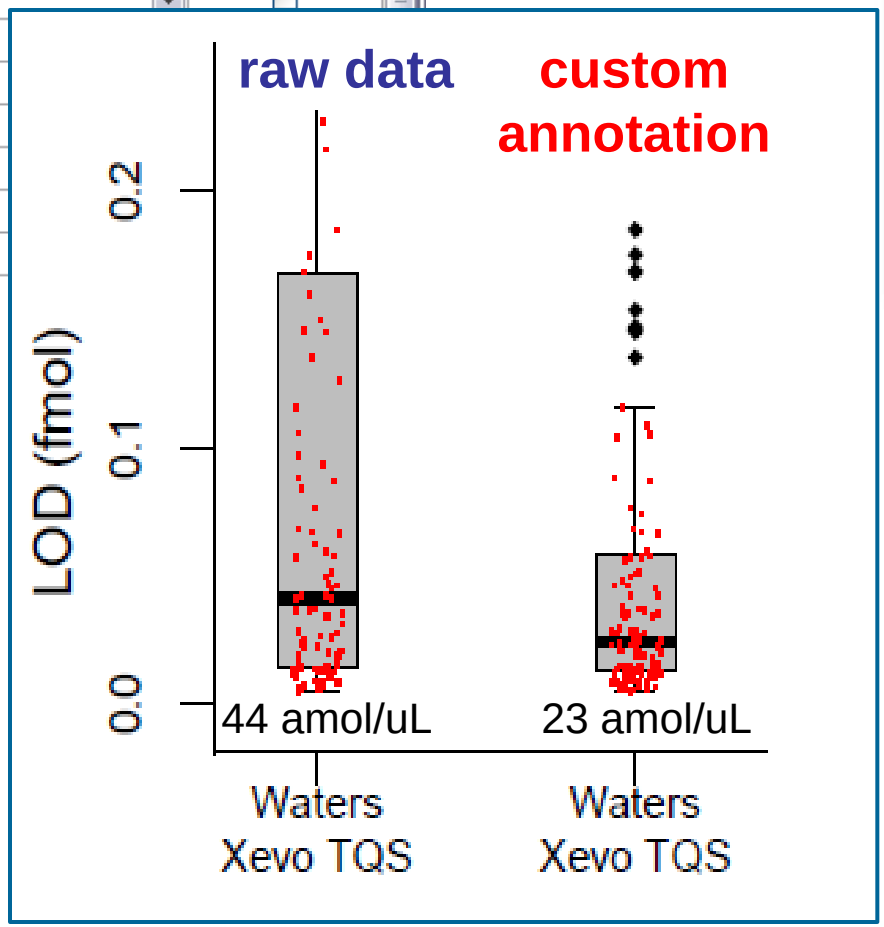
Data Quality Filtering and Custom Annotation by Operators for Data Sets Improves LOD

Results Grid

Total Area	signal quality	RT scheduling problem	general remarks	do not use
110472	peak tailing / poor peak shape			☑
125381				☐
127262		peak outside scheduling window.		☑
138775				☐
138484				
170053				
174726				
192156				
203884				
210323				

very weak signal
peak tailing / poor peak shape
interference
shoulder
transition missing
poor spray
narrow peak (not enough points)

Skyline



Automated version = "AuDIT"

Flags potentially bad transitions

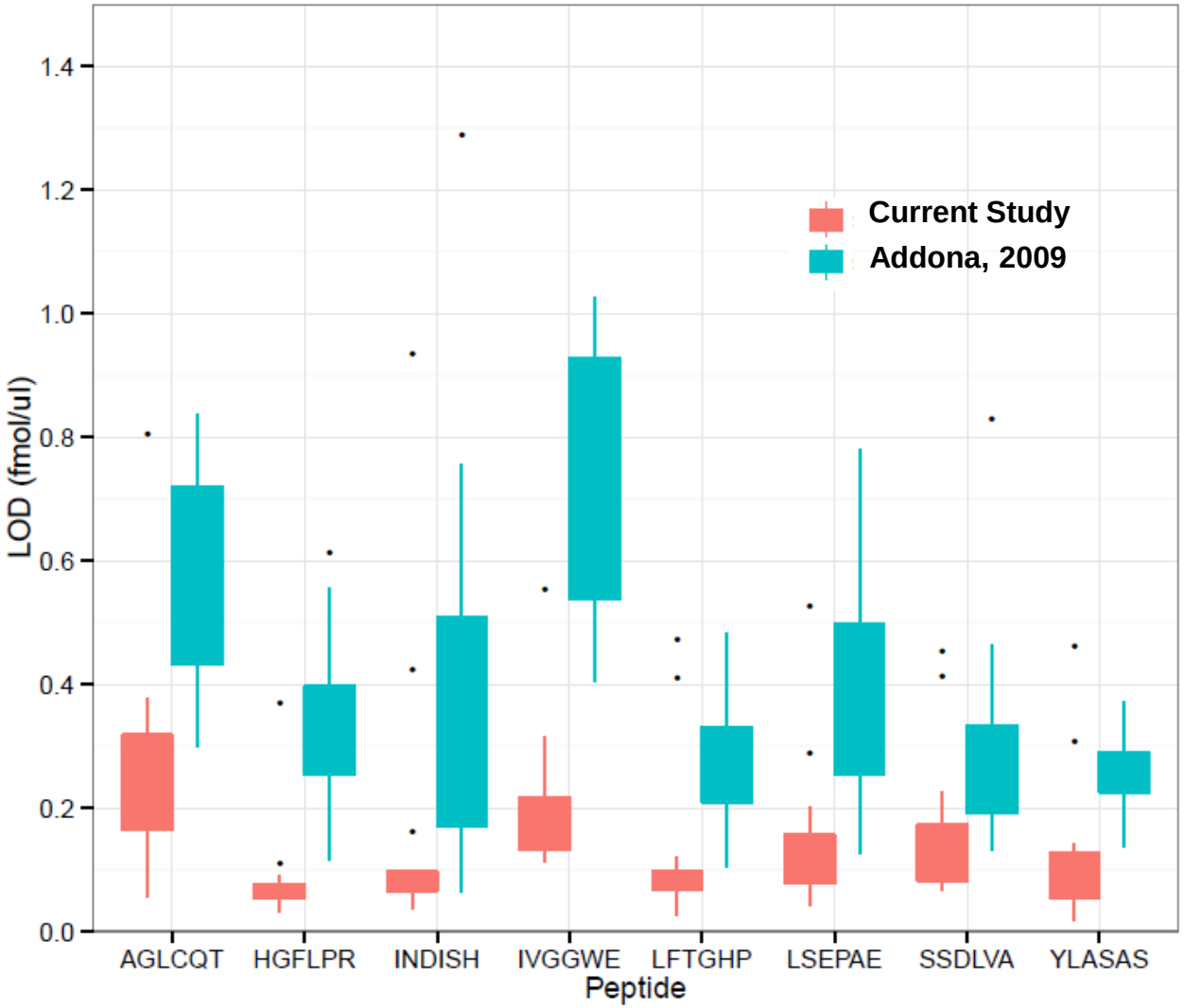
- **poor peak shape**
- **interferences**
- **missing data**

Reduces manual inspection to questionable data

Reduces subjectivity in data analysis

(Abbatiello, Mani et al. Clin. Chem. 2010)

Sensitivity Improvement due to Multiple Factors



- Sample enrichment
- Lower load amount
- Optimized gradient
- Experience!

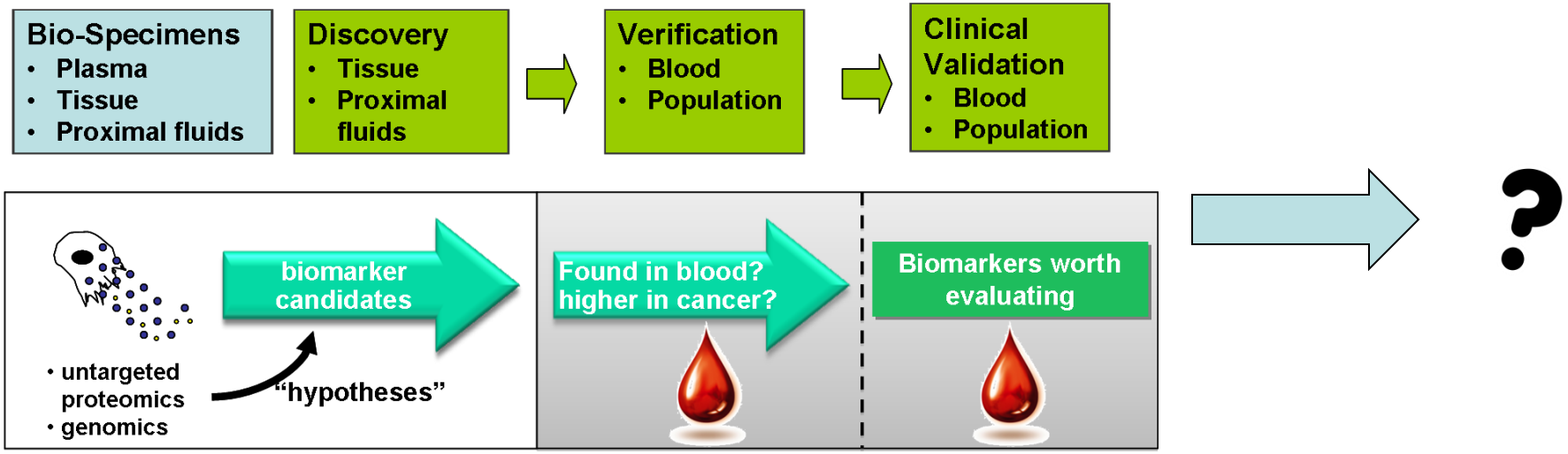
Summary

- First large-scale interlab study to include 15N protein reagents and >100 peptide targets (>350 peptide forms) on 13 different instruments
- Sensitivity improvement from previous study by using depleted plasma
- Retention time scheduling improves reproducibility
- Software and computational methods need to catch up with instrument capabilities and assay development throughput
- Data quality filtering helps remove subjectivity of data evaluation and increases data analysis throughput

Request To Vendors

- ✓ Data-dependent MRM-MS (to eliminate “night watch” of scheduled runs)
- ✓ RT Updating or prediction of systematic drift
- ✓ Column heaters for nanoflow columns
- Interference and Autointerference detection for data analysis and method development

What is the end goal of peptide quantitation?



- × **Biopsies**
- × **Invasive sample acquisition**
- ✓ **Evolution of method to point-of-care or at-home tests**
- ✓ **Earlier detection**
- ✓ **Better control**
- ✓ **Improved quality of life**

A Fluid Based Biomarker Success Story…

Blood Glucose Monitoring for Diabetes

What was once a test run only at hospitals (1970's)...

Is now available for <$1/test at home

And now capable of real-time monitoring every 5 min

www.minimed.com

CPTAC VWG Participants & Acknowledgements

Broad Institute: Susan Abbatiello, Terri Addona, Steven A. Carr, Hasmik Keshishian, D.R. Mani, Michael Burgess, James Markell
Buck Institute for Age Research: Michael P. Cusack, Bradford W. Gibson. Jason M. Held, **Birgit Schilling**
Fred Hutchinson Cancer Research Center: Amanda G. Paulovich, Jeffrey R. Whiteaker, Shucha Zhang
Indiana University: Mu Wang, Jong-Won Kim, Jimsan You
Massachusetts General Hospital: Steven J. Skates
Memorial Sloan-Kettering Cancer Center: Paul Tempst, Mousumi Ghosh
National Cancer Institute: Emily Boja Tara Hiltke, Christopher Kinsinger, Mehdi Mesri, Henry Rodriguez, Robert Rivers
NISS: Xingdong Feng, Nell Sedransk, Jessie Xia
NIST: Paul Rudnick
New York University: Thomas A. Neubert, Åsa Wahlander, Sofia Waldemarson, Pawel Sadowski
Plasma Proteome Institute: N. Leigh Anderson
Purdue University: Charles Buck, Fred Regnier, Dorota Inerowicz, Vicki Hedrick
University of California, San Francisco: Simon Allen, Susan J. Fisher, **Steven C. Hall**,
University of North Carolina: David Ransohof
University of Victoria: Christoph H. Borchers, Angela Jackson, Derek Smith
University of Washington: Michael MacCoss, Brendan MacLean, Daniela Tomazela
Vanderbilt University: Daniel Liebler, Kent Shaddox, Corbin Whitwell, Lisa Zimmerman

Funding: National Cancer Institute