Opening a SWATH Window on Posttranslational Modifications: Automated Pursuit of Modified Peptides*\textsuperscript{§}

Andrew Keller\textsuperscript{§}, Samuel L. Bader\textsuperscript{§}, Ulrike Kusebauch\textsuperscript{§}, David Shteynberg\textsuperscript{§}, Leroy Hood\textsuperscript{§}, and Robert L. Moritz\textsuperscript{§‡}

Posttranslational modifications of proteins play an important role in biology. For example, phosphorylation is a key component in signal transduction in all three domains of life, and histones can be modified in such a variety of ways that a histone code for gene regulation has been proposed. Shotgun proteomics is commonly used to identify posttranslational modifications as well as chemical modifications from sample processing. However, it favors the detection of abundant peptides over the repertoire presented, and the data analysis usually requires advance specification of modification masses and target amino acids, their number constrained by available computational resources. Recent advances in data independent acquisition mass spectrometry technologies such as SWATH-MS enable a deeper recording of the peptide contents of samples, including peptides with modifications. Here, we present a novel approach that applies the power of SWATH-MS analysis to the automated pursuit of modified peptides. With the new SWATHProphet\textsuperscript{PTM} functionality added to the open source SWATHProphet software, precursor ions consistent with a modification are identified along with the mass and localization of the modification in the peptide sequence in a sensitive and unrestricted manner without the need to anticipate the modifications in advance. Using this method, we demonstrate the detection of a wide assortment of modified peptides, many unanticipated, in samples containing unpurified synthetic peptides and human urine, as well as in phosho-enriched human tissue culture cell samples. *Molecular & Cellular Proteomics* 15: 10.1074/mcp.M115.054478, 1151–1163, 2016.

Posttranslational modifications play an important role in many cellular processes, including the regulation of transcript

© 2016 by The American Society for Biochemistry and Molecular Biology, Inc. This paper is available online at http://www.mcponline.org

Technological Innovation and Resources

From the §Institute for Systems Biology, 401 Terry Avenue North, Seattle, WA 98109

Received August 10, 2015, and in revised form, December 18, 2015

Published, MCP Papers in Press, December 24, 2015, DOI 10.1074/mcp.M115.054478

Author contributions: A.K., S.L.B., U.K., and R.L.M. designed the research; A.K., S.L.B., U.K., D.S., and R.L.M. performed the research; A.K., S.L.B., U.K., D.S., and R.L.M. contributed new reagents or analytic tools; A.K., S.L.B., U.K., D.S., L.H., and R.L.M. analyzed data; and A.K., S.L.B., U.K., D.S., L.H., and R.L.M. wrote the paper.
Data-independent acquisition mass spectrometry technologies, such as SWATH-MS, enable a deeper recording of the peptide contents of samples, including peptides with modifications. Using a spectral library generated beforehand with assays for sets of precursor ions, one can query SWATH-MS data in a targeted manner to identify and quantify peptides in a sample (21–26). Peptides with modifications, for which assays are available, can be analyzed in this manner using software such as open source SWATHProphet (27). Although the greater sensitivity of SWATH-MS approaches enables detection of modified peptides at a lower abundance (21), an assay, and thus the prior identification, is required.

Here, we present new functionality added to SWATH-Prophet that ventures beyond the spectral library by automatically pursuing modifications to library peptides that explain lower ranking results. With SWATHProphet<sup>TM</sup>, precursor ions consistent with a modification are identified, along with the mass and localization of the modification in the peptide sequence. In this manner, peptide modifications and amino acid substitutions can be identified, even if they are not themselves present in the spectral library and possibly have never before been observed. We demonstrate the detection of unanticipated modified peptides in a sample of synthetic peptides spiked into a background of human urine, as well as in phospho-enriched samples derived from a human cancer cell line.

**EXPERIMENTAL PROCEDURES**

**Synthetic Peptides**—1,055 synthetic nonhuman peptides were purchased as unpurified PEPotec peptides (Thermo Fisher Scientific, Huntsville, AL) with N- and C termini as free amine and carboxylic acid, respectively, and C-terminally heavy labeled as either K[13C<sub>6</sub>], H11002] or R[13C<sub>6</sub>, 15N4], both indicated by an asterisk in the peptide sequence. Cysteine residues were carboxamidomethylated. Peptides were analyzed in water/0.1% formic acid (v/v) as neat solution or spiked into samples.

**Preparation of Human Urine Samples**—Urine from a healthy human donor was desalted using a HiPrep 26/10 column (GE Healthcare, Pittsburgh, PA). The protein amount was determined by BCA assay (Thermo Fisher Scientific). The sample was diluted to 1 mg/ml protein, reduced with 10 mM dithiothreitol (Sigma Aldrich, St. Louis, MO) for 25 min at 56 °C, alkylated with 14 mM iodoacetamide (Sigma Aldrich) for 30 min in the dark at room temperature and digested overnight with a 1:100 ratio of trypsin (Promega, Madison, WI) to protein at 37 °C. Digestion was stopped by lowering the pH below 2, and the resulting digest was purified using 1C18 SepPak solid-phase extraction cartridges (waters, milford, MA). One thousand fifty-five synthesized, digested tryptic peptides were pooled at roughly equimolar concentrations of 60 nmol/µl and diluted 1:1 with 0.1% formic acid in water (v/v) or with 1 mg/ml of trypsin digested urine.

**Preparation of Phospho-Enriched and Diphosphorylated Human Tissue Culture Sample**—Human metastatic breast cancer T47D cells (HTB-133, ATCC, Manassas, VA) were grown in RPMI 1640 medium (ATCC) supplemented with 10% FBS (Atlanta Biologicals, Norcross, GA) and 0.2 units/ml bovine insulin (Sigma Aldrich) in a 5% CO<sub>2</sub> environment at 37 °C to 70% confluence. T47D cells were harvested using trypsin with 0.25% EDTA. Cell pellets were washed three times with dPBS before being snap-frozen in dry ice/ethanol and stored at −80 °C until cells were lysed in 8 M urea, 100 mM NH₄HCO₃, and 0.1% RapiGest SF Surfactant (Waters). Proteins were reduced with 5 mM tris(2-carboxyethyl)phosphine hydrochloride (1h, 37 °C), alkylated with 10 mM iodoacetamide (30 min, room temperature, darkness), and digested with trypsin (1:50 enzyme:substrate ratio, 37 °C, overnight). The digest was acidified with 2 M HCl to a final concentration of 50 mM. Trifluoroacetic acid (TFA) was added to 0.5% (v/v) followed by incubation for 15 min at 37 °C and centrifugation for 15 min at 4,000 rpm. The supernatant was dried under centrifugal evaporation (Savant, thermo Fisher Scientific, Huntsville, AL) to minimal liquid content, then diluted in 0.1% trifluoroacetic acid in water, acidified with trifluoroacetic acid (pH<sub>H</sub>−3), and desalted using TC18 SepPak cartridges (Waters). The eluate was dried under centrifugal evaporation. Immobilized metal affinity chromatography was performed to enrich for phosphopeptides using PHOS-Select Iron Affinity Gel (Sigma-Aldrich) following a similar protocol (28) and then the Titansphere<sup>TM</sup> Phos-TiO<sup>®</sup> kit (GL Sciences Inc., Tokyo, Japan) was used according to the manufacturer’s instructions. Peptides were desalted using TC18 SepPak cartridges prior to MS analysis. To dephosphorylate the peptides, an aliquot of the Fe(III)- and TiO<sub>2</sub>-enriched sample were mixed and spiked with retention time peptides, dried under centrifugal evaporation, and resolubilized in 125 mM NH₄HCO₃, 2 mM MgCl<sub>2</sub>, Alkaline phosphatase (1U) (Roche, Indianapolis, IN) was added and the sample incubated at 37 °C for 40 min. The sample was acidified prior to MS analysis.

**LC-MS/MS and SWATH-MS Analysis**—The peptide samples were analyzed using the Eksigent Ekspekt<sup>®</sup> nanoLC 425 System combined with the chipLC<sup>®</sup> system in Trap-Elute mode. The samples were first loaded on the chipLC trap (200 µm ×500 µm ChromXP C18-CL, 3 µm, 120 Å) and washed for 10 min at 2 µl/min and eluted with a 120 min linear gradient from 3–35% acetonitrile in water with 0.1% formic acid (v/v) at 300 nl/min using a nano chipLC column (75 µm × 15 cm ChromXP C18-CL, 3 µm, 120 Å).

Peptides eluting from the column were analyzed on a TripleTOF<sup>®</sup> 5600+ system (Sciex, Framingham, MA) equipped with a Nanospray-III<sup>®</sup> Source. To generate libraries, the data were acquired in data-dependent mode. MS1 spectra were collected in the range 100–2,000 Da for 250 ms. The 20 most intense precursor ions in the mass range of 400–1,250 Da with a charge state 2–4 were selected for fragmentation with a rolling collision energy and a collision energy spread of ±15 V, and MS/MS fragment spectra were collected in the range of 100–2,000 Da for 200 ms.

SWATH-MS data were acquired with an MS/MS<sup>TM</sup> with SWATH<sup>TM</sup> Acquisition method, where Q1 was scanned from 350–1,200 Da and MS/MS was acquired from 300–1,500 Da. Q1 transmission window was 27.56 Da wide for the synthetic peptide samples and 25 Da for the phospho-enriched samples, each with a 1 Da overlap with the previous window. Thirty-two steps were used with a 100 ms MS/MS accumulation time on each step for a total cycle time of 3.2 s. At the beginning of each cycle, a survey scan from 200–1,500 Da was acquired with an accumulation time of 50 ns. Data used in this study are deposited at the PeptideAtlas online repository at http://www.peptideatlas.org/PASS/PASS00732.

**Spectral Library and SWATH-MS Assay Construction**—Profile-mode wiff files from data-dependent acquisition were centroided and converted to mzML format (29) using msconvert of ProteoWizard version 3.0.4624 (30) selecting manufacturer’s peak picking algorithm for all spectra as the only filter. The raw data were searched with XTandem against either the sequences of synthetic peptides or the UniProt human varsplice database (March 2012, www.uniprot.org) supplemented with common contaminant proteins and decoy sequences. Search results using a more recent version of the human varsplice database (October 2015) gave nearly identical results. A total of 71,764 total combined entries was used as the database. Carboxamidomethylation of cysteine (+57.021464) was set as a fixed modification.
modification. Oxidation on methionine (+15.9949 Da) and for samples with synthetic peptides, heavy lysine (+8.014199 Da) and arginine (+10.008269 Da) were set as variable modifications. Up to two missed cleavages were allowed. The precursor and fragment ion accuracy were set to 300 ppm and 30 ppm, respectively. Search results were statistically validated with the Trans-Proteomic Pipeline (v4.6 OCCUPY rev 3) using PeptideProphet (31) and iProphet (32).

A raw spectral library was built based on the identifications filtered for an FDR of 1% using SpectraST (33). A consensus spectrum was constructed from all redundant spectra acquired for each precursor. Targeted SWATH-MS assays were generated with all b and y ions annotated in the consensus spectrum with a charge state of 1 or 2. The retention time of each precursor ion was assigned the median retention time of its spectra with the highest number of assigned fragment ions.

Data Analysis—Raw SWATH-MS files can be converted to mzML or to legacy mzXML (34) using msconvert of ProteoWizard and analyzed by SWATHProphet using the constructed spectral libraries. The library assays direct the extraction of ion chromatograms from the raw data and the identification of co-eluting peaks or peak groups in those chromatograms. Peak groups are evaluated with many scores reflecting chromatography, agreement between observed and expected fragment ion peak intensities, mass accuracy, and other features to help identify which likely correspond to their intended precursor ion. Spectral libraries for use with SWATHProphet included assays with only the six most intense fragment ions of each precursor ion. Employing a uniform number of fragment ions in all assays ensures that peak group scores whose values are sensitive to that number, e.g. peak intensity correlation with library, are unbiased. Because many precursor ions, especially shorter ones, have limited numbers of available fragment ions of high intensity, using assays with six fragment ions enables specificity of detecting the precursor ions of interest without excluding a large fraction of precursor ions.

Initially SWATH-MS data were analyzed by SWATHProphet in a conventional manner (27) to identify unmodified sample peptides. Library assays for a set of retention time normalization peptides were first extracted over the entire chromatogram and analyzed to correlate their predicted normalized and observed retention time values. Fragment and parent ions of all library assays were then extracted from the raw data with an m/z tolerance of ±0.05 Da and a retention time tolerance of ±7.5 min centered on the elution time corresponding to their assay normalized retention time.

Unmodified peptides identified with probability 0.9 or greater and with no missing assay fragment ion peaks were then pursued for modifications with SWATHProphet™ by re-extracting their library assays over the entire chromatography run in precursor windows corresponding to modification masses in the user-specified range ±205 to ±205 Da for the synthetic peptide and urine data set and ±25 to ±205 Da for the phosphopeptide-enriched samples. Modification mass spectra (see “Results” and Fig. 1B) were generated at 0.1 Da resolution by binning to infer the modification mass for low ranking peak groups with missing fragment ion peaks consistent with a localized modification. The putative modified peptides were then automatically pursued by re-extracting the missing fragment and parent ion traces of such peak groups from the raw data using m/z values updated for the modification and scoring the resulting “modified peptide” peak groups. Three user-specified modifications (deamidation of glutamine and asparagine, oxidation of methionine, and pyroglutamation) were pursued for identified modifications satisfying their defined criteria for modification mass, localized amino acid residues, and normalized retention time difference with respect to the unmodified peptide (see “Results”). Decoy precursor ion assays were generated for the pursued modified peptides by creating modifications at different locations in the peptides and with random modification masses that maintain the precursor window.

A linear combination discriminant score is learned based on the modified peptide peak groups, with the decoys serving as known incorrect results, using eight relevant scores: intensity correlation with assay, peak shape, co-elution, fraction of missing fragment ion peak intensities, parent peak shape, parent co-elution, parent m/z deviation, and parent isotope peak correlation. The last four scores were computed with the MS1 trace, though could rely on the MS2 in cases in which MS1 is not acquired. Probabilities that peak groups are correctly assigned to their corresponding modified peptide are then computed based on the discriminant score, as previously described (27), as well as the type of modification (i.e. one specifically pursued based on satisfying criteria or pursued on the basis of modification mass alone), and the presence or absence of matches to the Unimod database. Alternative modifications based on the same unmodified peptide peak group are grouped together, and their probabilities penalized to ensure they do not sum to more than unity. Results of multiple samples and replicates were combined together with iProphet as previously described (27) but run with the SHARPNS (number of sibling experiments model based on the normalized retention time difference between modified peptides identified in the same precursor windows in multiple experiments, weighted by the maximum probability observed among the experiment’s replicates), The No Number of Sibling Modified (NONSM) Peptides model and No Number of Sibling Peptides (NONS) model options exclude the use of numbers of alternative modifications of the same peptide model and numbers of alternative peptides corresponding to the same protein model, respectively. That information is not applicable in this case where we seek to validate identified modified versions of peptides we already know are present in the sample.

It is possible that a detected modified version of one library peptide actually corresponds to another library peptide with a similar sequence, for example differing by one or two amino acids. Such cases are detected as co-eluting peak groups of the same charge and precursor ion m/z that share a majority of b and y fragment ion masses within a specified mass tolerance and flagged as possible identical precursors. Such “modified” peptides can optionally be filtered out of the reported results.

RESULTS

Quantitative analysis of SWATH-MS data in a targeted manner requires precursor ion assays compiled in a spectral library. A high-quality precursor ion assay consists of a signature set of fragment ions with accurate expected relative peak intensities, including relative isotopic peak intensities, and retention time. Using SWATHProphet software, traces for library assay fragment and parent ions are extracted within a specified mass tolerance from MS2 spectra of the appropriate precursor window. One or more groups of co-eluting peaks in the extracted fragment ion traces, referred to as peak groups, are identified for each assay and assigned an overall score reflecting its chromatography, agreement between observed and expected fragment ion peak intensities, mass accuracy, and other features. The top-ranking peak groups corresponding to each assay are
then assigned probabilities that they are correctly assigned to their library precursor ion.

Identification of Modified Peptides Corresponding to Lower-Ranking Peak Groups—Since modified versions of peptides, including amino acid substitutions, often share many fragment ions with their unmodified siblings, it is sometimes possible to detect them in a sample even when they are not themselves represented in the spectral library. Candidates are lower-ranking peak groups assigned to the unmodified peptide that are missing peaks for one or more fragment ions consistent with a localized modification (Fig. 1A). The missing fragment ions and the precursor ion are presumed to be absent because they contain the modification and hence have different m/z values from those of the unmodified peptide. The possible location of the putative modification can extend across one or more amino acid positions, depending on the particular fragment ions in the library assay for the unmodified peptide and which of those are undetected. The fragment ions in a library assay constrain the amino acid positions at which a modification can be detected.

Based on this theoretical framework, we developed a novel approach that infers the mass of the putative modification from shifts in m/z exhibited by the precursor and missing fragment ions in lower-ranking peak groups (Fig. 1B). The modification mass, in the range consistent with the m/z boundaries of the precursor window, is determined by generating charge state (z) deconvoluted “modification mass spectra” of the parent and missing fragment ions through a linear transformation of either the MS1 spectrum near the peak group apex time, in the case of the parent ion, or the MS2 spectrum at the peak group apex time, in the case of missing fragment ion i:

\[
\text{ModificationMass} = (\text{Modified}_{m/z} - \text{Unmodified}_{m/z})^z
\]

\[
\text{ModificationMass}_{\text{parent}} = (\text{MS1}_{m/z} - \text{parent}_{m/z})^z
\]

\[
\text{ModificationMass}_{\text{fragment}} = (\text{MS2}_{m/z} - \text{fragment}_{m/z})^z
\]

(Eq. 1)

The MS2 spectrum can also be used for the unfragmented parent ion if MS1 data are not collected during the acquisition. Since the modification mass spectra are charge state deconvoluted, the modification monoisotopic mass is identified from 1 Da spaced peaks present in all modification mass spectra, the intensities of which correlate with the expected isotopic peak relative abundances.

Modification mass spectra can additionally be used to further localize the modification in the peptide sequence (Fig. 1C).
For example, if the +80 Da modification in the peptide FESPKEPEQLR is known to reside at the third or fourth amino acid residue because the assay y7 fragment ion peak is present and the y9 fragment ion peak is absent, the modification mass spectrum for the y8 fragment ion that is not present in the library assay is generated and queried for the presence of either the modification mass (+80 Da) or 0 Da peak. The former indicates that the modification resides at the proline residue since the y8 fragment ion contains the modification. The latter indicates that the modification resides at the serine residue since the y8 fragment ion lacks the modification. The analogous steps can be taken with respect to b ions when the end position of the localized modification sequence coincides with a missing b ion peak at that position. This approach is effective only if the discriminating fragment ion is detectable, despite its exclusion from the library assay.

**Automated Pursuit of Modified Peptides**—We added functionality, termed SWATHProphetPTM, to the SWATHProphet software to automatically pursue modified versions of library peptides, including biological posttranslational modifications, which could correspond to lower-ranking peak groups with missing fragment ion peaks. Modification mass spectra are generated for the peak groups with respect to their parent and missing fragment ions in order to infer the mass and sequence localization of their modifications. These putative modified peptides are then automatically pursued by re-extracting their peak group missing fragment and parent ion traces from the raw data using assays with m/z values updated for the modification and scoring and validating the resulting modified peptide peak groups (see Experimental Procedures). The updated assays, by necessity, employ the predicted fragment ion peak intensities and relative isotope peak intensities of the unmodified peptides. Nevertheless, they are expected to be effective as long as their fragment ion peak intensities do not differ severely from those of the modified peptides being pursued. In order to reduce the chance of false positives, detection of fragment ion peak(s) containing the modification is required. Modified peptides are reported displaying their localized modification sequence, the one or more adjacent amino acid residues that could host the modification, in lowercase letters followed by the modification mass in brackets. When an identified modification is not fully localized to a single amino acid residue, it could actually be composed of multiple modifications at distinct residues in the localized sequence, their masses summing to the reported modification mass.

Results are automatically mapped to the Unimod database for known modifications with matching mass and target amino acids. For example, an identified peptide ATNESelPQLVPIGK with a +80 Da modification localized to its fifth or sixth amino acid residue would be mapped to Unimod entry for phosphorylation at serine with modification mass +79.97 Da (ATNES[Pho]EDEIPQLVPIGK), consistent with the +80 Da modification localized to the serine residue of the peptide. Explanations of amino acid substitutions and N-terminal or C-terminal additional amino acids are also included and flagged with a tilde when found consistent with sequences in a specified protein sequence database. An identified human peptide aqy[+129.1]SVQPNFQQDK with a +129.1 Da modification localized to its first three amino acids would be mapped to the entry for N-terminal addition of E~ with modification mass +129.04 Da, the tilde indicating that the peptide EAQSVQPNFQQDK with an added glutamic acid residue at its N terminus is present in the specified human protein sequence database. The Unimod information must be manually reviewed to assess which if any matched entries are likely to explain the identification.

The manual step of reviewing matches to Unimod database entries can be bypassed for specific modifications that have a high likelihood of being present. Such modifications can be directly pursued based on defined criteria, including the mass of the putative modification, the localized modification sequence, and the normalized retention time difference with respect to the unmodified peptide ($\Delta r_{\text{norm}}$). For example, oxidized methionine might be pursued when the modification mass is +16 Da, the possible modified amino acids include methionine, and $\Delta r_{\text{norm}}$ is between -30 and -0.5 min, reflecting earlier observed elution times on reversed-phase liquid chromatography for peptides with that modification due to the more polar nature of the sulfoxide. When all such criteria are met, peak group ion traces are re-extracted using library assays for the peptide with the oxidized methionine, having the same fragment ions and predicted fragment peak intensities as the unmodified peptide assay, but with the precursor and fragment ion m/z values and relative isotope peak intensities of the modified peptide. Currently, up to two modifications per precursor are considered for automated pursuit. For example, two oxidation moieties could be pursued for a peak group when the modification sequence contains two methionine residues and the modification mass is +32 Da. Only putative modified peptides not already present in the spectral library are pursued in this manner. If the criteria of no specified modification are met, the modified peptide can be pursued on the basis of its modification mass alone.

Extracting library assay traces in adjacent precursor windows can extend the mass range of identified modifications. One can specify a desired modification mass range to direct extraction of each library assay in the corresponding precursor windows. A good strategy for detecting modifications is to first analyze the data in a conventional manner and then pursue modifications of the library precursors identified with high confidence and with detection of all assay fragment ion peaks, by re-extraction with a specified modification mass range and time tolerance (Supplemental Fig. S1).

**Analysis of Synthetic Peptides Spiked into Human Urine**—To test the SWATHProphetPTM automated pursuit of modified peptides, we analyzed samples with 1,055 heavy-labeled synthetic peptides spiked into either neat solvent or...
trypsinized human urine backgrounds. The synthesized peptides were unpurified and hence may contain uncharacterized peptide variants arising from the chemical synthesis process. Such examples of unanticipated modifications are a good test of the method. Both samples were acquired in SWATH-MS mode (three replicates each) and analyzed with a spectral library generated from shotgun MS data acquired from the same samples, containing 1,316 synthetic and 1,092 human precursor ion assays, all composed of the six most intense fragment ions. Library assays corresponding to oxidized methionine and pyroglutamation modified peptides (118 synthetic and 45 human) were excluded from this analysis; of interest is whether any of those modifications can be identified on the basis of unmodified peptide assays alone. SWATHProphet was initially run in a conventional manner to identify high-confidence unmodified sample peptides with no missing assay fragment ion peaks (~1,000 synthetic and 600 human urine peptides). Assay traces of those peptides were then extracted in precursor windows to allow detection of modifications with masses ranging from −205 to +205 Da in peak groups identified with missing fragment ion peaks. This wide range was chosen in order to identify as many modifications as possible, including insertion and deletion of amino acids (see Experimental Procedures). If a particular set of modifications is of interest, this range can be appropriately narrowed. Criteria were specified for the automated pursuit of deamidation (mass +1 Da, localized to a sequence containing glutamine or asparagine), oxidized methionine (mass +16 Da, localized to a sequence containing methionine), and pyroglutamation (mass −17 Da, localized to a sequence containing N-terminal glutamine). In all cases, no advance restriction was set for Δmnorm in order to learn the appropriate range from the data. Results were mapped against modifications in the UniMod database, as well as N- and C-terminal additions and deletions of amino acids. Identified modifications in the neat solvent and human urine background replicate runs were combined by iProphet to appropriately penalize results identified in few replicates and reward those identified in multiple replicates.

In total, 209 modified synthetic peptides and 65 modified human urine peptides were identified at a predicted 1% FDR (Supplemental Table S1). Eighty-seven of the modified peptides were obtained using specified criteria for automated pursuit: five deamidation identities, 67 oxidized methionine residues (Fig. 2A), and seven pyroglutamation identities were detected among the synthetic peptides, of which 27 were not present in the spectral library. Fewer human urine peptides were likewise identified: two deamidation identities, five oxidized methionine residues, and one pyroglutamation, of which four were not present in the spectral library. The remaining 187 identified modified peptides were pursued on the basis of the modification mass alone, of which 136 (73%) were annotated with the help of the Unimod database. In total, chemical explanations were obtained for 223 (81%) of the 274 total reported peptide modifications (Table I). The different types of modifications identified among synthetic peptides versus human urine peptides are a testament to the unrestricted nature of the SWATHProphet™ automated pursuit of modified peptides. The detection of modified peptides in biological samples in this study range over three orders of magnitude (~2,000-fold) from the lowest concentration of modified synthetic peptides doped into human urine compared with the most abundant endogenous human peptides recovered and from two to three orders of magnitude (~700-fold) for modified endogenous human peptides identified in human urine (see Table S1).

Several identified “modified” peptides correspond to amino acid substitutions in the peptide sequence. Cases in which the “modified” peptide likely corresponds to an alternative variant peptide sequence in the spectral library are automatically flagged. For example, a “modified” synthetic peptide AQAgl[-12]LEAEHQRANR* was a two amino acid substitution (GL→AS with a modification mass of −12 Da), corresponding to another peptide AQAASLEAEHQRANR* present in the spectral library. Four “modified” urine peptides were amino acid substitutions, including v[+14]SGGGLVOPGGSLR, with a +14 Da modification localized to its first or second amino acid, consistent with a V→L substitution to the homologous peptide LESGGGLVOPGGSLR of the human IgG variable region protein variant that was not present in the spectral library.

Thirteen modified synthetic peptides correspond to a missing N-terminal amino acid, presumably resulting from incomplete C' to N' terminal synthesis. No modified synthetic peptides correspond to a missing C-terminal amino acid nor to an extra amino acid added to either terminus. In contrast, 29 modified human urine peptides had termini different from the unmodified sequence, with deleted or additional 1–2 amino acids at either terminus (Fig. 2B). Twelve of these peptides were unanticipated by virtue of their absence in the spectral library. A relatively large number of semi-tryptic peptides were generally present in the urine sample. Interestingly, 11 and 25 modified synthetic peptides are consistent with having missing or duplicated internal amino acids, respectively. These may result from incomplete deblocking or washing during the individual steps of the chemical synthesis of the peptide. Additional identified modifications of synthetic peptides include 11 cases of oxidation of tryptophan (+4 or +32 Da), 10 cases of piperidine transamidation of aspartic acid (+67.1 Da), a side reaction with piperidine for the removal of the Fmoc protecting group during peptide synthesis (35), and 19 cases of dehydration at aspartic acid (−18 Da). Modifications of human urine peptides include 11 cases of carbamidomethylation at noncysteine N-terminal residues (+57 Da) resulting from treatment with iodoacetamide. Two modifications of a semi-tryptic human peptide were identified with masses −30 Da.
and –48 Da, consistent with conversion of the C-terminal methionine residue to homoserine and homoserine lactone, respectively (Supplemental Fig. S2). This unexpected finding suggests that such conversions, commonly observed when proteins are cleaved by CNBr, can occur spontaneously in urine. Homoserine itself has been detected in urine samples from neuroblastoma patients (36).

The remaining 51 identified modifications remain without chemical explanation. A major challenge to this analysis is that once modifications with specific masses and localized amino acid sequences are observed, it is still necessary to determine the chemical composition of the modification. Plotting modification masses localized to each amino acid versus \( \Delta r_{\text{norm}} \) can help characterize known modifications and their target amino acids, as well as reveal new ones. In fact, this approach enabled us to annotate piperidine transamidation, an unexpected modification with mass of +67.1 Da observed in several synthetic peptides with similar \( \Delta r_{\text{norm}} \) values, often localized to a single aspartic acid in the peptide (Fig. 3). Synthetic peptide modifications of mass 188.1 Da localized to tyrosine

### Table I

| Modification type                  | Synthetic | Urine |
|------------------------------------|-----------|-------|
| Deamidation                        | 5         | 2     |
| Oxidation of methionine            | 67        | 5     |
| Pyroglutamation                    | 7         | 1     |
| Amino acid substitution            | 1         | 4     |
| N-terminal deletion                | 13        | 6     |
| N-terminal addition                | 0         | 8     |
| C-terminal deletion                | 0         | 7     |
| C-terminal addition                | 0         | 8     |
| Missing internal amino acid        | 11        | 0     |
| Duplicated internal amino acid     | 25        | 0     |
| Oxidation of tryptophan            | 11        | 0     |
| Piperidine trans-amidation of aspartic acid | 10        | 0     |
| Non-cysteine carbamidomethylation  | 0         | 11    |
| Dehydration of aspartic acid       | 19        | 0     |
| Conversion of methionine to homoserine | 0         | 2     |
| **Total annotated**                | 169       | 54    |
| **Total identified**               | 209       | 65    |
| **Total unannotated**              | 40        | 11    |
and mass −45 Da localized to serine were characterized in a similar manner, though remain without chemical explanation. The restricted ranges of Δt_{norm} encountered for deamidation, oxidation of methionine, and pyroglutamation (Supplemental Table S2) are characteristic properties that can be used as criteria for automated pursuit of those modifications in the future. We will store information regarding observed modifications in a ModificationAtlas as part of the SWATHAtlas online resource (S. Bader, unpublished) where peptide modifications identified from SWATH-MS data can be queried.

The majority (149, or 67%) of the 223 identified modified peptides that were successfully annotated were not present in the spectral library. Many were not anticipated and hence not included in the shotgun MS search parameters during construction of the library. Others may have been at too low abundance to be detected by shotgun MS. These results demonstrate that SWATHProphetPTM can identify modified peptides and peptides with missing, additional, or substituted amino acids not previously observed. Doing so, it provides functional SWATH-MS assays for the modified peptides that can be added to the spectral library for future analyses. The automated pursuit of modified peptides can thereby expand the spectral library in the process of using it to analyze data sets.

Independent validation of the 74 identified modified peptides that were present in the spectral library was pursued by analyzing each synthetic peptide both by SWATH and shotgun MS, individually in the presence of only 14 retention time normalization peptides to reduce the sample complexity. The 14 modified synthetic peptides include five with N-terminal amino acid deletions ranging in probability from 0.26 to 1, one with an extra internal glycine with a probability of 0.94, and eight with pyroglutaminated aspartic acid ranging in probability from 0.87 to 1. In these samples with only 15 synthetic peptides (one peptide with a putative modified form and 14 retention time peptides), it is very unlikely to encounter co-eluting fragment ions contributed by multiple parents in the SWATH-MS data and easier to detect low-abundance modified peptides in the shotgun MS. Using SWATHProphet to analyze the SWATH-MS data in a conventional manner with a spectral library containing only the 14 retention time normalization peptides and 14 modified peptides along with their unmodified counterparts, validation was based on the identification at estimated 1% FDR of only the expected modified peptide and its unmodified counterpart, aside from the normalization peptides. In this manner, 13 of the 14 modified peptides were validated (Table II). Only dataset-[87]FLIATR* with reported possible N-terminal deletion of serine, a low-confidence result with probability 0.34, was not. Identical validation results were obtained from the shotgun data of each sample searched with the modification in question as variable, filtered for estimated 1% FDR. Of interest is the identified modification tygga+[57]YIAMNSR*, consistent with an unanticipated additional glycine, most likely inserted adjacent to the two glycine residues during synthesis. When the peptide sequence TYGGGAYIAMNSR* with the extra glycine was added to the search database, it matched three MS2 spectra in that sample’s shotgun data with high confidence. Figure 4 shows consecutive matching b and y ions that strongly support the identification. The shotgun data also confirm the amino acid localization of several of the validated modifications. For example, four of the peptides with pyroglutaminated aspartic acid residue, and in each case shotgun results are consistent with the reported localization of the +67.1 Da modification.

Analysis of a Phosphopeptide-Enriched Biological Sample—To demonstrate that this novel approach can identify biological posttranslational modifications in an unrestricted manner without library assays for the modified peptides, similar to the identification of amino acid substitutions and modifications due to sample preparation as shown above in the synthetic peptide and urine data set, we applied the SWATHProphetPTM automated pursuit of modified peptides to SWATH-MS data acquired from a human metastatic breast cancer T47D cell line lysate enriched for phosphopeptides by tandem immobilized metal-affinity chromatography using Fe(III) and subsequently TiO2. An aliquot of the Fe(III)– and TiO2–enriched samples were mixed together, treated with a phosphatase to dephosphorylate the peptides, and used to acquire both shotgun and SWATH-MS data. The shotgun
The modified peptide sequences indicate the modification mass in brackets and the possible host amino acids in lower case letters.

**TABLE II**

| Sample peptide | Identified modified peptide | Modification description | Probability | SWATH-MS Validation | Shotgun Validation |
|----------------|-----------------------------|--------------------------|-------------|---------------------|--------------------|
| TTHQVYQDELK* 3+ | tthq[-101]VQDELK* 3+       | N-terminal deletion of T  | 0.26        | Yes                 | Yes                |
| ENPHLTSPPILLEK* 3+ | enphl[-129]LTSPPILLEK* 3+ | N-terminal deletion of E | 0.99        | Yes                 | Yes                |
| SDSETFLIATR* 3+ | sdset[-87]FLIATR* 3+       | N-terminal deletion of S  | 0.34        | No                  | No                 |
| AAGLOSYEIPR* 2+ | aag[-71]OSYEIPR* 2+       | N-terminal deletion of A  | 1           | Yes                 | Yes                |
| GGGVALLTNIR* 2+ | gggv[-57]LLTNIR* 2+       | N-terminal deletion of G  | 1           | Yes                 | Yes                |
| TYGGAYIAMNSR* 2+ | tyggq[-57]YIAMNSR* 2+     | Duplicated internal G    | 0.94        | Yes                 | Yes                |
| WGAFTITNGVSIAS* 2+ | WGAFTITNg[-67.1]GVSIAK* 2+ | Piperidine trans-amidation at D | 1 | Yes                 | Yes                |
| LHEELLAVPDGR* 3+ | LHEELLAvpdg[-67.1]R* 3+  | Piperidine trans-amidation at D | 1 | Yes                 | Yes                |
| VTAEPFTDGQR* 2+ | VTAEPFTg[-67.1]GQR* 2+    | Piperidine trans-amidation at D | 1 | Yes                 | Yes                |
| LPDQTLGDIARY* 2+ | LPDQTLGDIAR[-67.1]Y* 2+  | Piperidine trans-amidation at D | 1 | Yes                 | Yes                |
| ETVDNQLTIPVR* 2+ | ETVDNQLTIPV[-67.1]R* 2+  | Piperidine trans-amidation at D | 1 | Yes                 | Yes                |
| NSDLGNPNTQGALIR* 2+ | NSDLGNPNTG[-67.1]ALIR* 2+ | Piperidine trans-amidation at D | 1 | Yes                 | Yes                |
| FADGSLTNPSADEVK* 2+ | FADGSLTNPSA[-67.1]DEVK* 2+ | Piperidine trans-amidation at D | 0.87 | Yes                 | Yes                |
| SGADVDFGCGSQR* 2+ | SGADVDFG[-67.1]CGSQR* 2+  | Piperidine trans-amidation at D | 0.90 | Yes                 | Yes                |

| SWATH Window on Posttranslational Modifications |

**File: IDA_7-1678-G8.03832.03832.2, Exp. m/z: 685.822, Charge: 2**

![Fig. 4. MS2 spectrum of synthetic peptide TYGGAYIAMNSR* with an extra internal glycine residue. Consecutive matching b and y ions to peptide TYGGAYIAMNSR* lend support for this identified “modification” of the synthesized peptide.](Image)
–25 to +205 Da to simulate having no advance knowledge of the expected modifications. Results of replicate runs of both phospho-enriched samples were combined together using iProphet.

In total, 210 modified peptides identified with a probability of 0.5 or greater likely contain a phosphorylated residue. One hundred ninety-one peptides had a modification mass of 80 Da localized to a sequence containing serine (Fig. 5), threonine, or tyrosine, consistent with a single phosphorylation event. Of those, 171 were observed in the shotgun MS analysis of the enriched samples, corresponding to a success rate of 53% for the 323 unmodified peptide assays that were deemed suitable for the automated pursuit of their phosphorylation. Two modified peptides with \( m/z \) values in the overlap range of adjacent precursor windows, \( 3^{+} \) and \( 3^{+} \), were independently identified for the \( 2^{+} \) and \( 3^{+} \) precursor ions. Assay fragment ions affected by the modification include the y11 and y13 for the \( 2^{+} \) parent ion, and b9 for the \( 3^{+} \) parent ion. The parent ion traces shown were scaled to the intensity range of the fragment ions.

Seven peptides had a modification mass of +159.9 Da localized to a sequence containing multiple serine/threonine residues, consistent with two phosphorylated amino acids, and 12 had modification masses consistent with a phosphorylated amino acid in conjunction with a nearby additional modification, including one pyroglutamation of glutamic acid, two oxidations of methionine, five additional, and four missing terminal amino acids (Supplemental Table S3). Importantly, 32 phosphopeptides were unexpected, not having been observed in the shotgun MS analysis of either the Fe(III)- or TiO\(_2\)-enriched sample. These results demonstrate that SWATHProphet\textsuperscript{TM} can identify previously undetected biological posttranslational modifications such as phosphorylation events without anticipating them on the basis of the unmodified peptide assays alone.

**DISCUSSION**

Comprehensive quantitative evaluation of posttranslational modification events provides valuable data that can be used to interpret biological functions of proteins and applied in a temporal context to elucidate cellular response at a systems level.
level. Identification of modified peptides by shotgun MS is limited by the great increase in computational database search time required to allow for possible modifications. Furthermore, it favors detection of high abundance peptides. SWATHProphetPTM overcomes these limitations by leveraging rich SWATH-MS data to identify modified peptides in a sensitive and unrestricted manner, without the need to anticipate them. The software automatically pursues lower-ranking peak groups with missing fragment ion peaks that may correspond to modified versions of the library peptide, including amino acid substitutions. Modification mass spectra are generated for the parent and missing fragment ions, enabling inference of the modification mass. Automated pursuit of a modified peptide can occur based on user-defined criteria for specific modifications or based on the modification mass alone. The peak group missing fragment and parent ion traces are re-extracted with m/z values updated for the modification, and the resulting modified peptide peak group is scored and validated. This enables detection in a sample of modified peptides that are not themselves represented in the spectral library and, perhaps, have never before been observed. The identification itself furthermore provides a functional library assay for the modified peptide that can immediately be used to detect and quantify it in additional samples. With SWATHProphetPTM, the automated pursuit of modified peptides can be applied in a straightforward manner to the majority of existing SWATH-MS data sets without the need to reacquire the data.

Quantitative analysis of SWATH-MS data in a targeted manner requires a spectral library containing assays for all peptides that are queried in a sample (21). These libraries are limited to include peptides that have already been detected by database search with MS2 spectra generated from shotgun analysis of a sample or with pseudo MS2 spectra extracted from SWATH-MS data (37). Over time, more and more precursor ions will be added to available spectral libraries, increasing the number of peptides that can be identified and quantified in SWATH-MS analysis of samples. However, it is unlikely that spectral libraries will ever become complete in this manner given the large numbers of modified peptides and peptides translated from genes with polymorphisms or edited mRNAs that are not present in protein sequence databases. For this reason, the SWATHProphetPTM automated pursuit of modified peptides can play an important role in supplementing spectral libraries with assays for modified peptides, and even peptides translated from alleles with nonsynonymous SNPs, initially identified on the basis of unmodified peptide assays alone.

Identification of modified peptides as lower-ranking peak groups requires unmodified peptide assays with specific properties: Of the fragment ions in the assay, all must be detected in the peak group of the unmodified peptide, yet at least one and not more than half, not detected in the peak group of the modified peptide due to the presence of the modification. It may also be possible, however, to pursue modifications in peak groups with no missing fragment ion peaks on the basis of a missing parent ion peak alone, localized to a residue contained in no assay fragment ions. Approximately one-third of the assays in the human metastatic breast cancer T47D cell line unmodified peptide library composed of the six most intense fragment ions had properties suitable for SWATHProphetPTM’s identification of their phosphorylation in the phospho-enriched samples. Of those assays, 53% were employed successfully; that number limited in part by differing observed peak intensities of fragment ions with and without the phosphorylation. It would be useful to explore strategies to generate precursor ion assays that can be used to detect modifications at as many amino acid positions as possible. For example, an assay of high-intensity complementary b and y ion pairs that together cover the entire peptide sequence would enable detection of a modification at any residue position. The use of such assays, when available, could help achieve overall detection of corresponding sample peptide modifications with sensitivity on the order of 53%, as observed for phosphopeptides. The sensitivity could potentially be further improved by using assays customized for specific modifications. For example, if one were interested in querying a sample for previously unidentified phosphopeptides for which no assays were available, one could design the unmodified peptide assays using fragment ions predicted to be optimal for detection of phosphorylation at their serine, threonine, and tyrosine residues. Additionally, when updating the assay for a putative phosphopeptide, one could employ fragment ion peak intensities predicted specifically for that modified peptide. Finally, rather than restrict all assays to a uniform number of fragment ions such as six, it may be beneficial to include as many as are available for each precursor ion. Toward that aim, we are exploring new peak group scores that are independent of the number of fragment ions.

Databases with known protein modifications play a critical role to facilitate chemical explanations for modified peptides encountered in samples. Explanations can range from modifications at a single amino acid to combinations of modifications at multiple neighboring residues, as observed in the phospho-enriched samples. We intend to utilize as much available information as possible in search for plausible matches with our results. In addition, we look forward to making contributions to the ModificationAtlas database in the future with discoveries made by applying the automated pursuit of modified peptides to SWATH-MS data. The ability of SWATHProphetPTM to identify peptide modifications in an unrestricted manner presents an opportunity to discover unanticipated posttranslational modifications and modifications due to sample handling, as well as amino acid substitutions such as those translated from genes with unknown SNPs or RNA editing events, as might occur in cells undergoing tumorigenesis (38, 39).
Software Availability and License—SWATHProphet with its new functionality to automatically pursue modified peptides (SWATHProphet™) is available at http://tools.proteomecenter.org/software/SWATHProphet/ and is released under a dual license. For academic, noncommercial use of the software, the GNU General Public License (GPLv3) open source license may be used. Other users who wish to use SWATHProphet in ways that are not compatible with open source licenses can contact the authors at the Institute for Systems Biology for licensing.

Acknowledgments—We thank Dr. Michael Hoopmann for advice; David S. Campbell, Barbara Grimes, and Sarah Li for technical assistance; and all members of the Moritz lab for helpful discussions. We thank Dr. Aaron Hudson, Mark Cafazzo, and Dr. Christie Hunter (Sciex) for access to a 5600+ SWATH-MS-enabled TripleToF MS instrument.

This work was funded in part by the American Recovery and Reinvestment Act (ARRA) funds through National Institutes of Health from the National Human Genome Research Institute grant No. RC2 HG005805; the National Science Foundation MRI Grant No. 0923536 and MCB Grant No. 1330912, the National Institute of General Medical Sciences under Grant Nos. 2PS0 GM076547/Centers for Systems Biology, and GM087221; and Grant Nos. OPP1039684 and OPP1115430 from the Bill & Melinda Gates Foundation. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

† To whom correspondence should be addressed: Institute for Systems Biology, 401 Terry Avenue North, Seattle, WA 98109, USA, Tel: 1-206-732-1244; Email: robert.moritz@systemsbiology.org

REFERENCES

1. Gurard-Levin, Z. A., and Almouzni, G. (2014) Histone modifications and a choice of variant: A language that helps the genome express itself. F1000Prime Reports 6, 76–85

2. Bonnet, A., and Palancae, B. (2014) Regulation of mRNA trafficking by nuclear pore complexes. Genes 5, 767–791

3. Filtz, T. M., Vogel, W. K., and Leid, M. (2014) Regulation of transcription factor activity by interconnected post-translational modifications. Trends Pharmacol. Sci. 35, 76–85

4. Wang, Y. C., Peterson, S. E., and Loring, J. F. (2014) Protein post-translational modifications and regulation of pluripotency in human stem cells. Cell Res. 24, 143–60

5. Hitosugi, T., and Chen, J. (2014) Post-translational modifications and the Warburg effect. Oncogene 33, 4279–4285

6. Merbi, Y., Refour, P., Patel, H., Springer, M., and Kirschner, M.W. (2013) Profiling of ubiquitin-like modifications reveals features of mitotic control. Cell 152, 1160–1172

7. Boeckmann, B., Bairoch, A., Apweiler, R., Blatter, M. C., Estreicher, A., Gasteiger, E., Martin, M. J., Michoud, K., O’Donovan, C., Phan, I., Pibout, S., and Schneider, M. (2003) The SWISS-PROT protein knowledgebase and its supplement TREMBL in 2003. Nucleic Acids Res. 31, 365–70

8. Khoury, G. A., Ballabin, R. C., and Floudas, C. A. (2011) Proteome-wide post-translational modification statistics: Frequency analysis and curation of the Swiss-Prot database. Sci. Rep. 13, 1–5

9. Creasy, D. M., and Cottrell, J. S. (2004) Unimod: Protein modifications for mass spectrometry. Proteomics 4, 1354–1358

10. Garavelli J.S. (2003) The RESID database of protein modifications: 2003 developments, Nucleic Acids Res. 31, 499–501

11. Olsen, J. V., and Mann, M. (2013) Status of large-scale analysis of post-translational modifications by mass spectrometry. Mol. Cell. Proteomics 12, 3444–3452

12. Witze, E. S., Old, W. M., Resing, K. A., and Ahn, N. G. (2007) Mapping protein post-translational modifications with mass spectrometry. Nat. Methods 4, 798–806

13. Craig, R., and Beavis, R. C. (2004) TANDEM: Matching proteins with mass spectra. Bioinformatics 20, 1466–1467

14. Zhang, J., Xin, L., Shan, B., Chen, W., Xie, M., Yuan, D., Zhang, W., Zhang, Z., Lajoie, G. A., and Ma, B. (2012) PEAKS DB: De novo sequencing assisted database search for sensitive and accurate peptide identification. Mol. Cell. Proteomics 11, M111.010587

15. Ballabin, R. C., DiMaggio, P. A., Piazas-Mayoraca, M. D., Young, N. L., Garcia, B. A., and Floudas, C. A. (2010) A novel approach for untargeted post-translational modification identification using integer linear optimization and tandem mass spectrometry. Mol. Cell. Proteomics 9, 764–779

16. Tsur, D., Tanner, S., Zandi, E., Bafna, V., and Pevzner, P. A. (2005) Identification of post-translational modifications via blind search of mass-spectra. Proc. IEEE Comput. Syst. Bioinform. Conf. 2005, 157–166

17. Shteynberg, D., and Savitski, M. M. (2010) Unbiased detection of posttranslational modifications using mass spectrometry. Meth. Mol. Biol. 673, 203–210

18. Ye, D., Fu, Y., Sun, R. X., Wang, H. P., Yuan, Z. F., Chi, H., and He, S. M. (2010) Open MS/MS spectral library search to identify unanticipated post-translational modification and increase spectral identification rate. Bioinformatics 26, i399–i406

19. M. C. W., and Lam, H. (2014) Hunting for unexpected post-translational modifications by spectral library searching with tier-wise scoring. J. Proteome Res. 13, 2262–2271

20. Chick, J. M., Kolippakkam, D., Nusinow, D. P., Zhai, B., Rad, R., Huttlin, E. L., and Gygi, S. P. (2015) A mass-tolerant database search identifies a large proportion of unassigned spectra in shotgun proteomics as modified peptides. Nature Biotechnol. 33, 743–749

21. Gill, L. C., Navarro, P., Tass S., Rist, H., Seleevek H., Reiter, L., Bonner, R., and Aebersold R. (2012) Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis. Mol. Cell. Proteomics 11, O111.016717

22. Panchaud, A., Scherl A., Shaffer S. A., von Haller P. D., Kulasekara H. D., Miller S. I., and Goodlett D. R. (2009) Precursor acquisition independent from ion count: How to dive deeper into the proteomics ocean. Anal. Chem. 81, 6481–6488

23. Plumb, R. S., Johnson K. A., Rainville P., Smith B. W., Wilson I. D., Castro-Perez J. M., and Nicholson J. K. (2006) UPLC/MSE: A new approach for generating molecular fragment information for biomarker structure elucidation. RCM 20, 1989–1994

24. Venable, J. D., Dong, M. Q., Wolhlsiegel, J., Dillin, A., and Yates, J. R. (2004) Automated approach for quantitative analysis of complex peptide mixtures from tandem mass spectra. Nat. Methods 1, 39–45

25. Egerton, J. D., Kuehn A, Merrihew G. E., Bateman N. W., MacLean B. X., Ting Y. S., Canterbury J. D., Marsh D. M., Kellrnann M., Zabrouskov V., Wu C. C., and MacCoss M. J. (2013) Multiplexed MS/MS for improved data-independent acquisition. Nat. Methods 10, 744–746

26. Chapman, J. D., Goodlett, D. R., and Masselon, C. D. (2013) Multiplexed and data-independent tandem mass spectrometry for global proteome profiling. Mass Spec. Rev. 33, 452–470

27. Keller, A., Bader, S. L., Shteynberg, D., Hood, L., and Moritz, R. L. (2015) Automated validation of results and removal of fragment ion interference in targeted analysis of data independent acquisition MS by SWATHProphet. Mol. Cell. Proteomics 14, 1411–1418

28. Villen, J., and Gygi, S. P. (2006) The SCX/IMAC enrichment approach for global phosphorylation analysis by mass spectrometry. Nat. Protocols 3, 1630–1638

29. Martens, L., Chambers, M., Sturm, M., Kessner, D., Levander, F., Schofstatt, J., Tang, W. H., Römm, A., Neumann, S., Pizarro, A.D., Montecchi-Palazzi, L., Tasman, N., Coleman, M., Reisinger, F., Soupa, P., Hermjakob, H., Bintz, P.A., Deutsch, E.W. (2011) mzML—A community standard for mass spectrometry data. Mol. Cell. Proteomics 10, R110.00133

30. Kessner, D., Chambers, M., Burke, R., Agus, D., and Mallick, P. (2008) ProteoWizard: Open source software for rapid proteomics tools development. Bioinformatics 24, 2534–2536

31. Keller, A., Nesvizhskii, A. I., Kolker, E., and Aebersold, R. (2002) Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. Anal. Chem. 74, 5383–5392

32. Shteynberg D., Deutsch E. W., Lam H., Eng J. K., Sun Z., Tasman N., Mendoza L., Moritz R. L., Aebersold R., and Nesvizhskii A. (2011)
iProphet: Multi-level integrative analysis of shotgun proteomic data improves peptide and protein identification rates and error estimates. Mol. Cell. Proteomics 10, M111.007690

33. Lam, H., Deutsch, E. W., Eddes, J. S., Eng, J. K., Stein, S. E., and Aebersold, R. (2008) Building consensus spectral libraries for peptide identification in proteomics. Nat. Methods 5, 873–875

34. Pedrioli, P. G., Eng, J. K., Hubley, R., Vogelzang, M., Deutsch, E. W., Raught, B., Pratt, B., Nilsson, E., Angeletti, R. H., Apweiler, R., Cheung, K., Costello, C. E., Hermjakob, H., Huang, S., Julian, R. K., Kapp, E., McComb, M. E., Oliver, S. G., Omenn, G., Paton, N. W., Simpson, R., Smith, R., Taylor, C. F., Zhu, W., and Aebersold, R. (2004) A common open representation of mass spectrometry data and its application in a proteomics research environment. Nature Biotechnol. 22, 1459–1466

35. Grant, G. A., editor. (2002) Synthetic Peptides a User’s Guide, 2nd ed., pp. 229–231, Oxford University Press, Oxford UK

36. von Studnitz, W. (1965) Homoserine in urine in patients with neuroblastoma. Scand. J. Clin. Lab. Invest. 17, 558–564

37. Tsou, C. C., Avtonomov, D., Larsen, B., Tucholska, M., Choi, H., Gingras, A. C., and Nesvizhskii, A. I. (2015) DIA-Umpire: Comprehensive computational framework for data-independent acquisition proteomics. Nat. Methods 12, 258–264

38. Gallo, A. (2013) RNA editing enters the limelight in cancer. Nat. Med. 19, 130–1

39. Watson, I. R., Takahashi, K., Futreal, P. A., and Chin, L. (2013) Emerging patterns of somatic mutant ions in cancer. Nature Rev. Genetics 14, 703–718