APPLICATION NOTE

Improved Sample Preparation Workflows for Protein Discovery and Quantification in Pathway Analysis

Xiaoyue Jiang[1], Ryan Bomgarden[2], Sergei Snovida[2], Julian Saba[1], Devin Drew[1], John Rogers[2], Rosa Viner[1], and Andreas Huhmer[1]
[1]Thermo Fisher Scientific, San Jose, CA; [2]Thermo Fisher Scientific, Rockford, IL

Key Words
TMT, identification, quantification, peptide concentration assay, reproducibility, high-PH reversed-phase fractionation, Proteome Discoverer 2.1

Goal
To develop a sensitive and reproducible sample preparation workflow for proteome discovery and quantification. The sample loss in the preparation is also discussed.

Introduction
New advances in mass spectrometry (MS) enable comprehensive characterization and accurate quantification of complete proteomes, facilitating the classification of the protein expressions and regulations in signaling pathways. However, careful consideration must be undertaken to ensure changes observed in the biological samples arise from the biology rather than the analytical techniques employed in the analysis. For example, in a liquid chromatography coupled to mass spectrometry (LC-MS) approach, the variations observed in peptide/protein identifications and quantification measurements among replicates can come from both the sample preparation and the analysis itself. In recent years, advances in instrumentation have improved LC reproducibility and MS speed/sensitivity, greatly reducing the variations from the instrument platform. This leaves sample preparation as the main source of variation preventing reproducible protein identification and quantification.

Isobaric mass tagging (e.g., Tandem Mass Tag™ (TMT™) reagents[1]) has become a common technique for relative quantification of proteins.[2–5] TMT-based multiplexed relative quantification has been shown to have lower experimental variance and fewer missing quantitative values among samples compared to label-free approaches.[6] Because the labeling steps are straightforward, this has been widely and successfully applied to the deep quantification of complex proteomes. However, due to the high dynamic range of the protein concentration in some organism proteomes, low abundant proteins may only be identified by one or a few unique peptides, which typically have low signal-to-noise ratios, making quantification difficult. Precise quantification can be hindered from differences in sample handling, which results in higher variation among replicates.

To improve the quantification of low abundant proteins, we optimized a complete quantitative bottom-up workflow by incorporating an offline high pH reversed-phase peptide fractionation step and a novel peptide concentration assay. High pH reversed-phase fractionation has been shown to be an excellent method to improve the protein coverage as it is a separation method that is highly orthogonal to low pH separations used for LC-MS.[7] As the amount of peptide is different in the various fractions, a quantitative colorimetric peptide assay was used to determine the peptide concentration to normalize the amount loaded on column for LC-MS analysis, thereby minimizing variability that might have occurred during sample preparation. A peptide assay was also used to normalize peptide concentrations prior to TMT-based isobaric labeling and sample multiplexing. Overall, this modified workflow significantly improves the sensitivity and reproducibility of deep proteome profiling.

Methods

Sample Preparation

A549 human cells were serum starved overnight and stimulated for 15 minutes with insulin or IGF-1. Cells of control, insulin-treated, and IGF-1-treated conditions were lysed, and protein concentrations were determined using the Thermo Scientific™ Pierce™ BCA Protein Assay Kit. Lysed cells were further digested using the Thermo Scientific™ Pierce™ Mass Spec Sample Prep Kit for Cultured Cells in triplicate for each treatment, followed by peptide concentration determination using the Thermo Scientific™ Pierce™ Quantitative Colorimetric Peptide Assay.[8] Specifically, the colorimetric reaction product formed by the modified BCA reagents exhibits a strong absorbance at 480 nm and can be used to measure peptide concentrations.

Concentrations were normalized based on peptide assay before labeling samples with Thermo Scientific™ TMTsixplex™ labeling reagents. Labeled samples were fractionated using the Thermo Scientific™ Pierce™ High pH Reversed-Phase Peptide Fractionation Kit.[9] Specifically, labeled digests were loaded onto an equilibrated, high-pH, reversed-phase fractionation spin column. Peptides were bound to the hydrophobic resin under aqueous conditions and desalted by washing the column with water by low-speed centrifugation. A step gradient of increasing acetonitrile concentrations in a volatile high pH elution solution was then applied to the columns to elute bound peptides into eight different fractions collected by centrifugation. The peptide concentrations were measured using the colorimetric peptide assay before and after the fractionation to evaluate the sample loss. Labeled samples were further enriched using Thermo Scientific™ Pierce™ Fe-NTA Phosphopeptide Enrichment Kit for pathway analysis (Figure 1).

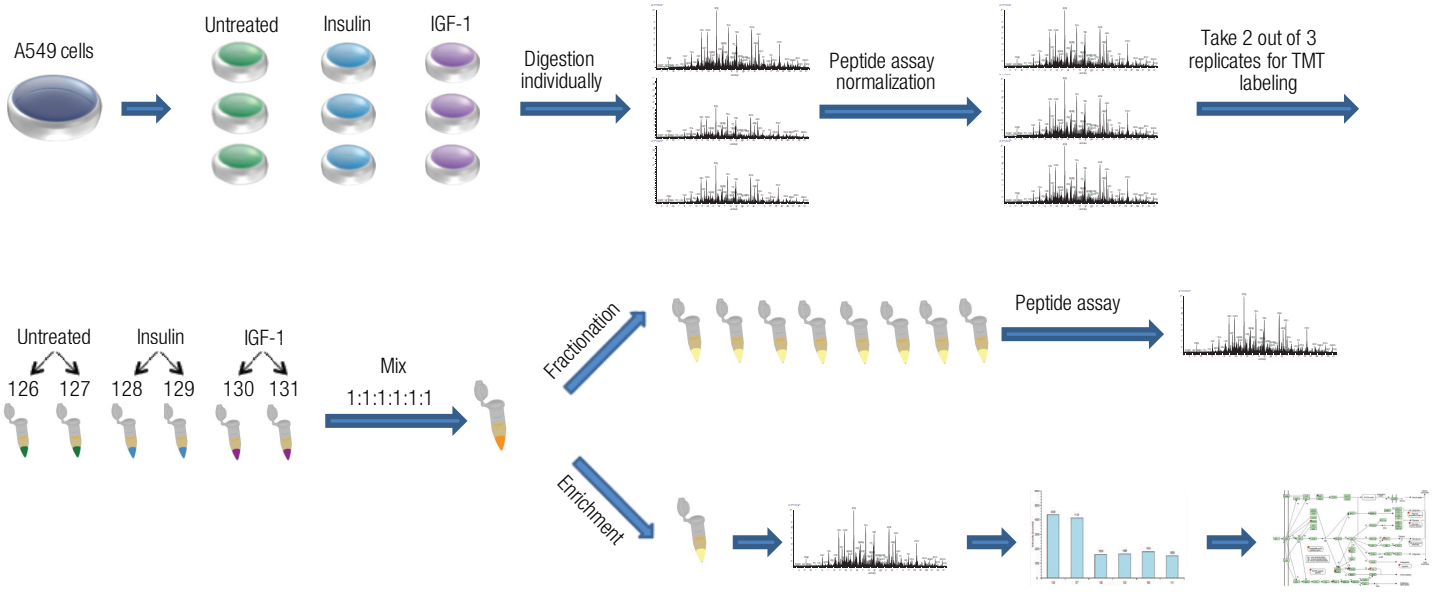

Figure 1. Experimental workflow.

Liquid Chromatography and Mass Spectrometry

A Thermo Scientific™ EASY-nLC™1000 UHPLC system and Thermo Scientific™ EASY-Spray™ source with 50 cm Thermo Scientific™ EASY-Spray™ LC Column was used to separate peptides with 30% acetonitrile gradient over 120 min, at a flow rate of 300 nL/min. The A549 cell lysate of control, insulin-treated, and IGF-1 treated conditions (100 ng) and TMT-labeled lysate fractions (500 ng) were analyzed on a Thermo Scientific™ Orbitrap Fusion™ Tribrid™ mass spectrometer. Phosphopeptide enriched TMT-labeled sample (1 µg) was analyzed on a Thermo Scientific™ Q Exactive™ Plus mass spectrometer. LC and MS settings are shown in Table 1.

Data Analysis

The LC-MS data for protein identification were analyzed using Thermo Scientific™ Proteome Discoverer™ software v.2.1 with the SEQUEST® HT search engine. Data were searched against a UniProt® human database with a 1% FDR criteria using Percolator®.

Table 1. LC and MS settings.

	Unlabeled Sample on Orbitrap Fusion MS	TMT6-Labeled Fractions on Orbitrap Fusion MS	TMT6-Labeled Phosphopeptides on Q Exactive Plus MS
LC Gradient	5–25% B in 90 min 25–35% B in 10 min	10–25% in 85 min 25–60% in 20 min	3–22% in 165 min 22–35% in 20 min
Full MS	Orbitrap	Orbitrap	Orbitrap
Resolution	120K	120K	70K
Target Value	4e5	4e5	3e6
Max Injection Time	50 ms	50 ms	50 ms
Top N	Speed 3 s	Speed 3 s	15
MS^2	Ion trap HCD	Ion trap CID	OT HCD
Isolation Mode	Quadrupole	Quadrupole	Quadrupole
Isolation Width	1.6	1.2	0.7
NCE	28	35	32
Resolution	Rapid	Turbo	35K
Target Value	1e4	1e4	1e5
Max Injection Time	70 ms	70 ms	120 ms
First Fixed Mass	110	100	100
$SPS\ MS^3$		OT HCD	
Isolation Width		2	
NCE		65	
Resolution		60K	
Target Value		1e5	
Max Injection Time		120 ms	
Search Parameters	SequestHT	SequestHT	SequestHT
Precursor Tolerance	10 ppm	10 ppm	10 ppm
Fragment Tolerance	0.6 Da	0.6 Da	0.02 Da
Static	Carbamidomethyl(C)	Carbamidomethyl(C) TMT6 (K, N term)	Carbamidomethyl(C) TMT6 (K, N term)
Dynamic	Oxidation (M)	Oxidation (M) Phospho (S,T,Y)	Oxidation (M) Phospho (S,T,Y)

Results

Peptide Quantification Before LC-MS Improves Data Reproducibility

In a typical bottom-up proteomics experiment, protein concentrations are measured prior to digestion. Unfortunately, after these measurements proteins undergo multiple sample preparation steps such as reduction, alkylation, digestion, and C18 cleanup, resulting in changes to the final peptide concentration that is loaded on to the column. To confirm this, we analyzed triplicate samples (100 ng) using column loads estimated from the original BCA protein assay concentration. We observed up to 30% variation in peptide identification and 15% variation in protein identifications among the replicate LC-MS runs (Figure 2A). These variations in peptide identification highlight the need to monitor protein and peptide concentration before and after each sample preparation step.

Using the new quantitative colorimetric peptide assay, we measured the concentration of the final digested samples used in LC-MS analysis. Our measurements revealed up to 75% difference in peptide concentration among the triplicate replicates. In order to minimize the variation we observed, we normalized the injection amount based on the peptide assay results. After normalization, less than 2% variability in peptide and protein identifications were observed among replicates (Figure 2B). These results demonstrate the need for accurate measurement of peptide concentration before LC-MS analysis.

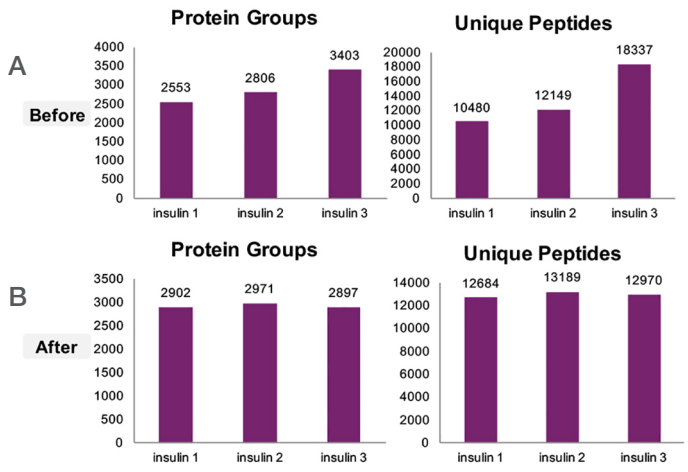

Figure 2. The number of unique protein groups and peptides identified when injecting 100 ng of insulin-treated triplicates on an Orbitrap Fusion MS before and after normalization.

Peptide Assay Optimizes the Sample Loading on Mass Spectrometer

In our next set of experiments, we focused on identifying the stages in sample preparation where the sample loss was occurring. The ability to accurately monitor the peptide concentration using the colorimetric peptide assay facilitated the process. The initial focus was on the digestion step. We compared the results of the BCA protein assay before and the colorimetric peptide assay after tryptic digestion. The sample loss was observed to be less than 10% during the digestion procedure.

Although fractionation increases the dynamic range of peptides and improves protein sequence coverage, this step presents another case for sample loss. Previously, the injected amount for each fraction was based on a theoretical average concentration per fraction, without adjusting for sample loss or the actual peptide concentration among different fractions. While the theoretical concentration for each fraction is 350 ng/µL, the actual loading concentration varied from 117 ng/µL to 561 ng/µL across the eight fractions (Figure 3). One interesting observation using the colorimetric assay was that the measurement for the first fraction was much higher than expected result, suggesting a potential interference from excess, unreacted TMT reagent. To eliminate this interference, we included additional cleanup step using 5% of acetonitrile before fractionation. Overall, we observed the high pH reversed-phase fractionation procedure recovered approximately 70% of the loaded peptides (Figure 3), suggesting that sample cleanup is the main source of sample loss.

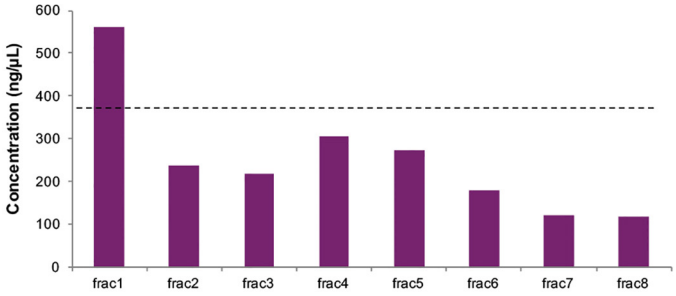

Figure 3. The peptide concentration distribution across eight fractions. The dotted line is the theoretical peptide concentration, assuming the peptide is equally eluted into each fraction. The overall sample recovery is ~70%.

Taking account of the measured peptide concentration in each fraction from the colorimetric peptide assay, we adjusted injection volume and re-ran our samples. Overall, we observed increased peptide identification for seven out of the eight fractions, with the last two fractions showing the largest increase (Figure 4). In total, 38,133 peptide groups and 5,059 protein groups were identified relative to 36,207 peptide groups and 4,839 protein groups without peptide concentration correction.

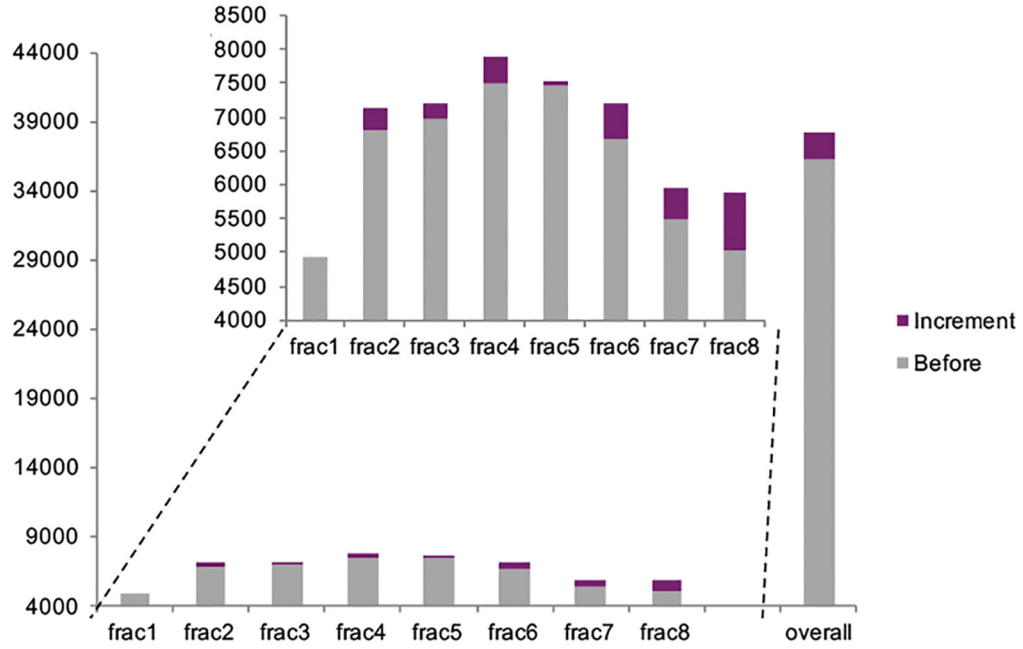

Figure 4. (A) Numbers of peptides identified when injecting 500 µg each fraction and overall combined results. (B) Magnification of (a) for fractions. Original identification numbers are shown in grey. After normalization, the identification increments are highlighted in purple.

TMT6plex Quantification and Pathway Analysis of Enriched Sample

Even though TMT6 labeling enabled quantification of > 99% of identified peptides and proteins, less than 1% of peptides are identified as phosphopeptides. Therefore, we used a FE-NTA phosphopeptide enrichment kit to enrich TMT6plex-labeled samples. We identified over 2,570 protein groups, of which 2,277 were phosphoproteins, resulting in an 88% enrichment rate using the Q Exactive Plus mass spectrometer. After 15 min stimulation with insulin or IGF-1, many phosphoproteins were shown to be regulated as shown by changes in protein phosphorylation relative to unstimulated controls. This enabled the partial mapping of numerous important KEGG signaling pathways, which were stimulated including DNA replication, RNA splicing, protein synthesis, and cell division (Table 2).

Table 2. Functional annotated pathways found in DAVID for up-regulated and down-regulated proteins. A 1.25-fold change up or down is used as the threshold for differences in regulation.

	Pathway Mapped	Protein Count	P-Value
1	Ribosome	7	7.3E-3
2	Spliceosome	8	1.2E-2
3	Cell cycle	7	3.7E-2
4	Insulin signaling pathway	7	5.1E-2
5	Wnt signaling pathway	7	7.9E-2
6	Purine metabolism	7	8.3E-2
7	mTOR signaling pathway	4	8.4E-2

In addition, we confirmed that insulin and mTOR signaling pathways are activated. For example, proline-rich AKT 1 substrate 1 was found to be up-regulated for both treatments (Figure 5A). In contrast, reticulon-4 show increased relative abundance after IGF-1 treatment compared to insulin treatment (Figure 5B).

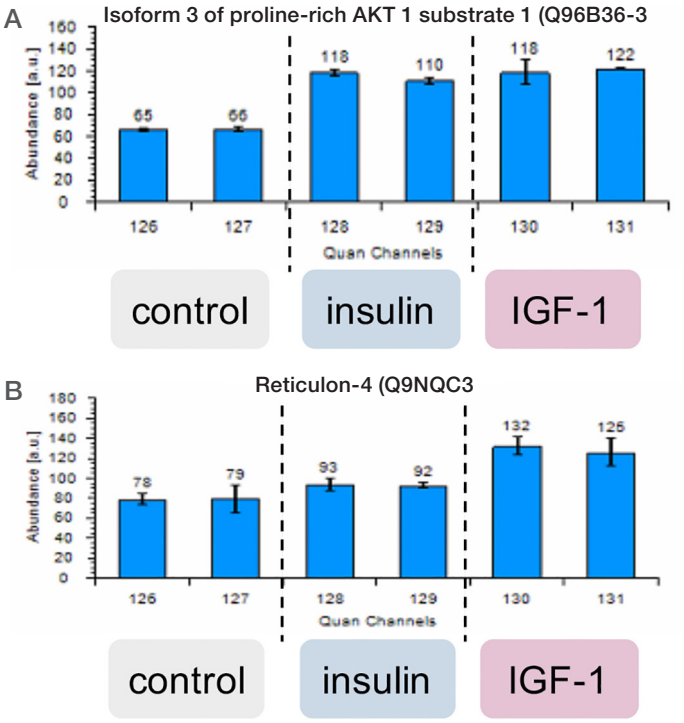

Figure 5. Differential expression of key proteins in A549 cell lines determined by TMT reporter ion based quantification. Results are presented as normalized weighted average protein abundances for runs of two replicates.

Dihydropyrimidinase-related protein 2 was down-regulated for cells treated with insulin and IGF-1 (Figure 6A). For this protein, peptide TVTPASSAKTSPAK[512-525] with S517, S518, and S522 triple phosphorylated significantly decreased in intensity upon treatment, while the same peptide sequence TVTPASSAKTSPAK with S517 and S522 double phosphorylated did not show a change in abundance (Figure 6B and 6C), indicating the site S518 may play an important role in regulating the pathway. Although this S518 site has been identified in a few publications before,[10–11] our data is the first to show this particular site is down regulated after both insulin and IGF-1 treatment.

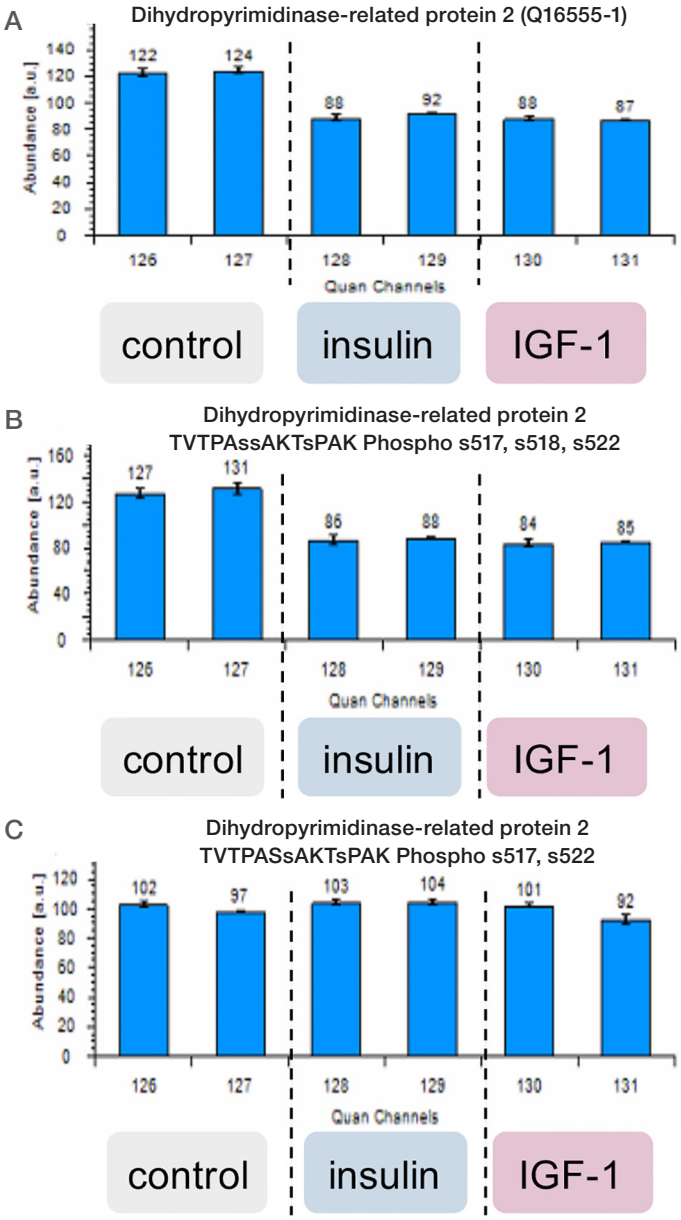

Figure 6. Differential expression of phosphosites in protein dihydropyrimi-dinase-related protein 2 (Accession Q16555-1) upon stimulation. (A) Overall proteins was down-regulated; (B) Abundance of triple phosphorylated peptide[512- 525] got down-regulated; (C) Abundance of double phosphorylated form of the same peptide was unaffected.

thermo scientific

Conclusion

We have shown that accurate peptide concentration measurement can improve biological replicate reproducibility, the number of peptide identifications, and the quantitative accuracy of low abundant peptides. Using a peptide assay to monitor sample preparation, we determined the amount of sample loss for each sample preparation step, which highlighted the need to normalize final peptide concentrations before MS analysis. Implementation of a high pH reversed-phase fractionation of peptide samples resulted in an approximately 300% increase in the number of proteins/peptides identified and quantified compared to direct analysis. Finally, TMT-based relative protein quantification and pathway analysis revealed differences in the regulation and of low abundant insulin pathway signaling proteins after insulin and IGF-1 treatment.

References

1. Schäfer, T. *et al. Anal. Chem.* **2003,** *75*(8),1895.
2. Savitski, M. *et al. Science* **2014,** *346*(6205), 1255784.
3. Weekes, M. *et al. Cell* **2014,** *157*(6), 1460.
4. Rauniyar, N. *et al. J. Proteome Res.* **2014,** 13(12), 5293.
5. Christoforou, A. *et al. Anal Bioanal. Chem.* **2012,** *404*(4), 1029.
6. Li, Z *et al. J. Proteome Res.* **2012,** *11*(3), 1582.
7. Gilar, M. *et al. Anal. Chem.* **2005,** *77*(19), 6426.
8. Haney, P. *et al.* ASMS poster W121, 2014.
9. Snovida, S. *et al.* Thermo Scientific Poster Note 64606, 2015.
10. Rigbolt, K. *et al. Sci. Signal* **2011,** *4*(164), rs3.
11. Gu, Y. *et al. Biochemistry* **2000,** *39*(15), 4267.

Find out more at **www.thermofisher.com**

ThermoFisher SCIENTIFIC