Large Scale Mass Spectrometry-based Identifications of Enzyme-mediated Protein Methylation Are Subject to High False Discovery Rates*

Gene Hart-Smith‡§¶, Daniel Yagoub‡§, Aidan P. Tay‡, Russell Pickford¶, and Marc R. Wilkins‡

All large scale LC-MS/MS post-translational methylation site discovery experiments require methylpeptide spectrum matches (methyl-PSMs) to be identified at acceptably low false discovery rates (FDRs). To meet estimated methyl-PSM FDRs, methyl-PSM filtering criteria are often determined using the target-decoy approach. The efficacy of this methyl-PSM filtering approach has, however, yet to be thoroughly evaluated. Here, we conduct a systematic analysis of methyl-PSM FDRs across a range of sample preparation workflows (each differing in their exposure to the alcohols methanol and isopropyl alcohol) and mass spectrometric instrument platforms (each employing a different mode of MS/MS dissociation). Through \(^{13}\)CD\(_3\)-methionine labeling (heavy-methyl SILAC) of \(S\)accharomyces \(c\)erevisiae cells and in-depth manual data inspection, accurate lists of true positive methyl-PSMs were determined, allowing methyl-PSM FDRs to be compared with target-decoy approach-derived methyl-PSM FDR estimates. These results show that global FDR estimates produce extremely unreliable methyl-PSM filtering criteria; we demonstrate that this is an unavoidable consequence of the high number of amino acid combinations capable of producing peptide sequences that are isobaric to methylated peptides of a different sequence. Separate methyl-PSM FDR estimates were also found to be unreliable due to prevalent sources of false positive methyl-PSMs that produce high peptide identity score distributions. Incorrect methylation site localizations, peptides containing cysteinyl-S-\(\beta\)-propionamide, and methylated glutamic or aspartic acid residues can partially, but not wholly, account for these false positive methyl-PSMs. Together, these results indicate that the target-decoy approach is an unreliable means of estimating methyl-PSM FDRs and methyl-PSM filtering criteria. We suggest that orthogonal methylpeptide validation (\(e.g\). heavy-methyl SILAC or its offshoots) should be considered a prerequisite for obtaining high confidence methyl-PSMs in large scale LC-MS/MS methylation site discovery experiments and make recommendations on how to reduce methyl-PSM FDRs in samples not amenable to heavy isotope labeling. Data are available via ProteomeXchange with the data identifier PXD002857.

Molecular & Cellular Proteomics 15: 10.1074/mcp.M115.055384, 989–1006, 2016.

Post-translational methylation is a widespread protein modification, which predominantly occurs on lysine and arginine residues (1). Protein-lysine methyl/transferases catalyze the methylation of lysine residues; these enzymes facilitate the incorporation of methyl groups into the \(N\) atoms of lysine residues to produce either mono-, di-, or tri-methyllysine (MML,\(^1\) DML, and TML, respectively). Protein-arginine methyltransferases catalyze the methylation of arginine residues; these enzymes primarily act upon \(N\) atoms to produce mono, asymmetric di-, or symmetric di-methylarginine, although the enzyme-mediated modification of \(N\) atoms to produce \(\delta\)-MMA has also been reported in \(S\)accharomyces \(c\)erevisiae (2).

Traditionally, lysine and arginine methylation have been closely associated with histone proteins, and their crucial roles in modifying chromatin structure have been extensively studied (3). In recent years, however, a growing number of large scale methylation site discovery experiments have indi-

---

\(^1\) The abbreviations used are: MML, mono-methyllysine; DMA, di-methylarginine; DML, di-methyllysine; ETD, electron-transfer dissociation; FDR, false discovery rate; HCD, higher energy collision dissociation; heavy-methyl SILAC, heavy-methyl stable isotope labeling by amino acids in cell culture; HILIC, hydrophilic interaction liquid chromatography; MMA, mono-methylarginine; methyl-PSM, methylpeptide spectrum match; MS\(_n\), multiple-stage mass spectrometry; PSM, peptide spectrum match; AdoMet, S-adenosyl-L-methionine; sDMA, symmetric di-methylarginine; TML, tri-methyllysine; XIC, extracted ion chromatogram; CID, collision-induced dissociation.
cated that methylation is also widespread among non-histone proteins (4–16). These studies have associated methylation with a diverse range of cellular processes, including RNA processing, DNA repair and splicing, translation, helicase activity, ATPase activity, and spindle assembly checkpoints (4, 17, 18).

Liquid chromatography-tandem mass spectrometry (LC-MS/MS) has been at the core of these large scale methylation site discovery experiments. Specifically, these studies have made use of state-of-the-art mass spectrometric instrumentation (e.g. Thermo Scientific Q-Exactive, LTQ Orbitrap Elite, and Velos instruments), often in conjunction with novel methylpeptide enrichment techniques. Demonstrations of significantly enhanced methylation site discovery have, for example, been reported from samples generated via pan-specific antibody (8, 13) and methyl-lysine binding domain-based (19) pulldowns of methylpeptides, analyzed on Orbitrap Elite and Q Exactive or Orbitrap Velos instruments, respectively, and from samples enriched for arginine-methylated peptides using hydrophilic interaction liquid chromatography (HILIC), analyzed on a Q Exactive instrument (6). Together these contemporary instrument platforms and analytical workflows have enabled thousands of novel methylation sites to be identified from hundreds of proteins in the human proteome (4, 6, 7, 12, 13, 15), whilst large scale LC-MS/MS characterizations of methylation in other organisms (5, 8–11, 14, 16) have reinforced the notion that these modifications are widespread and sometimes conserved in eukaryotes (summarized in Table I).

In interpreting any LC-MS/MS-derived data for the purposes of methylation site discovery, there is a common requirement that must be met: methylpeptide spectrum matches (methyl-PSMs) must be identified at acceptably low false discovery rates (FDRs) following sequence database searching. The standard method of removing probable false positive peptide identifications involves performing searches against reversed or decoy databases to estimate FDRs (target-decoy approach) (20). Based on these estimates, peptide spectrum matches (PSMs) are then filtered to meet an estimated FDR threshold. When attempting to identify peptides of a particular subgroup, such as peptides containing a specific post-translational modification, the application of <1% FDR thresholds determined from global FDR estimates (i.e. FDR estimates made using all subgroup and non-subgroup PSMs) are often used to produce the final outputs for subgroup PSMs (21, 22). Recent studies have, however, indicated that obtaining separate estimates for subgroup FDRs may provide more appropriate subgroup score thresholds (23, 24).

Whether the target-decoy approach is applied globally or to peptide subgroups, it remains possible that searches against reversed or decoy databases may not provide accurate FDR estimates for methyl-PSMs. One proposed reason for this lies in the fact that the mass differences between numerous amino acids are identical to those observed for methylation (e.g. the mass differences between serine and threonine or leucine and valine). It is therefore feasible that misidentifications of methylpeptides can occur when peptides associated with single amino acid substitutions (or combinations of these substitutions) are subjected to MS/MS, or when an organism’s proteome otherwise produces related proteolytic peptides that differ in mass by the equivalent of a single methyl group (19, 25). Another potential reason for this relates to the fact that glutamic acid and aspartic acid residues have been shown to undergo esterification reactions in sample preparation protocols that feature methanol (26, 27) or ethanol (28). These reactions produce artifactual methylation or ethylation of these amino acid residues, which can be misidentified as enzyme-mediated mono- or di-methylation, respectively, on proximal arginine or lysine residues.

To account for such potential issues, orthogonal methylpeptide validation techniques—that is, independent forms of methylpeptide validation applied in conjunction to MS/MS and sequence database searches—can be of value. The most widely adopted orthogonal methylpeptide validation strategies involve isotopically labeling enzyme-mediated methylation sites; this is usually achieved through heavy methyl Stable Isotope Labeling by Amino Acids in Cell Culture (heavy-methyl SILAC) (25). Heavy-methyl SILAC involves growing cells in a medium containing $^{13}$CD$_3$-labeled methionine. As methionine is the precursor to S-adenosyl-L-methionine (AdoMet), the methyl group donor employed by all known methyltransferases, isotopically labeled methylpeptides are produced, which exhibit mass shifts that are diagnostic for the number of incorporated methyl groups. These mass shifts can aid in the validation of enzyme-mediated methylation sites identified from sequence database searches.

Despite the prospective issues associated with sequence database search-derived methylpeptide identifications, the use of orthogonal methylpeptide validation in large scale methylation site discovery studies remains sporadic. Although several such studies have employed heavy-methyl SILAC (6, 7, 14, 16) or other closely related methylpeptide-specific labeling techniques (12, 13) to validate methylation sites, others have chosen to bypass orthogonal validation and to instead predominantly rely on the target-decoy approach to provide estimates for high stringency methylpeptide filtering criteria (4, 5, 8, 9, 15) or to inform manual data curation (see Table I) (10, 11). This irregular use of orthogonal methylpeptide validation is a reflection of the fact that in-depth studies into methylpeptide FDRs have yet to be performed. Several studies have, however, indicated that for particular experimental workflows, methylpeptide FDRs can indeed be substantially higher than those estimated using the target-decoy approach (6, 13, 19). This suggests that systematic investigations into methylpeptide FDRs, the efficacy or the target-decoy approach, and likely sources of false positive methyl-PSMs are required.
Here, we provide the first systematic investigation of methylpeptide FDRs across a range of sample preparation workflows and mass spectrometric instrument platforms (see Fig. 1). Specifically, we investigate data obtained from whole cell lysates from a model organism, S. cerevisiae, grown in media containing either unlabeled or $^{13}$CD$_3$-labeled methionine; lysates were mixed and prepared for LC-MS/MS analysis using a variety of commonly employed sample preparation workflows, each differing in their use or non-use of the alcohols methanol and isopropyl alcohol. Samples were subjected to LC-MS/MS analysis using the following three mass spectrometric instrument platforms, each employing a different MS/MS dissociation method: LTQ Orbitrap Velos Pro (collision-induced dissociation (CID)), LTQ Orbitrap Velos Pro ETD (electron-transfer dissociation (ETD)), and Q Exactive Plus (higher energy collision dissociation (HCD)). By making use of the isotopic labeling of enzyme-mediated methylation, in-depth automated and manual inspections of LC-MS/MS data were performed to accurately determine true positive methyl-PSMs following sequence database searches. These lists of true positive methyl-PSMs were then used to accurately determine methylpeptide FDRs in datasets produced using traditional data filtering methods, such as target-decoy approach-based score thresholding, and to assess the validity of these data filtering methods for methylpeptides. (See under “Materials and Methods” for further details.) Together, these data provide new insights into methylpeptide FDRs, the efficacy of the target-decoy approach, sources of false positive methyl-PSMs, and the necessity of orthogonal methylpeptide validation in large scale LC-MS/MS analyses.

MATERIALS AND METHODS

Yeast Strains and Culture Conditions—For heavy-methyl SILAC, wild-type yeast (BY4741 strain, Open Biosystems) cells were cultivated in synthetic complete media: 2 g/liter histidine and methionine drop-out mix (D9537–10, US Biological), 1.7 g/liter yeast nitrogen base without amino acids or ammonium sulfate (BD Biosciences), 5 g/liter ammonium sulfate, 20 g/liter glucose, 82 mg/liter histidine, with 82 mg/liter unlabeled (light) or $^{13}$CD$_3$-labeled (heavy) methionine (299154, Sigma). Cells were harvested at an OD$_{600}$ of 0.7–1.0.

Sample Preparation—Three different workflows were used to prepare samples for LC-MS/MS analysis: in-solution digestion and HILIC; SDS-PAGE, Coomassie staining, and in-gel digestion; and SDS-PAGE (unstained) and in-gel digestion. In all samples, light peptides were used to identify PSMs and methyl-PSMs; heavy peptides were solely used to validate true positive methyl-PSMs.

For HILIC-separated samples, upon harvest, cells were washed three times in ice-cold PBS and resuspended in lysis buffer (50 mM HEPES, pH 7.5, 100 mM NaCl, 2 mM EDTA, 0.5% (v/v) Triton X-100) with protease inhibitors (11873580001, Roche Applied Science). Cells were disrupted, protein concentrations determined, and lysates mixed following the procedure described above. Gel electrophoresis was performed according to standard methods (17). Gels were fixed in 10% (v/v) acetic acid and 25% (v/v) isopropyl alcohol. For samples subjected to gel staining, Biosafe Coomassie G-250 (1:1–1.0% (v/v) methanol; Bio-Rad) was used. Gel lanes were excised into 28 slices according to protein mass, which were destained (when required), reduced, and alkylated following standard procedures (29). In-gel trypsic digestions and peptide extractions were performed following procedures described previously (30). Peptide extraction solutions were dried in a SpeedVac and reconstituted in 20 μl of 0.1% (v/v) formic acid.

Mass Spectrometry—For each proteolytic peptide sample, up to four technical replicate injections were subjected to LC-MS/MS analysis on each of the three mass spectrometric instrument platforms utilized in this study, i.e. the LTQ Orbitrap Velos, LTQ Orbitrap Velos Pro ETD, and Q Exactive Plus mass spectrometers (Thermo Scientific, Bremen, Germany). Each mass spectrometer was interfaced with an UltiMate 3000 HPLC and autosampler system (Dionex, Amsterdam, The Netherlands). Proteolytic peptides were separated by nano-LC, and eluting peptides were ionized using positive ion mode nano-ESI following experimental procedures described previously (31). For LTQ Orbitrap Pro analyses, survey scans m/z 350–1750 were acquired in the Orbitrap (resolution = 30,000 at m/z 400) with an initial accumulation target value of 1 × 10$^6$ ions in the linear ion trap; lock mass was applied to polycyclodimethylsiloxane background ions of exact m/z 445.1200 and 429.0887. The instrument was set to operate in data-dependent acquisition mode, and up to the 10 most abundant ions (>5000 counts) with charge states of +2 were sequentially isolated and fragmented via CID with an activation q = 0.25, an activation time of 30 ms, normalized collision energy of 30%, and at a target value of 10,000 ions. Dynamic exclusion was enabled (exclusion duration = 45 s), and fragment ions were mass analyzed in the linear ion trap.

LTQ Orbitrap Pro ETD analyses were performed as above, with the following exception: precursor ions were fragmented via ETD rather than CID, using parameters described previously (32). For Q Exactive Plus analyses, survey scans m/z 300–1750 (MS automatic gain control = 3 × 10$^5$) were recorded in the Orbitrap (resolution = 70,000 at m/z 200). The instrument was set to operate in data-dependent acquisition mode, and up to the 12 most abundant ions with charge states of +2 were sequentially isolated and fragmented via HCD using the following parameters: normalized energy 30, resolution = 17,500, maximum injection time = 125 ms, and MS$_{i}$, upon which trypsin (V5111, Promega) was added at a 100:1 ratio (w/w) and the digestion was carried out overnight at 37 °C. A C18 clean-up using a Sep-Pak column (WAT051910, Waters) was performed according to the manufacturer’s instructions. Eluted peptides were evaporated to dryness in a SpeedVac (Savant SPD1010, ThermoFisher Scientific), reconstituted in 0.1% (v/v) formic acid, 95% (v/v) acetonitrile (buffer A), and applied to a HILIC column (ZIC-HILIC PEEK HPLC column, 3.5-μm particle size, 150-mm length, 150447.0001, Merck). Following Uihmann et al. (6), peptides were eluted with a shallow gradient of 0.1% (v/v) formic acid (buffer B) up to 80% buffer B (v/v) at a flow rate of 300 μl/min. Fractions (600 μl) were collected every 2 min over a 50-min period. Fractions were then individually evaporated to dryness in a SpeedVac and reconstituted in 40 μl of 0.1% (v/v) formic acid for subsequent LC-MS/MS analysis.

For samples separated by SDS-PAGE, upon harvest, cells were washed three times in ice-cold PBS and resuspended in lysis buffer (50 mM HEPES, pH 7.5, 100 mM NaCl, 2 mM EDTA, 0.5% (v/v) Triton X-100) with protease inhibitors (11873580001, Roche Applied Science). Cells were disrupted, protein concentrations determined, and lysates mixed following the procedure described above. Gel electrophoresis was performed according to standard methods (17). Gels were fixed in 10% (v/v) acetic acid and 25% (v/v) isopropyl alcohol. For samples subjected to gel staining, Biosafe Coomassie G-250 (1:1–1.0% (v/v) methanol; Bio-Rad) was used. Gel lanes were excised into 28 slices according to protein mass, which were destained (when required), reduced, and alkylated following standard procedures (29). In-gel trypsic digestions and peptide extractions were performed following procedures described previously (30). Peptide extraction solutions were dried in a SpeedVac and reconstituted in 20 μl of 0.1% (v/v) formic acid.

Molecular & Cellular Proteomics 15.3

991
automatic gain control = $1 \times 10^5$. Dynamic exclusion was enabled (exclusion duration = 30 s).

**Sequence Database Searches**—Sequence database searches were performed using the Proteome Discoverer mass informatics platform (version 1.4, Thermo Scientific), using the search program Mascot (versions 2.3–5, Matrix Science). Peak lists derived from LC-MS/MS were searched using the following parameters: instrument type was ESI-TRAP for LTQ Orbitrap Velos Pro and Q Exactive Plus derived data and ETD-TRAP for LTQ Orbitrap Velos Pro ETD derived data; precursor ion and fragment mass tolerance ranges were set as the observed m/z values of the peptide features of interest from 2 to 5 ppm and ±0.4 Da, respectively, for LTQ Orbitrap Velos Pro and LTQ Orbitrap Velos Pro ETD derived data and ±5 ppm and ±0.02 Da, respectively, for Q Exactive Plus derived data; variable modifications included in each search were carbamidomethyl (Cys) and oxidation (Met); additional variable modifications included in separate searches were methyl (Lys), dimethyl (Lys), and trimethyl (Lys) or methyl (Arg) and dimethyl (Arg), or methyl (Asp/Glu), or ethyl (Asp/Glu) (i.e. ethylation of glutamic acid or aspartic acid, manually defined in Mascot as the following chemical addition: CO$_2$H$_2$CH$_2$COCH$_3$) or isopr (DE) (i.e. isopropylation of glutamic acid or aspartic acid, manually defined in Mascot as the following chemical addition: CO$_2$H to CO$_2$H$_2$CH$_2$COCH$_3$) or propionamide (Cys); enzyme specificity was trypsin with up to two missed cleavages; the Swiss-Prot database (July, 2013 release, 540,732 sequence entries) was searched using sequences and other selected methylpeptides of interest (China-Mass Peptides, Shanghai, China; see supplemental Table S1). Methyl-PSMs determined to have MS/MS spectra matching their synthetic counterpart were designated as true positives.

When following the target-decoy approach, sequence database searches against target and decoy databases were conducted upon batches of LC-MS/MS outputs associated with a given sample preparation workflow, instrument type, and technical replicate. When filtering datasets using global FDR estimates (i.e. FDR estimates obtained using all PSMS from target and decoy databases), Proteome Discoverer q-values were determined via the Percolator algorithm (35) for individual batches. PSMS with Proteome Discoverer q-value of $>0.01$ were removed to yield datasets with estimated global peptide FDRs of $<1%$.

**RESULTS**

To evaluate the validity of the target-decoy approach for filtering methyl-PSMs, for each of the nine sample preparation and mass spectrometric instrumentation combinations in Fig. 1, methyl-PSM FDRs were determined for peptide datasets produced via two methods of peptide confidence thresholding. Specifically, methyl-PSM FDRs were determined for datasets produced via Percolator filtering using global FDR estimates and datasets produced from Mascot Ion Score thresholding. These FDRs were calculated as shown in Equation 1,

$$\text{FDR} = \left(1 - \frac{TP}{P}\right) \times 100$$

where for datasets filtered to an estimated $<1\%$ global FDR via Percolator, $TP = \text{the number of remaining non-redundant true positive methyl-PSMs}$ (where redundant methyl-PSMs refer to methylpeptide identifications of identical amino acid sequence and modification state, regardless of charge state), and $P = \text{the number of remaining non-redundant methyl-PSMs}$.

For datasets filtered using Mascot Ion Score thresholds, $TP = \text{the number of true positive methyl-PSMs of Mascot}$.
Expect value < 0.05 above the applied Mascot Ion Score threshold, and $P$ = the number of methyl-PSMs of Mascot Expect value < 0.05 above the applied Mascot Ion Score threshold.

In obtaining correct TP values it was crucial to (i) accurately determine true positive methyl-PSMs and (ii) minimize false negatives in each dataset. Results pertaining to i and ii were derived from the workflow illustrated in Fig. 1B and are presented below. Methylpeptide FDRs, characterizations of sources of false positive methyl-PSMs, and evaluations of the efficacy of the target-decoy approach are then presented for each dataset.

**Depth of Proteome Coverage and True Positive Methylpeptide Spectrum Matches**—To effectively evaluate the efficacy of the target-decoy approach for filtering methyl-PSMs in large scale datasets, in addition to obtaining correct TP values, datasets featuring sufficiently deep proteome coverage and total numbers of true positive methyl-PSMs are required. Results pertaining to both the depth of proteome coverage and the confidence of heavy-methyl SILAC-validated true positive methyl-PSMs are described below.

The present data were collected from 368 LC-MS/MS experiments (95 Orbitrap Velos Pro, 131 Orbitrap Velos Pro ETD, and 142 Q Exactive Plus experiments). After filtering PSMs to an estimated <1% global FDR via Percolator, 576,152 total PSMs and 57,343 non-redundant PSMs were identified (excluding methylpeptides and non-*S. cerevisiae* contaminants). A total of 3459 *S. cerevisiae* proteins were observed when considering only proteins identified from ≥2 peptides.

From these LC-MS/MS data, 59 non-redundant true positive methyl-PSMs, associated with 34 distinct methylation sites on 13 methylated proteins, were identified (summarized in supplemental Tables SII and SIII). These true positive methyl-PSMs are all associated with arginine or lysine methylation (35 lysine and 24 arginine methylpeptides, 13 lysine and 21 arginine methylation sites, and 5 lysine and 8 arginine methylated proteins); no evidence for AdoMet-derived methylation of glutamic or aspartic acid residues was uncovered. All true positive di- or tri-methylation sites were observed on the internal or N-terminal lysine or arginine residues of tryptic peptides (as opposed to C-terminal residues), which is consistent with these modifications inhibiting tryptic cleavage. In addition, all true positive methyl-PSMs reported here identify either known *S. cerevisiae* methylation sites (14, 16, 30, 32, 34, 36–39) or previously unreported arginine methylation sites on known substrates of the protein-arginine methyltransferase HMT1 (34, 40), with the exception of one methylpeptide identifying Arg-60 di-methylation on the eukaryotic initiation factor 4F.
subunit p150. Evidence used to validate this methyl-PSM, specifically a high quality ETD MS/MS spectrum with neutral loss-derived product ions associated with arginine di-methylation and XICs supporting the presence of a co-eluting heavy labeled partner, are presented in supplemental Fig. S6. This methylpeptide identifies arginine methylation on a known protein-arginine methyltransferase substrate motif, RGG (6, 8, 18, 34), further supporting the designation of this match as a true positive. In addition, three methylpeptides designated as true positives identified two alternative forms of lysine methylation on known elongation factor 1-α methylation sites: di-methylation on Lys-390 (previously only characterized as MML) and tri-methylation on Lys-316 (previously only characterized as MML and DML) (30, 37). XIC and synthetic peptide-derived MS/MS data used to validate these methyl-PSMs are presented in supplemental Figs. S7 and 8.

False Negative Methylpeptide Spectrum Matches—To identify the presence of possible false negatives in the workflow illustrated in Fig. 1B, methylpeptides associated with previously reported S. cerevisiae methylation sites, specifically methylation sites annotated in Uniprot or additional literature sources (10, 14, 16, 17, 30, 32, 41–43), were identified in unfiltered sequence database search outputs and tracked. In the instances that these Uniprot/literature-derived methylpeptides were removed as candidate true positive matches, removal almost exclusively occurred during the Mascot Expect value or Proteome Discoverer q-value thresholding employed in step 1 of Fig. 1B, revealing poor fragmentation of methylpeptide precursor ions as the predominant source of false negatives in each dataset. However as these likely false negatives were also removed from the datasets to which Equation 1 is applied, they do not impact upon the FDRs calculated here.

Three methyl-PSMs were found to be exceptions to the above: KGGNIPMIPGWVMD*FPTGK (putatively derived from Asp-72 mono-methylation on hexokinase-2, as reported by Wang et al. (10)); KLIEAFNEIAEDSEQFDK* (putatively derived from Lys-412 mono-methylation on ATP-dependent molecular chaperone HSC82, as reported by Wang et al. (10)); and VIND*AFGIE*E*GLM^oxTVHSLATQK (putatively derived from Glu-169 and Glu-170 mono-methylation on glyceraldehyde-3-phosphate dehydrogenase 3, as reported from Sprung et al. (43), in addition to Asp-164 mono-methylation and Met-173 oxidation), where * denotes mono-methylation and M^ox denotes methionine oxidation. In this study, each of these three methyl-PSMs were identified with high Mascot Ion Scores across multiple datasets, but in each instance they were removed as candidate true positive matches during step 2 of Fig. 1B. Specifically co-eluting heavy labeled partners for these peptides were unambiguously absent, as illustrated in the XICs of supplemental Fig. S9. The MS/MS spectrum identifying VINDAFGIE*E*GLM^oxTVHSLATQK presented by Sprung et al. (43) closely matches the present MS/MS spectra identifying VIND*AFGIE*E*GLM^oxTVHSLATQK (data not shown), suggesting that both studies have identified peptides of the same sequence in different methylation states. The identity of the methylpeptide reported by Sprung et al. (43) was validated by the authors using a synthetic counterpart, but orthogonal validation via heavy isotope labeling of methylated residues was not attempted; the present data therefore suggest that these glyceraldehyde-3-phosphate dehydrogenase 3 mono-methylation sites are not AdoMet-derived. In the study described by Wang et al. (10), the specific XIC and MS/MS data used to validate the methylpeptides KGGNIPMIPGWMD*FPTGK and KLIEAFNEIAEDSEQFDK* were not reported, and thus comparisons between these previously described data and those of this study cannot be made. Altogether, the present data indicate that, in the samples analyzed here, these three methyl-PSMs cannot be considered AdoMet-derived, and their removal during step 2 of Fig. 1B does not point toward a significant source of false negatives during this step of the workflow.

Of all the Uniprot/literature-derived methylpeptides unable to be designated as true positives from the present data, only the three methylpeptides described above were observed among the lysine, arginine, and glutamic or aspartic acid methyl-PSMs remaining after Percolator filtering or the application of Mascot Expect value <0.05 thresholds (e.g. 771, 289, and 1458 total non-redundant lysine, arginine, and glutamic or aspartic acid methyl-PSMs, respectively, following Percolator filtering); i.e. no additional evidence for false negatives was observed when considering steps 3–5 of Fig. 1B. The Uniprot/literature-derived methylpeptides unable to be uncovered in this study were previously described from either enriched or overexpressed methylprotein samples (14, 16, 17, 41, 42), including samples of overexpressed poly(A)-binding protein, which featured methylated glutamic acid residues following Coomassie staining (42), or from sequence database search outputs derived from atypically broad search parameters (i.e. methyl-PSMs derived from ±50 ppm precursor ion mass tolerances for LTQ Orbitrap XL-derived data (10)); it is therefore unsurprising that these methylpeptides were not identified from the data described here.

Although the Uniprot/literature-derived methylpeptides tracked above identify the Mascot Expected value or Proteome Discoverer q-value thresholding employed during step 1 of Fig. 1B as the predominant source of false negatives, it remains feasible that false negatives associated with previously unknown methylation sites may have been produced during steps 2–5. However, care was taken to ensure that such false negatives were minimized. Specifically, if the filtering criteria imposed during any of steps 2–4 (i.e. identification of heavy-labeled partner peptides; manual interrogation of MS/MS spectra; and replicate analyses) were deemed to be ambiguous for any given methyl-PSM, the match was preserved for further analysis (see for example supplemental Fig. S3). Two ambiguous methyl-PSMs remained following step 4: LR*CEPAK (putatively derived from Arg-166 mono-methyla-
tion on meiotic activator RIM4) and QLRDAELK** (putatively derived from Lys-462 di-methylation on protein SEY1), where * denotes mono-methylation and ** denotes di-methylation. Both of these methyl-PSMs were ruled out as true positives during step 5, i.e. during synthetic peptide validation. The MS/MS and synthetic peptide-derived MS/MS data associated with these methyl-PSMs are presented in supplemental Figs. S10 and 11.

Methylpeptide False Discovery Rates—Fig. 2 shows results from peptide datasets filtered to estimated <1% FDRs using the global target-decoy approach. True and false positive methyl-PSMs, and corresponding methyl-PSM FDRs, are illustrated for S. cerevisiae cell lysate-derived datasets from nine sample preparation and MS instrumentation combinations. HCD, CID, and ETD data are derived from Q Exactive Plus, LTQ Orbitrap Velos Pro, and LTQ Orbitrap Velos Pro ETD instruments, respectively. Total non-redundant true and false positive arginine and lysine methyl-PSMs are presented as percentages of the total non-redundant PSMs observed in each dataset. Where relevant, error bars represent standard deviations observed across technical replicate LC-MS/MS experiments.

Strikingly, these data show that methyl-PSM FDRs substantially exceed the <1% FDRs estimated by the global target-decoy approach, with methyl-PSM FDRs typically exceeding 80% for each combination of sample preparation and MS instrumentation employed here. High methyl-PSM FDRs are observed for both lysates exposed and not exposed to alcohols during sample preparation. Moreover, these high methyl-PSM FDRs are observed across a range of MS instruments of different sensitivity to methylated and unmethylated peptides (see supplemental Table SIV for absolute numbers of
non-redundant methyl-PSMs and PSMs observed in each dataset; it is likely that these varying instrument sensitivities are influenced by both MS/MS dissociation methods and instrument duty cycles.

The datasets described in Fig. 2, and in subsequent figures, are derived from sequence database searches against *S. cerevisiae*-specific sequences. Datasets derived from searches against all taxonomies in the Swiss-Prot database produce qualitatively similar methyl-PSM FDR results (together with losses in true positive methyl-PSM sensitivity), indicating that these high methyl-PSM FDRs cannot be attributed to non-*S. cerevisiae* contaminants (see supplemental Table SIV for non-taxonomy-specific sequence database search-derived data). Together, the results described in Fig. 2 indicate that for the samples and MS data collection methods studied here, the global target-decoy approach produces dramatically unsuitable methyl-PSM filtering criteria.

Several recent investigations have drawn attention to the fact that when applying the target-decoy approach, global FDR estimates may not reflect those of specific peptide subgroups (23, 24, 44, 45). The results described in Fig. 2 indicate that methylpeptides represent one such peptide subgroup. For these peptide subgroups, separate FDR estimates may provide more suitable peptide filtering criteria to yield datasets of <1% FDR (24). Figs. 3 and 4 provide insights into the feasibility of applying separate methyl-PSM FDR estimates to produce high confidence methyl-PSM datasets.

Fig. 3 illustrates Mascot Ion Score distributions for true and false positive methyl-PSMs and associated methyl-PSM FDRs and true positive rates (sensitivities) across varying Mascot Ion Score thresholds for samples from unstained SDS-PAGE. Critically, these results indicate that even high identity score thresholds are incapable of reducing methyl-PSM FDRs when applied to the present HCD- and CID-derived methyl-PSM datasets. For the ETD-derived datasets, the high Mascot Ion Score thresholds (>80 for lysine methyl-PSMs and >59 for arginine methyl-PSMs) required to produce datasets with <10% methyl-PSM FDRs also result in extremely low methyl-PSM sensitivity (i.e. true positive rates of <1 and <17% for lysine and arginine methyl-PSMs, respectively). Samples produced from Coomassie-stained SDS-PAGE and HILIC display qualitatively similar results, as illustrated in supplemental Figs. S13 and S14. These results show that, for the datasets studied here, high quality outputs of methyltransferase-derived methyl-PSMs cannot be produced using identity score-based thresholding as a standalone method of data filtering, rendering obsolete any form of identity score-based filtering derived from target-decoy approach methyl-PSM FDR estimates.

These findings are reinforced by the results described in Fig. 4, which show the individual FDRs observed for MML, DML, TML, MMA, and DMA methyl-PSMs for the datasets of Fig. 3, alongside separate methyl-PSM FDR estimates for these methylpeptide subgroups. It can be seen that for each methylpeptide subgroup (i.e. mono-, di-, or tri-), high identity score thresholds are either incapable of reducing FDRs or produce substantial losses in overall methyl-PSM sensitivity (i.e. true positive methyl-PSM rates of 0–15%, as per Fig. 3) in the instances when FDRs can be reduced to <10%. Interestingly, these results also show that when employing the target-decoy approach, separate methyl-PSM FDR estimates substantially exceed global FDR estimates for all methylpeptide subgroups. This suggests that high methyl-PSM FDRs relative to unmodified PSM FDRs are an inherent aspect of sequence database searching. This is likely due to the high number of amino acid combinations capable of producing peptide sequences that are isobaric to methylated peptides of a different sequence; these findings are elaborated upon below. Nonetheless, for the datasets studied here, FDRs for each methylpeptide subgroup typically exceed even separate methylpeptide subgroup FDR estimates as score thresholds are increased; this is, for example, particularly pronounced in the lysine methyl-PSM dataset derived from Q Exactive Plus instrumentation (discussed below). Together, these results suggest that high methyl-PSM FDRs relative to unmodified PSM FDRs are an unavoidable consequence of sequence database searching and that methyl-PSM FDRs are further increased by false positive methyl-PSMs that are unable to be predicted by the target-decoy approach.

Characterized Sources of False Positive Methylenpeptide Identifications—The abovementioned results show that many more false positive methyl-PSMs are produced when conducting sequence database searches against target databases relative to decoy databases. This highlights the fact that false positive methyl-PSMs can be split into two categories: those that can be predicted by decoy database searches (e.g. false positives derived from unmodified peptides that are isobaric to methyl-PSMs but of different sequence), and those that cannot (e.g. false positives derived from peptides that are isobaric to methyl-PSMs with uncharacterized modifications). It is conceivable that separate methyl-PSM FDR estimates could prove accurate if these latter sources of false positive methyl-PSMs are characterized and removed from datasets prior to applying the target-decoy approach. This would allow separate methyl-PSM FDR estimates to be used to produce reliable methyl-PSM filtering thresholds.

To gain insight into the possible sources of these false positive methyl-PSMs, sequence database searches against *in vitro* modified peptides capable of producing false positive methyl-PSMs were first analyzed. Sequence database searches against the products of methyl and ethyl esterification reactions were specifically considered (26–28). As isopropyl alcohol was employed in the preparation of SDS-PAGE samples, sequence database searches for putatively isopropylated glutamic or aspartic acid residues were also considered. In addition, we note that cysteiny1-S-α-propionamide, the by-product of acrylamide adduct formation in
SDS-PAGE samples (46–48), produces a mass shift relative to unmodified cysteine (71.0371 Da) that is equivalent to the mass shift associated with cysteine alkylation plus mono-methylation on a proximal amino acid. Searches against peptides containing cysteinyl-S-propionamide were therefore also analyzed.

Fig. 5A illustrates, for peptide datasets filtered to estimated 1% FDRs using the global target-decoy approach, the relative proportions of non-redundant PSMs containing the abovementioned putative esterification and acrylamide adduct products for each employed sample preparation method and MS instrument platform. Relative proportions of non-redundant false positive mono-, di-, and tri-methylated methyl-PSMs are also illustrated. In interpreting these data, it must be noted that PSMs with the abovementioned glutamic and aspartic acid modifications are likely to have high FDRs relative to unmodified PSM FDRs (for the same reasons that the separate methyl-PSM FDR estimates described above are higher than global FDR estimates). It is therefore probable that a high percentage of the PSMs with methylated, ethylated, or

Molecular & Cellular Proteomics 15.3

997
isopropylated glutamic or aspartic acid residues shown in Fig. 5A are false positives.

The present results indicate that methyl esterification reactions are prevalent when samples are prepared via SDS-PAGE and Coomassie staining. The average proportions of PSMs with methylated glutamic or aspartic acid residues in datasets derived from Coomassie-stained SDS-PAGE samples are significantly higher than equivalent datasets derived from samples not exposed to methanol, *i.e.* HILIC and unstained SDS-PAGE samples (when comparing non-methanol-exposed datasets together against methanol-exposed datasets using two-tailed *t*-tests, *p* = 7.8 × 10⁻⁶, 1.7 × 10⁻³, and 3.2 × 10⁻³ for HCD, CID, and ETD-derived datasets respectively). In addition, the present results also confirm that *in vitro* isopropylation, the low proportions of PSMs with ethylated glutamic or aspartic acid residues shown in Fig. 5A, which are similar to the proportions of false positive di-methylated methyl-PSMs, provide no evidence to suggest that these reactions occur in the samples analyzed here. This is not surprising given that none of the employed sample preparation methods exposed cell lysates to ethanol. Regarding the possibility for *in vitro* isopropylation, the average proportions of PSMs with isopropylated glutamic or aspartic acid residues in datasets derived from SDS-PAGE samples (i.e. samples exposed to isopropyl alcohol) are higher than in equivalent datasets derived from HILIC samples (i.e. samples not exposed to isopropyl alcohol); however, these differences are not statistically significant. It is therefore unlikely that sizeable numbers of these *in vitro* modifications exist in the samples studied here.

Intriguingly, careful inspection of the datasets derived from HILIC and unstained SDS-PAGE samples (i.e. samples not exposed to methanol) suggests that not all of the PSMs with methylated glutamic and aspartic acid residues in these datasets are false positives. Numerous such PSMs are, for example, identified together with otherwise equivalent unmethylated PSMs; inspections of the MS/MS spectra associated with these unmethylated and putatively methylated PSM pairs...
frequently reveal closely matching spectra, differing only in product ion mass shifts consistent with methylation localized to the putatively modified glutamic or aspartic acid residue(s) (data not shown). The identity of one such PSM with a methylated aspartic acid residue, KQDFD*AAK (putatively derived from Asp-145 methylation on reduced viability upon starvation protein 161, where * denotes mono-methylation), which was identified from an unstained SDS-PAGE-derived sample via CID, was unambiguously confirmed using MS/MS data derived from a synthetic peptide counterpart (see supplemental Fig. S12). These results are consistent with data reported by Sprung et al. (43), who unambiguously identified glutamic and aspartic acid methylation from cell lysates that were not exposed to methanol during sample preparation. Together, these results indicate that peptides with non-enzyme-mediated methylation of glutamic or aspartic acid residues may act as possible sources of false positive lysine or arginine methyl-PSMs even in samples that have not been exposed to methanol.

The proportions of false positive lysine and arginine methyl-PSMs that can be explained by equal or higher scoring PSMs with alternative sites of arginine or lysine methylation, cysteinyl-S-, propionamide, or methylated glutamic or aspartic acid residues, shown for combined HCD-, CID-, and ETD-derived datasets. Remaining false positive methyl-PSMs are considered “uncharacterized.”

Fig. 5. False positive methyl-PSMs in S. cerevisiae cell lysate samples are produced from the following: the mis-assignment of methylation sites during sequence database searching; acrylamide adduct-derived products in SDS-PAGE samples; methyl esterification products in Coomassie-stained SDS-PAGE samples; and non-methyl esterification-derived PSMs with methylated glutamic or aspartic acid residues in all samples. A, total non-redundant false positive mono-, di-, and tri-methylated methyl-PSMs and modified PSMs capable of producing false positive methyl-PSMs are presented as percentages of the total non-redundant PSMs observed in each dataset. Peptide datasets are filtered to estimated <1% FDRs using the global target-decoy approach. Acrylamide adduct-derived products (**) are significantly more abundant in SDS-PAGE samples. Methyl esterification-derived products (***') are significantly more abundant in methanol-exposed samples. Non-methyl esterification-derived PSMs with methylated glutamic and aspartic acid residues (‡) are confirmed to exist. Where relevant, error bars represent standard deviations observed across technical replicate LC-MS/MS experiments. B, proportions of false positive methyl-PSMs derived from incorrect lysine or arginine site localizations are also given. These results confirm that cysteinyl-S- propionamide formation acts as a notable source of false positive methyl-PSMs in SDS-PAGE samples. These results also confirm that false positive methyl-PSMs derived from methyl esterification reactions are particularly pronounced in Coomassie-stained SDS-PAGE samples; however, as predicted, methylated glutamic or aspartic acid residues can also explain a substantial number of false positive methyl-PSMs in HILIC and unstained SDS-PAGE samples. The supplemental Figs. S15–20 illustrate, using multiple sequence alignment and iceLogo (49), the relative frequencies of amino acids proximal to false positive methylated lysine and arginine residues. These results reveal that, for all sample
preparation methods, glutamic acid residues are among the significantly \( p < 0.05 \) over-represented amino acids proximal to false positive methylated residues. This supports the hypothesis that methylated glutamic or aspartic acid residues, as shown for combined SDS-PAGE (Coomassie stained and unstained)- and HILIC-derived datasets. Remaining false positive methyl-PSMs are considered uncharacterized. Proportions of uncharacterized false positive methyl-PSMs predicted by the target-decoy approach (using separate methyl-PSM FDR estimates) are shown. B, average amino acid compositions of decoy lysine and arginine methyl-PSMs relative to respective high confidence unmethylated lysine- and arginine-containing PSMs (above), and the difference between the average amino acid compositions of decoy methyl-PSMs and uncharacterized false positive PSMs from target database searches (below). For each amino acid, numbers of mass differentials between single amino acids that match to mass differentials associated with mono-, di-, or tri-methylation (14.0157, 28.0314, and 42.0470 Da, respectively) are listed. Amino acids with no such mass differentials are under-represented in decoy methyl-PSMs (light gray boxes). ETD data were from an LTQ Orbitrap Velos Pro ETD instrument; CID data were from an LTQ Orbitrap Velos Pro instrument; HCD data were from a Q Exactive Plus instrument. All data are from PSMs of Mascot Except value of \(<0.05\).

Although the above analyses reveal some sources of false positive methyl-PSMs, methyl-PSM FDRs still exceed separate methyl-PSM FDR estimates after these characterized false positive methyl-PSMs are removed from peptide datasets (illustrated in supplemental Figs. S21–23). This indicates the existence of additional uncharacterized false positive methyl-PSMs, which are not predicted by decoy database searches. These remaining uncharacterized false positive methyl-PSMs are discussed below.

**Uncharacterized Sources of False Positive Methy1peptide Identifications**—To gain insight into the remaining uncharacterized false positive methyl-PSMs, ETD-, CID-, and HCD-derived datasets were analyzed separately. Fig. 6A illustrates the relative proportions of decoy database search-predicted and non-decoy database search-predicted false positive methyl-PSMs in these datasets. Fig. 6B illustrates the average amino acid compositions of decoy methyl-PSMs (above) and the differences between the average amino acid compositions of decoy methyl-PSMs and uncharacterized false positive PSMs from target database searches (below). In addition, Fig. 6B shows, for each listed amino acid, the numbers of mass differentials between single amino acids (including methylated lysine or arginine residues and oxidized methionine residues) that match the mass differentials associated with mono-, di-, or tri-methylation (25). For example glycine (exact mass = 75.032 Da) features two such mass differentials: its mass differential with alanine (exact mass = 89.048 Da) is 14.016 Da, which matches the mass differential associated with mono-methylation; and its mass differential with valine (exact mass = 117.078 Da) is 42.047 Da, which matches the mass differential associated with tri-methylation. It can be predicted that amino acids with fewer such mass differentials should be under-represented in false positive methyl-PSMs derived from misidentifications of isobaric unmethylated peptide sequences (i.e. false positive methyl-PSMs capable of being detected in decoy databases).

Inspection of Fig. 6 reveals two particularly noteworthy results. First, amino acids with zero of the abovementioned mass differentials are over-represented in decoy methyl-PSMs (light gray boxes) and under-represented in decoy methyl-PSMs (light gray boxes).
decoy methyl-PSMs (light gray boxes of Fig. 6B), indicating that the numbers of methyl-PSMs identified in decoy databases are influenced by the high number of amino acid combinations capable of producing peptide sequences isobaric to methylated peptides of a different sequence. This implies that the target-decoy approach should, in all sequence database searches, predict high methyl-PSM FDRs relative to global FDRs.

Second, Fig. 6A shows that, after removal of methyl-PSMs that can be explained by equal or higher scoring PSMS with alternative sites of arginine or lysine methylation, cysteiny1-S-β-propionamide, or methylated glutamic or aspartic acid residues, decoy database searches predict a large (but incomplete) proportion of the remaining false positive methyl-PSMs in ETD- and CID-derived datasets. In contrast, HCD-derived datasets contain proportionately fewer decoy database search-predicted false positive methyl-PSMs. These decoy database search predictions are corroborated by Fig. 6B, i.e. the average amino acid compositions of uncharacterized false positive methyl-PSMs and decoy methyl-PSMs closely match the average amino acid compositions of uncharacterized false database search-predicted false positive methyl-PSMs. These decoy database search predictions are corroborated by Fig. 6B, i.e.

The above findings reflect the fact that high mass accuracy HCD MS/MS spectra generate proportionately fewer spurious PSMS relative to the lower mass accuracy ETD and CID MS/MS spectra produced in this study, and thus fewer total decoy methyl-PSMs (data not shown). Interestingly, however, relative to ETD- and CID-derived datasets, HCD-derived datasets typically display high methyl-PSM FDRs as Mascot Ion Score thresholds are increased (see Figs. 3 and 4 and supplemental Figs. S13 and S14 and S21–S23). This stems from the fact that HCD produces relatively high Mascot Ion Score distributions for false positive methyl-PSMs. Given that high scoring HCD-derived PSMS are generally accurate (see for example Fig. 4D), the amino acid sequences of the remaining (non-decoy database search-predicted) uncharacterized false positive methyl-PSMs are therefore likely to closely match those of the isobaric peptides from which they are misidentified. This in turn suggests that, for all dissociation methods, false positive methyl-PSMs that are unable to be predicted by decoy database searches should be observed with high peptide identity score distributions.

The above finding implies that, for the samples analyzed here, attempts to reduce methyl-PSM FDRs using peptide identity score thresholds will be compromised even if the majority of (non-decoy database search-predicted) false positive methyl-PSMs can be characterized and removed from peptide datasets (see for example the ETD- and CID-derived datasets of supplemental Figs. 21–23). Any form of identity score-based filtering derived from target-decoy approach methyl-PSM FDR estimates will therefore remain problematic.

**ETD-derived Neutral Losses Can Improve the Accuracy of Arginine Methylpeptide Identifications**—The results described above point toward the necessity of validating methyl-PSMs using information not accessed by standard (target or decoy) sequence database searches. To date, most sequence database search algorithms have yet to incorporate methylation-specific neutral losses or product ions as a standard method of increasing the confidence of methyl-PSMs. A number of reports have, however, suggested that such information may be diagnostic for methylation (32, 34, 50–55). From the manual data curation undertaken for this study, we find that MMA- and DMA-associated neutral losses from charge-reduced precursor ions in ETD spectra aid in the differentiation of true and false positive arginine methyl-PSMs.

The ETD experiments conducted upon Coomassie-stained SDS-PAGE samples provide a case in point. These experiments identified a total of 59 non-redundant arginine methyl-PSMs of Proteome Discoverer q-value of <0.01; 36 with MMA (including four true positives), and 27 with DMA (including 11 true positives). Manual inspections of the highest scoring of these MMA methyl-PSMs reveal three spectra displaying evidence for losses of mono-methylamine (31.042 Da), i.e. an MMA methyl-PSM true positive rate of 75%, false negative rate of 25%, and FDR of 0% (illustrative spectra are shown in supplemental Fig. S24). In addition 13 spectra display evidence for neutral losses of mono-methylguanidine (73.064 Da), which have previously also been associated with MMA, i.e. an MMA true positive rate of only 25%, false negative rate of 75% and a FDR of 33%. These particular neutral losses therefore do not appear to be specific to or selective for MMA. Regarding the highest scoring DMA methyl-PSMs, manual inspections of these data reveal 19 spectra displaying evidence for losses of di-methylamine (45.058 Da), di-methylguanidine (87.087 Da), or di-methylcarbodiimide (70.053 Da), i.e. a DMA methyl-PSM true positive rate of 100%, false negative rate of 0%, and FDR of 30%.

Taken together, these neutral losses produce a methyl-PSM true positive rate of 93%, false negative rate of 7%, and FDR of 14% (when disregarding the nonspecific mono-methylguanidine neutral losses). This indicates that evidence for methylarginine-associated neutral losses in ETD spectra can increase the confidence of methyl-PSMs, and in particular MMA methyl-PSMs, either in collaboration with or independent of heavy-methyl SILAC validation.

**DISCUSSION**

**The Target-decoy Approach Produces Unreliable Estimates of Methylpeptide False Discovery Rates**—The target-decoy approach, applied either as a stand-alone technique or in conjunction with other peptide validation procedures (e.g. together with manual data curation or with orthogonal methyleptide validation), remains a highly popular method of filtering methyl-PSMs (see Table I). When applied as a stand-alone technique, the most common application of the approach involves obtaining methyl-PSM thresholding criteria based on global FDR estimates (4, 5, 8, 9); two recent studies have, however, made use of separate methyl-PSM FDR esti-
Arginine and lysine methylation are abbreviated metR and metK, respectively.

| Organism                        | Sample preparation                                      | MS instrument platform(s) | Methyl-PSM data filtering                        | No. of methylation sites reported | Ref. (year) |
|---------------------------------|---------------------------------------------------------|---------------------------|--------------------------------------------------|-----------------------------------|-------------|
| Human (Jurkat cells)            | Peptide separation (HILIC)                              | Q Exactive                | Heavy-methyl SILAC                                 | 249 metR                          | 6 (2012)    |
| Trypanosoma brucei              | Peptide separation (SCX)                                | LTQ Orbitrap ETD (alternating CID and ETD) | Thresholding to meet global target-decoy FDR estimates | 333 metR                          | 5 (2012)    |
| *T. brucei*                     | Peptide separation (SCX)                                | LTQ Orbitrap ETD (alternating CID and ETD) | Thresholding to meet global target-decoy FDR estimates | 1332 metR                         | 9 (2013)    |
| Human (HeLa cells)              | Antibody-based peptide immunoprecipitation and peptide separation (SCX) | LTQ Orbitrap XL           | Thresholding to meet global target-decoy FDR estimates | 552 metK                          | 4 (2013)    |
| Human (HeLa cells)              | Antibody-based protein immunoprecipitation and SDS-PAGE, SDS-PAGE, or peptide separation (IEP) | LTQ Orbitrap Velos        | Heavy-methyl SILAC                                 | 397 metR and metK                 | 7 (2013)    |
| Human (HeLa cells)              | Antibody-based peptide immunoprecipitation             | Orbitrap Elite and Q Exactive | Thresholding to meet global target-decoy FDR estimates | >1000 metR and ~160 metK          | 8 (2013)    |
| Arabidopsis thaliana            | SDS-PAGE<sup>b</sup>                                    | Q-TOF Ultima and LTQ-FT<sup>b</sup> | Separate target-decoy FDR estimates, in-depth manual inspection and curation of MS/MS data | 31 metR and metK                  | 11 (2014)   |
| Human (HEK 293T cells)          | Antibody-based peptide immunoprecipitation and peptide separation (SCX) | Q Exactive                | Thresholding to meet separate target-decoy FDR estimates | 1027 metR                         | 15 (2014)   |
| Human (T cells)                 | Antibody-based peptide immunoprecipitation             | Q Exactive                | Isomethionine-methyl SILAC                         | 2,502 metR                        | 13 (2015)   |
| Human (HeLa, K562, SW620, A649, SMM7721 cells) | Antibody-based derivatized peptide immunoprecipitation | Orbitrap Elite            |<sup>c</sup> <sup>13</sup>C<sub>D</sub><sub>3</sub>-methionine labeling with derivatized MML | 446 metK                           | 12 (2015)   |
| S. cerevisiae                   | In-solution digestion of cell lysates                  | LTQ Orbitrap XL           | Global target-decoy FDR estimates and arbitrary thresholding (heavy AdoMet labeling)<sup>c</sup> | 64 (various)<sup>d</sup>           | 10 (2015)   |
| S. cerevisiae                   | Antibody-based peptide immunoprecipitation             | Q Exactive and Orbitrap Elite (alternating CID and ETD) | Heavy-methyl SILAC                                 | 41 metR                           | 14 (2015)   |
| S. cerevisiae                   | Antibody-based peptide immunoprecipitation             | LTQ Orbitrap Velos Pro ETD | Heavy-methyl SILAC                                 | 21 metR                           | 16 (2015)   |

<sup>a</sup> Samples from <sup>13</sup>C<sub>D</sub><sub>3</sub>-methionine labeled cells were analyzed but not used to validate methyl-PSMs, i.e. analysis of heavy samples was performed separately to light samples.

<sup>b</sup> Data were obtained from the AT_CHOLO database [57].

<sup>c</sup> Isotopic labeling employed to identify putative heavy/light pairs in first parse filtering; in-depth validation using heavy/light peptide data not performed.

<sup>d</sup> Including redundant identifications, i.e. multiple methyl-PSMs associated with methylation on different amino acid residues identified from individual MS/MS spectra.
mates in their methyl-PSM filtering procedures (11, 15). The data described here indicate that global FDR estimates drastically differ from observed methyl-PSM FDRs and are therefore an unreliable method for obtaining appropriate methyl-PSM thresholding criteria. We also find that separate methyl-PSM FDR estimates, although potentially capable of producing appropriate methyl-PSM filtering thresholds, also dramatically differ from observed methyl-PSM FDRs.

In considering the global target-decoy approach, two main sources for its ineffectiveness can be identified. Foremost are the marked differences between global FDR and separate methyl-PSM FDR estimates, which indicate that high methyl-PSM FDRs relative to unmethylated PSM FDRs are a fundamental aspect of sequence database searching. This can be related to the high number of amino acid combinations capable of producing peptide sequences isobaric to methylated peptides of another sequence, as evidenced by the underrepresentation in decoy databases of the amino acids histidine, proline, phenylalanine, tryptophan, and tyrosine (i.e. the five amino acids without single amino acid mass differentials that correspond to methylation-related mass differentials). Crucially, these findings can be generalized to all experiments aiming to uncover methyl-PSMs from LC-MS/MS data and sequence database searches. For example, in the results reported here, although the differences between global FDR and separate methyl-PSM FDR estimates differ from dataset to dataset, they remain consistently and dramatically high across instrument-specific datasets of different size and across equivalently sized datasets derived from different MS/MS dissociation methods (see supplemental Table SIV).

False positive methyl-PSMs that are unable to be predicted by decoy databases further undermine the effectiveness of the global target-decoy approach in estimating methyl-PSM FDRs. The sources of these false positive methyl-PSMs are likely to be sample-specific and are discussed in relation to separate methyl-PSM FDR estimates below. In sum, the present findings strongly reject the global target-decoy approach as an effective means of estimating methyl-PSM filtering thresholds.

In considering separate methyl-PSM FDR estimates as a means of generating methyl-PSM filtering thresholds, the results reported here suggest that comprehensive characterizations of sources of false positive methyl-PSMs (beyond the misidentifications predicted by decoy database searches) are required for this approach to be effective. Without such characterizations, the mismatches between separate methyl-PSM FDR estimates and methyl-PSM FDRs can be pronounced, as observed in the present datasets. In practice, when analyzing an unknown peptide population, the comprehensiveness of such characterizations will be difficult to assess. This is because numerous potential sources of false positive methyl-PSMs exist; for example, the modified peptides considered in Fig. 5; peptides containing unannotated single amino acid substitutions (e.g. aspartic acid to glutamic acid substitutions); the existence of separate proteolytic peptides with mass differentials equivalent to those derived from methylation; peptides with as yet uncharacterized in vitro or in vivo modifications capable of being misidentified as methylated peptides of similar sequence; and via the mis-assignment of methylation sites (e.g. the mis-assignment of unconsidered N-terminal methylation or methylation on amino acids other than arginine and lysine). The difficulties in fully characterizing these sources of false positive methyl-PSMs are underscored by the finding that for each sample subjected to analysis in this study, increases in the depth of proteome coverage lead to concomitant increases in detections of uncharacterized sources of false positive methyl-PSMs, which in turn reduce the accuracies of separate methyl-PSM FDR estimates. Together, the findings reported here emphasize that the vast majority of the sources of false positive methyl-PSMs must be characterized and removed from datasets prior to applying the target-decoy approach. As it is unlikely that such criteria can be met with confidence for unknown peptide samples, the filtering of methyl-PSM datasets using thresholds determined from separate methyl-PSM FDR estimates should, in general, not be used as a stand-alone method of quality control.

Methylpeptide False Discovery Rates Can Be Expected to Remain High Following Methylpeptide Enrichment—The S. cerevisiae samples analyzed in this study contain low proportions of true positive methylpeptides (~0.3% of total PSMs). Samples enriched for arginine or lysine methylation, for example via antibody-based immunoprecipitations, can be expected to contain higher proportions of true positive methylpeptides than those reported here. It can also be expected that certain methylpeptide enrichment procedures should diminish (or fail to produce) potential sources of false positive methyl-PSMs (e.g. in-solution digests of antibody-based immunoprecipitations should not produce cysteinyl-S-β-propionamide-containing peptides). Together, these considerations suggest that, for samples prepared using methylpeptide enrichment strategies, the discrepancies between methyl-PSM FDRs and target-decoy estimated methyl-PSM FDRs may be lower than those observed in this study. The findings reported here nonetheless suggest that even when methylpeptide enrichment is performed, methyl-PSM filtering solely via the target-decoy approach will remain problematic. It is likely that global FDR estimates will remain substantially higher than separate methyl-PSM FDR estimates for the reasons described earlier; methyl-PSM filtering based on global FDR estimates therefore remains highly unreliable. In addition, the uncharacterized sources of false positive methyl-PSMs observed in this study appear, in large part, to be inherent to the S. cerevisiae proteome (as opposed to in vitro modifications), as they are ubiquitous across the three employed sample preparation methods. Thus for analyses aiming to maximize depth of sample coverage and total methyl-PSM detections following methylpeptide enrichment, it can be expected that inherently proteome-derived sources of false pos-
itive methyl-PSMs are likely to be detected in most cases, even if only in residual quantities. The following implication therefore still holds true for enriched methylpeptide samples: unless sources of false positive methyl-PSMs can be confidently and comprehensively characterized, separate methyl-PSM FDR estimates should be avoided as a means of determining methyl-PSM filtering thresholds.

These deductions are reinforced by the datasets generated by Uhlmann et al. (6) and Geoghegan et al. (13). Both studies aimed to enrich for arginine methylpeptides in human T cells; in addition, both employed heavy-methyl SILAC to validate methyl-PSMs, which allowed observed methyl-PSM FDRs to be compared with FDRs estimated using traditional methods. The samples generated by Uhlmann et al. (6) contained up to 4.52% arginine-methylated peptides, i.e. the arginine methylpeptide proportions were >100-fold higher than those reported here. After filtering their datasets to estimated <1% global FDRs using the target-decoy approach, these authors reported an observed methyl-PSM FDR of 67% (6). In the study described by Geoghegan et al. (13), a >500-fold enrichment of arginine methylpeptides relative to unenriched samples was described following antibody-based peptide immunoprecipitations. In the resultant dataset, methyl-PSM FDRs estimated at iProphet probabilities of 1.00 were 1 order of magnitude higher than observed methyl-PSM FDRs (13). These studies therefore strongly support the implications reported here; even after methylpeptide enrichment, it is likely that observed methyl-PSM FDRs will remain higher than methyl-PSM FDRs estimated using the target-decoy approach.

Conclusions and Recommendations—The present findings, derived from S. cerevisiae samples, describe consistently high methyl-PSM FDRs relative to the methyl-PSM FDRs estimated using the target-decoy approach. These specific FDRs are influenced by various factors; for example, the proportions of true positive methyl-PSMs observed, the employed MS/MS dissociation parameters, and potentially sources of false positive methyl-PSMs that are particular to the S. cerevisiae proteome. It can therefore be expected that LC-MS/MS datasets produced via different analytical workflows and from different organisms may produce dissimilar methyl-PSM FDRs. Nonetheless, these results point toward universal pitfalls in some of the traditional methods of filtering methyl-PSM data. Specifically when applying the target-decoy approach, global FDR estimates should be considered a highly unreliable means of estimating methyl-PSM FDRs, and separate methyl-PSM FDR estimates should be applied with a considerable degree of caution. Furthermore, even if reliable methyl-PSM filtering thresholds can be confidently determined using separate methyl-PSM FDR estimates, it can be expected that the thresholds required to produce low FDRs should generally result in sizeable losses in methyl-PSM sensitivity.

These findings suggest that to obtain reliable and sensitive methyl-PSMs in large scale LC-MS/MS methylation site discovery experiments, orthogonal methylpeptide validation should, in the vast majority of cases, be considered a prerequisite. Heavy-methyl SILAC, or any of its offshoots, is an obvious and versatile choice for such orthogonal methylpeptide validation. Specifically, if the retention times and peak areas of putative light and heavy methyl-PSM pairs can be reliably compared, the present results confirm that heavy-methyl SILAC can allow true and false positive methyl-PSMs to be accurately discriminated without losses in methyl-PSM sensitivity. Software has been designed to automate this process (e.g. MethylQuant (13, 14) and the in-house perl scripts described here); such software can be expected to be indispensable to future investigations. In addition, we find that one potential drawback of heavy-methyl SILAC, the misidentification of unmethylated methionine-containing peptides as methyl-PSMs with heavy labeled partners, is rare and should typically have near-negligible effects on methyl-PSM FDRs following careful heavy-methyl SILAC validation (see also the data presented by Geoghegan et al. (13)). Nevertheless, labeling strategies have been developed to bypass this issue entirely (i.e. iMethyl-SILAC (13) and MILS (10)).

For samples not derived from cell cultures (e.g. tissue samples for clinical investigations), the isotopic labeling of enzyme-mediated methylation is not yet possible. Alternative strategies to reduce methyl-PSM FDRs must therefore be adopted if large scale methylation site discovery experiments are to be undertaken. In this regard the propionylation of MML residues, as reported by Wu et al. (12), may prove to be particularly beneficial. This is because the limitations of the target-decoy approach identified in this study are directly related to the specific mass shifts imparted by methylation (and are therefore not relevant to other amino acid modifications of different mass). By propionylating MML in the manner reported by Wu et al. (12), and thereby altering the mass shifts associated with these modifications, the methylation-specific drawbacks of the target-decoy approach no longer apply to MML residues when they are identified in their derivatized form. With regard to the identification of arginine methylation, the results reported here suggest that FDRs can be reduced by removing methyl-PSMs lacking evidence for methylarginine-associated neutral losses in ETD spectra. For experiments in which the abovementioned strategies are not feasible, we recommend, at minimum, the following: (i) separate methyl-PSM FDR estimates should be employed when filtering datasets using the target-decoy approach; (ii) sources of false positive methyl-PSMs likely to be present in the samples of interest should be identified, and the peptides giving rise to these false positive methyl-PSMs should be characterized and removed from datasets; and (iii) tryptic methyl-PSMs with C-terminal di- or tri-methylation should also be removed from datasets. When interpreting datasets derived from these filtering criteria alone, we suggest that methylation sites of
particular interest should be independently validated; for example by comparing native peptide- and synthetic peptide-derived MS/MS spectra; through radiolabeling experiments employing putative methyltransferases, followed by in-depth LC-MS/MS analyses of purified putative methyltransferase substrates (17, 41).

The proteomics datasets described here have been deposited to the ProteomeXchange Consortium (56) via the PRIDE partner repository with the dataset identifier PXD002857.

Acknowledgments—We thank Dr. Ling Zhong, Sydney Liu Lau, and Associate Prof. Mark Raftery for their maintenance of the orbitrap mass spectrometers housed at the University of New South Wales Bioanalytical Mass Spectrometry Facility.

* This work was supported by the Australian Research Council (to G.H.-S. and M.R.W.) and University of New South Wales Early Career Researchers Grants Program (to G.H.-S.).
[S] This article contains supplemental materials.
§ These authors contributed equally to this work.
¶ To whom correspondence and reprint requests should be addressed: New South Wales Systems Biology Initiative, University of New South Wales, Sydney, New South Wales 2052, Australia. Tel.: 61-2-9385-3857; Fax: +61-2-9385-3950; E-mail: g.hart-smith@unsw.edu.au.

REFERENCES

1. Khoury, G. A., Baliban, R. C., and Floudas, C. A. (2011) Proteome-wide post-translational modification statistics: frequency analysis and curation of the Swiss-Prot database. Sci. Rep. 1, rep00090
2. Zobel-Thropp, P., Gary, J. D., and Clarke, S. (1998) N-Methylarginine is a novel posttranslational modification of arginine residues in yeast proteins. J. Biol. Chem. 273, 29283–29286
3. Lee, D. Y., Teyssier, C., Strahl, B. D., and Stallcup, M. R. (2005) Role of methylation in regulation of transcription. Endocr. Rev. 26, 147–170
4. Cao, X.-J., Arnaudo, A. M., and Garcia, B. A. (2013) Large scale global identification of protein lysine methylation in vivo. Epigenetics 8, 477–485
5. Fisk, J. C., Li, J., Wang, H., Aletta, J. M., Qu, J., and Read, L. K. (2013) Proteomic analysis reveals diverse classes of arginine methylproteins in mitochondria of trypanosomes. Mol. Cell. Proteomics 12, 302–311
6. Uhlimann, T., Geoghegan, V. L., Thomas, B., Ridiolva, G., Trudgian, D. C., and Acuto, O. (2012) A method for large scale identification of protein arginine methylation. Mol. Cell. Proteomics 11, 1489–1499
7. Bremaing, M., Cuomo, A., Agresta, A. M., Stugiewicz, M., Spadotto, V., and Bonaldi, T. (2013) Mass spectrometry-based identification and characterisation of lysine and arginine methylation in the human proteome. Mol. BioSyst. 9, 2231–2247
8. Guo, A., Gu, H., Zhou, J., Mulhem, D., Wang, Y., Lee, K. A., Yang, V., Aguilar, M., Kornhauer, J. A., Xia, X., Ren, J., Beausoleil, S. A., Silva, J. C., Vemulapalli, V., Bedford, M. T., and Comb, M. J. (2014) Immunofluorescence enrichment and mass spectrometry analysis of protein methylation. Mol. Cell. Proteomics 13, 372–387
9. Lott, K., Li, J., Fisk, J. C., Wang, H., Aletta, J. M., Qu, J., and Read, L. K. (2013) Global proteomic analysis in trypanosomes reveals unique proteins and conserved cellular processes impacted by arginine methylation. J. Proteomics 91, 210–225
10. Wang, K., Zhou, Y., Liu, H., Cheng, K., Mao, J., Wang, F., Liu, W., Ye, M., Zhao, Z. K., and Zou, H. (2015) Proteomic analysis of protein methylation in the yeast Saccharomyces cerevisiae. J. Proteomics 114, 226–233
11. Alban, C., Tardif, M., Mininno, M., Brugiere, S., Gilgen, A., Ma, S., Mazzoleni, M., Gigarel, O., Martin-Laffon, J., Ferro, M., and Ravanal, S. (2014) Uncovering the protein lysine and arginine methylation network in A. nidulans. PLoS One 9, e95512
12. Wu, Z., Zheng, Z., Sun, M., Wan, X., Liu, P., He, T., Tan, M., and Zhao, Y. (2015) A chemical proteomics approach for global analysis of lysine monomethyl profiling. Mol. Cell. Proteomics 14, 329–339
13. Geoghegan, V., Guo, A., Trudgian, D., Thomas, B., and Acuto, O. (2015) Comprehensive identification of arginine methylation in primary T cells reveals regulatory roles in cell signalling. Nat. Commun. 6, 6758
14. Plank, M., Fischer, R., Geoghegan, V., Charles, P. D., Konietzny, R., Acuto, O., Pears, C., Schofield, C. J., and Kessler, B. M. (2015) Expanding the yeast protein arginine methylome. Proteomics 15, 3232–3243
15. Sylvestersen, K. B., Horn, H., Jungmichel, S., Jensen, L. J., and Nielsen, M. L. (2014) Proteomic analysis of arginine methylation sites in human cells reveals dynamic regulation during transcriptional arrest. Mol. Cell. Proteomics 13, 2072–2088
16. Vaziri, D., Hart-Smith, G., Mocekling, J., Erce, M. A., and Wilkins, M. R. (2015) Yeast proteins Gar1p, Nop1p, Npl3p, Npr1p, and Rps2p are natively methylated and are substrates of the arginine methyltransferase Hmt1p. Proteomics 15, 3209–3218
17. Low, J. K., Hart-Smith, G., Erce, M. A., and Wilkins, M. R. (2013) Analysis of the Proteome of Saccharomyces cerevisiae for methylinagine. J. Proteome Res. 12, 3884–3899
18. Bedford, M. T., and Clark, S. (2009) Protein arginine methylation in mammals: who, what, and why. Mol. Cell. 33, 1–13
19. Carlson, S. M., Moore, K. E., Green, E. M., Martin, G. M., and Gozani, O. (2014) Proteome-wide enrichment of proteins by lysine methylation. Nat. Protoc. 9, 37–50
20. Elias, J. E., and Gygi, S. P. (2007) Target-decoy strategy for increased confidence in large scale protein identification by mass spectrometry. Nat. Methods 4, 207–214
21. Olsen, J. V., Vermeulen, M., Santamaria, A., Kumar, C., Miller, M. L., Jensen, L. J., Gnad, F., Cox, J., Jensen, T. S., Nigg, E. A., Brunak, S., and Mann, M. (2010) Quantitative phosphoproteomics reveals widespread full phosphorylation site occupancy during mitosis. Sci. Signal. 3, ra3-ra3
22. Lundby, A., Lage, K., Weinert, B. T., Bekker-Jensen, D. B., Secher, A., Skovgaard, T., Keirstup, C. D., Dmytryiev, A., Choudhary, C., Lundby, C., and Olsen, J. V. (2012) Proteomic analysis of lysine acetylation sites in rat tissues reveals organ specificity and subcellular patterns. Cell Rep. 2, 419–431
23. Marx, H., Lemeer, S., Schliep, J. E., Matheron, L., Mohammed, S., Cox, J., Mann, M., Heck, A. J., and Kuster, B. (2013) A large synthetic peptide and phosphopeptide reference library for mass spectrometry-based proteomics. Nat. Biotechnol. 31, 557–564
24. Fu, Y., and Qian, X. (2014) Transferred subgroup false discovery rate for rare post-translational modifications detected by mass spectrometry. Mol. Cell. Proteomics 13, 1359–1368
25. Ong, S. E., Mittler, G., and Mann, M. (2004) Identifying and quantifying rare post-translational modifications detected by mass spectrometry. Nat. Methods 1, 207–214
26. Jung, S. Y., Li, Y., Wang, Y., Chen, Y., Zhao, Y., and Qin, J. (2008) Complications in the assignment of 14 and 28 Da mass shift detected by mass spectrometry as in vivo methylation from endogenous proteins. Anal. Chem. 80, 1721–1729
27. Chen, G., Liu, H., Wang, X., and Li, Z. (2010) In vitro methylation by methanol: proteome screening and prevalence investigation. Anal. Chim. Acta 661, 67–75
28. Xing, G., Zhang, J., Chen, Y., and Zhao, Y. (2008) Identification of four novel types of in vitro protein modifications. J. Proteome Res. 7, 4603–4608
29. Shevchenko, A., Wilm, M., Vorm, O., and Mann, M. (1996) Mass spectrometric sequencing of proteins from silver-stained polyacrylamide gels. Anal. Chem. 68, 850–855
30. Hart-Smith, G., Chia, S. Z., Low, J. K., McKay, M. J., Molloy, M. P., and Wilkins, M. R. (2014) Stoichiometry of Saccharomyces cerevisiae lysine methylation: insights into non-histone protein lysine methyltransferase activity. J. Proteome Res. 13, 1744–1756
31. Hart-Smith, G., and Raftery, M. J. (2012) Detection and characterization of low abundance glycopeptides via higher-energy c-trap dissociation and orbitrap mass analysis. J. Am. Soc. Mass Spectrom. 23, 124–140
32. Hart-Smith, G., Low, J. K., Erce, M. A., and Wilkins, M. R. (2012) Enhanced methylinagine characterization by post-translational modification-specific targeted data acquisition and electron-transfer dissociation mass spectrometry. J. Am. Soc. Mass Spectrom. 23, 1376–1389
33. Cox, J., and Mann, M. (2008) MaxQuant enables high peptide identification...
rates, individualized ppb-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* 26, 1367–1372

34. Erce, M. A., Pang, C. N., Hart-Smith, G., and Wilkins, M. R. (2012) The methylproteome and the intracellular methylation network. *Proteomics* 12, 564–586

35. Käll, L., Canterbury, J. D., Weston, J., Noble, W. S., and MacCoss, M. J. (2007) Semi-supervised learning for peptide identification from shotgun proteomics datasets. *Nat. Methods* 4, 923–925

36. Couttas, T. A., Raftery, M. J., Padula, M. P., Herbert, B. R., and Wilkins, M. R. (2012) Methylation of translation-associated proteins in *Saccharomyces cerevisiae*: identification of methylated lysines and their methyltransferases. *Proteomics* 12, 960–972

37. Cavallius, J., Zoll, W., Chakraburtty, K., and Merrick, W. C. (1993) Characterization of yeast EF-1α: nonconservation of post-translational modifications. *Biochim. Biophys. Acta* 1163, 75–80

38. Itoh, T., and Wittmann-Liebold, B. (1978) The primary structure of protein 44 from the large subunit of yeast ribosomes. *FEBS Lett.* 96, 399–402

39. Webb, K. J., Laganowsky, A., Whitelegge, J. P., and Clarke, S. G. (2008) Identification of two SET domain proteins required for methylation of lysine residues in yeast ribosomal protein Rpl42ab. *J. Biol. Chem.* 283, 35561–35568

40. Low, J. K., and Wilkins, M. R. (2012) Protein arginine methylation in *Saccharomyces cerevisiae*. *FEBS J.* 279, 4423–4443

41. Erce, M. A., Abeygunawardena, D., Low, J. K., Hart-Smith, G., and Wilkins, M. R. (2013) Interactions affected by arginine methylation in the yeast protein-protein interaction network. *Mol. Cell. Proteomics* 12, 3184–3198

42. Low, J. K., Hart-Smith, G., Erce, M. A., and Wilkins, M. R. (2014) The *Saccharomyces cerevisiae* poly(A) binding protein is subject to multiple post-translational modifications, including the methylation of glutamic acid. *Biochim. Biophys. Res. Commun.* 443, 543–548

43. Sprung, R., Chen, Y., Zhang, K., Cheng, D., Zhang, T., Peng, J., and Zhao, Y. (2008) Identification and validation of eukaryotic aspartate and glutamate methylation in proteins. *J. Proteome Res.* 7, 1001–1006

44. Wilhelms, M., Schlegel, J., Hahne, H., Moghadass Gholami, A., Lieberenz, M., Savitski, M. M., Ziegler, E., Butzmann, L., Gessulat, S., Marx, H., Mathieson, T., Lemeer, S., Schnatbaum, K., Reimer, U., Wenschuh, H., Mollenhauer, M., Slotta-Huspenina, J., Bruley, T., and Kuster, B. (2014) Mass-spectrometry-based draft of the human proteome. *Nature* 509, 582–587

45. Fu, Y. (2012) Bayesian false discovery rates for post-translational modification proteomics. *Statistics Interface* 5, 47–59

46. Clauser, K. R., Hall, S. C., Smith, D. M., Webb, J. W., Andrews, L. E., Tran, H. M., Epstein, L. B., and Burlingame, A. L. (1995) Rapid mass spectrometry of peptide sequencing and mass matching for characterization of protein quantification. *Adv. Mass Spectrom.* 18, 786–787

47. Itoh, T., and Wittmann-Liebold, B. (1978) The primary structure of protein 44 from the large subunit of yeast ribosomes. *FEBS Lett.* 96, 399–402

48. Spring, R., Chen, Y., Zhang, K., Cheng, D., Zhang, T., Peng, J., and Zhao, Y. (2008) Identification and validation of eukaryotic aspartate and glutamate methylation in proteins. *J. Proteome Res.* 7, 1001–1006

49. Wilhelms, M., Schlegel, J., Hahne, H., Moghadass Gholami, A., Lieberenz, M., Savitski, M. M., Ziegler, E., Butzmann, L., Gessulat, S., Marx, H., Mathieson, T., Lemeer, S., Schnatbaum, K., Reimer, U., Wenschuh, H., Mollenhauer, M., Slotta-Huspenina, J., Bruley, T., and Kuster, B. (2014) Mass-spectrometry-based draft of the human proteome. *Nature* 509, 582–587