ThermoFisher SCIENTIFIC

Quantification with Proteome Discoverer

Bernard Delanghe

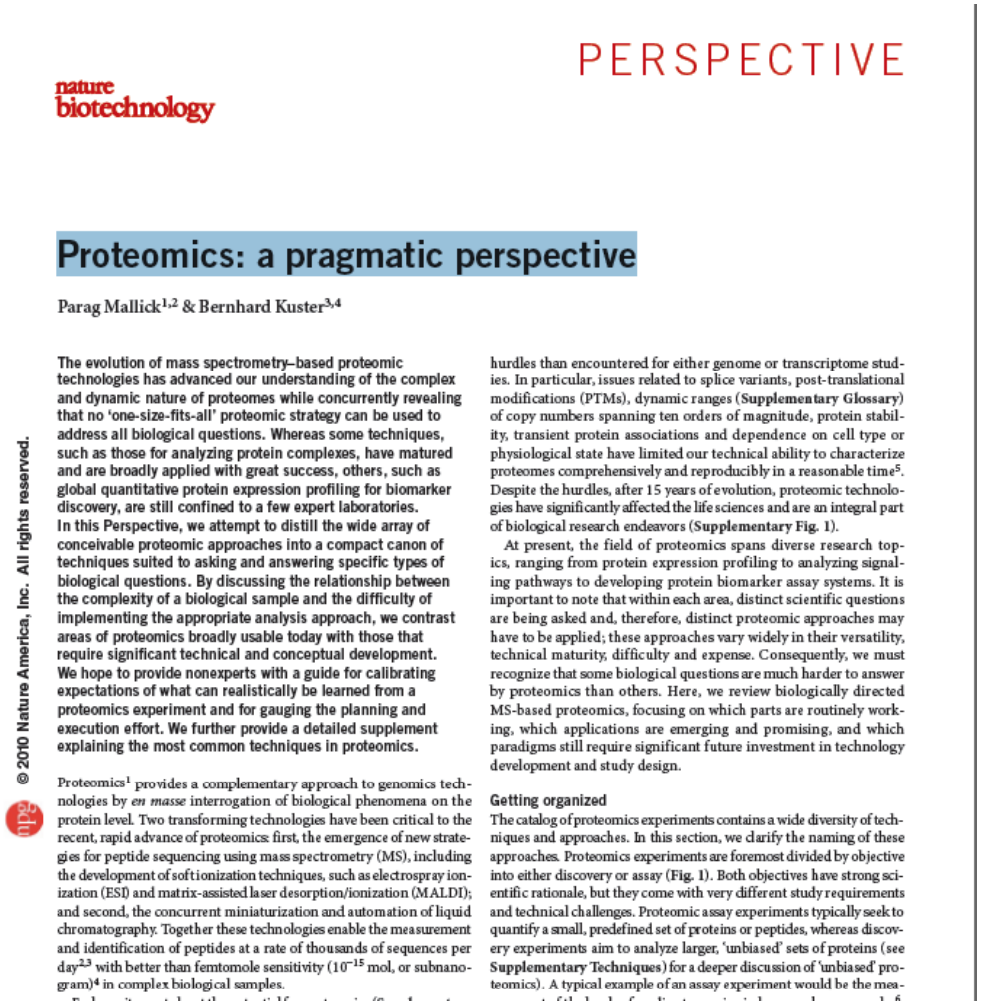
Overview: Which approach to use?

Proteome Discoverer

Quantification Method	What	When to use
Metabolic labeling	SILAC	Cell culture systems Small changes (10-50%)
Peptide labeling	Dimethylation, TMT, iTRAQ	Tissue proteins Multiplexing (time courses) Moderate changes (20-200%)
Label free using detector response	Extracted Ion chromatograms, area calculation	Many largely similar experiments Moderate changes (20-200%)
Label free using spectral counting	Number of spectra per protein	Many highly similar experiments Large changes (>100%)
Single or Multiple Reaction Monitoring (SMR or MRM)	Absolute quantification, spiked with standards (AQUA)	Complex biological matrix (Serum)

PinPoint

Proteomics and Quantification



nature biotechnology

PERSPECTIVE

Proteomics: a pragmatic perspective

Parag Mallick[1,2] & Bernhard Kuster[3,4]

The evolution of mass spectrometry–based proteomic technologies has advanced our understanding of the complex and dynamic nature of proteomes while concurrently revealing that no 'one-size-fits-all' proteomic strategy can be used to address all biological questions. Whereas some techniques, such as those for analyzing protein complexes, have matured and are broadly applied with great success, others, such as global quantitative protein expression profiling for biomarker discovery, are still confined to a few expert laboratories. In this Perspective, we attempt to distill the wide array of conceivable proteomic approaches into a compact canon of techniques suited to asking and answering specific types of biological questions. By discussing the relationship between the complexity of a biological sample and the difficulty of implementing the appropriate analysis approach, we contrast areas of proteomics broadly usable today with those that require significant technical and conceptual development. We hope to provide nonexperts with a guide for calibrating expectations of what can realistically be learned from a proteomics experiment and for gauging the planning and execution effort. We further provide a detailed supplement explaining the most common techniques in proteomics.

Proteomics[1] provides a complementary approach to genomics technologies by *en masse* interrogation of biological phenomena on the protein level. Two transforming technologies have been critical to the recent, rapid advance of proteomics: first, the emergence of new strategies for peptide sequencing using mass spectrometry (MS), including the development of soft ionization techniques, such as electrospray ionization (ESI) and matrix-assisted laser desorption/ionization (MALDI); and second, the concurrent miniaturization and automation of liquid chromatography. Together these technologies enable the measurement and identification of peptides at a rate of thousands of sequences per day[2,3] with better than femtomole sensitivity (10^{-15} mol, or subnanogram)[4] in complex biological samples.

hurdles than encountered for either genome or transcriptome studies. In particular, issues related to splice variants, post-translational modifications (PTMs), dynamic ranges (**Supplementary Glossary**) of copy numbers spanning ten orders of magnitude, protein stability, transient protein associations and dependence on cell type or physiological state have limited our technical ability to characterize proteomes comprehensively and reproducibly in a reasonable time[5]. Despite the hurdles, after 15 years of evolution, proteomic technologies have significantly affected the life sciences and are an integral part of biological research endeavors (**Supplementary Fig. 1**).

At present, the field of proteomics spans diverse research topics, ranging from protein expression profiling to analyzing signaling pathways to developing protein biomarker assay systems. It is important to note that within each area, distinct scientific questions are being asked and, therefore, distinct proteomic approaches may have to be applied; these approaches vary widely in their versatility, technical maturity, difficulty and expense. Consequently, we must recognize that some biological questions are much harder to answer by proteomics than others. Here, we review biologically directed MS-based proteomics, focusing on which parts are routinely working, which applications are emerging and promising, and which paradigms still require significant future investment in technology development and study design.

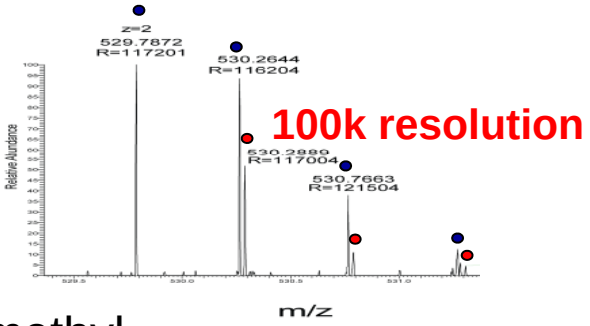
Getting organized

The catalog of proteomics experiments contains a wide diversity of techniques and approaches. In this section, we clarify the naming of these approaches. Proteomics experiments are foremost divided by objective into either discovery or assay (**Fig. 1**). Both objectives have strong scientific rationale, but they come with very different study requirements and technical challenges. Proteomic assay experiments typically seek to quantify a small, predefined set of proteins or peptides, whereas discovery experiments aim to analyze larger, 'unbiased' sets of proteins (see **Supplementary Techniques**) for a deeper discussion of 'unbiased' proteomics). A typical example of an assay experiment would be the mea-

ThermoFisher SCIENTIFIC

New feature for Proteome Discoverer 1.2

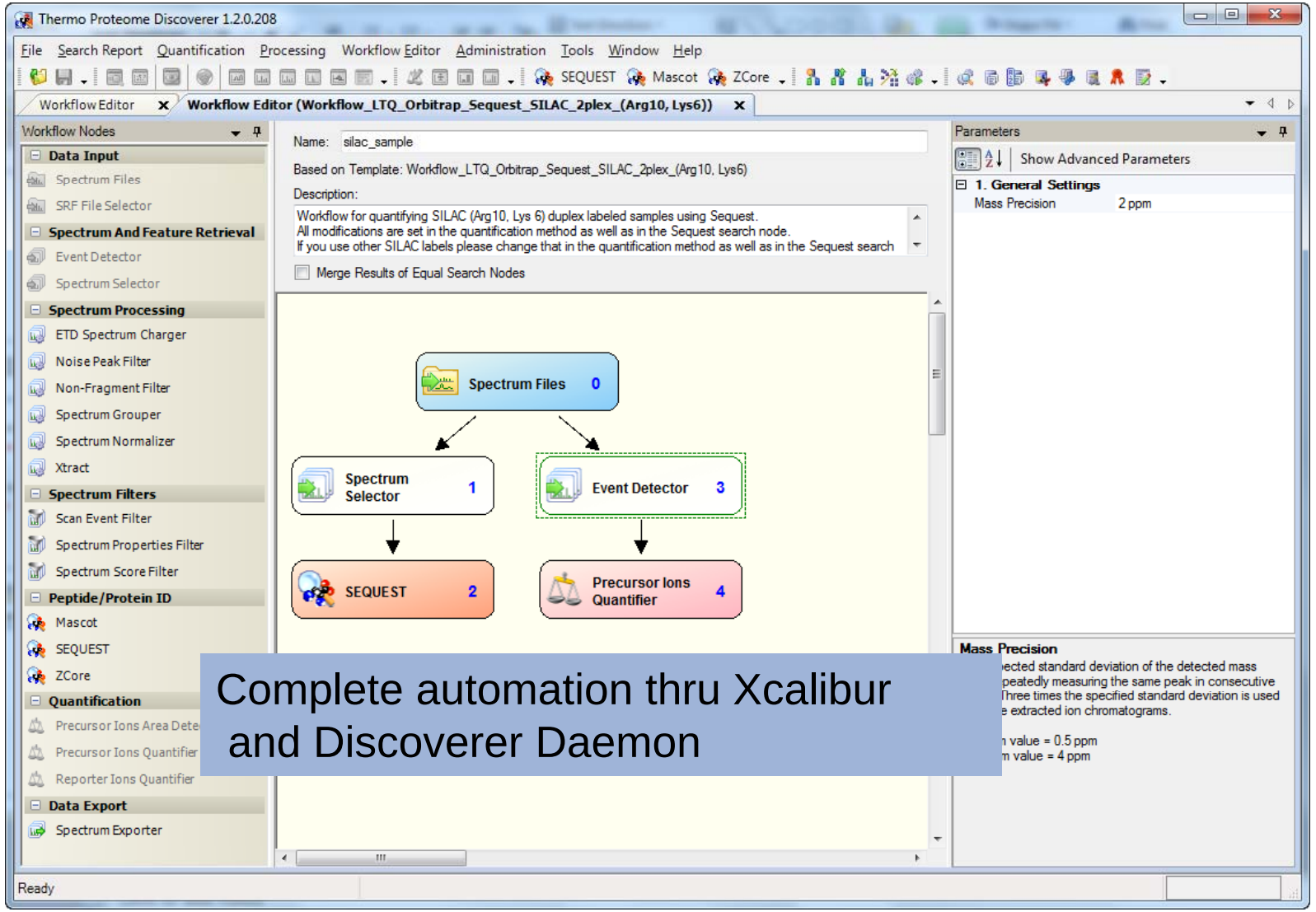
- Fast, Easy and automated Stable Isotope Precursor Ion Quantification
- Exploits the capabilities of high-resolution MS with precursor ion-based quantification



100k resolution

1. Classical SILAC with heavy Arginene and Lysine
2. SILAC with heavy Isoleucine
3. Peptides labeled with light, medium, and heavy dimethyl
4. Others like O16/O18, ICAT, ICPL
5. Doublets or Triplets
6. Using any enzyme
7. Any fragmentation technique (or a combinartion):CID, HCD, ETD, ECD
8. Any search engine (or a combinartion : Mascot, Sequest or Zcore

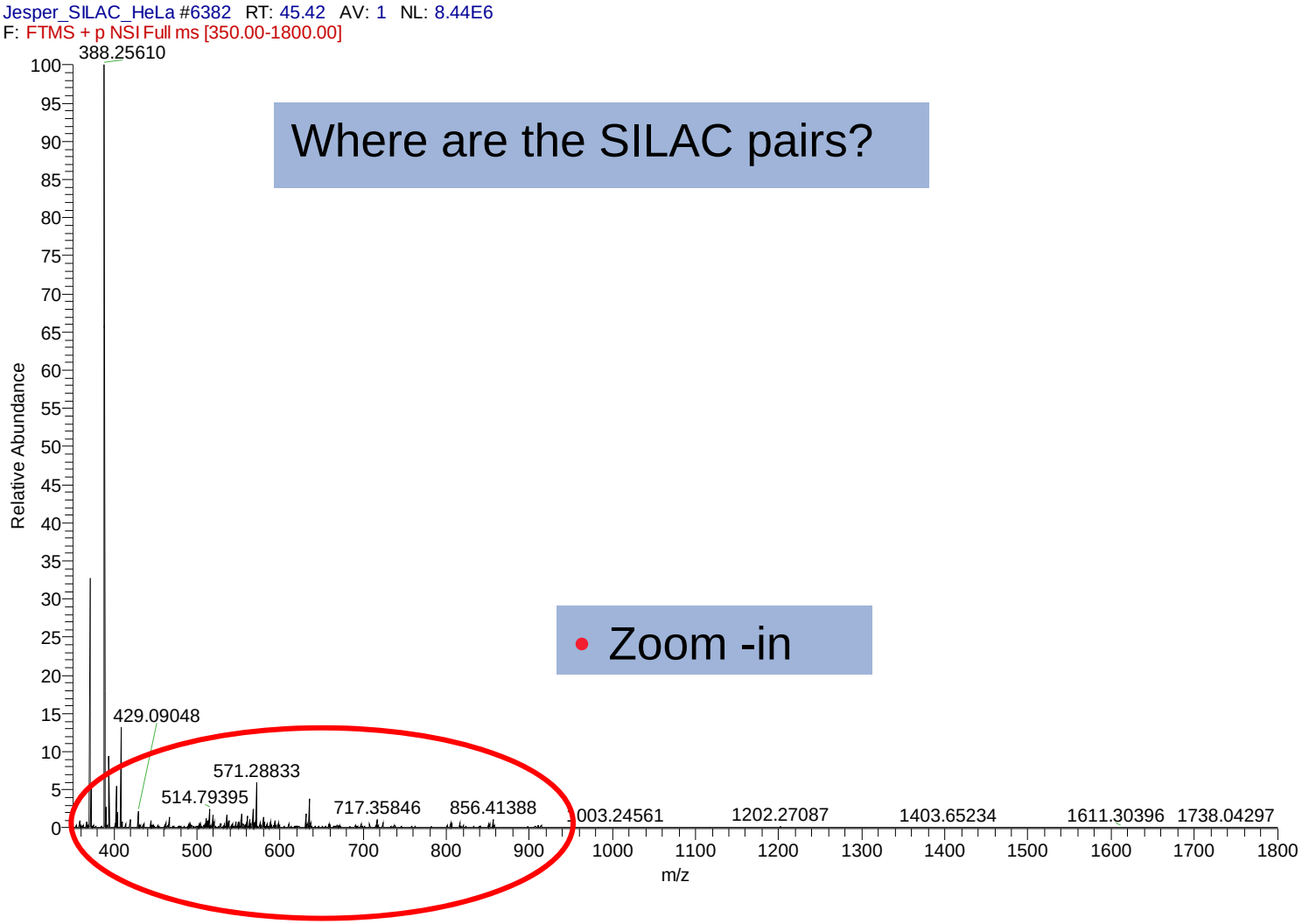
Preconfigured workflows



Complete automation thru Xcalibur and Discoverer Daemon

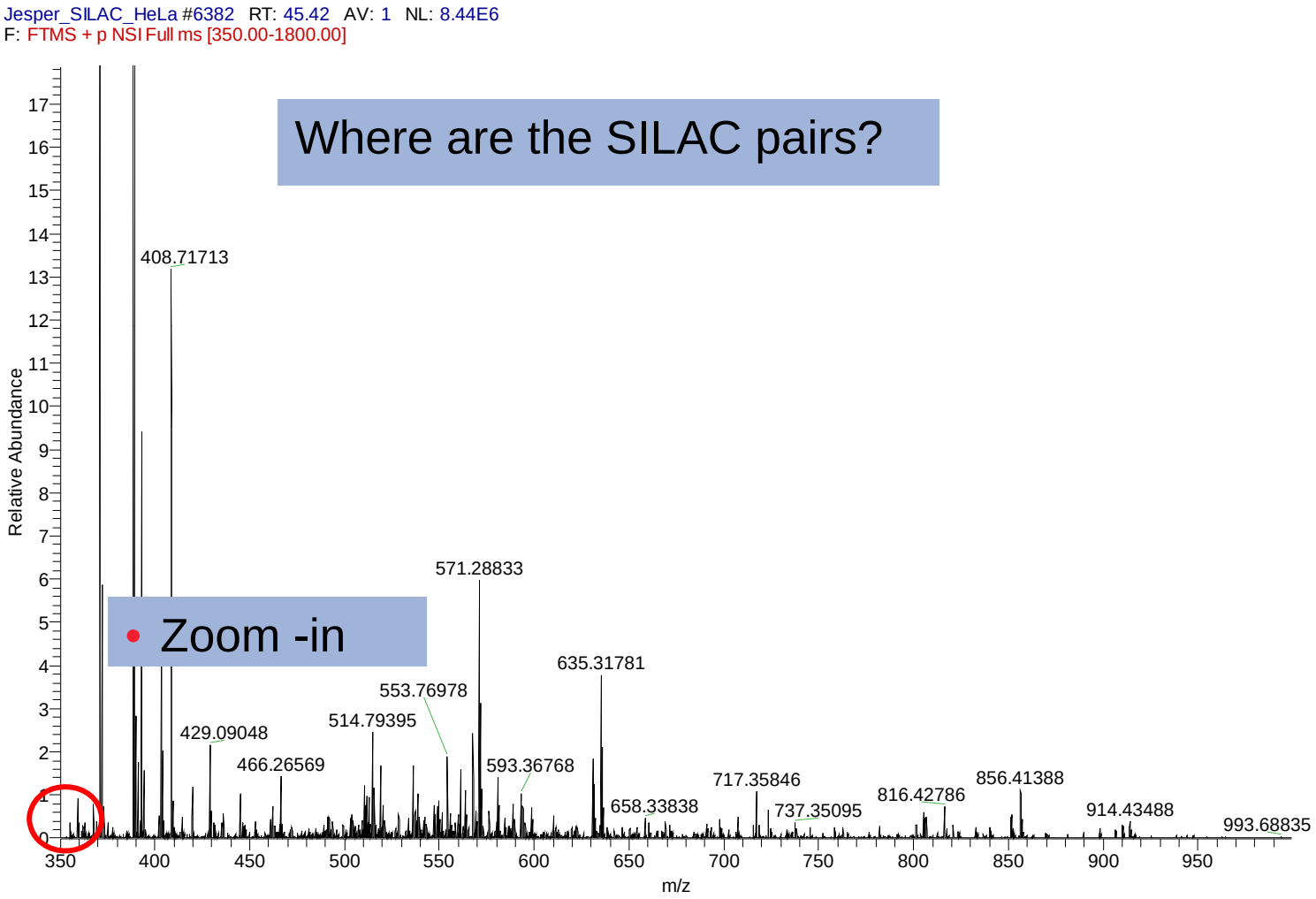
ThermoFisher SCIENTIFIC

Why are reliable and sensitive algorithms needed?



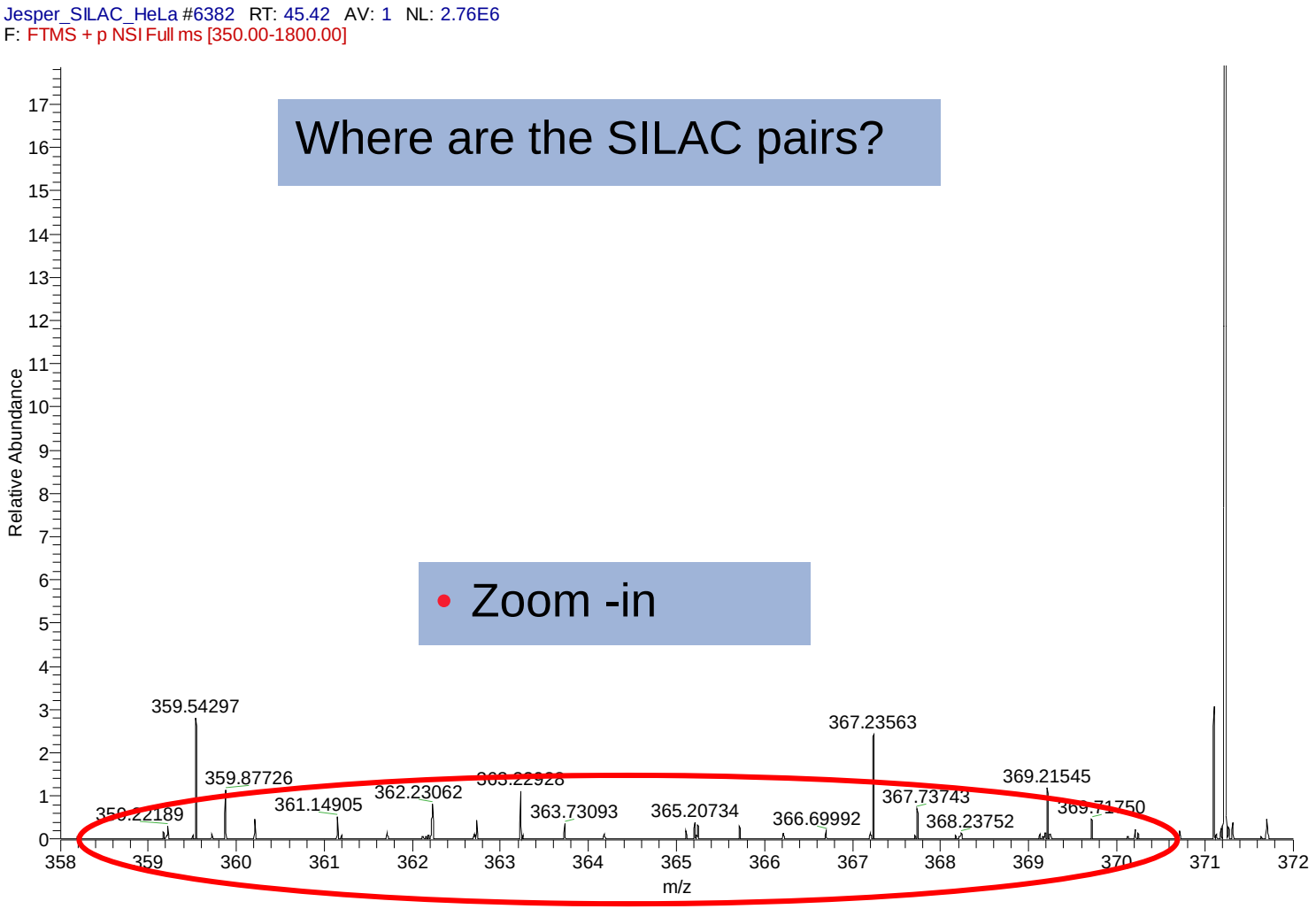
ThermoFisher SCIENTIFIC

Why are reliable and sensitive algorithms needed?



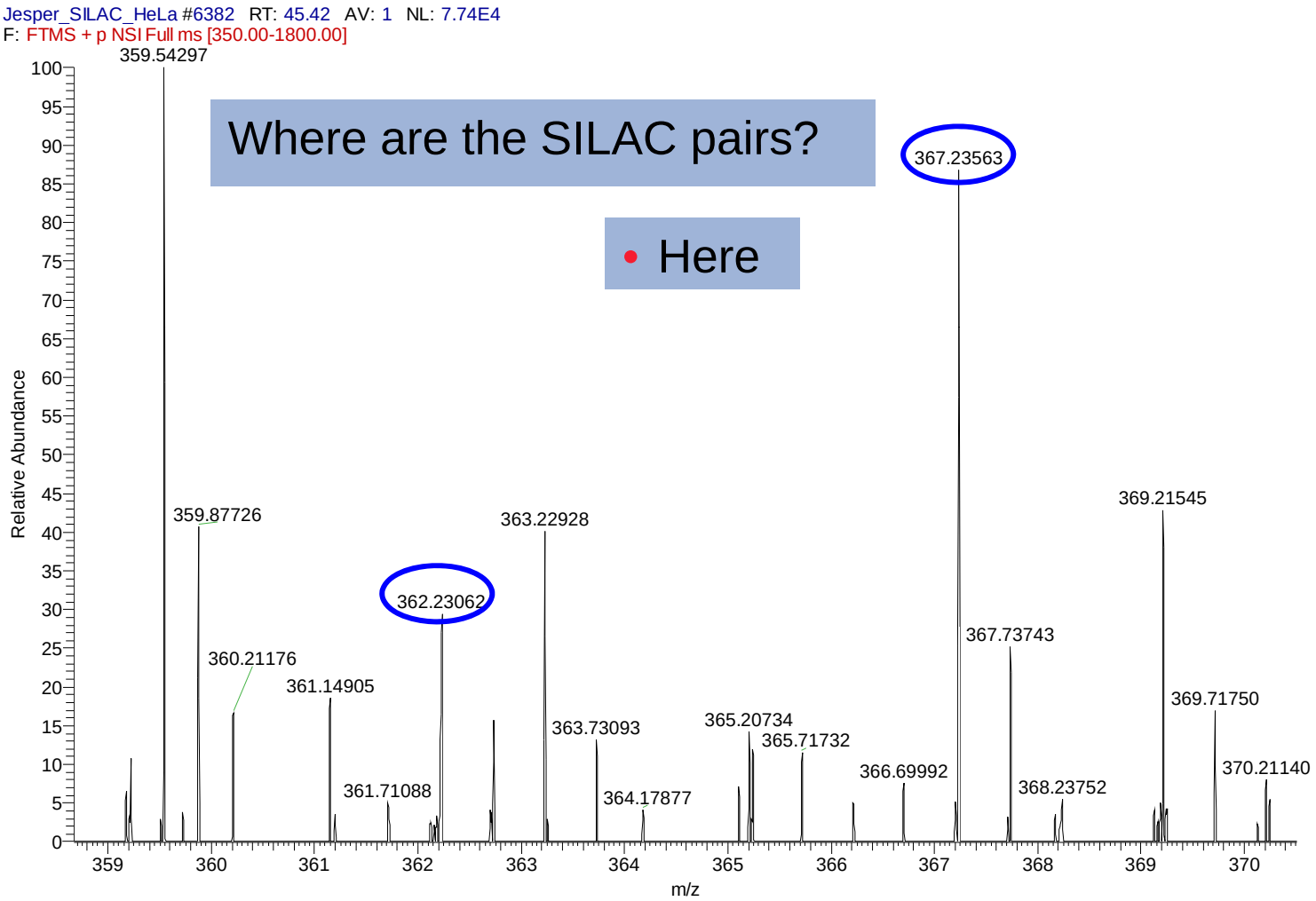
ThermoFisher SCIENTIFIC

Why are reliable and sensitive algorithms needed?



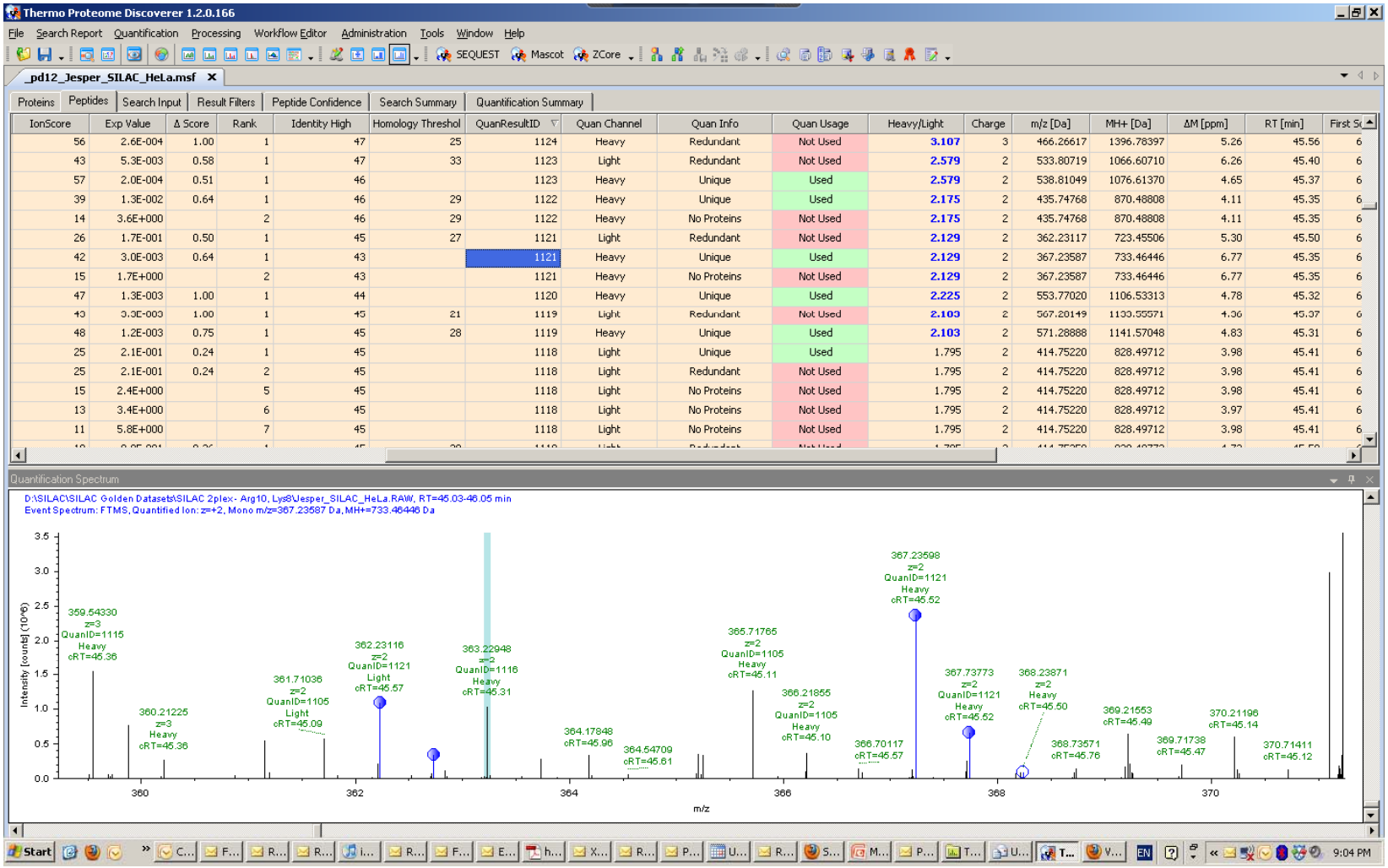
ThermoFisher SCIENTIFIC

Why are reliable and sensitive algorithms needed?



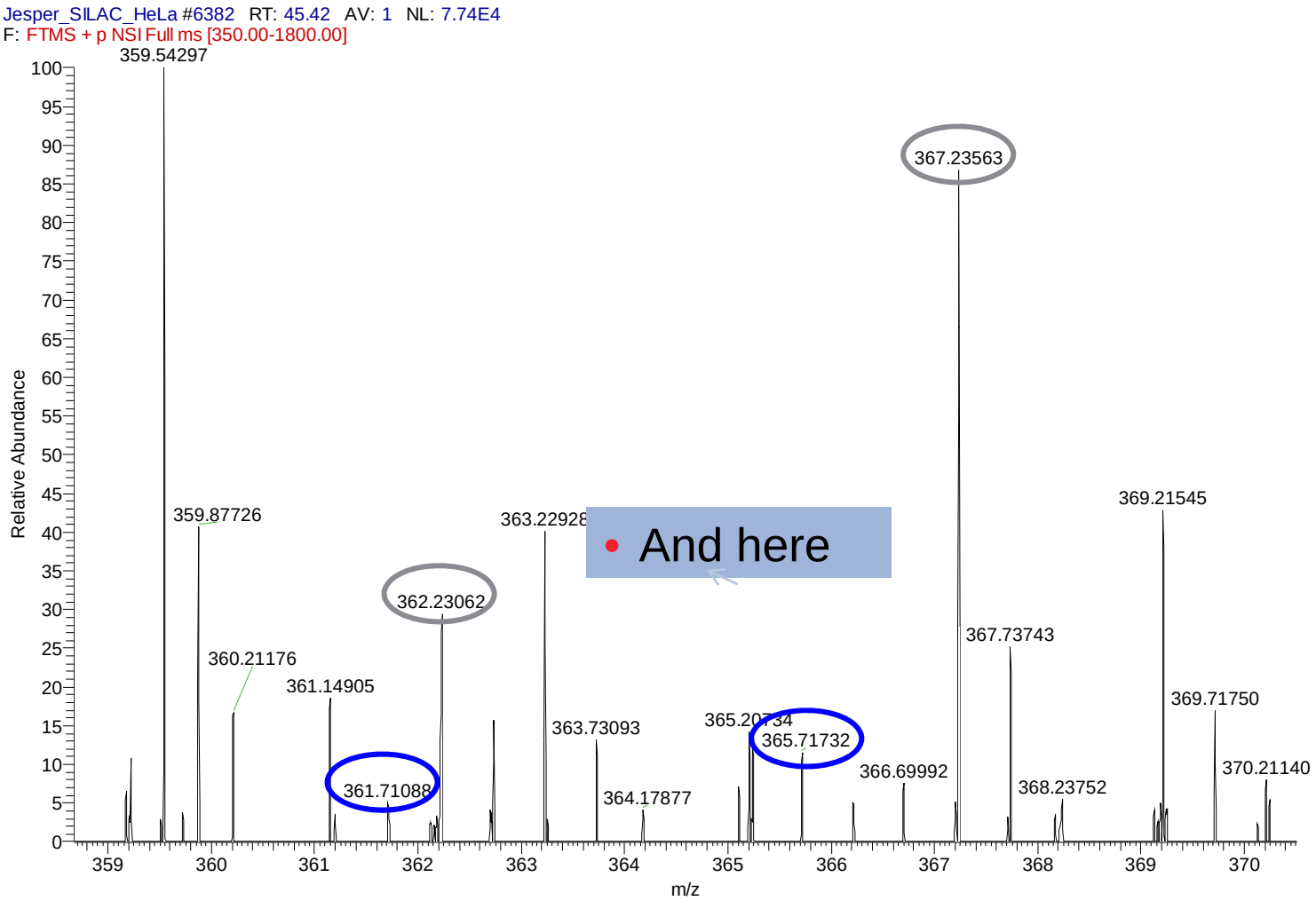
ThermoFisher SCIENTIFIC

Result in Proteome Discoverer



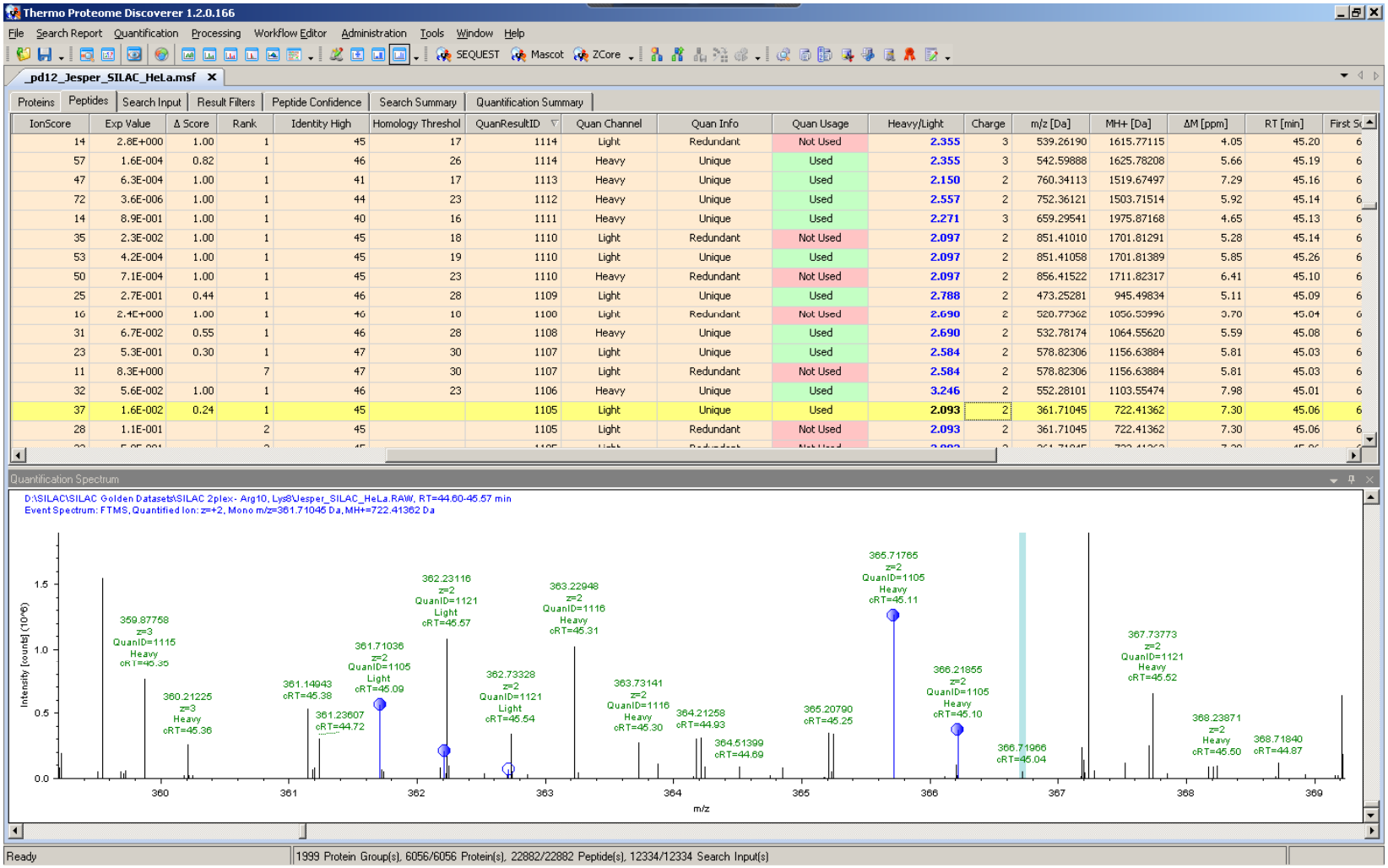
ThermoFisher SCIENTIFIC

Why are reliable and sensitive algorithms needed?



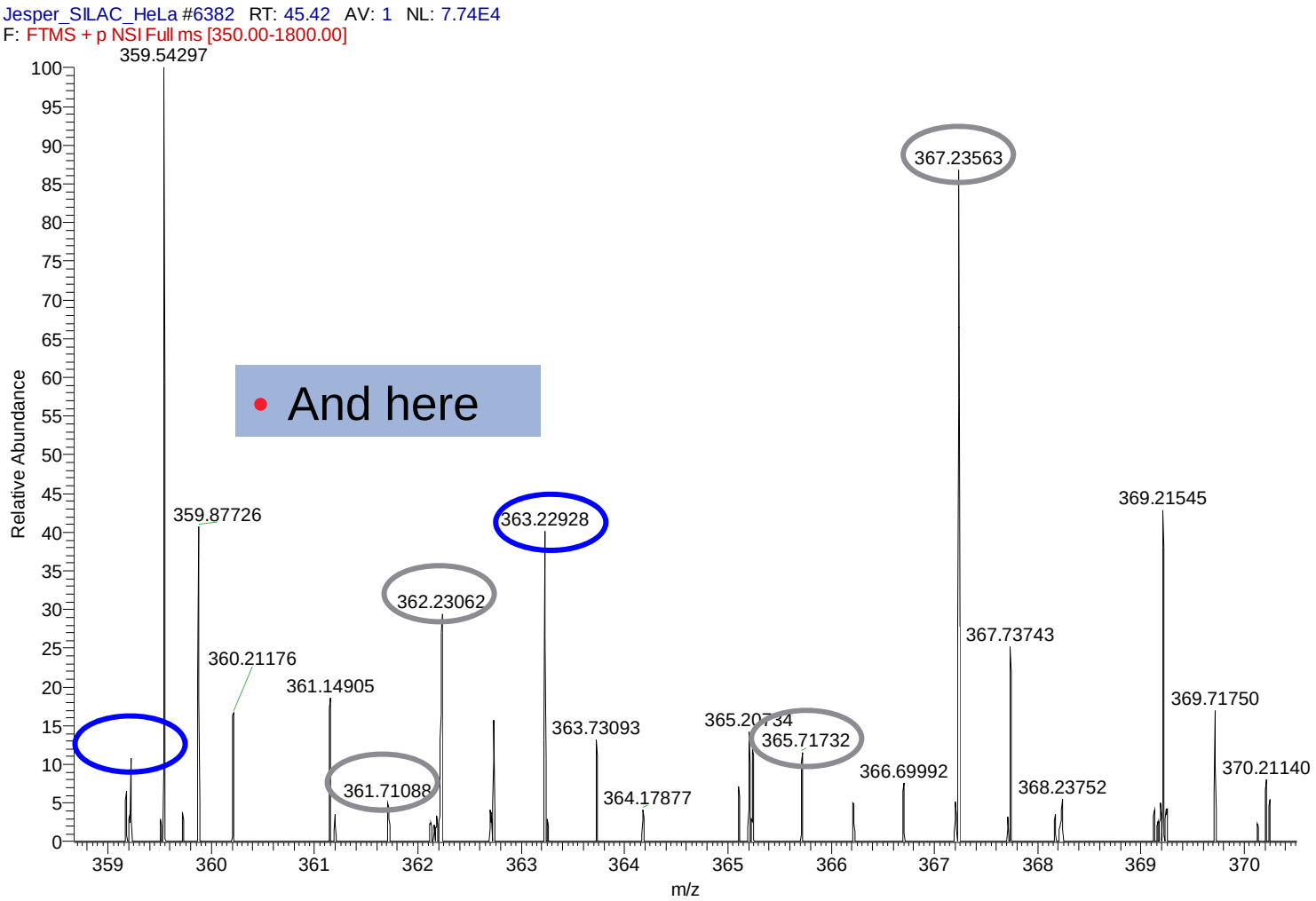
ThermoFisher SCIENTIFIC

Result in Proteome Discoverer



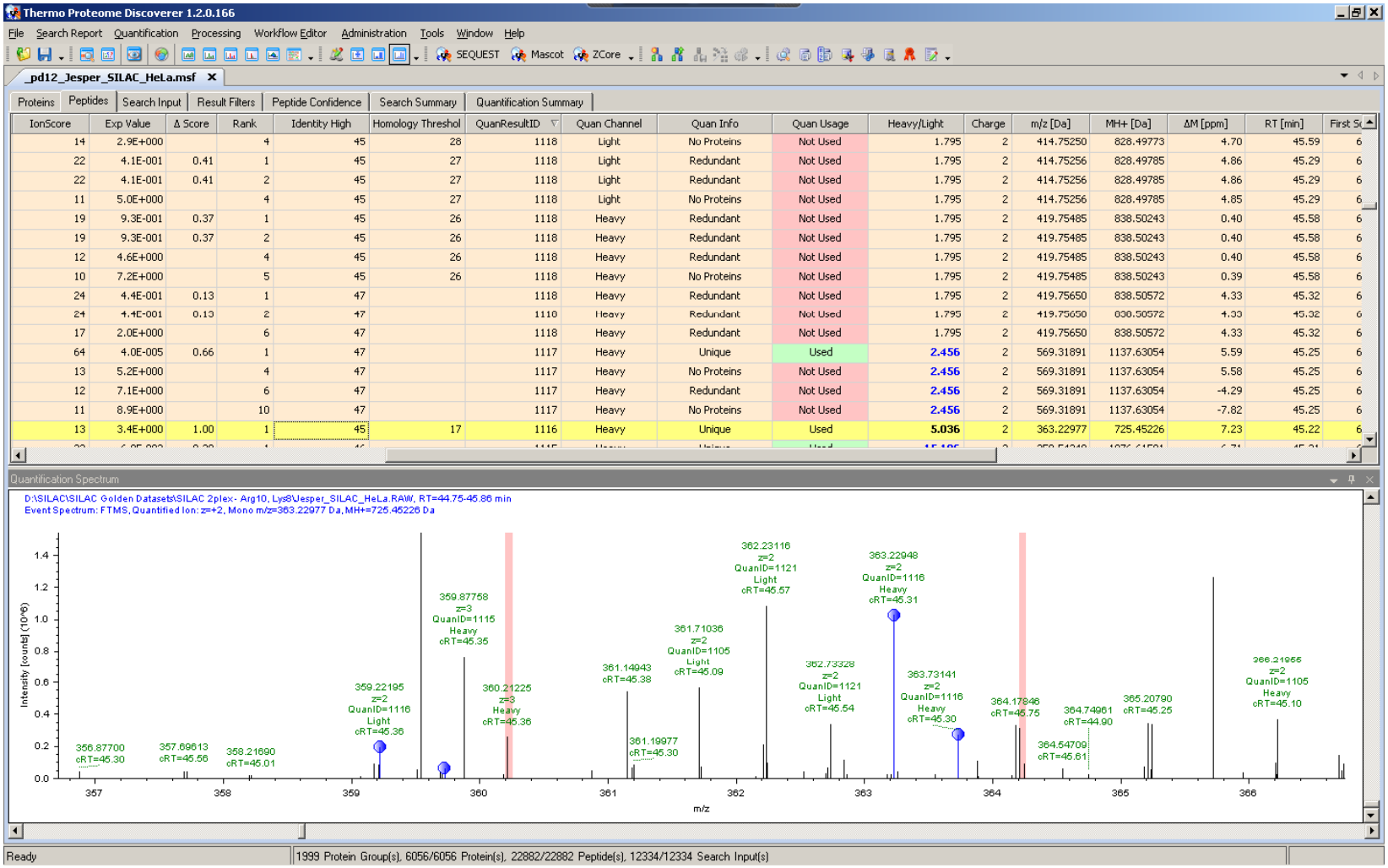
ThermoFisher SCIENTIFIC

Why are reliable and sensitive algorithms needed?



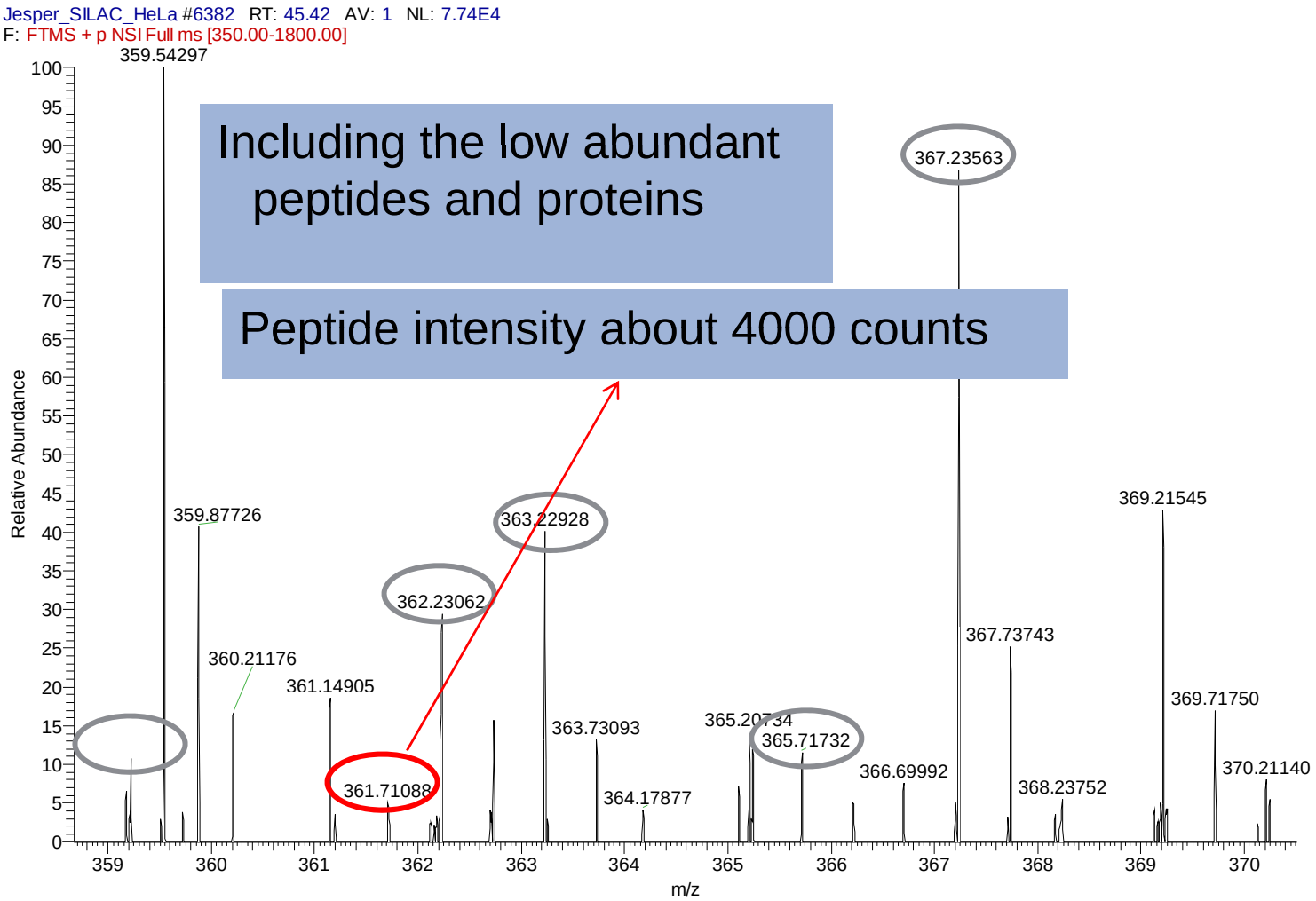
ThermoFisher SCIENTIFIC

Result in Proteome Discoverer



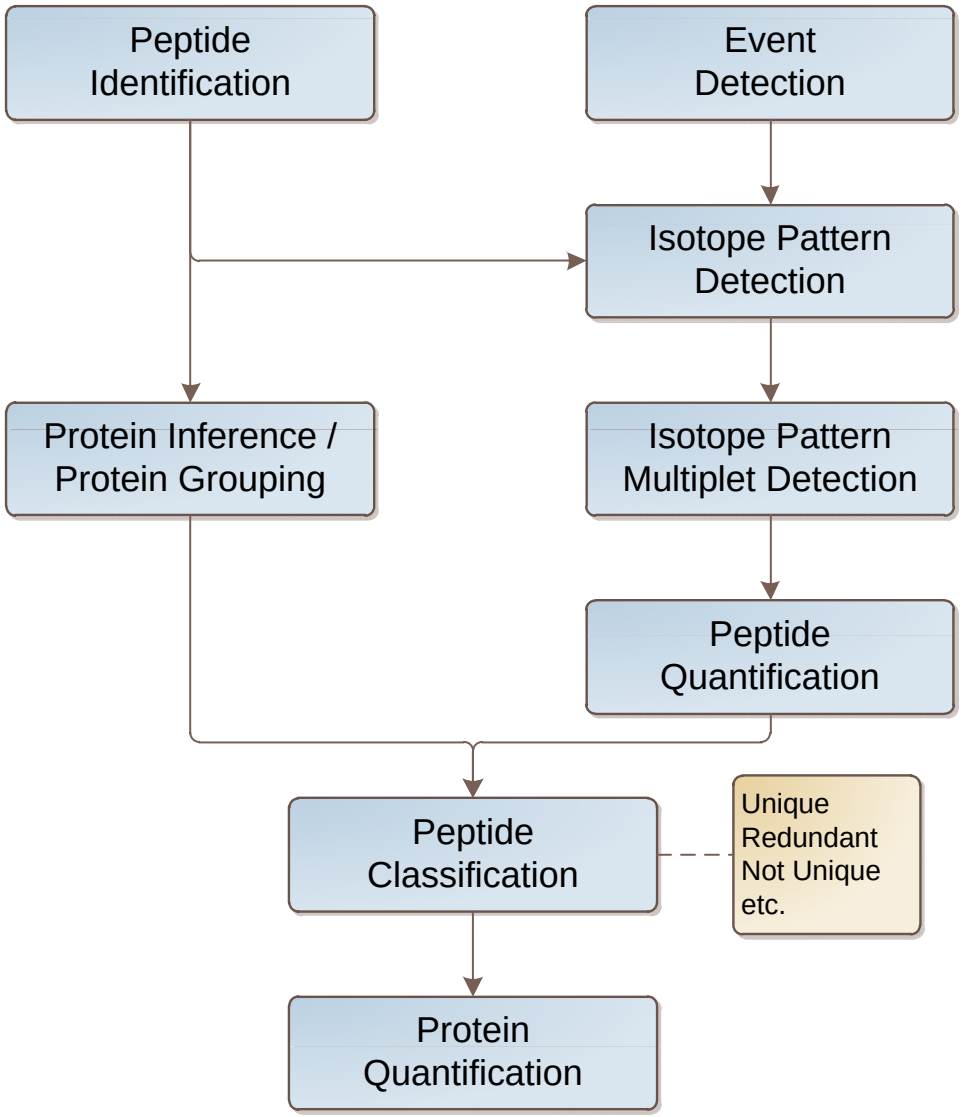
ThermoFisher SCIENTIFIC

Why are reliable and sensitive algorithms needed?

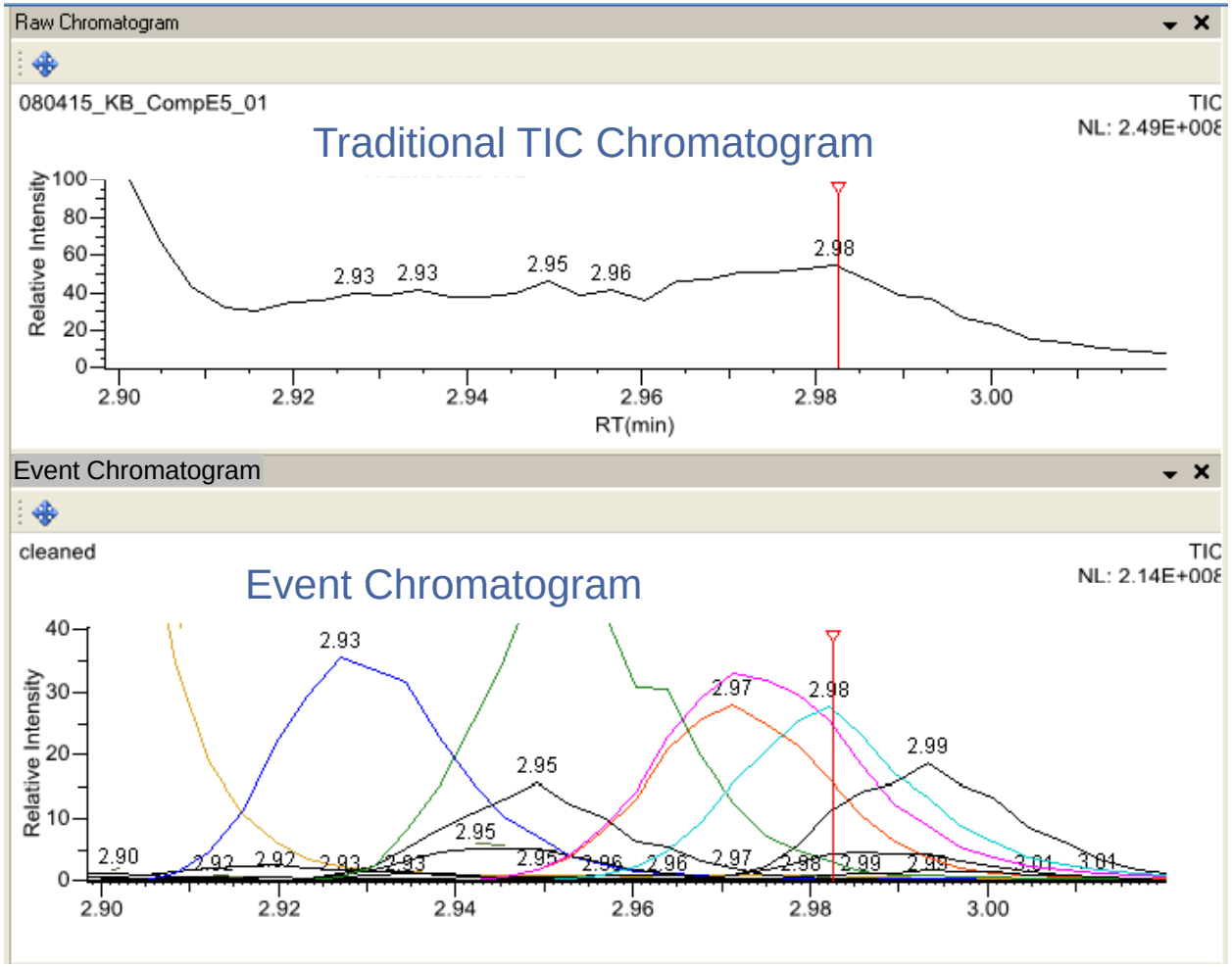


ThermoFisher SCIENTIFIC

Precursor Quan: How Does It Work?



Event Detection



Create XIC for every peak in every scan

Keep XICs with shape of eluting component

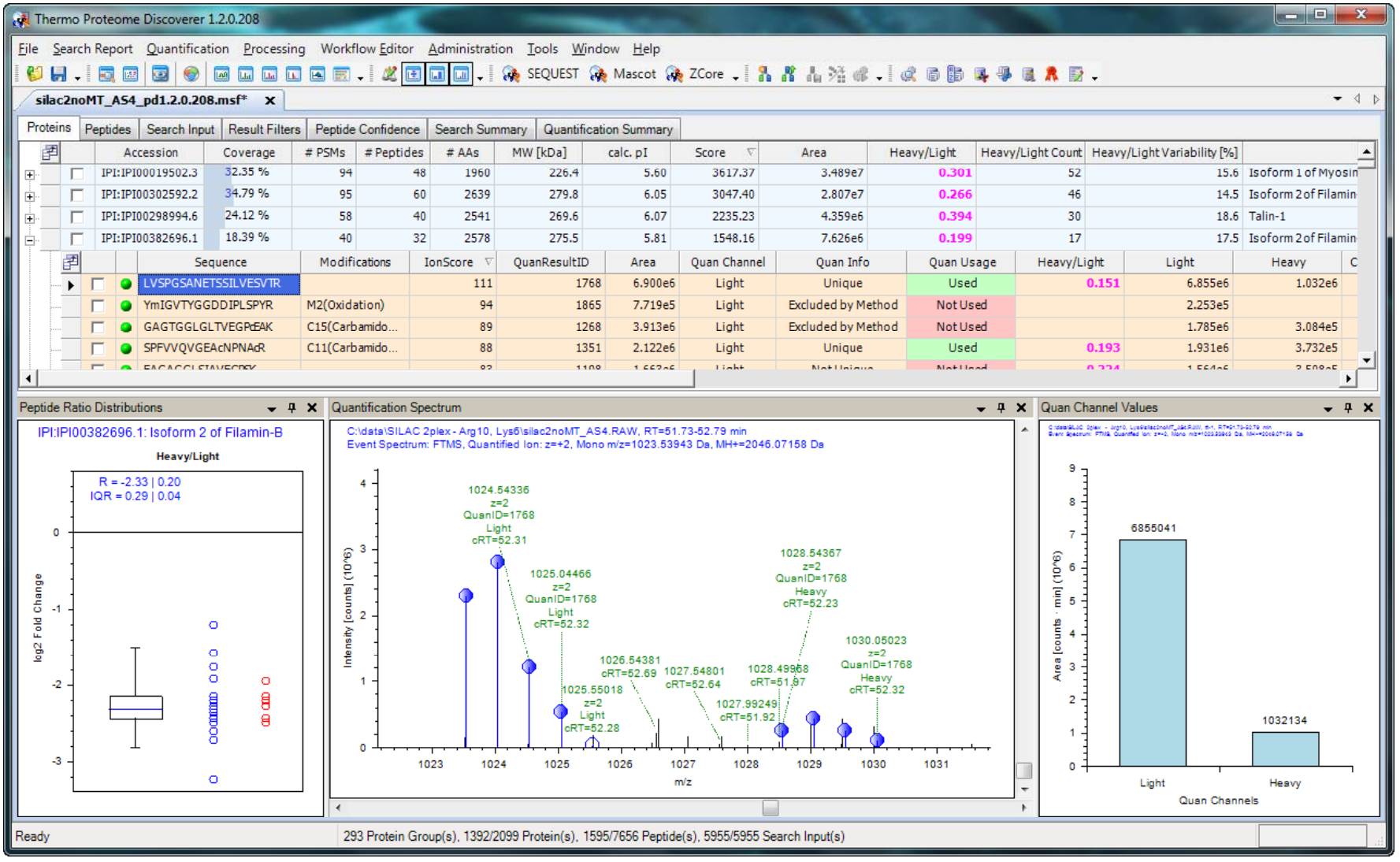
Peak-detection on every „good“ XIC

Create an „event“ for each „good“ XIC

- Mass
- Intensity
- Area
- RT
- Etc.

ThermoFisher SCIENTIFIC

Precursor Quan: Quan Result Display

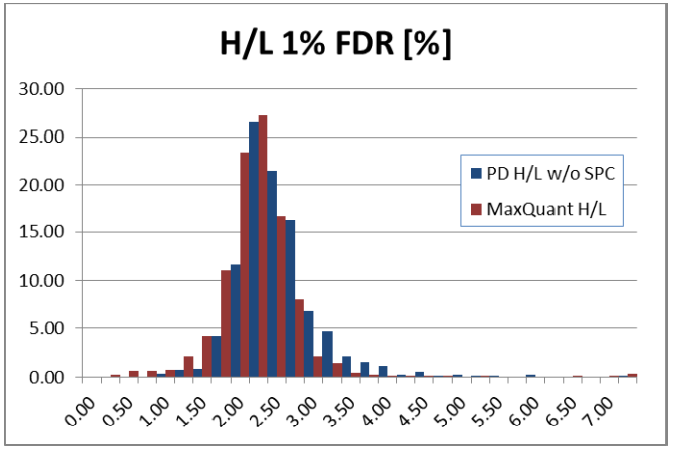


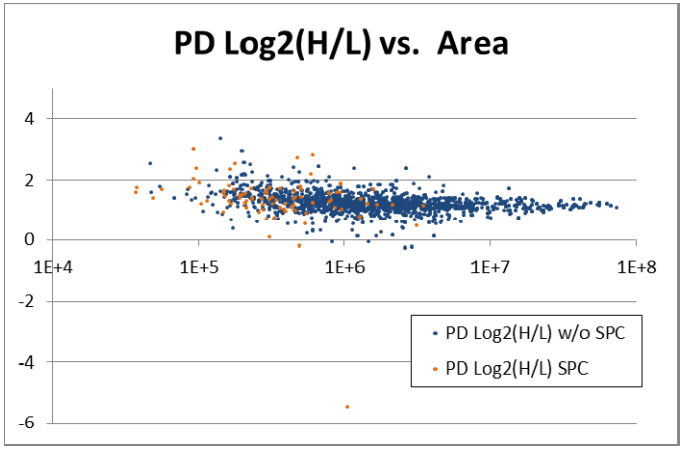
ThermoFisher SCIENTIFIC

Precursor Quan: Validation

HeLa Digest
(Arg10, Lys8)

Mixture
L : H = 1 : 3





Proteome Discoverer: Peptides at 1% FDR

	Count	%	Log2 based Median	Mean	σ*	68% Interval
Unique	1182	56.18	2.31	2.35	1.25	1.84 - 2.90
Redundant	681	32.37	2.27	2.28	1.19	1.91 - 2.70
Not Unique	50	2.38	2.23	2.16	1.19	1.87 - 2.65
Peptides with Single Peak Channels						
Unique	83	3.94	2.59	2.52	1.89	1.37 - 4.90
Redundant	17	0.81	2.93	3.10	1.34	2.19 - 3.93
Not Unique	0	0.00				
Indistinguishable Channels	3	0.14				
Inconsistently Labeled	7	0.33				
Excluded by Method	53	2.52				
No Quan Values	28	1.33				
Total	2104					
Quantified	2013	95.67				
Not Quantifiable	63	2.99				
No Quan Values	28	1.33				

Processing Time [min]

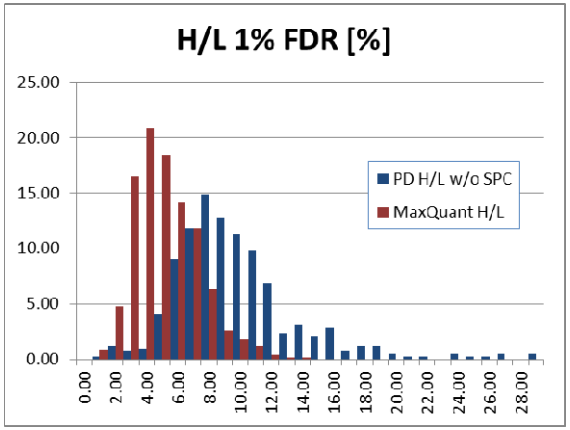
	Proteome Discoverer
Mascot Search	10
Event Detection	6
Quantification	2
Total	**18**

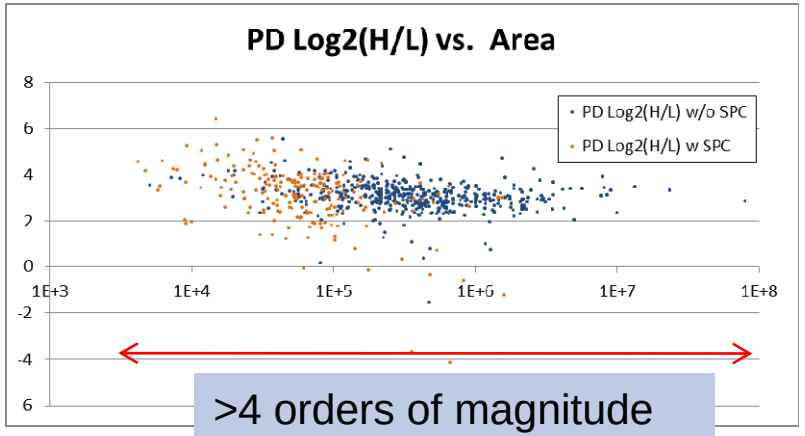
ThermoFisher SCIENTIFIC

Precursor Quan: Validation

HeLa Digest
(Arg10, Lys6)

Mixture
L : H = 1 : 10





>4 orders of magnitude

Proteome Discoverer: Peptides at 1% FDR

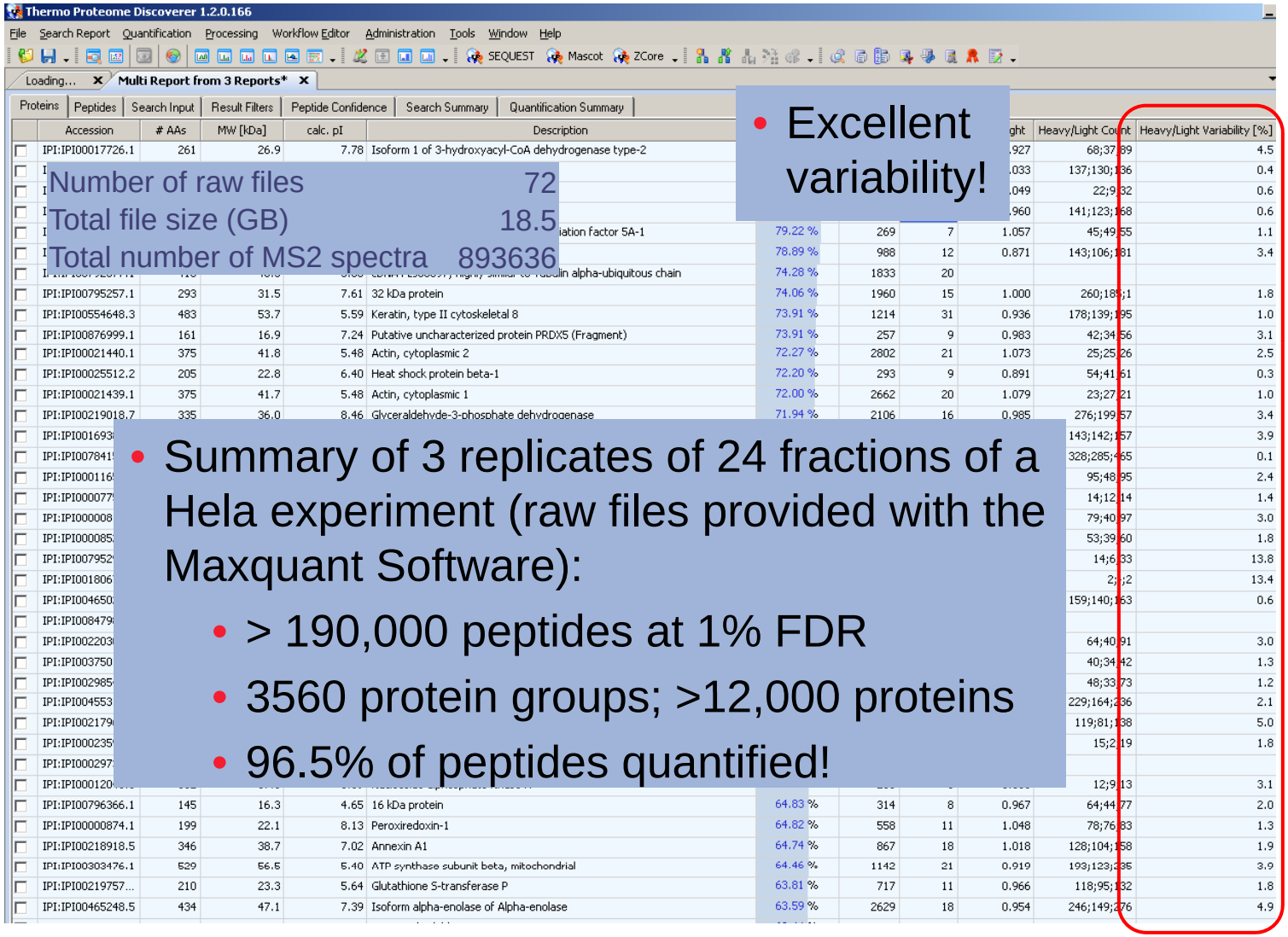
	Count	%	Log2 based Median	Mean	σ*	68% Interval
Unique	422	31.14	8.51	8.49	1.58	5.37 - 13.49
Redundant	133	9.82	8.44	8.72	1.43	5.91 - 12.03
Not Unique	46	3.39	8.32	8.15	1.29	6.44 - 10.76
Peptides with Single Peak Channels						
Unique	180	13.28	9.83	8.60	2.57	3.83 - 25.23
Redundant	8	0.59	9.75	11.99	1.81	5.39 - 17.64
Not Unique	15	1.11	10.93	7.98	2.62	
Indistinguishable Channels	2	0.15				
Inconsistently Labeled	4	0.30				
Excluded by Method	391	28.86				
No Quan Values	154	11.37				
Total	1355					
Quantified	804	59.34				
Not Quantifiable	397	29.30				
No Quan Values	154	11.37				

Processing Time [min]

	Proteome Discoverer
Mascot Search	8
Event Detection	1
Quantification	1
Total	**10**

ThermoFisher SCIENTIFIC

Robustness and quality of the algorithms



Accession	# AAs	MW [kDa]	calc. pI	Description					Heavy/Light Count	Heavy/Light Variability [%]
IPI:IPI00017726.1	261	26.9	7.78	Isoform 1 of 3-hydroxyacyl-CoA dehydrogenase type-2					68;37	4.5
IPI:IPI00795257.1	293	31.5	7.61	32 kDa protein	74.06 %	1960	15	1.000	260;18	1.8
IPI:IPI00554648.3	483	53.7	5.59	Keratin, type II cytoskeletal 8	73.91 %	1214	31	0.936	178;139;195	1.0
IPI:IPI00876999.1	161	16.9	7.24	Putative uncharacterized protein PRDX5 (Fragment)	73.91 %	257	9	0.983	42;34	3.1
IPI:IPI00021440.1	375	41.8	5.48	Actin, cytoplasmic 2	72.27 %	2802	21	1.073	25;25	2.5
IPI:IPI00025512.2	205	22.8	6.40	Heat shock protein beta-1	72.20 %	293	9	0.891	54;41	0.3
IPI:IPI00021439.1	375	41.7	5.48	Actin, cytoplasmic 1	72.00 %	2662	20	1.079	23;27	1.0
IPI:IPI00219018.7	335	36.0	8.46	Glyceraldehyde-3-phosphate dehydrogenase	71.94 %	2106	16	0.985	276;199	3.4
IPI:IPI00796366.1	145	16.3	4.65	16 kDa protein	64.83 %	314	8	0.967	64;44	2.0
IPI:IPI00000874.1	199	22.1	8.13	Peroxiredoxin-1	64.82 %	558	11	1.048	78;76	1.3
IPI:IPI00218918.5	346	38.7	7.02	Annexin A1	64.74 %	867	18	1.018	128;104;158	1.9
IPI:IPI00303476.1	529	56.5	5.40	ATP synthase subunit beta, mitochondrial	64.46 %	1142	21	0.919	193;123;336	3.9
IPI:IPI00219757...	210	23.3	5.64	Glutathione S-transferase P	63.81 %	717	11	0.966	118;95;132	1.8
IPI:IPI00465248.5	434	47.1	7.39	Isoform alpha-enolase of Alpha-enolase	63.59 %	2629	18	0.954	246;149;276	4.9

ThermoFisher SCIENTIFIC

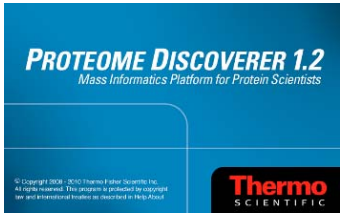
Complete SILAC Quantitation Workflow

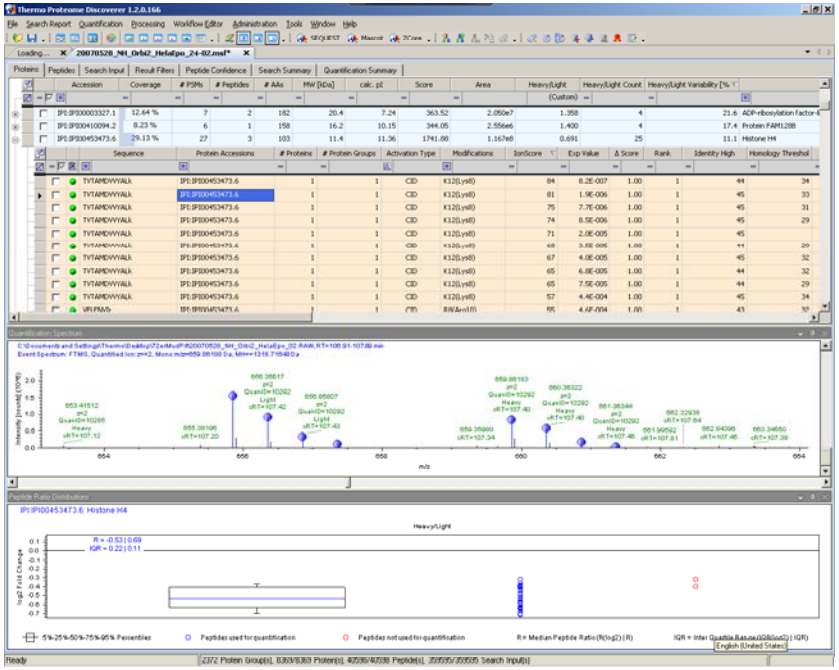
- **Pierce® SILAC Quantitation Kits**

  [500 ml media (2x), 50 ml dialyzed FBS (2x) and 50 mg of $^{13}C_6$ L-lysine (1x), L-lysine (1x), L-arginine (2x)]

  - RPMI 1640 kit, #89982
  - DMEM kit, #89983
  - Human Mesenchymal Stem Cell Kit #88200
  - Mouse Embryonic Stem Cell Kit #88206

- LTQ Orbitrap Velos
- Proteome Discoverer 1.2

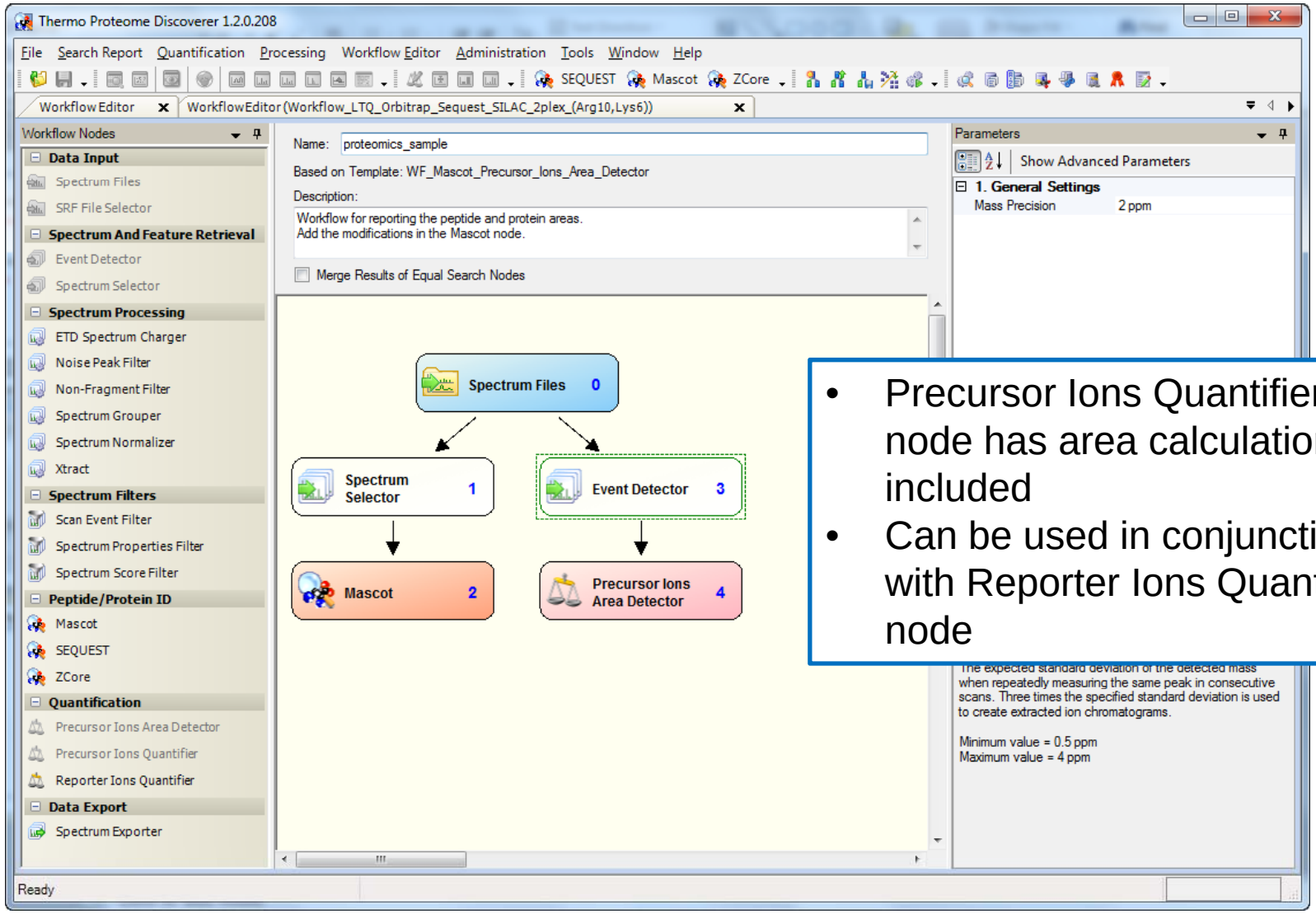




Comprehensive Quantitative Proteomics enabled by simultaneous MS and MS/MS

ThermoFisher SCIENTIFIC

Precursor Quan: Workflow for Unlabeled Peptides



- Precursor Ions Quantifier node has area calculation included
- Can be used in conjunction with Reporter Ions Quantifier node

ThermoFisher SCIENTIFIC

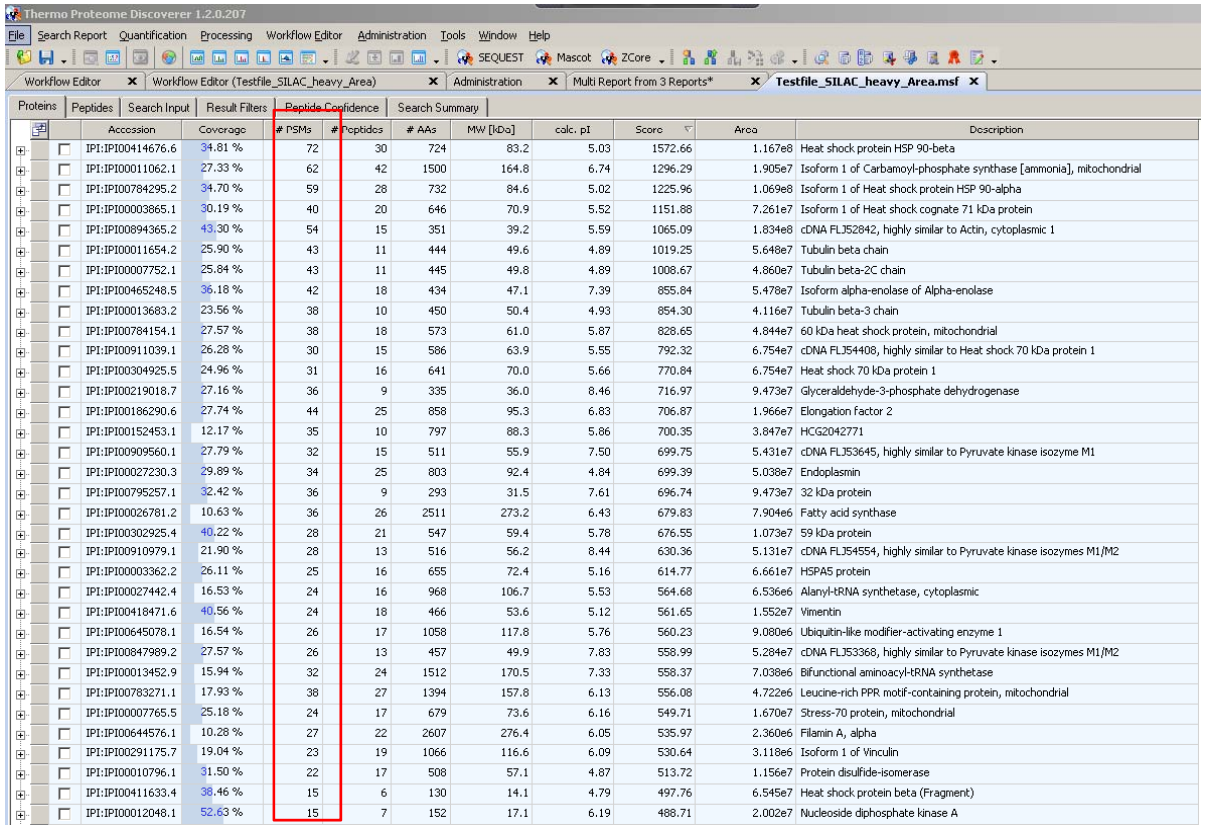
ThermoFisher SCIENTIFIC

Spectral Counting

Bernard Delanghe

Definition

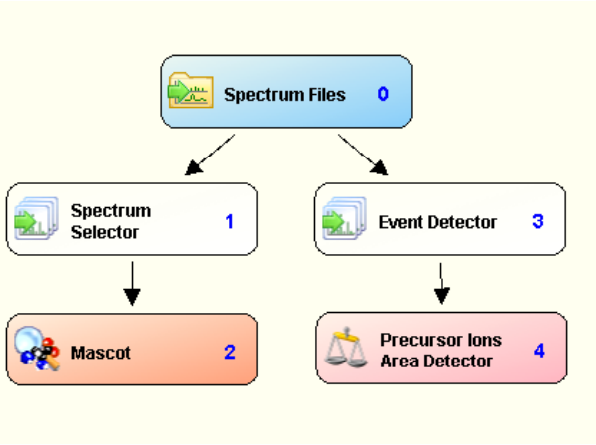
- The total number of identified peptide sequences (peptide spectrum matches) for the protein, including those redundantly identified.
- Standard feature in Proteome Discoverer



Accession	Coverage	# PSMs	# Peptides	# AAs	MW [kDa]	calc. pI	Score	Area	Description
IPI:IPI00414676.6	34.81 %	72	30	724	83.2	5.03	1572.66	1.167e8	Heat shock protein HSP 90-beta
IPI:IPI00011062.1	27.33 %	62	42	1500	164.8	6.74	1296.29	1.905e7	Isoform 1 of Carbamoyl-phosphate synthase [ammonia], mitochondrial
IPI:IPI00784295.2	34.70 %	59	28	732	84.6	5.02	1225.96	1.069e8	Isoform 1 of Heat shock protein HSP 90-alpha
IPI:IPI00003865.1	30.19 %	40	20	646	70.9	5.52	1151.88	7.261e7	Isoform 1 of Heat shock cognate 71 kDa protein
IPI:IPI00894365.2	43.30 %	54	15	351	39.2	5.59	1065.09	1.834e8	cDNA FLJ52842, highly similar to Actin, cytoplasmic 1
IPI:IPI00011654.2	25.90 %	43	11	444	49.6	4.89	1019.25	5.648e7	Tubulin beta chain
IPI:IPI00007752.1	25.84 %	43	11	445	49.8	4.89	1008.67	4.860e7	Tubulin beta-2C chain
IPI:IPI00465248.5	36.18 %	42	18	434	47.1	7.39	855.84	5.478e7	Isoform alpha-enolase of Alpha-enolase
IPI:IPI00013683.2	23.56 %	38	10	450	50.4	4.93	854.30	4.116e7	Tubulin beta-3 chain
IPI:IPI00784154.1	27.57 %	38	18	573	61.0	5.87	828.65	4.844e7	60 kDa heat shock protein, mitochondrial
IPI:IPI00911039.1	26.28 %	30	15	586	63.9	5.55	792.32	6.754e7	cDNA FLJ54408, highly similar to Heat shock 70 kDa protein 1
IPI:IPI00304925.5	24.96 %	31	16	641	70.0	5.66	770.84	6.754e7	Heat shock 70 kDa protein 1
IPI:IPI00219018.7	27.16 %	36	9	335	36.0	8.46	716.97	9.473e7	Glyceraldehyde-3-phosphate dehydrogenase
IPI:IPI00186290.6	27.74 %	44	25	858	95.3	6.83	706.87	1.966e7	Elongation factor 2
IPI:IPI00152453.1	12.17 %	35	10	797	88.3	5.86	700.35	3.847e7	HCG2042771
IPI:IPI00909560.1	27.79 %	32	15	511	55.9	7.50	699.75	5.431e7	cDNA FLJ53645, highly similar to Pyruvate kinase isozyme M1
IPI:IPI00027230.3	29.89 %	34	25	803	92.4	4.84	699.39	5.038e7	Endoplasmin
IPI:IPI00795257.1	32.42 %	36	9	293	31.5	7.61	696.74	9.473e7	32 kDa protein
IPI:IPI00026781.2	10.63 %	36	26	2511	273.2	6.43	679.83	7.904e6	Fatty acid synthase
IPI:IPI00302925.4	40.22 %	28	21	547	59.4	5.78	676.55	1.073e7	59 kDa protein
IPI:IPI00910979.1	21.90 %	28	13	516	56.2	8.44	630.36	5.131e7	cDNA FLJ54554, highly similar to Pyruvate kinase isozymes M1/M2
IPI:IPI00003362.2	26.11 %	25	16	655	72.4	5.16	614.77	6.661e7	HSPA5 protein
IPI:IPI00027442.4	16.53 %	24	16	968	106.7	5.53	564.68	6.536e6	Alanyl-tRNA synthetase, cytoplasmic
IPI:IPI00418471.6	40.56 %	24	18	466	53.6	5.12	561.65	1.552e7	Vimentin
IPI:IPI00645078.1	16.54 %	26	17	1058	117.8	5.76	560.23	9.080e6	Ubiquitin-like modifier-activating enzyme 1
IPI:IPI00847989.2	27.57 %	26	13	457	49.9	7.83	558.99	5.284e7	cDNA FLJ53368, highly similar to Pyruvate kinase isozymes M1/M2
IPI:IPI00013452.9	15.94 %	32	24	1512	170.5	7.33	558.37	7.038e6	Bifunctional aminoacyl-tRNA synthetase
IPI:IPI00783271.1	17.93 %	38	27	1394	157.8	6.13	556.08	4.722e6	Leucine-rich PPR motif-containing protein, mitochondrial
IPI:IPI00007765.5	25.18 %	24	17	679	73.6	6.16	549.71	1.670e7	Stress-70 protein, mitochondrial
IPI:IPI00644576.1	10.28 %	27	22	2607	276.4	6.05	535.97	2.360e6	Filamin A, alpha
IPI:IPI00291175.7	19.04 %	23	19	1066	116.6	6.09	530.64	3.118e6	Isoform 1 of Vinculin
IPI:IPI00010796.1	31.50 %	22	17	508	57.1	4.87	513.72	1.156e7	Protein disulfide-isomerase
IPI:IPI00411633.4	38.46 %	15	6	130	14.1	4.79	497.76	6.545e7	Heat shock protein beta (Fragment)
IPI:IPI00012048.1	52.63 %	15	7	152	17.1	6.19	488.71	2.002e7	Nucleoside diphosphate kinase A

Protein Area Calculation

- Alternative method in Proteome Discoverer: Protein area calculation
  - Average of the 3 highest peptide areas
  - Same algorithms as for Precursor Ion Quantification
  - Standard workflow



Molecular & Cellular Proteomics

Supplemental Material can be found at:
http://www.mcponline.org/cgi/content/full/M500230-MCP200/DC1

Research

Absolute Quantification of Proteins by LCMS^E

A VIRTUE OF PARALLEL MS ACQUISITION*[S]

Jeffrey C. Silva‡§, Marc V. Gorenstein‡, Guo-Zhong Li‡, Johannes P. C. Vissers¶, and Scott J. Geromanos‡

Relative quantification methods have dominated the quantitative proteomics field. There is a need, however, to conduct absolute quantification studies to accurately model and understand the complex molecular biology that results in proteome variability among biological samples. A new method of absolute quantification of proteins is described. This method is based on the discovery of an unexpected relationship between MS signal response and protein concentration: the average MS signal response for the three most intense tryptic peptides per mole of protein is constant within a coefficient of variation of less than ±10%. Given an internal standard, this relationship is used to calculate a universal signal response factor. The universal signal response factor (counts/mol) was shown to be the same for all proteins tested in this study. A controlled set of six exogenous proteins of varying concentrations was studied in the absence and presence of human serum. The absolute quantity of the standard proteins was determined with a relative error of less than ±15%. The average MS signal responses of the three most intense peptides from each protein were plotted against their calculated protein concentrations, and this plot resulted in a linear relationship with an R^2 value of 0.9939. The analyses were applied to determine the absolute concentration of 11 common serum proteins, and these concentrations were then compared with known values available in the literature. Additionally within an unfractionated *Escherichia coli* lysate, a subset of identified proteins known to exist as functional complexes was studied. The calculated absolute quantities were used to accurately determine their stoichiometry. ***Molecular & Cellular Proteomics 5:144–156, 2006.***

To date a majority of the quantitative proteomic analyses have been performed using stable isotope labeling strategies such as ICAT (3), iTRAQ™ (4), SILAC (stable isotope labeling by amino acids in cell culture) (5), and ^{18}O labeling (6, 7). These methodologies require complex, time-consuming sample preparation and can be relatively expensive.

Recently there have been numerous reports applying label-free methods to monitor the relative abundance of protein between different conditions (8–11). Relative quantification provides information regarding specific protein abundance changes between two conditions caused by an induced perturbation (environment-induced, drug-induced, and disease-induced). These studies require comparison of identical proteolytic peptides in each of the two experiments to accurately determine relative ratios of the particular protein(s) of interest. Relative abundance values for each peptide to a given protein can then be obtained to quantitatively characterize the differential expression of proteins between different sample states. Many of these methods are based on determining the ratios of the peak area of identical peptides between different conditions. One critical factor limiting the quantitative reproducibility of these methods includes the ability to efficiently cluster the detected peptides. This in turn relies on the accuracy of the mass measurement and the chromatographic reproducibility. Although relative quantification monitors changes in protein abundance between two conditions, it does not determine the absolute quantity of these proteins.

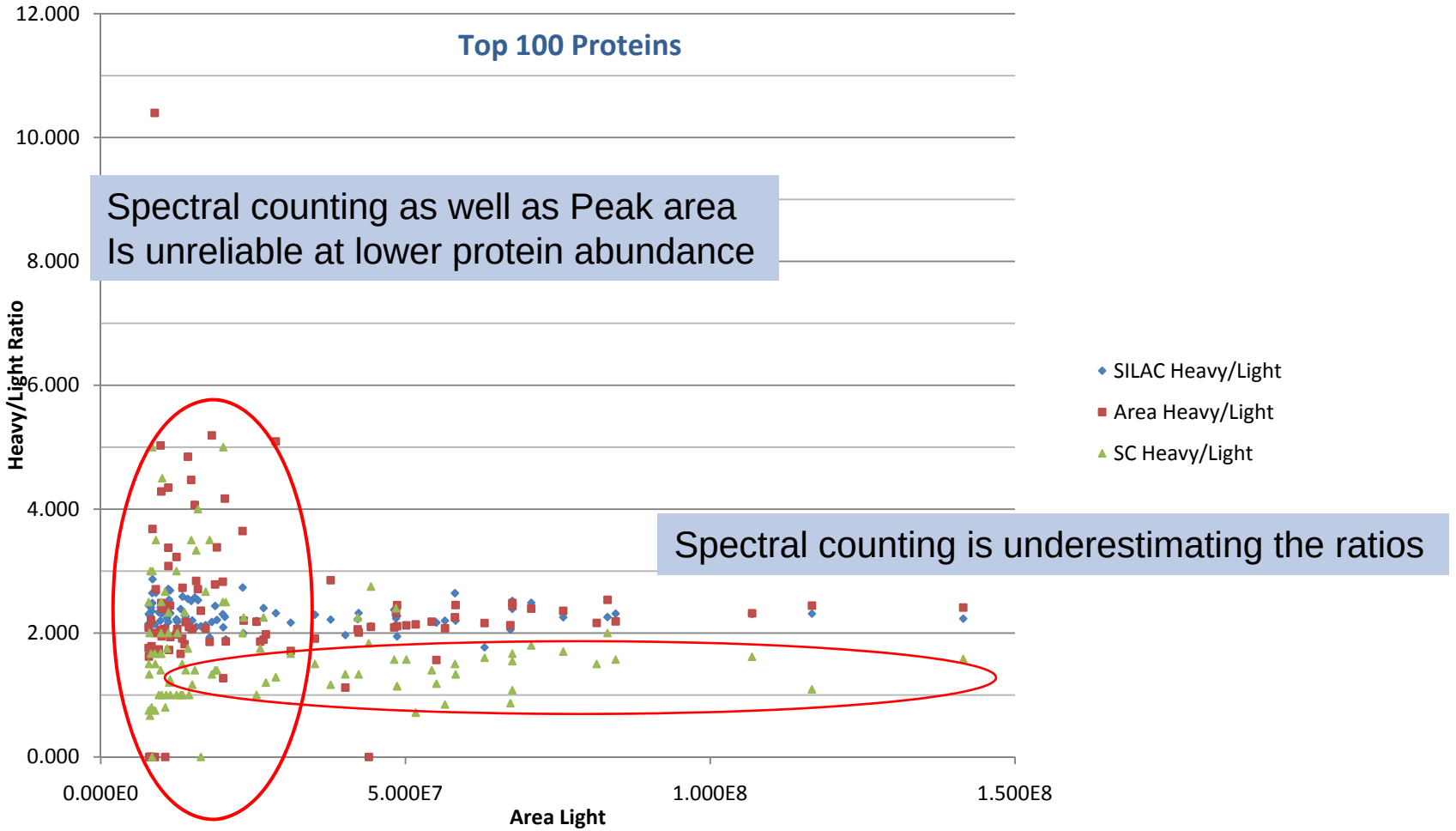
The ability to determine the absolute concentration of a protein (or proteins) present within a complex protein mixture

Comparison of different quantification methods

- LTQ Orbitrap Velos raw file, SILAC labeled (Lys 8, Arg 10) HELA cells
  - Fixed Heavy Light ratio 3/1
  - 686 proteins identified (peptides at 1% FDR)
- Quantification methods
  - SILAC: Heavy/Light ratio
  - Area: Protein Area Heavy labeled peptide\ Protein Area Light labeled peptide (area calculated as average of the 3 highest peptide areas)
  - Spectral Counting

	SILAC	Area	Spectral Counting
Median	2.3	2.1	1.4
Average	2.4	2.2	1.6
SD	0.7	1.6	1.2
% Quantified Proteins	82%	54%	51%

Distribution of Protein Heavy/Light ratios



ThermoFisher SCIENTIFIC

Conclusion

- Proteome Discoverer has fast, robust, sensitive and reliable algorithms for Precursor based quantification.
- The workflow can be completely automated.
- Protein Area calculation method is clearly better than the spectral counting method.
- Thermo Fisher Scientific offers a complete workflow: from reagents to results.