ThermoFisher SCIENTIFIC

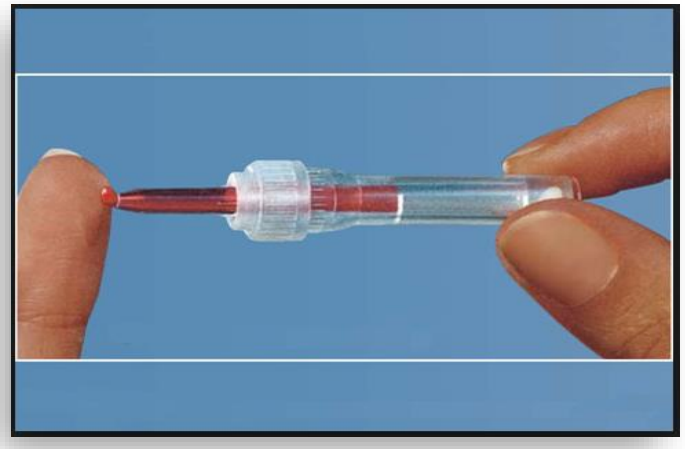
Plasma Proteomics – Next Generation Workflow Tools for Precision Medicine Research

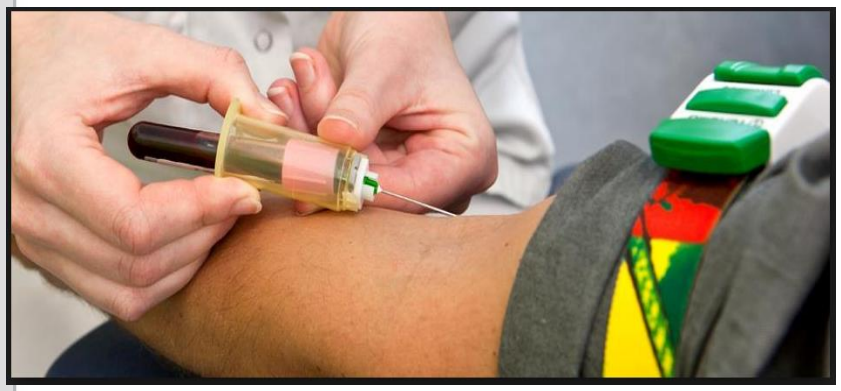
Scott Peterman Ph.D.

Special Edition Metabolomics and Proteomics Seminar
December 7, 2017

Renewed Interest in Plasma Proteomics

- Whole blood is the most common bio-specimen
- LDT have been developed for disease diagnosis or confirmation, risk prediction, prognosis monitoring, and evaluating treatment
- Readily available from clinical trials and epidemiological studies
- There are over 100 FDA-cleared or FDA-approved clinical plasma or serum tests
  - Ca 70 proteins in top 300 protein ranking, another 47 in next 1200
- Excellent source of cfDNA, exosomes, metabolomics, and proteomics
- Three primary classes of proteins in plasma used for specific tests
- Excellent source for additional assays used for biological status
- Changes in plasma proteome can be used determine phenotypes
- Whole blood is an excellent source for exosomes, Buffy coat (PBMCs), platelets, Ig's, and RBCs used for additional tests





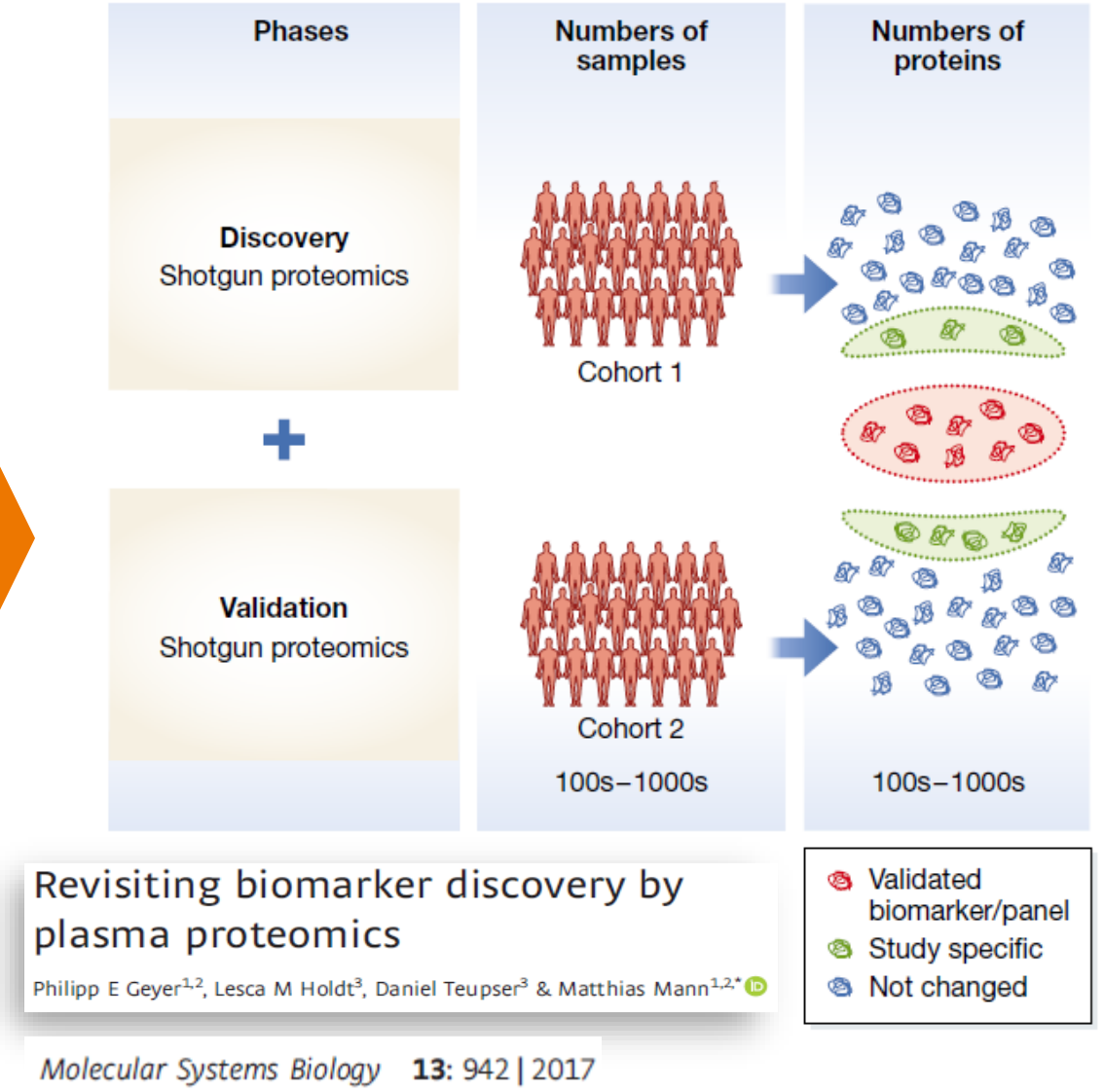
Global Proteome Profiling – Generating Routine Phenotyping Capabilities

Although DNA provides the blueprint, bodily house is built of and maintained by proteins

Desire to map person protein makeup (phenotyping) based on health, environment, genetics

Utilize proteome phenotyping to perform longitudinal (personal) and/or populational analysis

Requires "industrialized" plasma proteome profiling pipeline



Revisiting biomarker discovery by plasma proteomics

Philipp E Geyer[1,2], Lesca M Holdt[3], Daniel Teupser[3] & Matthias Mann[1,2,*]

Molecular Systems Biology **13**: 942 | 2017

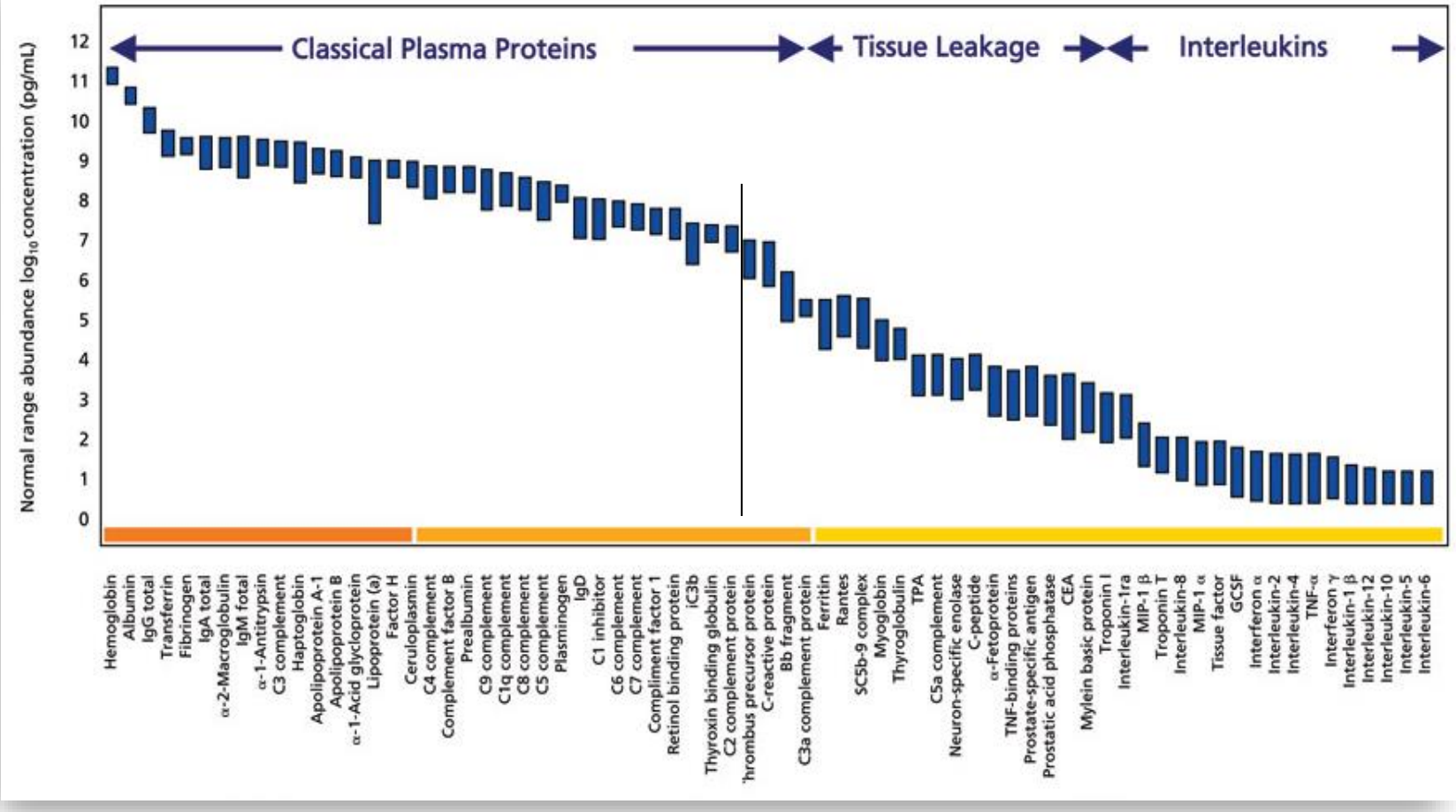
ThermoFisher SCIENTIFIC

Voice of Customer for What is Requested/Expected for Assays

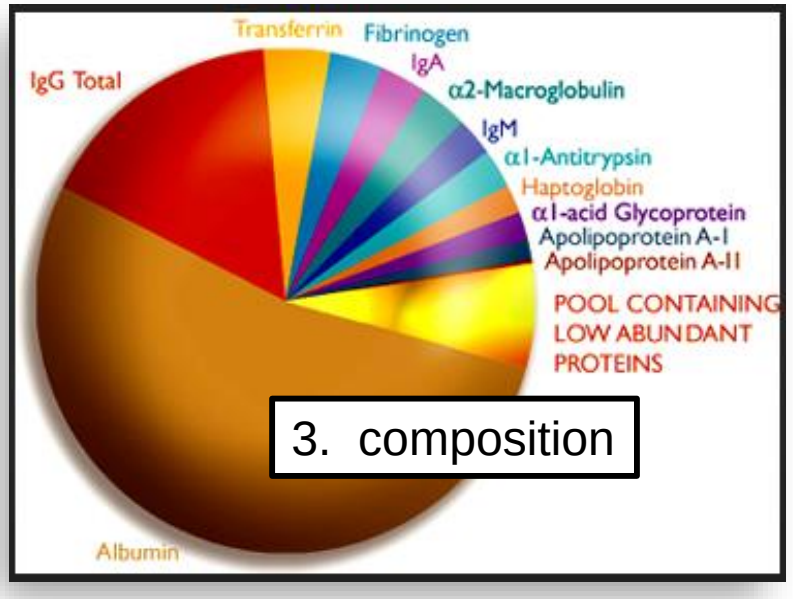
- What is the number of proteins routinely quantified per sample
  - Desired – 1000
  - Settled for – 500
  - Current levels – ca. 250-300 (for non-depleted plasma samples)
- Why the drive for global proteome profiling?
  - More interested in identifying protein patterns (panels) between biological cohorts
  - Translating highly-multiplexed panels to routine quantitative methods
  - Perform targeted data extraction to potential screen for multiple diseases with one study
- Avoid depletion if possible
  - Cost per sample is significantly increased $27/sample is high end
  - Additional levels of sample handling
  - Introduction of variance

Biological Challenges in Plasma Analysis

1. Dynamic range



2. Ig's contribution - sequences



3. composition

- Large dynamic range of plasma protein abundance
- Few proteins constitute the majority of the protein mass
- Performing tryptic digestion transfers large dynamic range issue from protein to peptide
- Co-elution of high and low abundant peptides reduces intra-scan dynamic range
- 1000's of Ig proteins each with different sequences can hinder sequence matching routines

ThermoFisher SCIENTIFIC

Developing a Highly Robust Workflow at BRIMS – Collaborator Support

- Robust, reproducible workflow designed to maximize proteome coverage while maintaining high-throughput

Sample preparation

- Lipids and salts
- Protein aggregation
- Incomplete and unpredictable digestion

BRIMS digestion protocols utilizing different buffers afforded by introduction of the trapping column and divert valves

Sample Loading

- Large dynamic protein expression
- The bulk of the protein weight is taken by few proteins
- Creates many issues with loading

- Leverages Vanquish UHPLC system and slightly wider bore columns in modular format
- Introduction of trapping column

Chromatographic separation

- UHPLC support
- Heated solvents
- Stable temperature control
- Additional divert valves

- Significantly increase peak capacity despite shorter gradients
- Options for alternative solvent blends
- System suitability implemented

Data acquisition

- Co-elution of peptides originating from top 20 proteins reduce intra-scan dynamic range
- Co-elution of Ig peptides can hamper data searching

- Increased peak capacity reduces co-elution
- HRAM MS and MS/MS possible with DDA concepts

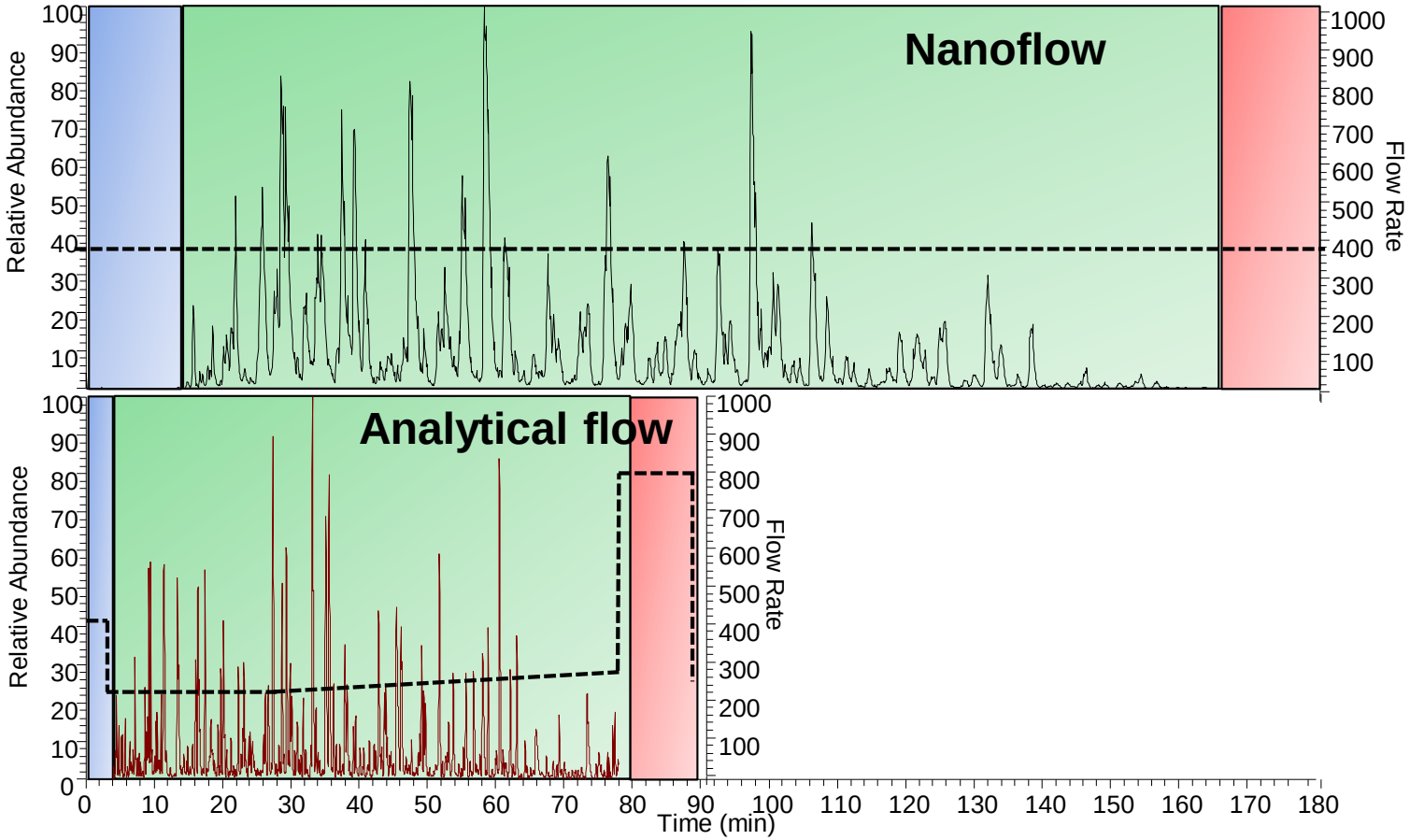
Data Interpretation

- DDA only searched using sequence matching
- Global spectral libraries omit realistic proteins/peptides and relative abundance
- Selectivity using DIA is reduced

- In-depth sample-specific spectral libraries provide greates confidence
- FDR routines
- Protein expression orders
- Increased confidence in PTM analysis

Leveraging High Resolution Analysis for Global Profiling

- Chromatographic resolution properties as defined by peak capacities
- Maximizing throughput for large-scale pilot and clinical studies



- **Maximum backpressure is 1500 bar enabling greater peak capacities (narrower peak widths across shorter gradients)**
- **Performs solvent heating to help loading rates**
- **Extra divert valve in the heated column chamber**
- **Reservoir for trapping column cleaning solvents**

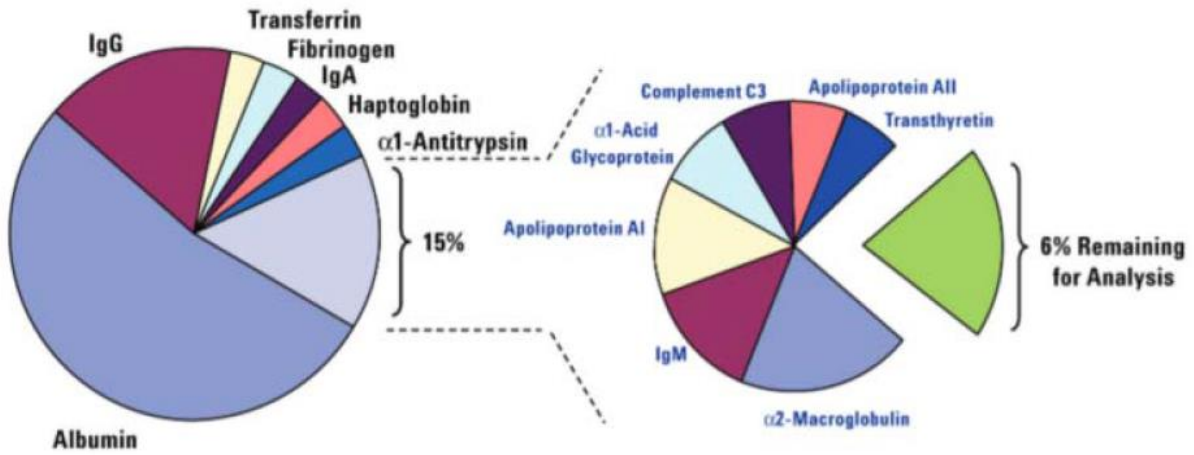
The following equation was used to calculate peak capacity:

$$P = 1 + \frac{t_G}{\frac{1}{n}\sum_{1}^{n} w_P}$$

Where, n is the number of peaks used for the calculation, t_G is the gradient time, w_P is the average peak width measured at 4σ peak height[1].

Understanding Plasma Complexity – Dynamic Range

The top 14 proteins make up ~94% of total protein in plasma which makes the identification/quantification of remaining 6% difficult



Injecting 1 µg on column using nanospray

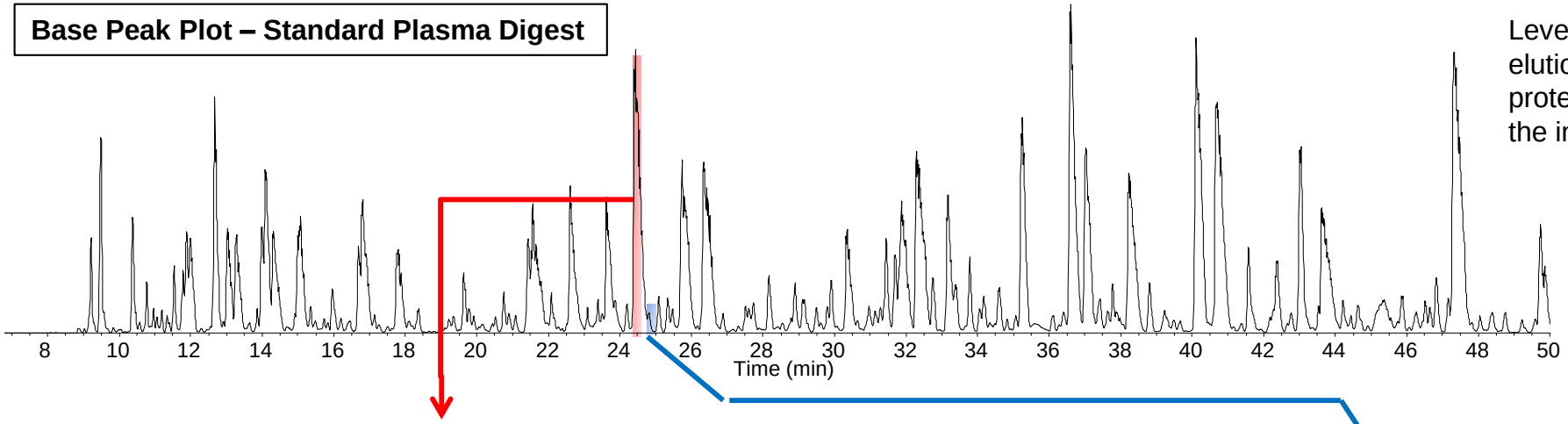
- 850 ng attributed to a few proteins
- 150 ng for all other proteins
- Example of 1 ng for a 30kDa protein = 33 fmol on column
- Extend that down for 1000-fold for 33 amol on column for detection

Injecting 100 µg on column using analytical flow

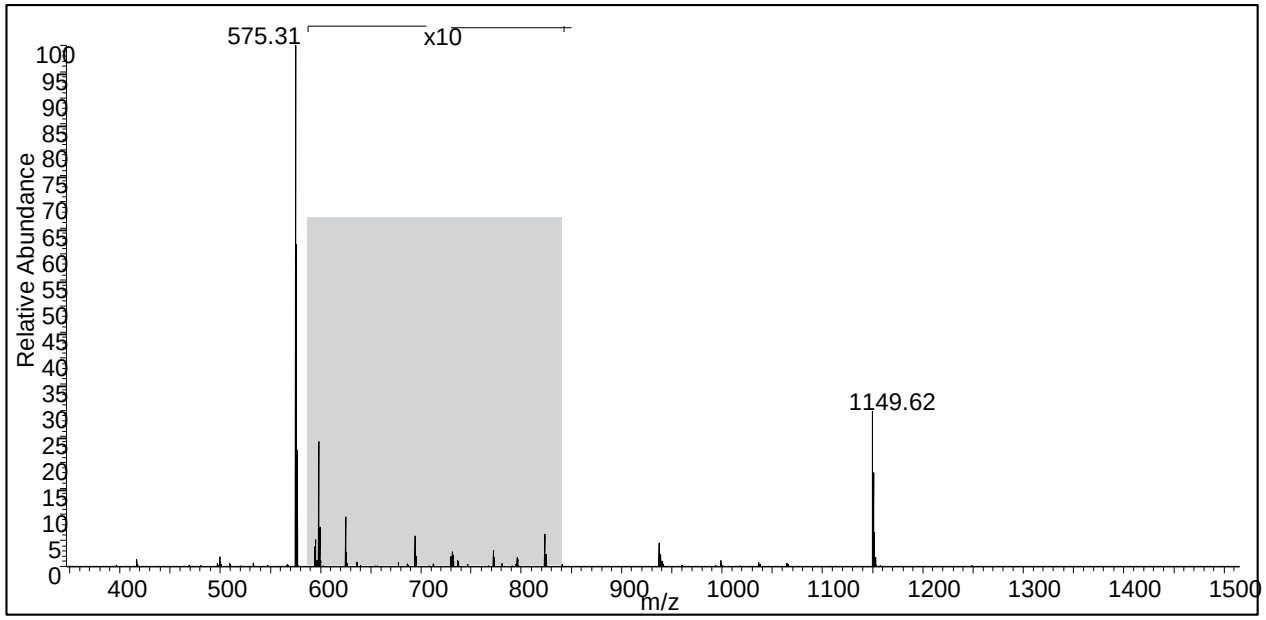
- 85 µg attributed to a few proteins
- 15 µg for all other proteins
- Example of 1 µg for a 30kDa protein = 33 pmol on column
- Extend that down for 1000-fold for 33 fmol on column for detection

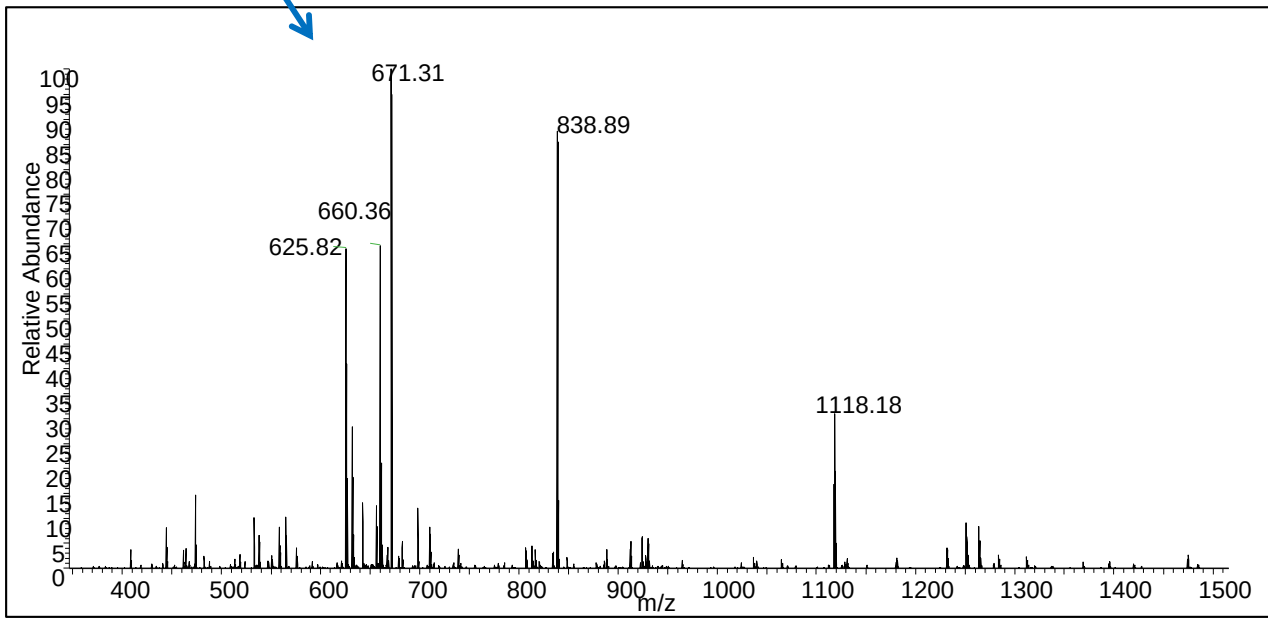
Comparative Full Scan Mass Spectral Analysis – UHPLC Separation

Base Peak Plot – Standard Plasma Digest



Leverage chromatographic resolution to minimize co-elution of highly abundant peptides from the top 14 proteins with remaining plasma proteins to increase the intra-scan dynamic range

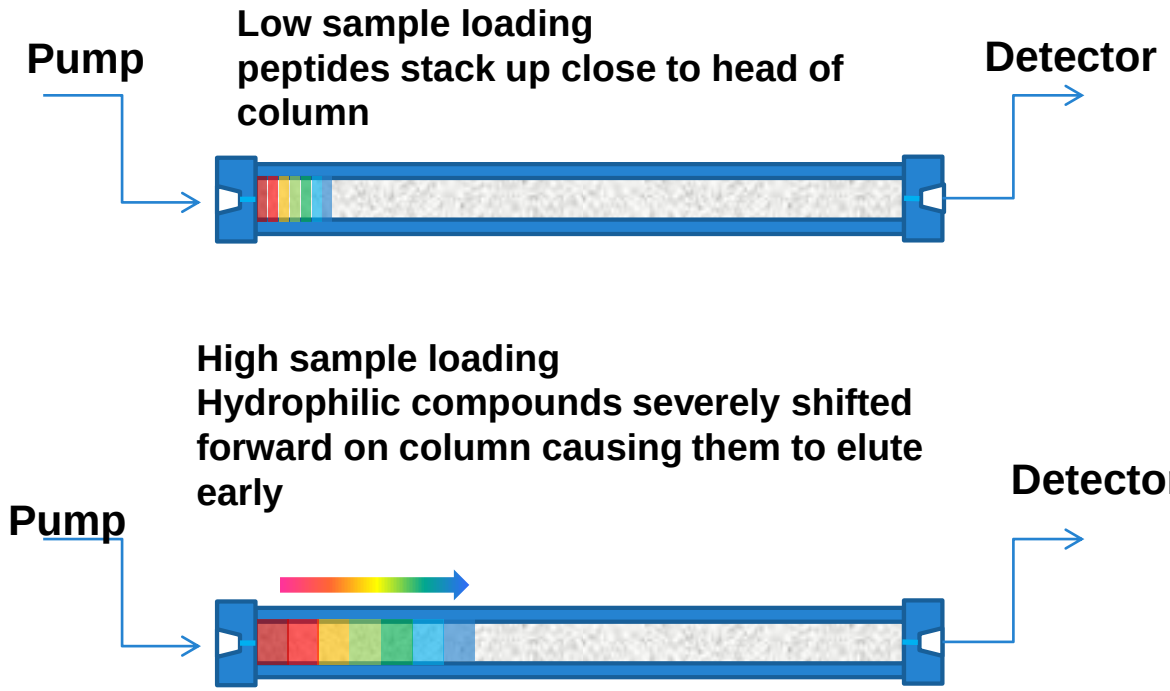


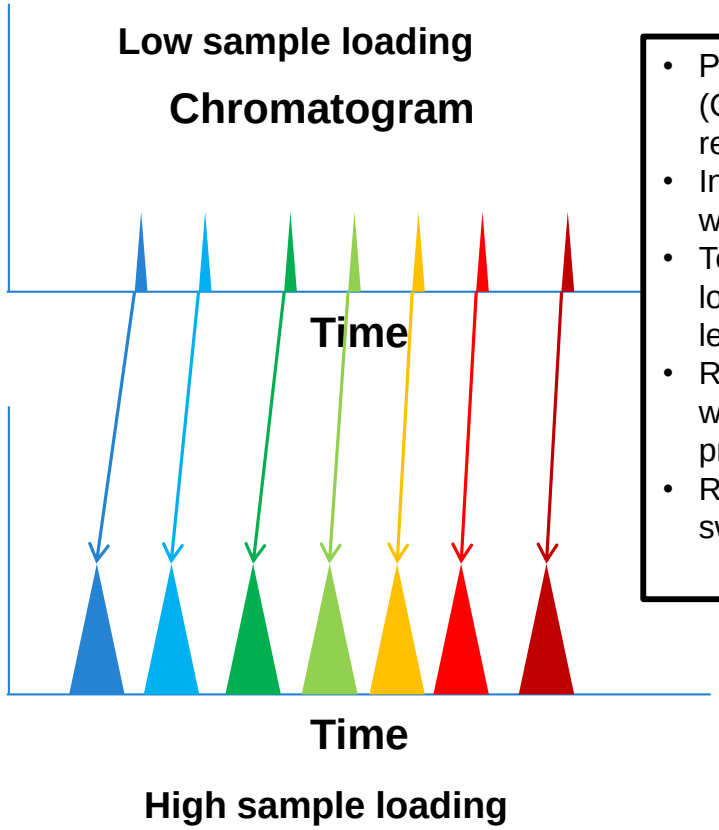


ThermoFisher SCIENTIFIC

Increasing the Analytical Stability and Robustness of Sample Loading

Single column (or 2 columns, trap and analytical, with the same chemistry). Displacement chromatography occurs

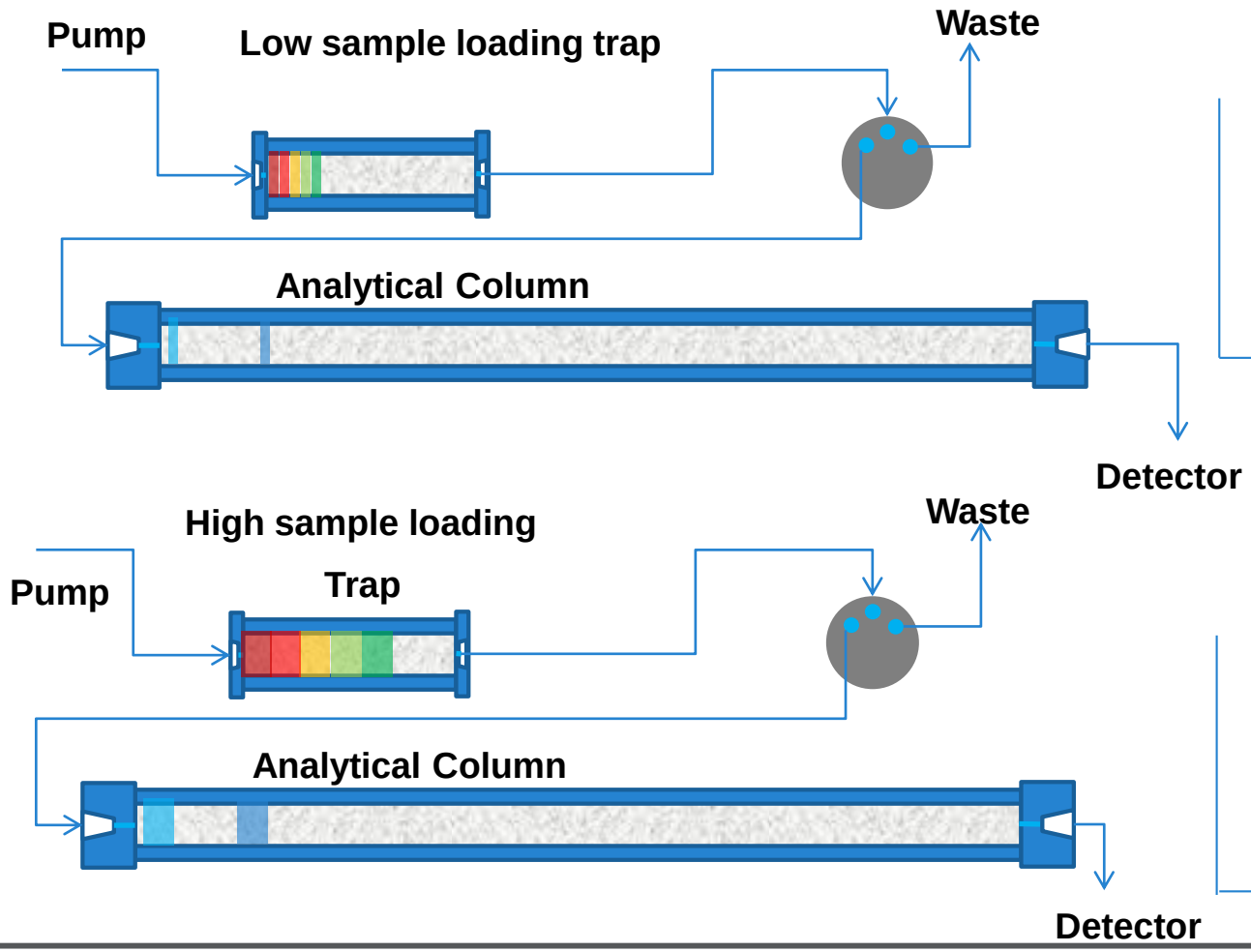


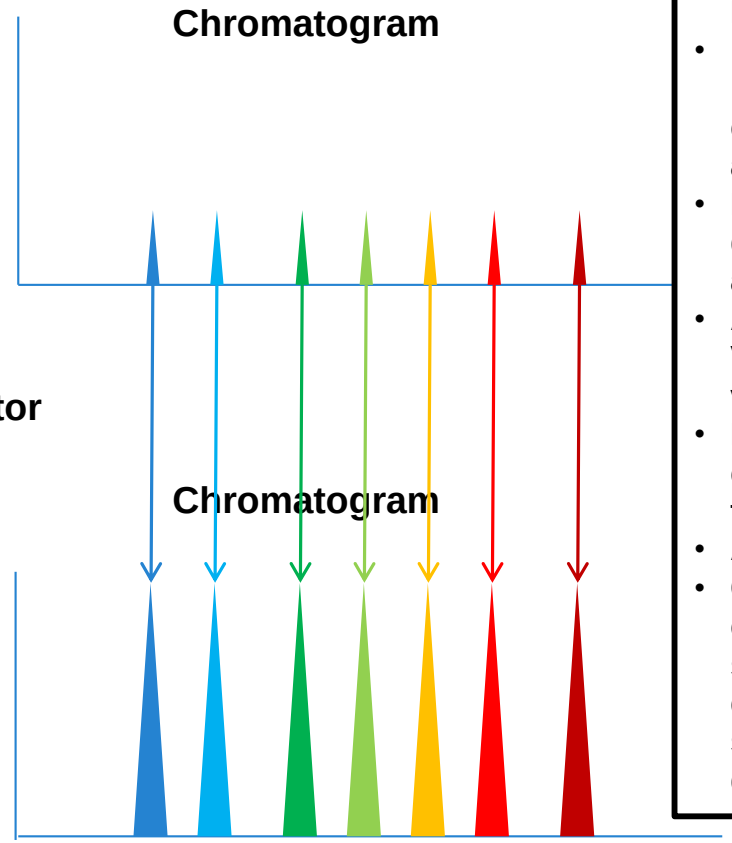


- Pairing columns with minimal k' values (C8 or C18 based packing chemistries) reduces benefits of trapping
- Increased loading amounts can result in wider chromatographic peaks
- To reduce RT shifting as a function of loading amounts, must increase column lengths
- Reduced RT reproducibility requires wider windows for automated data processing and/or scheduled windows
- Reduces column lifetime requiring switching during study – data interruption

Increasing Loading Capacity While Maintaining Robustness

- Maximizing k' differences between trapping and analytical columns
- Utilization of two different divert valves to increase robustness and column lifetimes

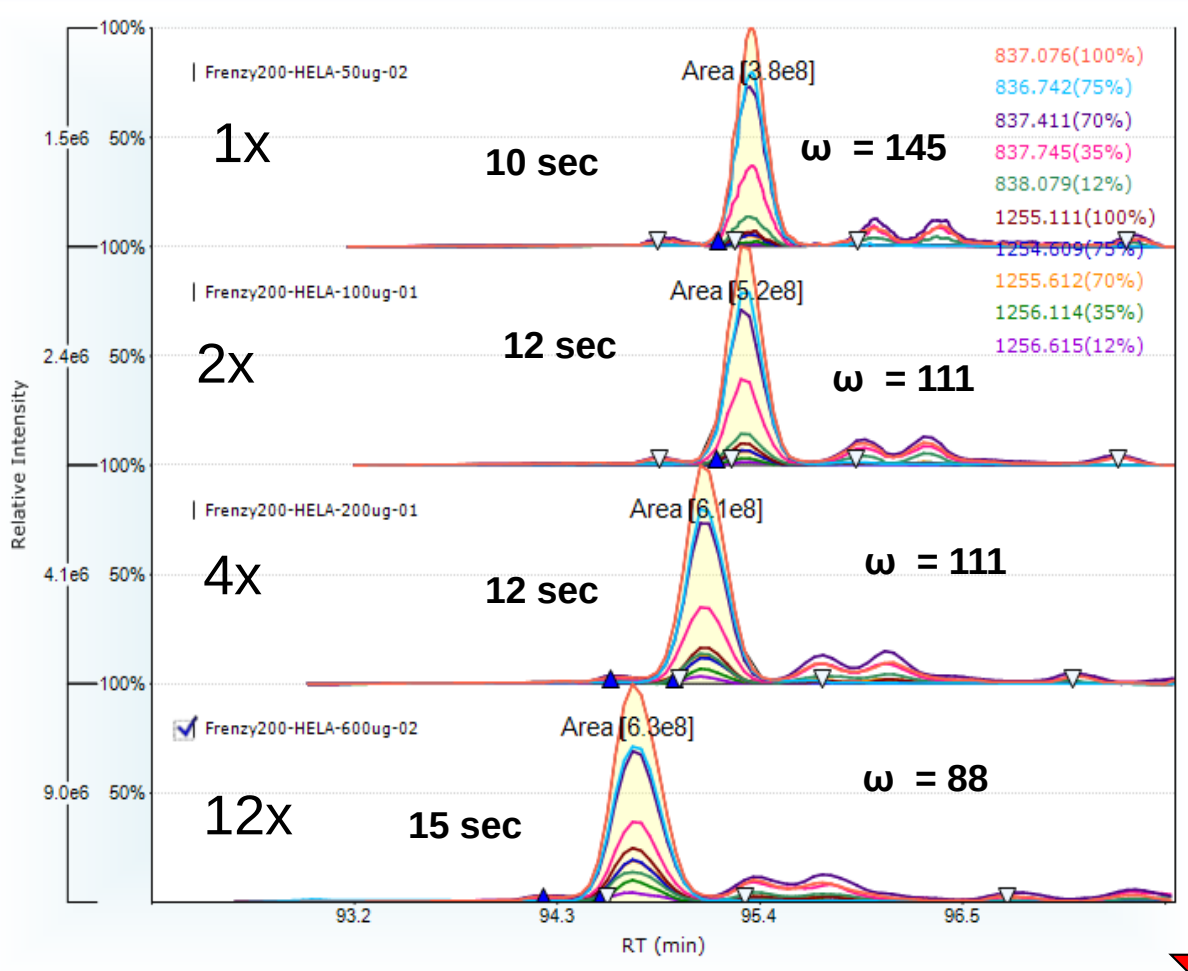




- Trapping columns comprised of PRLP-S packing material
- Large k' difference resulted in increased peak shapes and capacity – refocusing at the analytical column head
- Matching trapping column dimensions to fit expected load amounts
- Added extra divert valve on Vanquish and utilized divert valve on instrument
- Both trapping and analytical columns maintained at high temperature
- Added solvent heater
- Column and pump configuration enabled flow switching at the beginning and end of methods to wash salts, small molecules, and non-digested proteins to waste

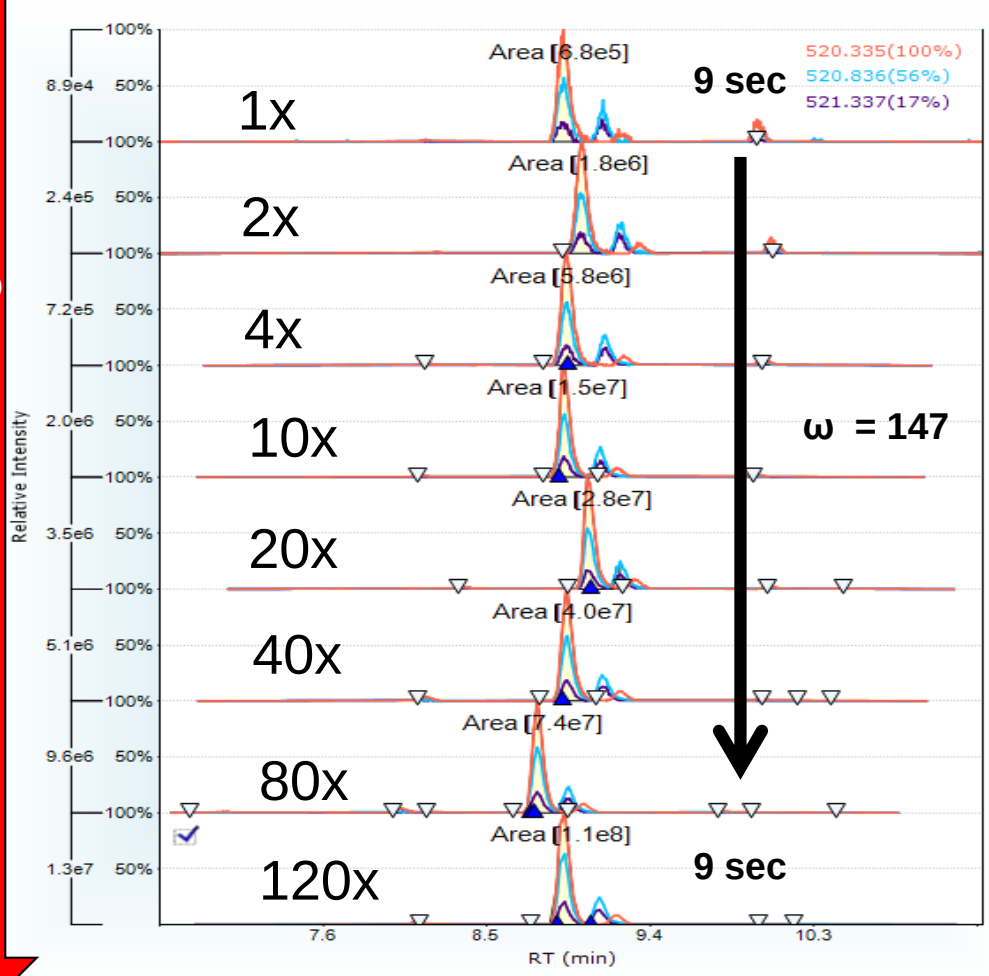
Examples – With and Without Optimal Trapping Columns

Without Trapping Column



Increased Loading Amounts

With Trapping Column

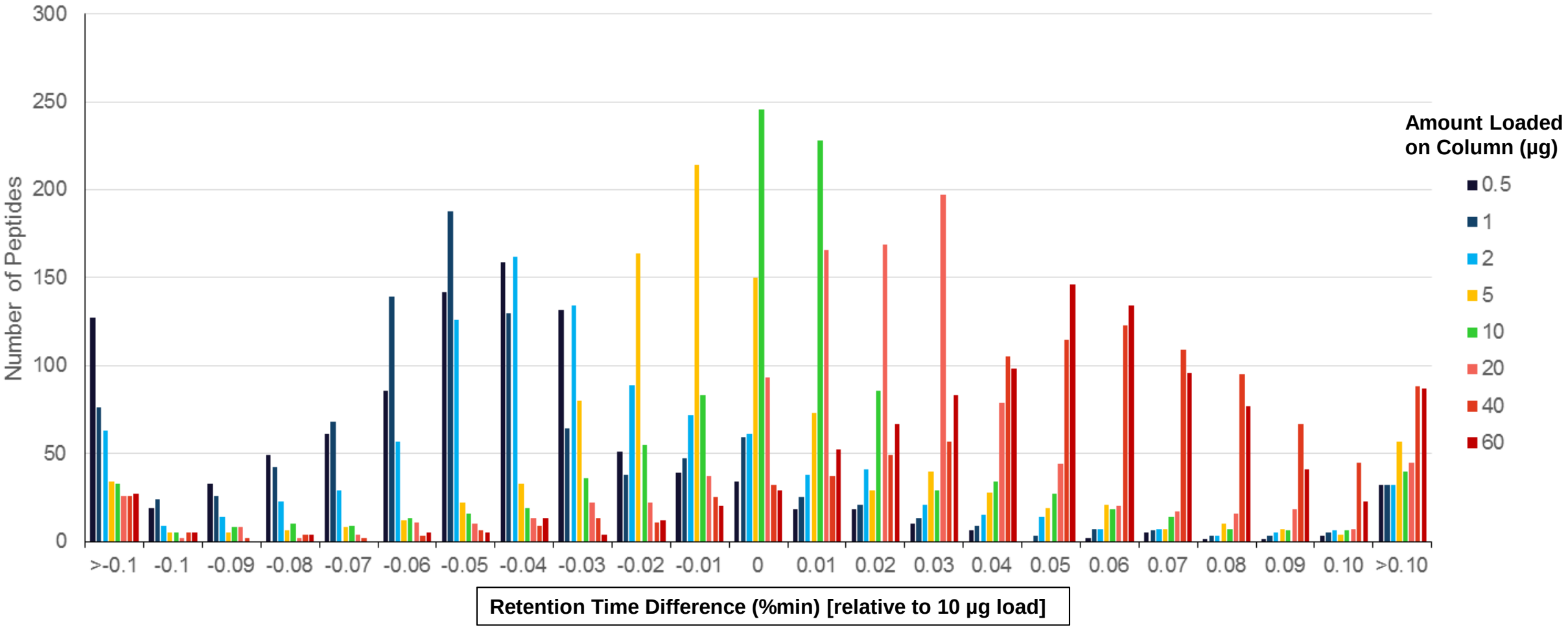


Also effects peak capacity and routine and comprehensive data acquisition

Additional experiments have extended the loading amounts to 220 µg

ThermoFisher SCIENTIFIC

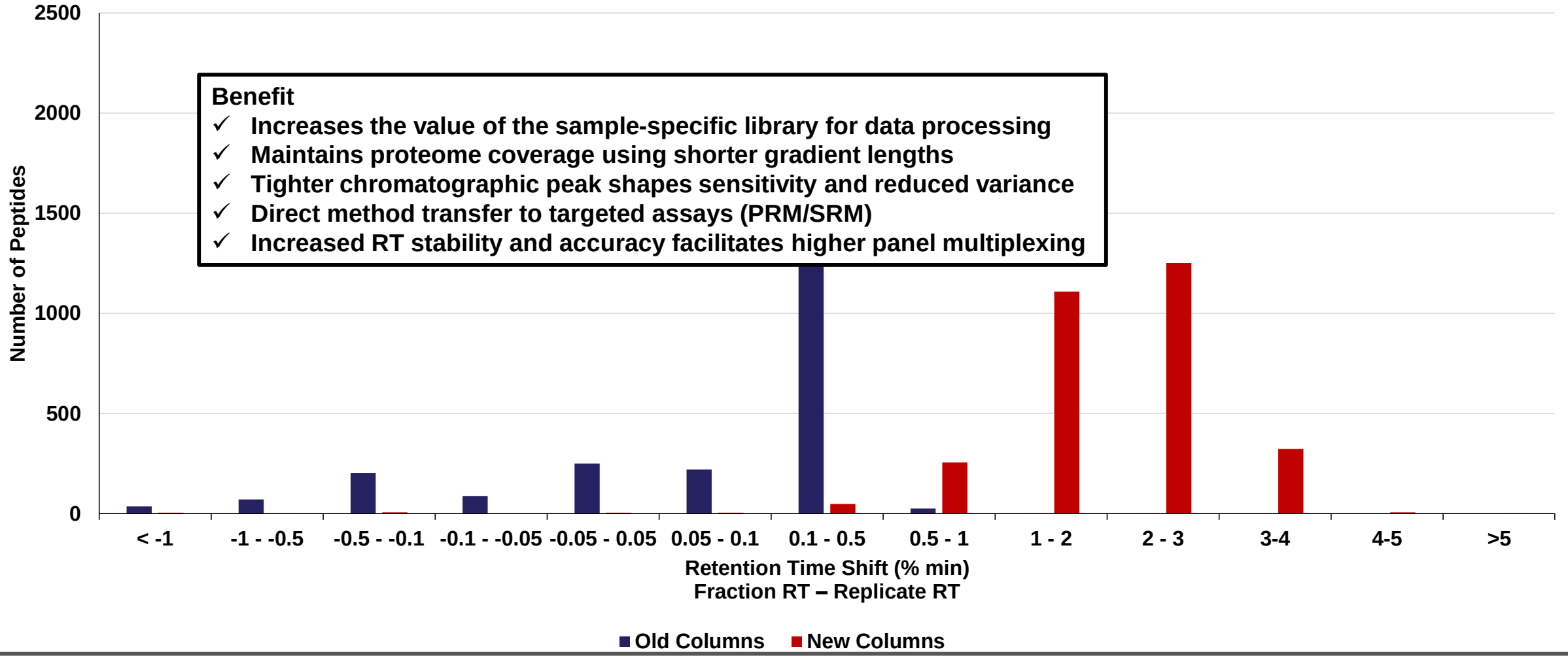
Retention Time Difference as a Function of Load Amount



ThermoFisher SCIENTIFIC

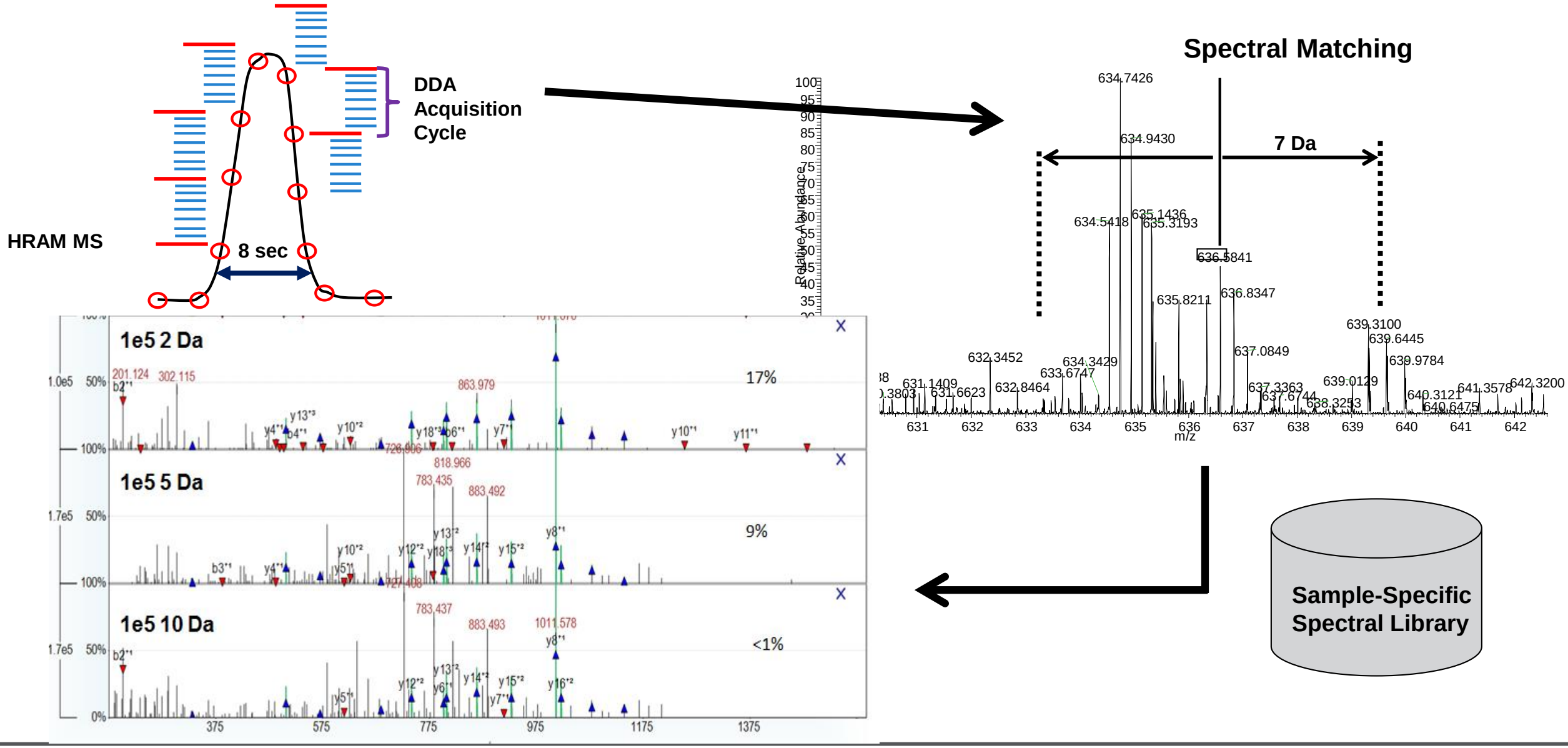
Retention Time Correlation for Matched Peptides: Fraction Libraries vs. Replicates

- 24 fraction libraries were acquired on the old columns
- The replicate (unfractionated) samples were analyzed using both old and new columns and searched against the same 24 fraction library



ThermoFisher SCIENTIFIC

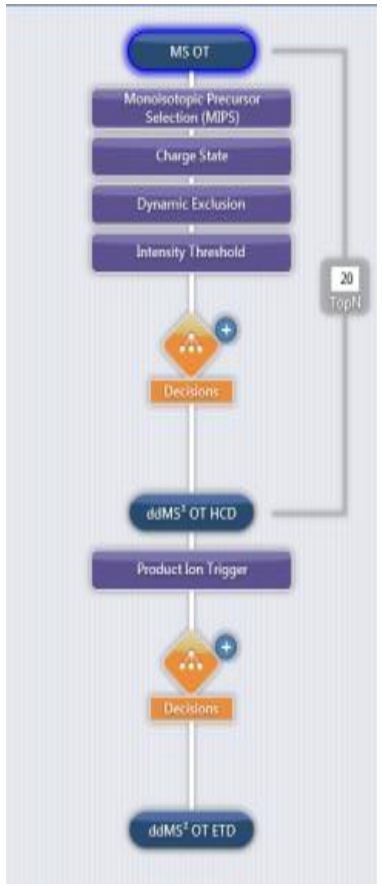
Tandem Spectral Acquisition Approaches

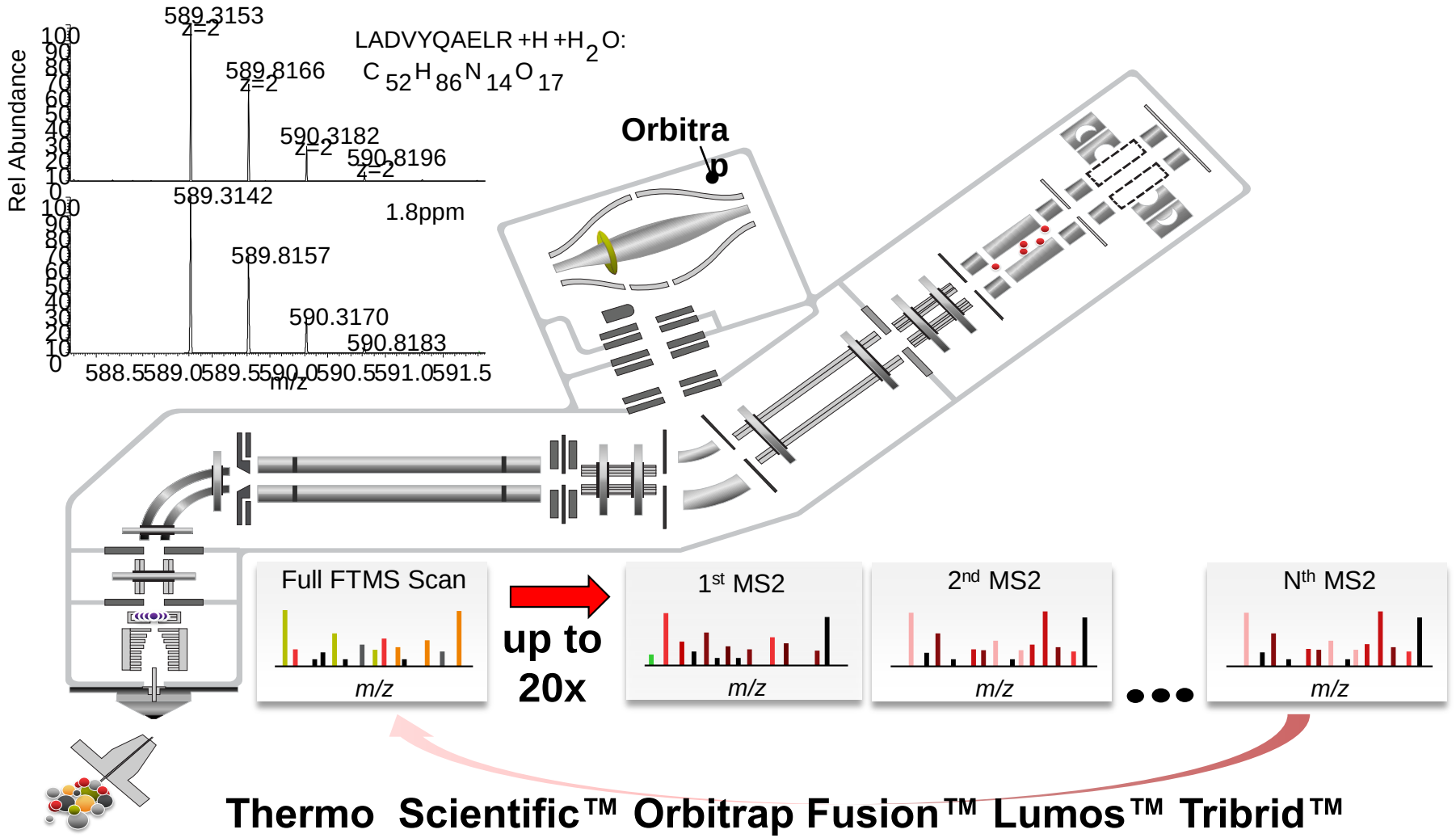


ThermoFisher SCIENTIFIC

Enabled by Accurate Mass, High Resolution, Intelligent MS

Selectivity and Instrumental Intelligence to Exhaustively Characterize Complex Samples

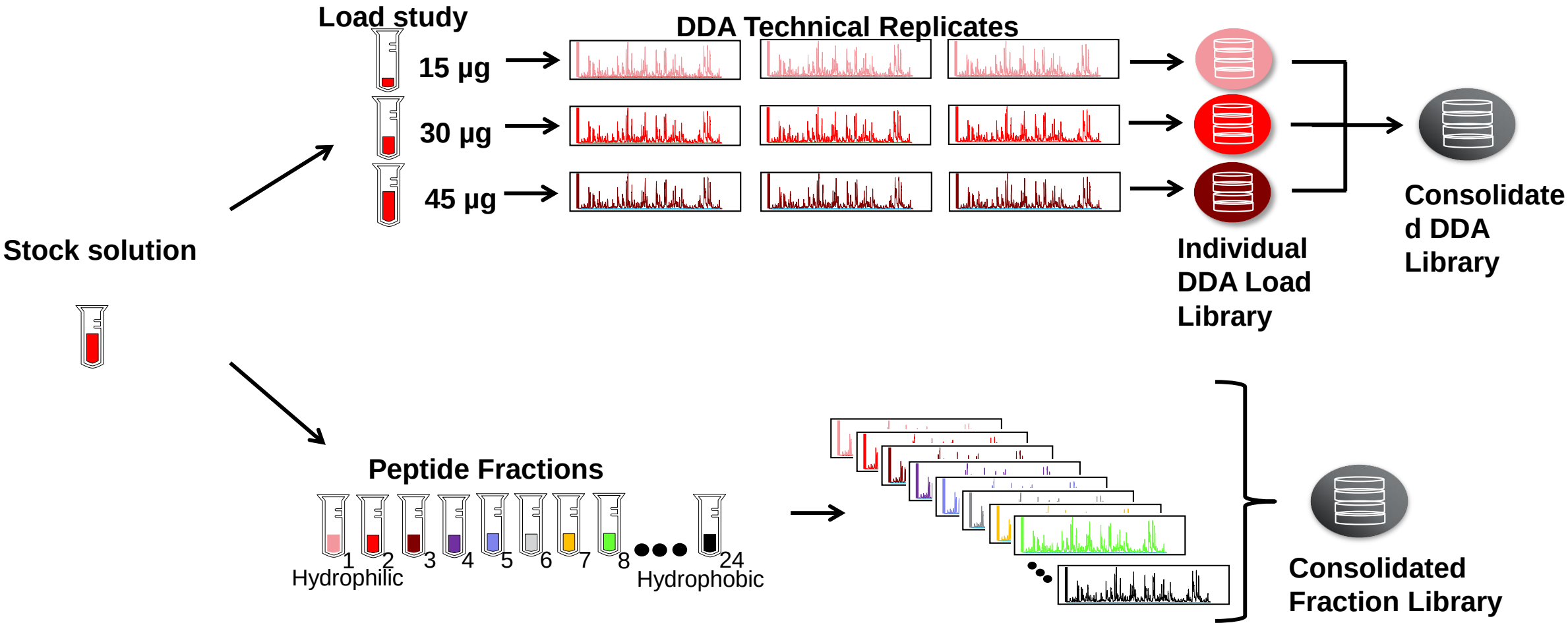




For research use only. Not for use in diagnostic procedures.

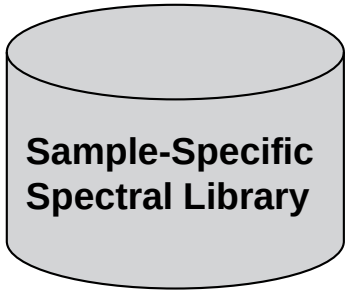
ThermoFisher SCIENTIFIC

Spectral Library Building Approach

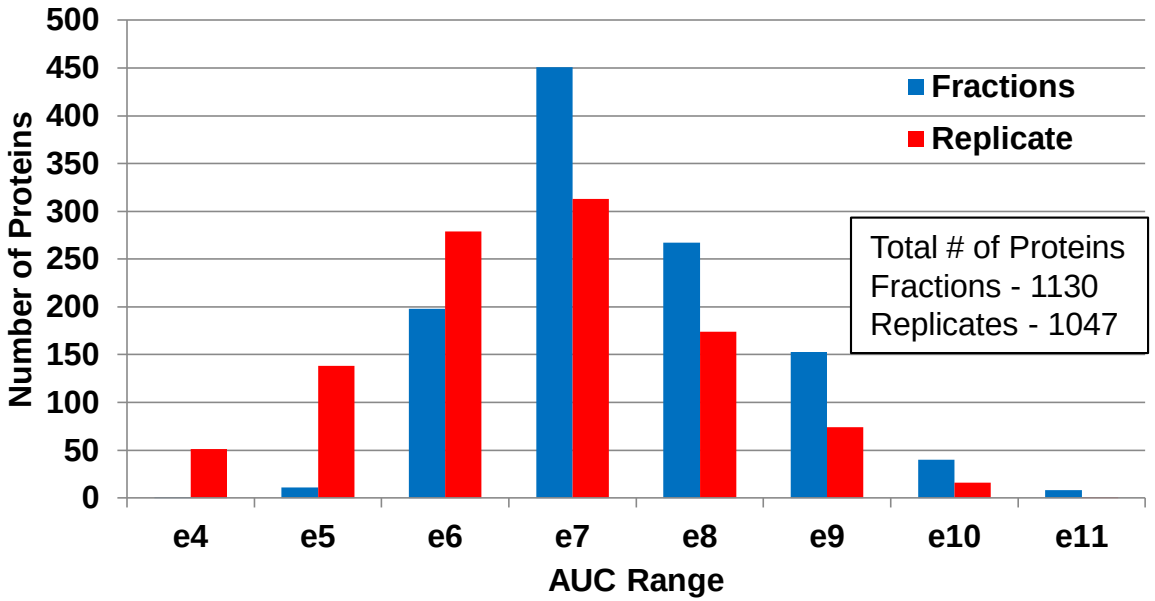


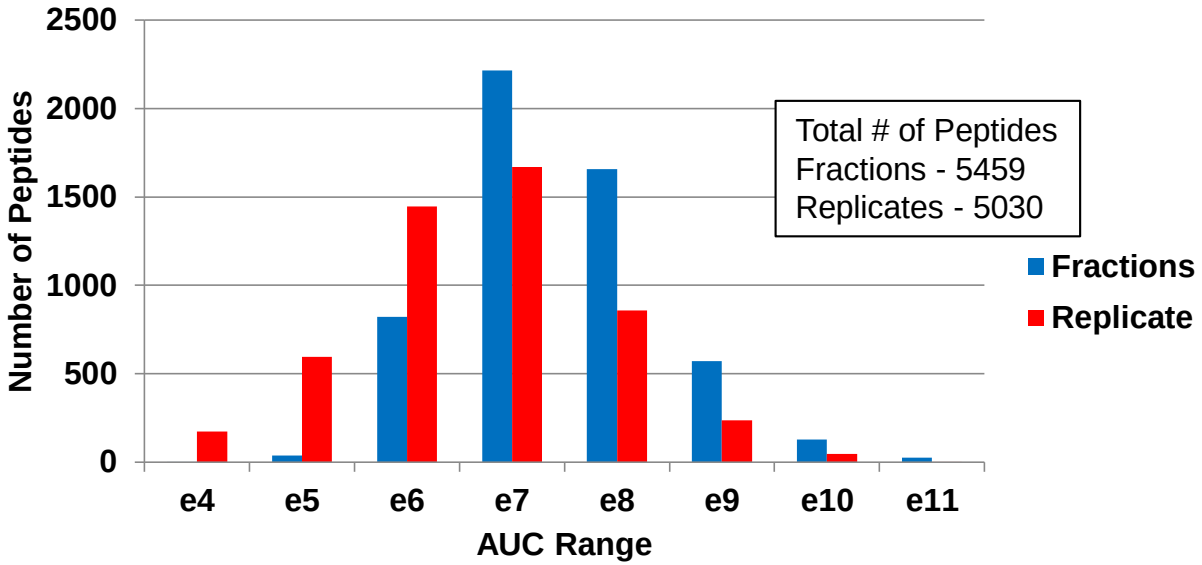
ThermoFisher SCIENTIFIC

Sample Specific Spectral Library Information



24 Fractions generated from reverse-phase high pH separation



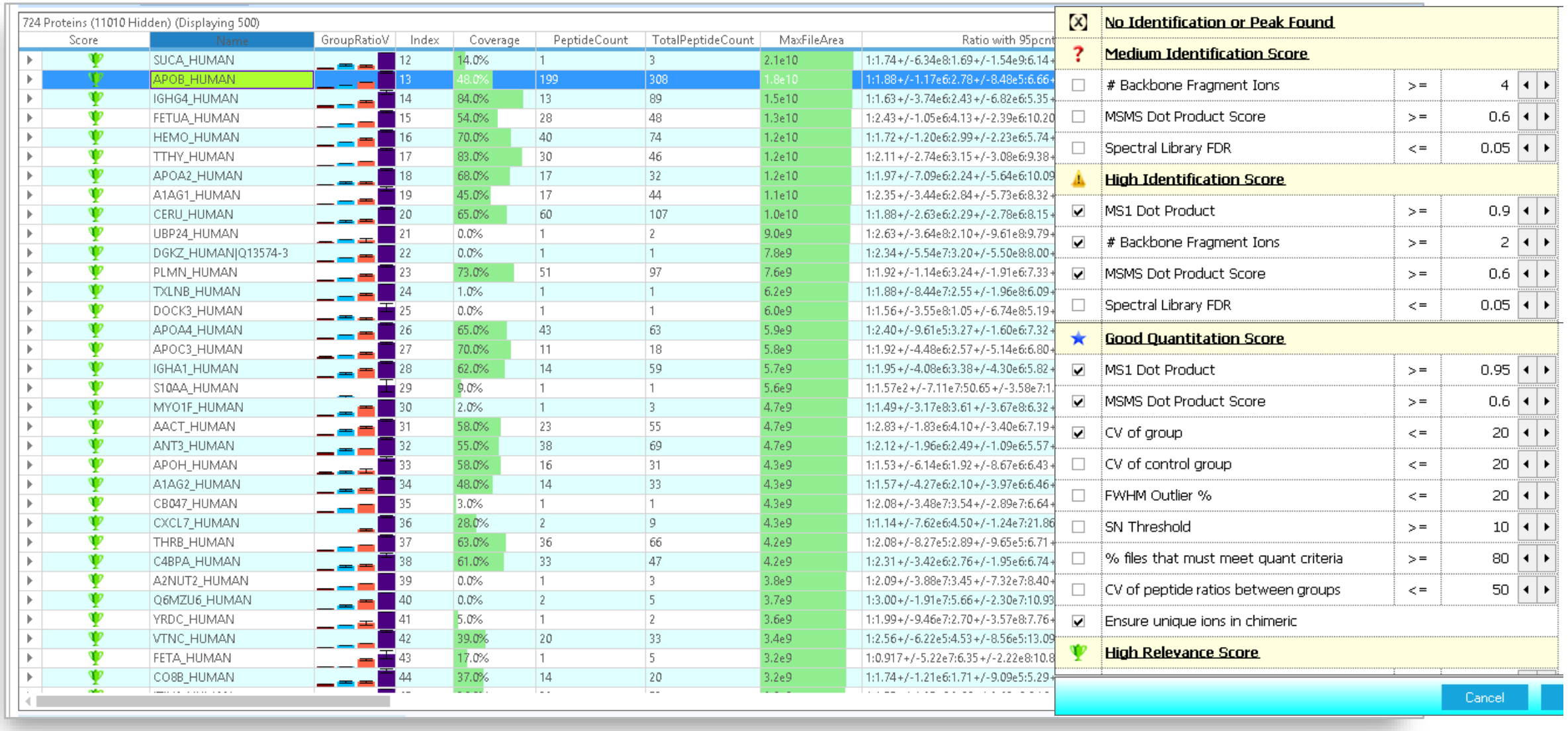


ThermoFisher SCIENTIFIC

Testing the BRIMS Workflow – Three Different Scenarios

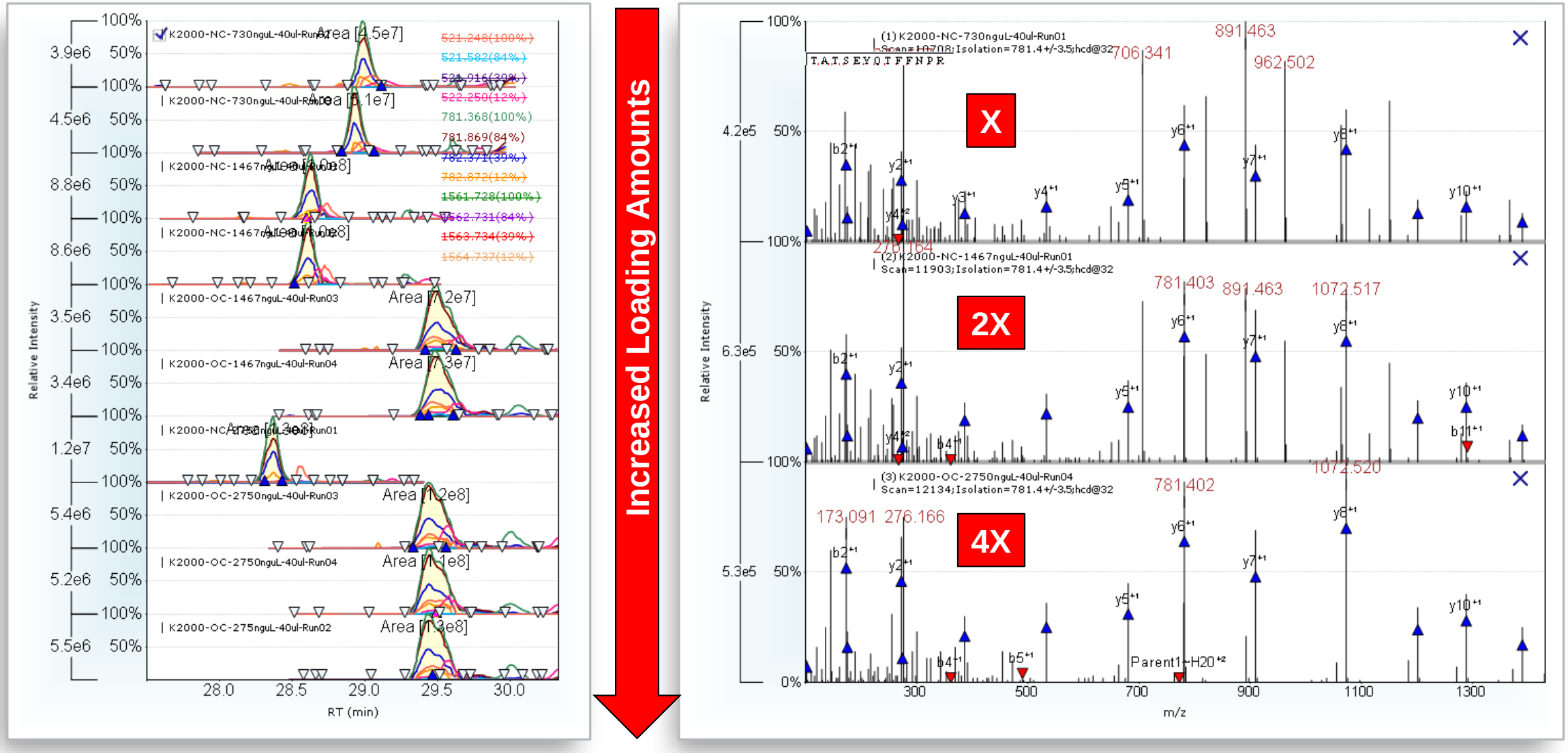
- Standard plasma samples from a single healthy donor
  - Plasma load study (29, 58, 110, and 220 µg) to evaluate loading capacity and replicate stability
    - Stock plasma at different concentrations through dilution factors
    - Two different sets of analytical columns (>1500 injections prior to study) and brand new columns
  - Large-scale sample analysis
    - 15 different draws and 3 different aliquots per tube
    - Each aliquot analyzed by 8 technical replicates for a total of 360 injections
    - 24 injections out of a single well containing a pooled sample acquired at the end of each row of wells
- Set of pooled plasma samples from six different donor groups (Cedars Sinai Medical Center and Uni. of Louisville Medical Center)
  - Each pool of plasma was sent following centrifugation at Louisville
  - Digestion was performed following the same protocols as that for the plasma standards above
  - Technical replicates were performed on each of the 6 pooled samples

Evaluation of the Loading Study



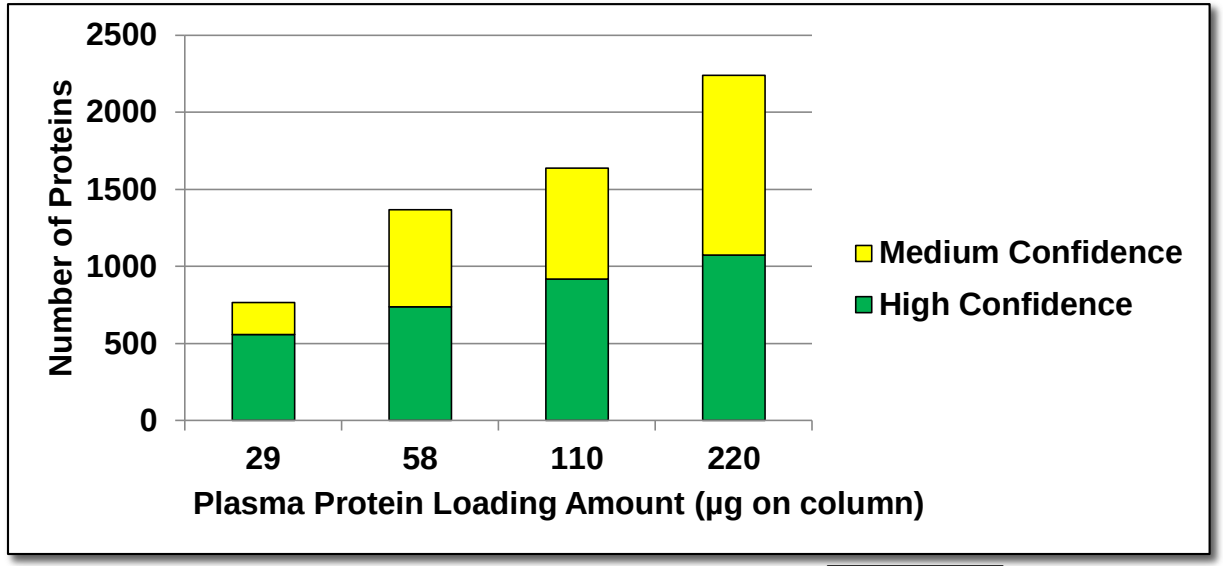
ThermoFisher SCIENTIFIC

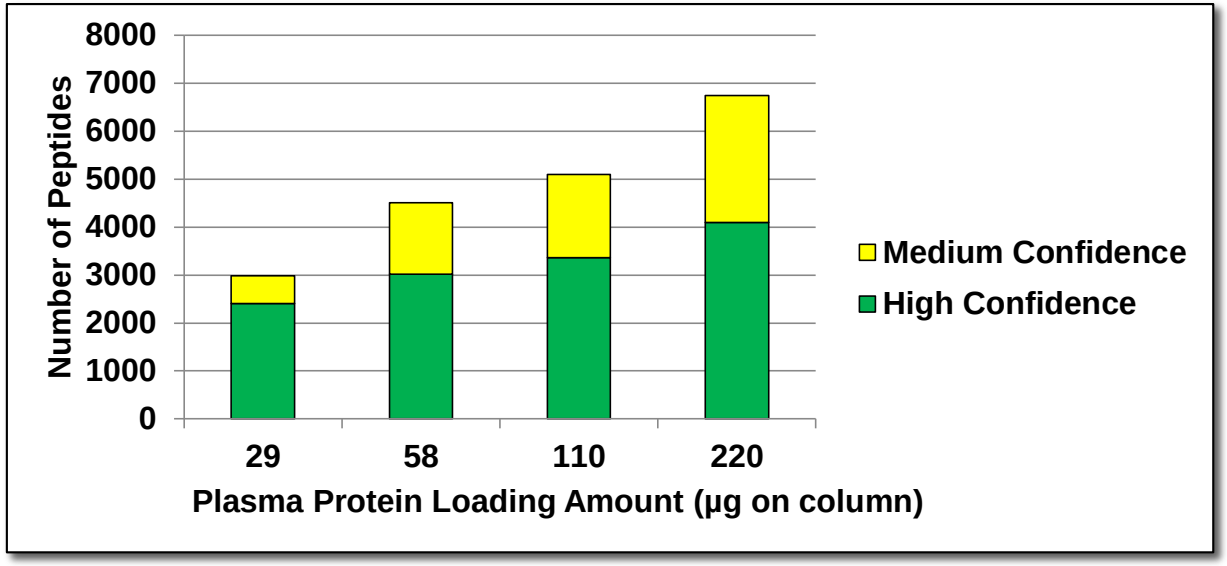
Supporting Peptide Information – Qualitative and Quantitative Analysis

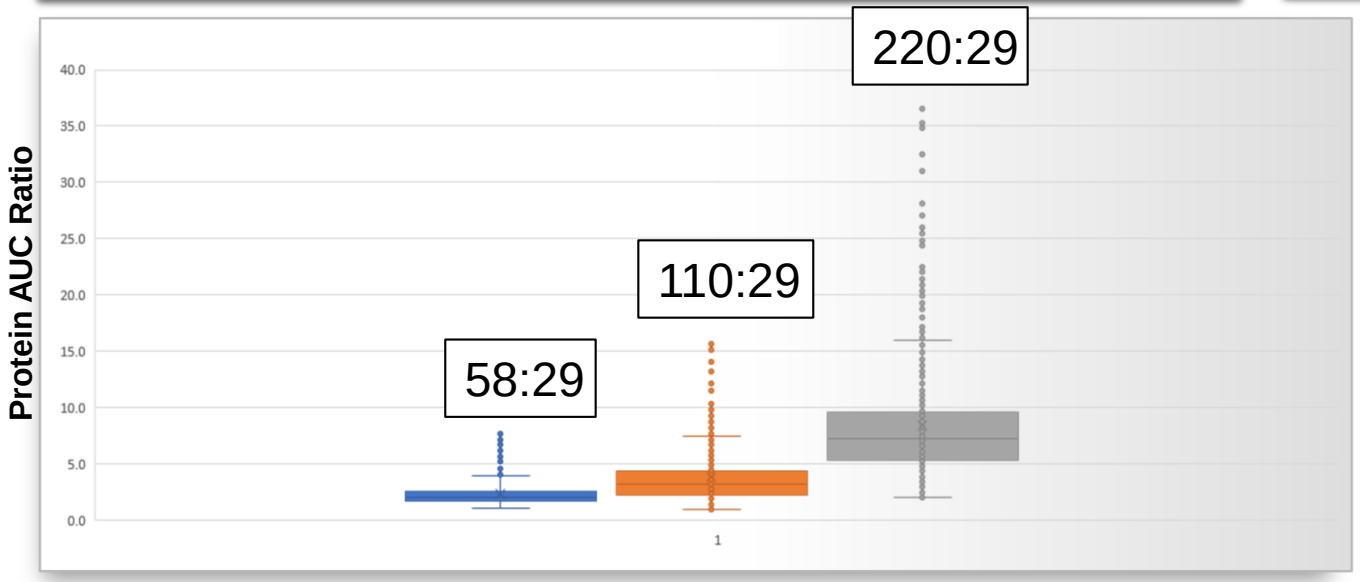


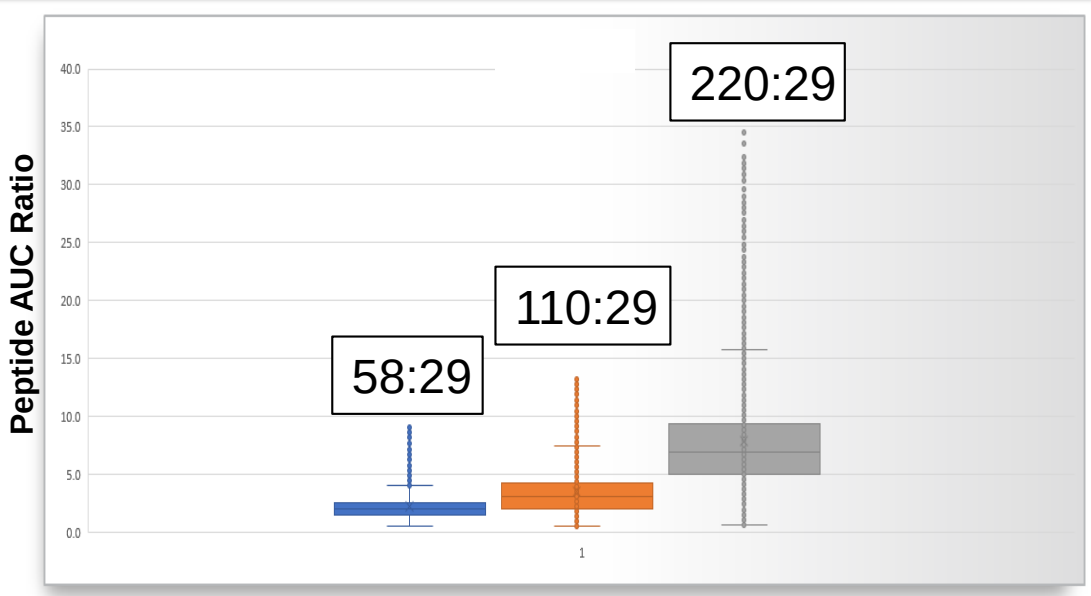
ThermoFisher SCIENTIFIC

Protein Loading Amounts (ug on column)









ThermoFisher SCIENTIFIC

Robustness Evaluation -

Sample Collection

16 bioloigcal samples

Sample Preparation

- **1 hr delay**
- **Centrifugation (2000 RCF for 30 min)**

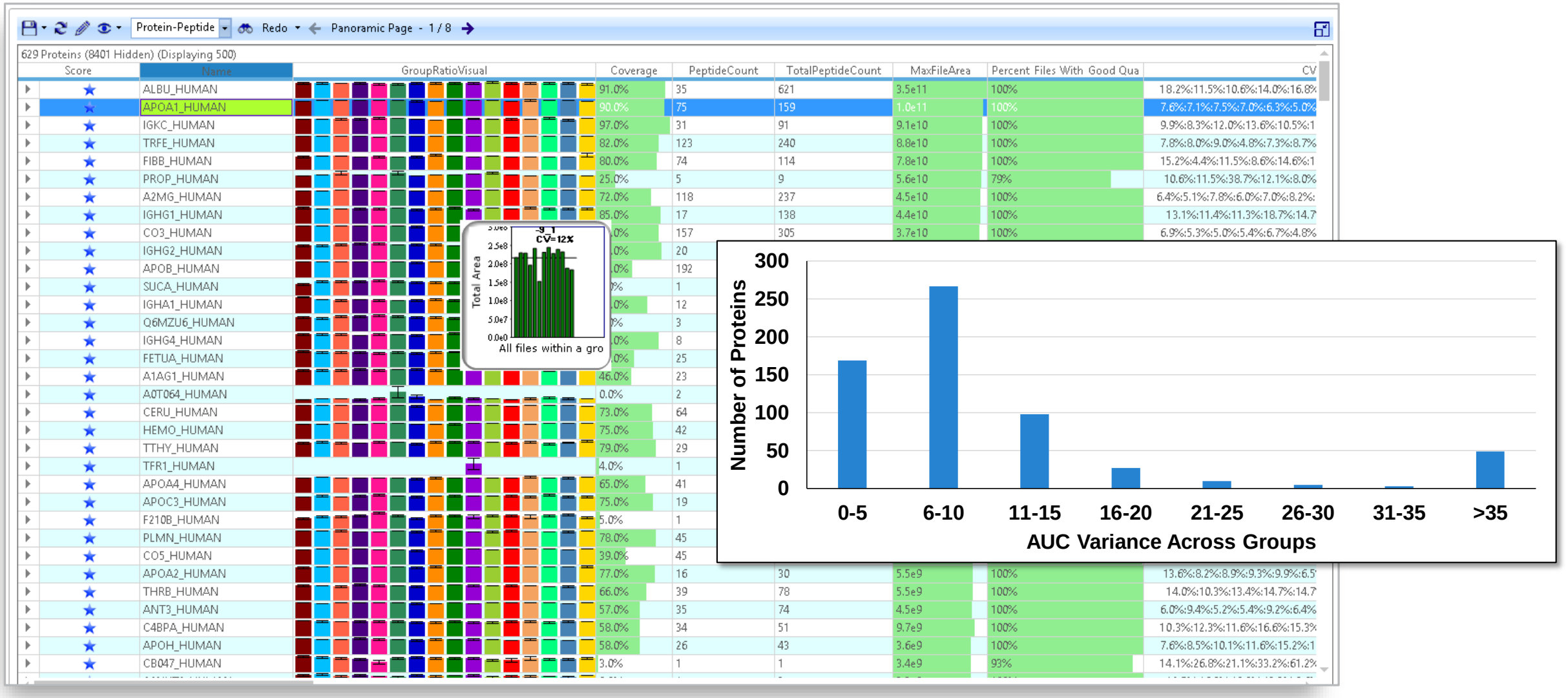
48 biological and technical replicates

- **Extract 3 different 100 uL aliquots**

336 biological and technical replicates

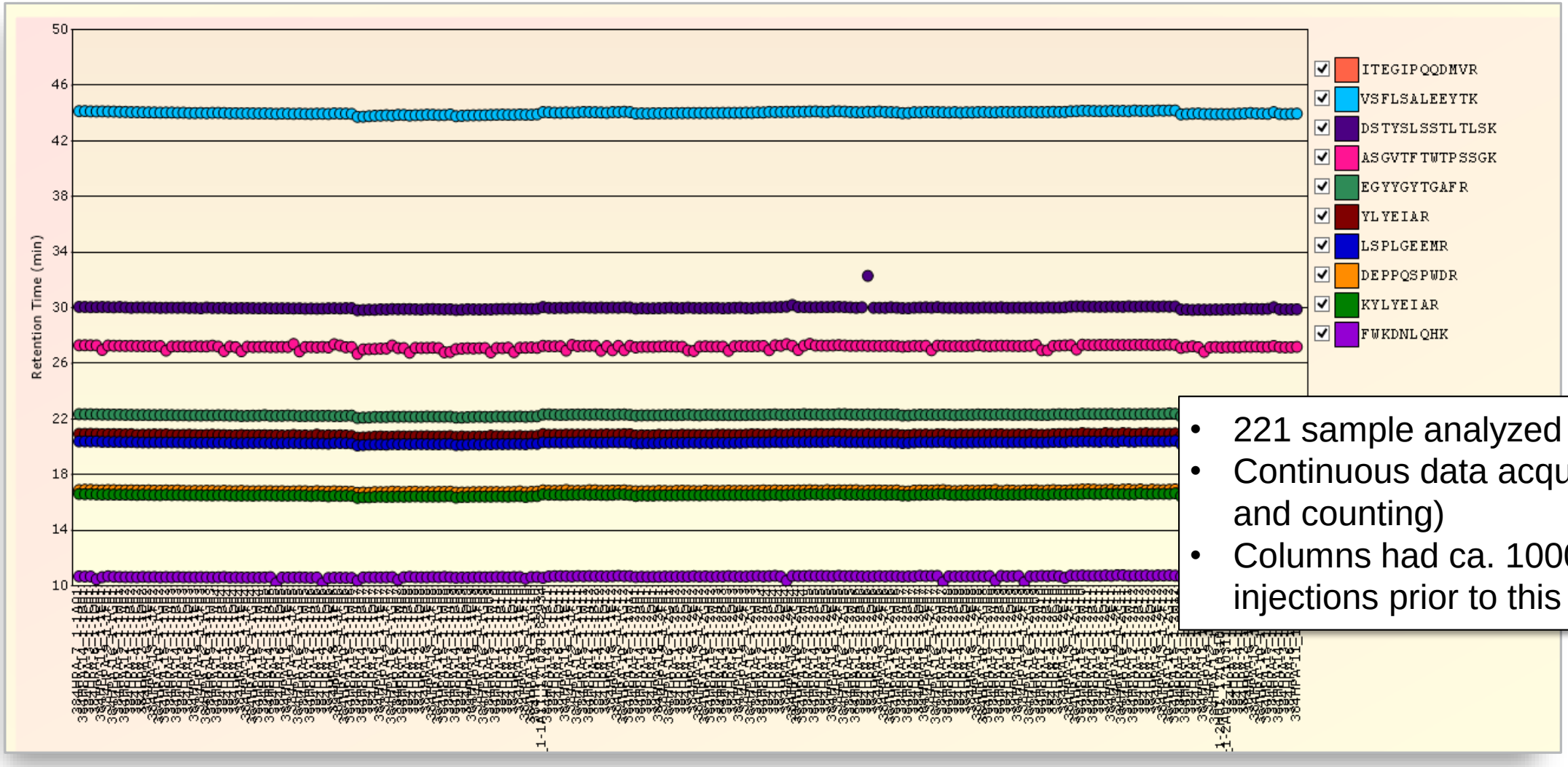
- **Extract 7 different 10 uL aliquots**
- **Spike with β-Gal (QC)**
- **Digest**
- **Spike digested samples with PRTC kit**
- **Analyze**

Global Analysis of 2.5 Plates of Plasma Replicates



ThermoFisher SCIENTIFIC

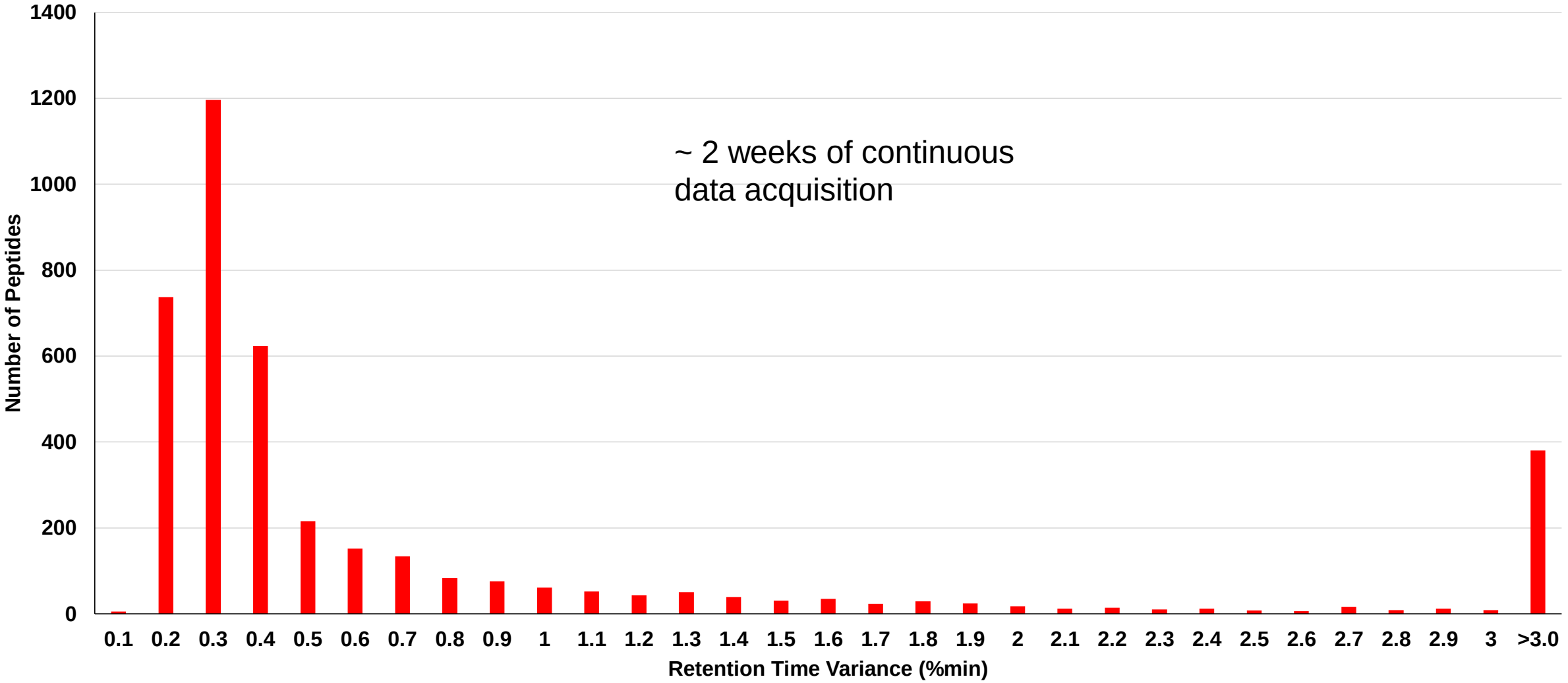
Large-scale reproducibility – Apo A1



- 221 sample analyzed (of 384)
- Continuous data acquisition (12 days and counting)
- Columns had ca. 1000 plasma injections prior to this data set

ThermoFisher SCIENTIFIC

Retention Time Stability Analysis for All Peptides

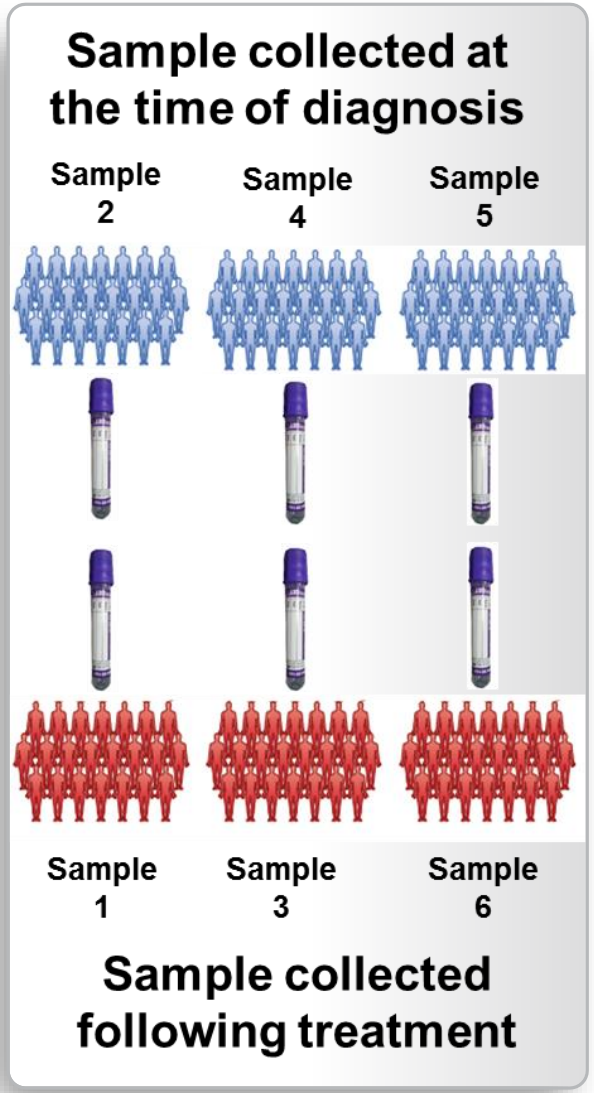


ThermoFisher SCIENTIFIC

Evaluating the Cedars Pooled Samples to Assess the Experimental Workflow

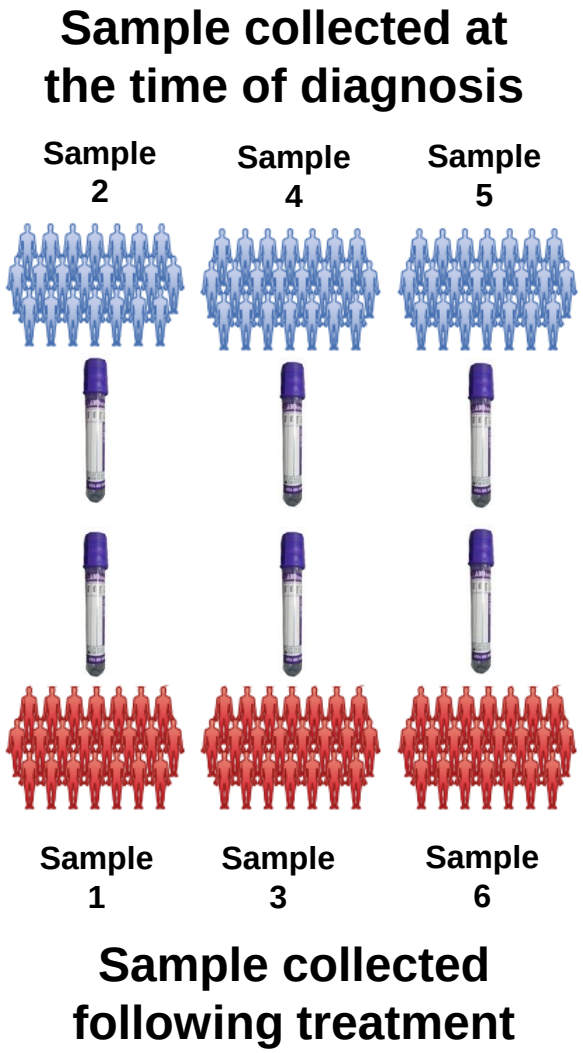
CEDARS-SINAI

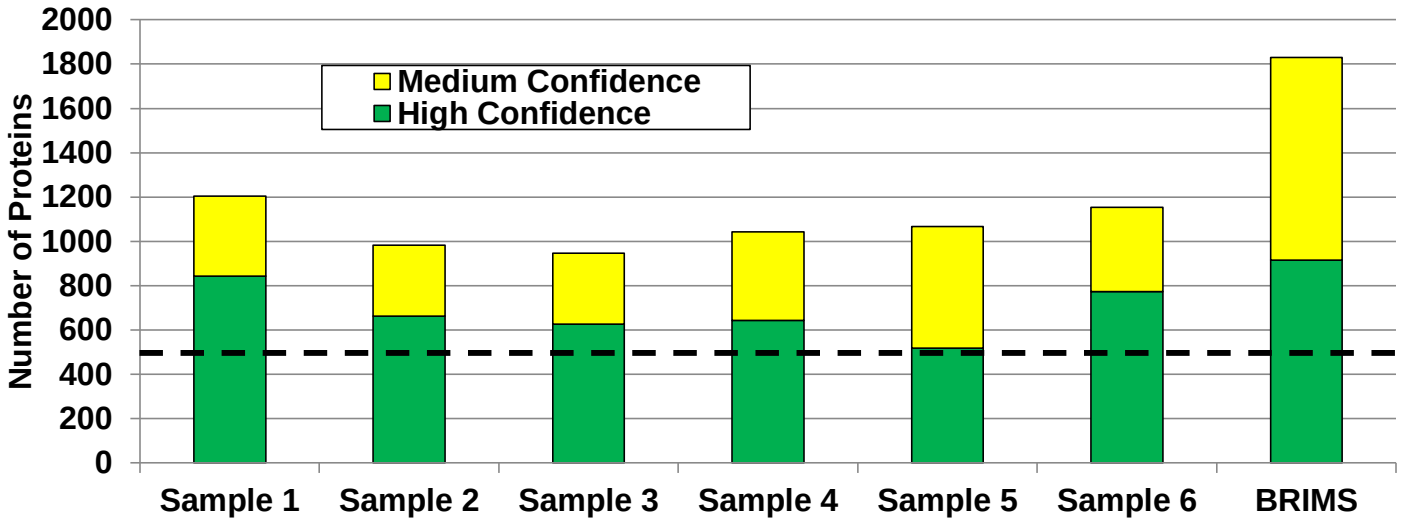
- Goals of the demonstration
  - Evaluate the degree of plasma proteome coverage based on quantifiable proteins
  - Evaluate the reproducibility (%CV) for the LC-MS method
  - Determine the overlap of targeted peptides quantified using HRAM MS and the targeted protein list previously done by SRM
- Instrumentation
  - UHPLC separations using analytical flow rates and 2.1 mm ID columns
  - Maintain HRAM quality data using (OT-OT)
- Data Processing
  - Sample spectral library generation (24 fractions)
  - Qualitative and quantitative assessment of peptides and proteins
  - Quality control and normalization strategies
  - Evaluate coverage of targeted protein panels (72) routinely analyzed using SRM experiments

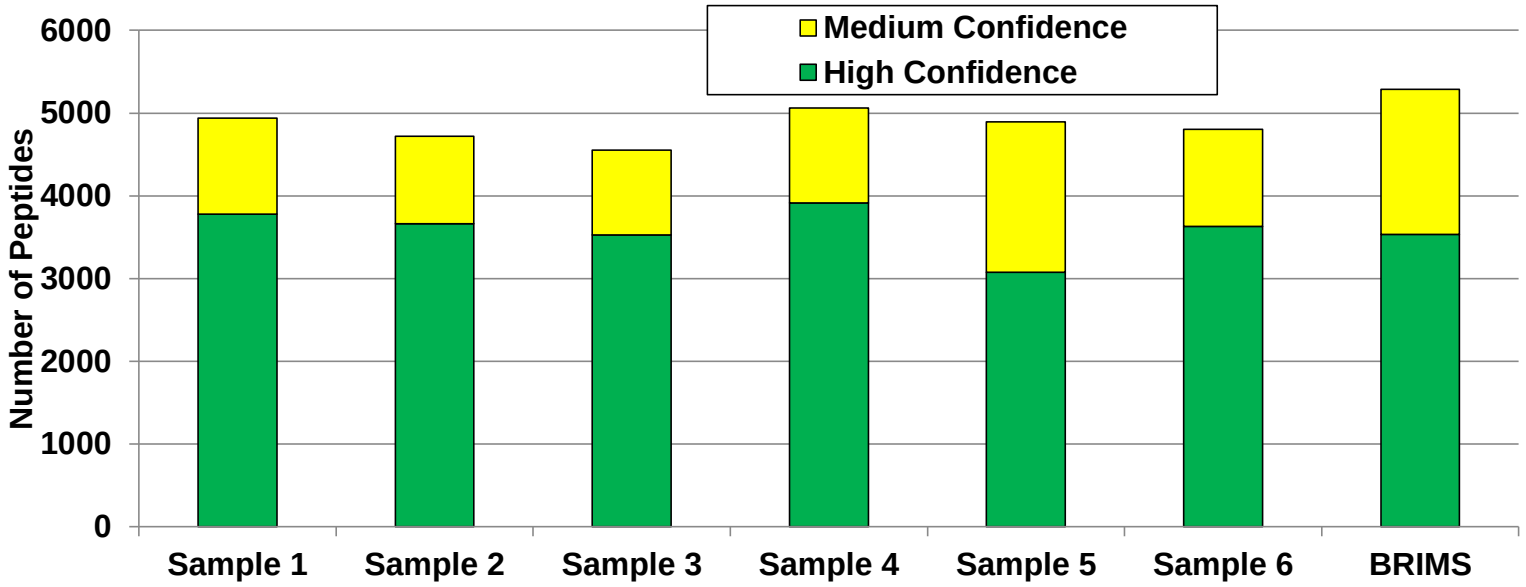


ThermoFisher SCIENTIFIC

Replicate Analysis



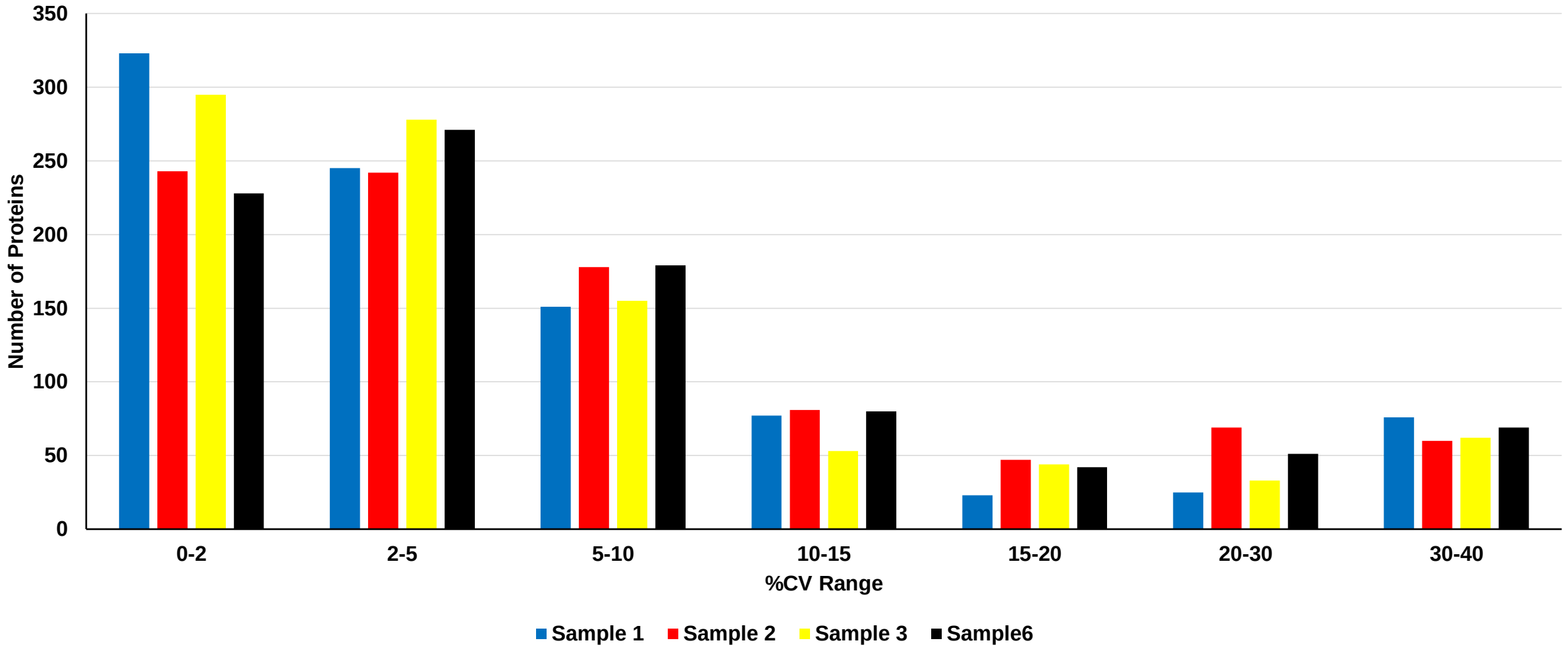




ThermoFisher SCIENTIFIC

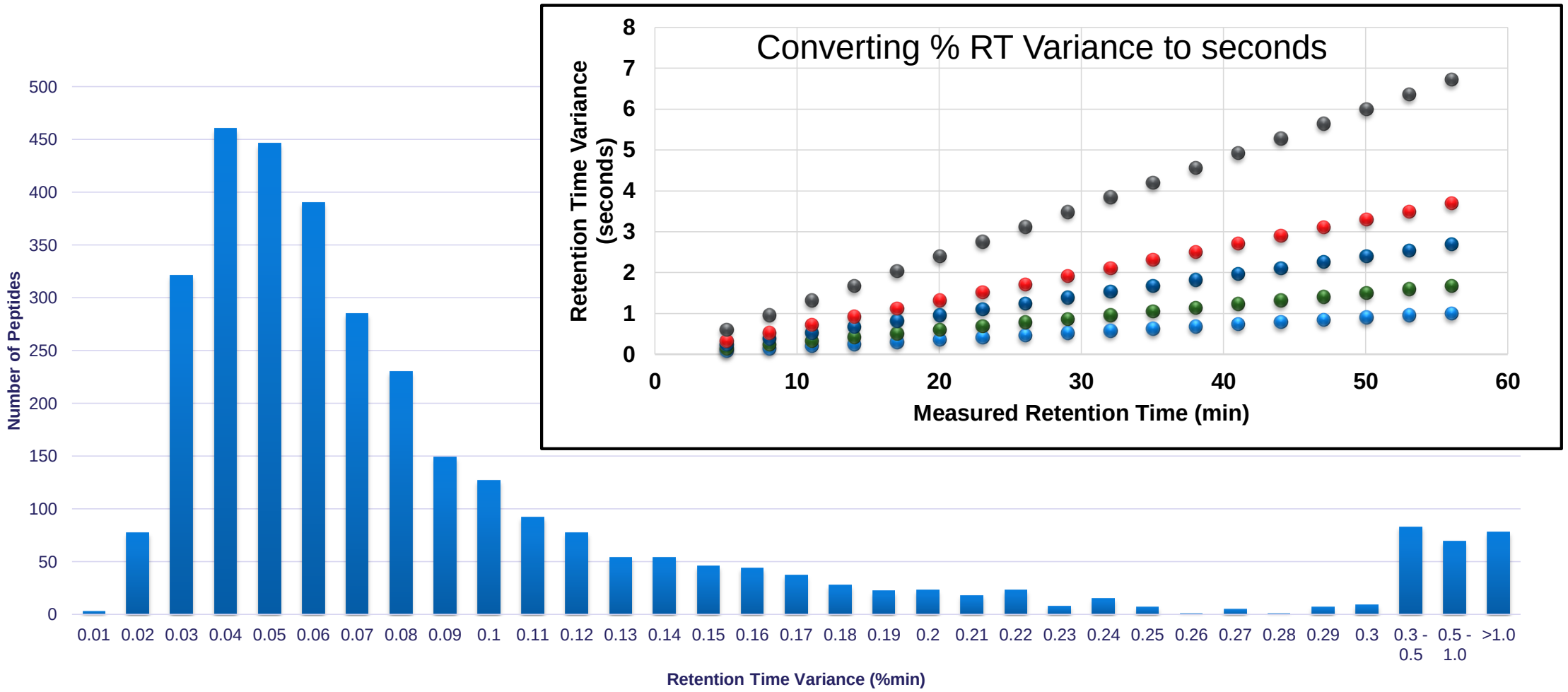
Comparative Analysis of Measure Protein AUC Variance

AUC values reported are determined based on median AUC values for high-quality peptides



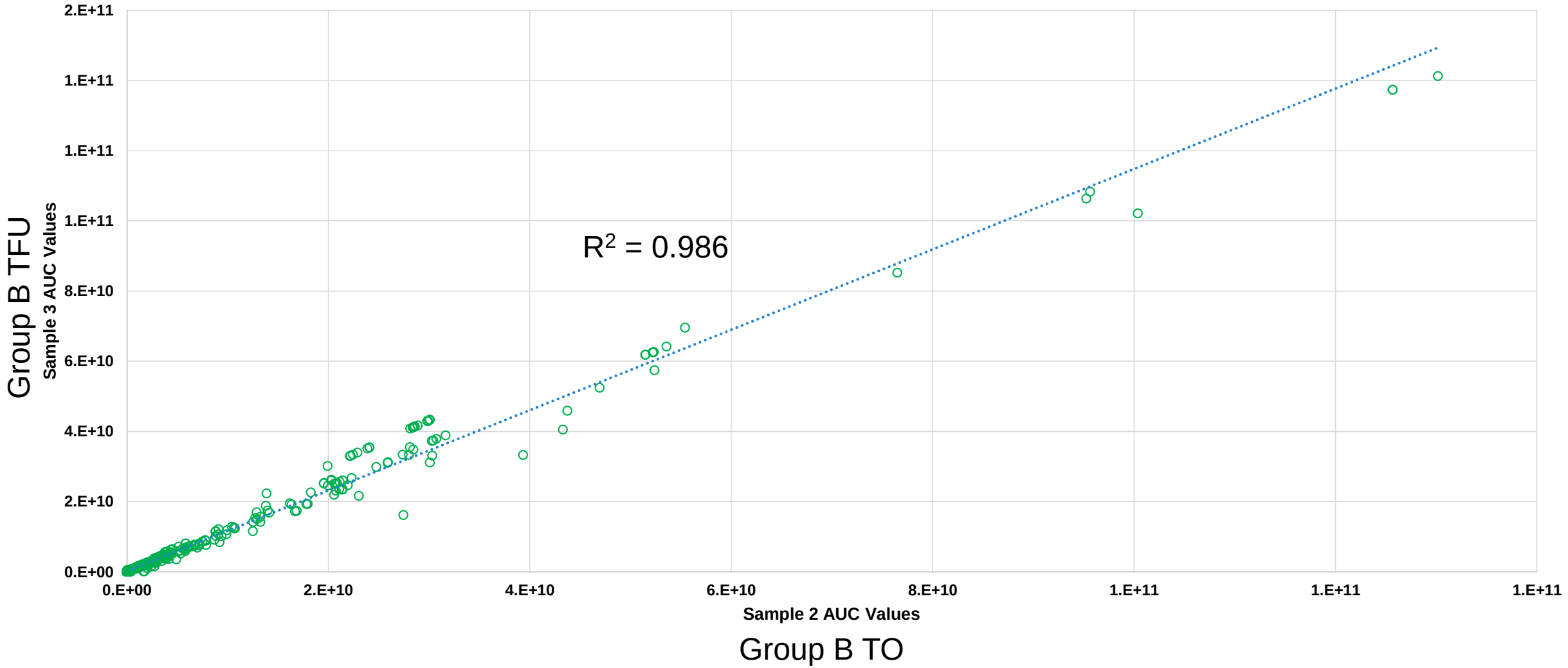
ThermoFisher SCIENTIFIC

Determination of Retention Time Stability (3 Days of Continuous Data Acquisition)



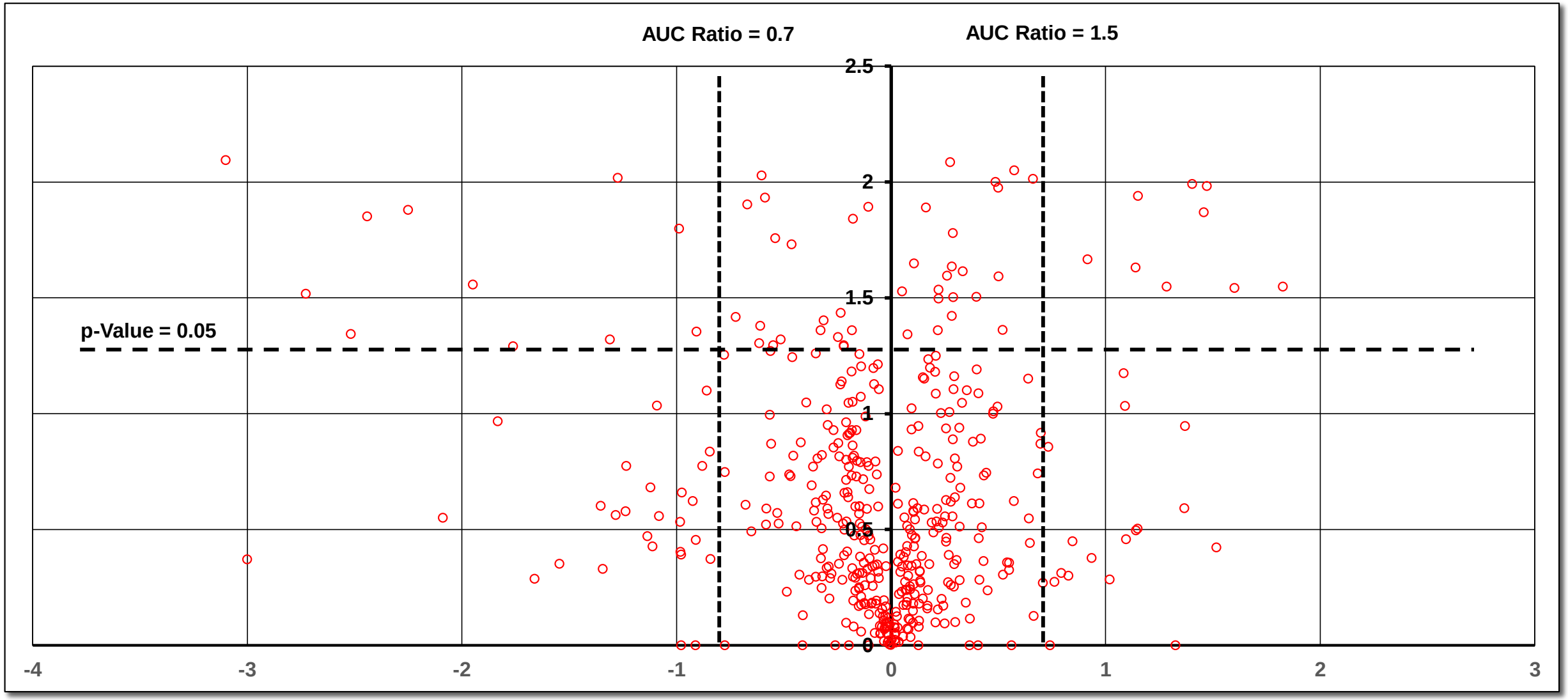
ThermoFisher SCIENTIFIC

Evaluating Proteins Levels Between Samples



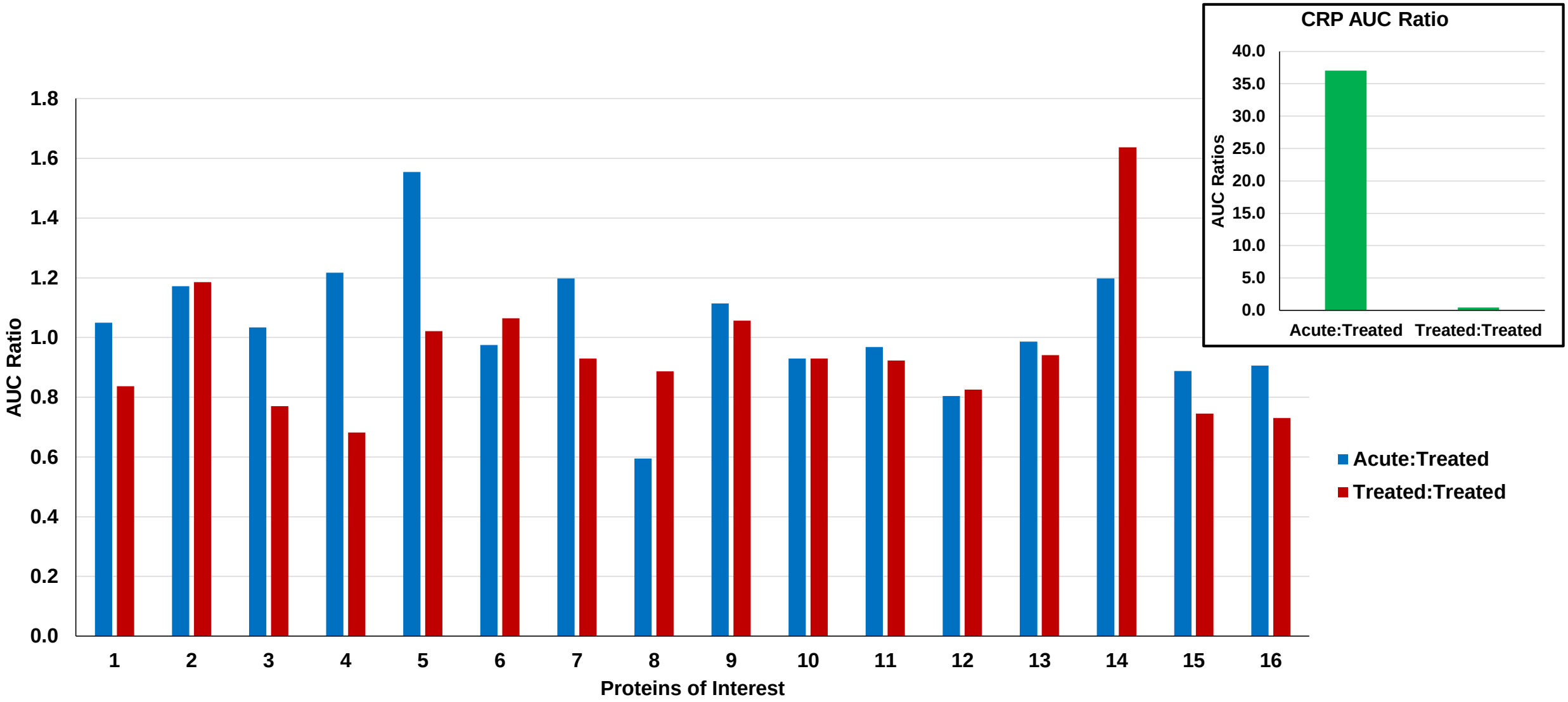
ThermoFisher SCIENTIFIC

Increasing the Graphical Value through Volcano Plots



ThermoFisher SCIENTIFIC

Comparative AUC Ratio Analysis for a Portion of the Cedars Targeted Panel



ThermoFisher SCIENTIFIC

Summary from Cedars Data

- Required over 500 proteins routinely detected and quantified
  - **Averaged ca. 678 highly confident proteins (as well as an additional 388 proteins identified) and ca. 3600 peptides in a 52 minute gradient**
- Maintain a 1 hour injection cycle
  - **Experimental method utilzies 52 minute gradient and 10 minutes of QC/column re-equilibration routien**
- Evaluate the detection capabilities for Cedars targeted proteins used for disease profiling – 72 medium- to high-level proteins currently being routinely quantified using Sciex QQQ on a 52 minutes gradient
  - **Was able to detect, verify, and quantify 73 proteins as well as detect the specific set of peptides targeted in the Cedars QQQ method**
- Data processing workflow is automated, exhaustive and requires little manual data interrogation
  - **Combination of sample-specific spectral libraries, wide DDA methods, and Pinnacle software enabled accurate data processing and interpretation as well as automate method reports**

Additional Conclusions – Overlap with Quantified Proteins and FDA Target Lists

Clinical Chemistry 56:2
177–185 (2010)

Mini-Reviews

The Clinical Plasma Proteome: A Survey of Clinical Assays for Proteins in Plasma and Serum

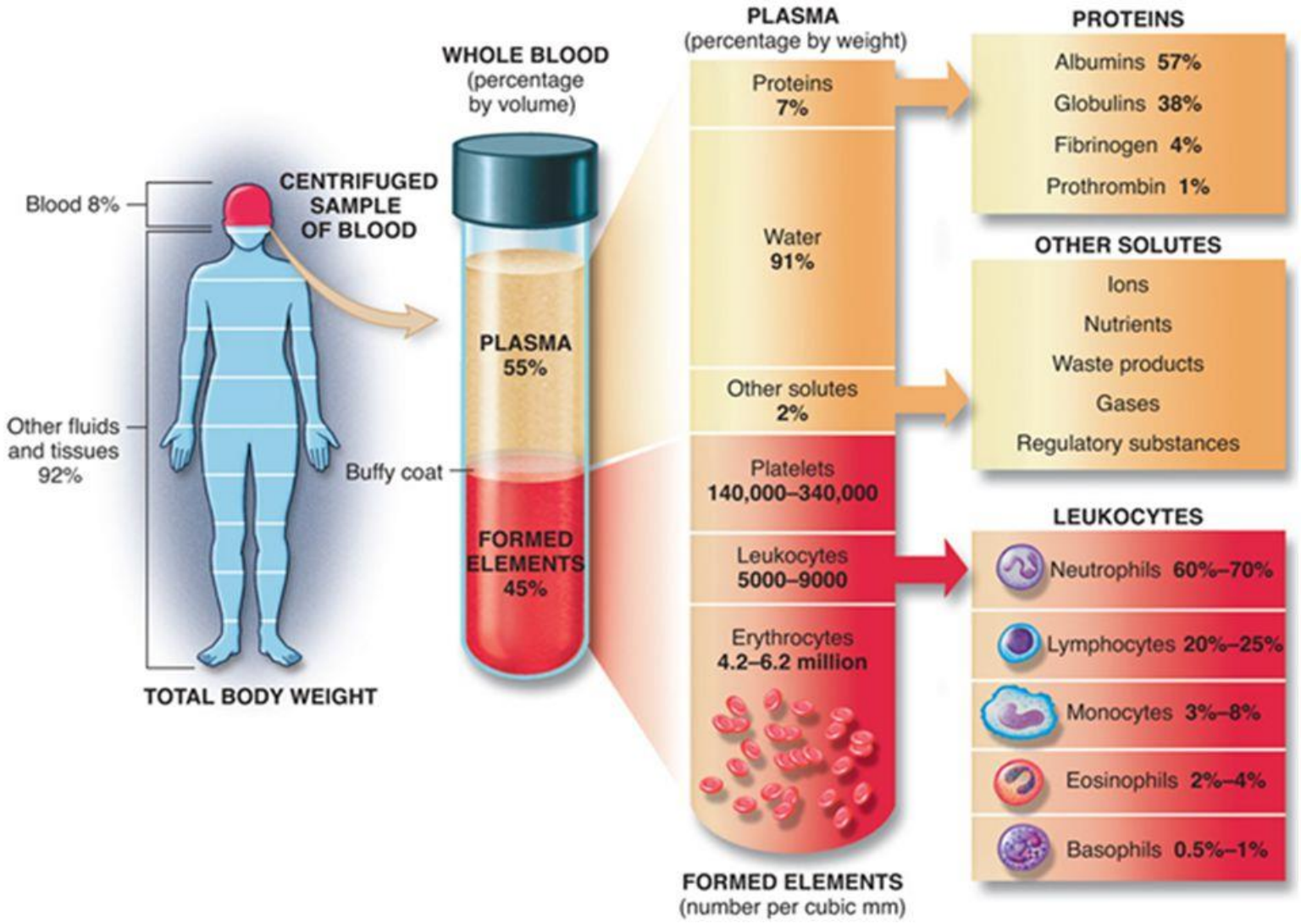
N. Leigh Anderson[1]

Table 1. FDA-cleared or -approved protein analytes assayed in serum or plasma.

1	Acid phosphatase	56	IgG
2	Alanine aminotransferase (ALT or SGPT)	57	IgM
3	Albumin	58	Inhibin-A
4	Aldolase	59	Insulin
5	Alkaline phosphatase (ALP)	60	Insulinlike growth factor-I (IGF-I)
6	α-1-Acid glycoprotein (orosomucoid)	61	Insulinlike growth factor-II (IGF-II)
7	α-1-Antitrypsin	62	IGFBP-1
8	α-2-Antiplasmin	63	IGFBP-3
9	α-2-HS-glycoprotein	64	Interleukin-2 receptor (IL-2R)
10	α-2-Macroglobulin	65	Isocitric dehydrogenase
11	α-Fetoprotein (tumor marker)	66	κ Light chains
12	Amylase	67	Lactate dehydrogenase heart fraction (LDH-1)
13	Amylase, pancreatic	68	Lactate dehydrogenase liver fraction (LLDH)
14	ACE	69	Lactoferrin
15	Antithrombin III (ATIII)	70	λ Light chains
16	Apolipoprotein A1	71	Lipase
17	Apolipoprotein B	72	Lp(a)
18	Aspartate aminotransferase (AST or SGOT)	73	Lipoprotein-associated phospholipase A2 (LP-PLA2)
19	β-2 Microglobulin	74	LH
20	β-Thromboglobulin	75	Lysozyme
21	Biotinidase	76	Myeloperoxidase (MPO)
22	Cancer antigen 125 (CA 125)	77	Myoglobin
23	Cancer antigen 15-3 (CA 15-3)	78	Osteocalcin
24	Cancer antigen, human epididymis protein 4 (HE4)	79	Parathyroid hormone, intact
25	Carcinoembryonic antigen (CEA)	80	Phosphohexose isomerase
26	Ceruloplasmin	81	Plasminogen
27	Cholinesterase	82	Plasminogen activator inhibitor (PAI)
28	Complement C1	83	Prealbumin
29	Complement C1 Inhibitor	84	NTproBNP
30	Complement C1Q	85	Procalcitonin (PCT)
31	Complement C3	86	Prolactin
32	Complement C4	87	Properdin factor B
33	Complement C5	88	Prostatic acid phosphatase (PAP)
34	CRP	89	Prostatic specific antigen (PSA)
35	Creatine kinase-BB (CKBB)	90	Protein C
36	Creatine kinase-MM (CKMM)	91	Protein S
37	Cystatin C	92	Pseudocholinesterase
38	Erythropoietin	93	Pyruvate kinase
39	Factor IX antigen	94	Renin
40	Factor X	95	Retinol binding protein (RBP)
41	Factor XIII	96	Sex hormone–binding globulin
42	Ferritin	97	Soluble mesothelin-related peptide
43	Fibrinogen	98	Sorbital dehydrogenase (SDH)
44	Fibronectin	99	Thyroglobulin
45	FSH	100	TSH
46	GGT	101	Thyroxine binding globulin (TBG)
47	Haptoglobin	102	Tissue plasminogen activator (T-PA)
48	Human chorionic gonadotropin (hCG), beta, serum, quantitative	103	Transferrin
49	Hemopexin	104	Transferrin receptor (TFR)
50	her-2/neu protein	105	Troponin T (TnT)
51	Human growth hormone (HGH)	106	TnI (cardiac)
52	Human placental lactogen (HPL)	107	Trypsin
53	IgA	108	Urokinase
54	IgD	109	Von Willebrand factor
55	IgE		

- **Identified and quantified peptides attributed to 48 proteins identified as FDA-cleared or approved**
- **Keep in mind the disease plasma evaluated was from donors suffering from cardiovascular disease and not cancer**
- **Efforts are underway to obtain plasma samples from cancer donors to generate a comprehensive spectral-library to complement the CVD spectral library**

Expanding Biomarker Potential in Whole Blood



- Primary biological fluid for clinical research
- Research focused on routine measurements of defined proteome
- Whole blood is still challenging for proteome coverage due to dynamic range and complexity
- Depletion strategies still face question and differing opinions on utility
- Leverage constituents of blood for more targeted analysis
- Other targeted extraction procedures
  - HDL/LDL
  - Exosomes
    - Targeted cellular extraction
  - Peptidome
  - Platelets
  - PBMCs

Conclusions and Projections

- Sample Preparation
  - Alternative buffers and denaturants increase digestion efficiency
  - Spinning plasma samples at 2000 RCF for 30 minutes reduces post-collection perturbation
- Introduction of the Trapping and Analytical Column Configuration
  - Handles increased loading amounts
  - Maximizes throughput
  - Facilitates harsh cleaning solvents to significantly reduce carry-over
  - Stabilizes retention times for direct method transfer to targeted studies
- Data Acquisition Strategies
  - Maintains quantitative accuracy
  - Exhaustive data extraction and reproducibility
- Data Processing using Spectral Libraries

Acknowledgements

- BRIMS Center
  - David Sarracino, Ph.D
- Cedars Sinai Medical Center
  - Jennifer Van Eyk, Ph. D.
  - Irina Tschernyshyov, Ph. D.
  - Qin Fu, Ph.D.
- Optys Technologies
  - Amol Prakash, Ph. D.

ThermoFisher SCIENTIFIC

Appendix

Experimental

- Autosampler and LC System
  - Vanquish UHPLC binary pump
- Column
  - Set of two Acclaim 120 columns with dimensions of 2.1 x 250 mm (for each column)
  - The two columns linked in series through a small stainless still tube and Viper fittings
  - Spectral libraries and initial data was acquired using old columns (1500-2000 injections on them)
  - All data was acquired on brand new columns
- Mass Spectrometer
  - Fusion Lumos Orbitrap mass spectrometer
  - Operated in OT-OT mode for HRAM measurements for all scan event cycles

- 60 min experiment

Time	%B
0	1
0.5	1
0.9	5
1.0	6
2.0	10
50	38
52	38

DDA/DE Settings for Fraction Library Acquisition

- HRAM MS
  - Resolution = 60k
  - AGC = 3e6
  - Max ion fill time = 50 msec
  - Mass Range = 350-1500 Da

- HRAM MS2
  - Resolution = 15k
  - AGC = 5e4
  - Max ion fill time = 75 msec
  - Acquisition Cycle time = 1 sec
  - Precursor isolation = 2.5 Da
  - Stepped CE starting at 30% and increasing 5%
  - Charge state (2-6) and isotope exclusion = On
- Dynamic Exclusion
  - Repeat count = 1
  - Exclusion duration = 6 sec

DDA/DE Settings for Routine Data Acquisition

- HRAM MS
  - Resolution = 60k
  - AGC = 3e6
  - Max ion fill time = 50 msec
  - Mass Range = 350-1500 Da

- HRAM MS2
  - Resolution = 15k
  - AGC = 1e5
  - Max ion fill time = 100 msec
  - Acquisition Cycle time = 1 sec
  - Precursor isolation = 8 Da
  - Stepped CE starting at 30% and increasing 5%
  - Charge state (2-6) and isotope exclusion = On
- Dynamic Exclusion
  - Repeat count = 1
  - Width = 4 Da
  - Exclusion duration = 6 sec

Spectral Library Generation

- Sample Preparation
  - 90 ug of pooled digest was collected across 6 samples sent
  - 24 fractions collected from RP high pH fractionation
  - Each fraction was loaded and analyzed using both 60 and 120 min experimental methods
  - LC-MS settings for fraction analysis (DDA/DE) shown in Slides 7 and 8
- LC-MS Data Acquisition
  - LC-MS settings for fraction analysis (DDA/DE) shown in Slides 7 and 8
- Unbiased Database searching
  - PD 2.1 using Uniprot human database
  - Mods incorporated into searching routines include: deamidation and oxidation
- Spectral Matching for Replicate Analysis
  - All replicate analysis performed in Pinnacle