Computational approaches to identify sites of phosphorylation

Alex Joyce[1] and Brian Searle[1]

[1]The Ohio State University Medical Center

June 15, 2023

Abstract

Due to their oftentimes ambiguous nature, phosphopeptide positional isomers can present challenges in bottom-up mass spectrometry-based workflows as search engine scores alone are often not enough to confidently distinguish them. Additional scoring algorithms can remedy this by providing confidence metrics in addition to these search results, reducing ambiguity. Here we describe challenges to interpreting phosphoproteomics data and review several different approaches to determine sites of phosphorylation for both data-dependent and data-independent acquisition-based workflows. Finally, we discuss open questions regarding neutral losses, gas-phase rearrangement, and false localization rate estimation experienced by both types of acquisition workflows and best practices for managing ambiguity in phosphosite determination.

Computational approaches to identify sites of phosphorylation

Alex W. Joyce[1,2] and Brian C. Searle[1,2,3 *]

[1]Department of Biomedical Informatics, The Ohio State University Medical Center, Columbus, OH, 43210, USA.
[2]Pelotonia Institute for Immuno-Oncology, The Ohio State University Comprehensive Cancer Center, Columbus, OH, 43210, USA.
[3]Department of Chemistry and Biochemistry, The Ohio State University, Columbus, OH, 43210, USA.

*Corresponding Address: Brian C. Searle, Ph.D., brian.searle@osumc.edu

Abstract:

Due to their oftentimes ambiguous nature, phosphopeptide positional isomers can present challenges in bottom-up mass spectrometry-based workflows as search engine scores alone are often not enough to confidently distinguish them. Additional scoring algorithms can remedy this by providing confidence metrics in addition to these search results, reducing ambiguity. Here we describe challenges to interpreting phosphoproteomics data and review several different approaches to determine sites of phosphorylation for both data-dependent and data-independent acquisition-based workflows. Finally, we discuss open questions regarding neutral losses, gas-phase rearrangement, and false localization rate estimation experienced by both types of acquisition workflows and best practices for managing ambiguity in phosphosite determination.

Introduction:

Post-translational modifications (PTMs) are a significant regulatory mechanism of protein activity in cells and can operate on a much faster time scale than gene regulation. The addition of phosphates,[1] glycans,[2] methyls,[3] acetyls,[4] and even other proteins[5,6] can act as signals, activate or change the function of specific proteins, or mark proteins for degradation. Phosphorylation is a common form of cell signaling PTM, where cascades of kinases phosphorylating other kinases amplify signals from the cell surface to the nucleus. There are estimated to be over 500 unique kinases in the human genome, which phosphorylate over 200,000 unique phosphosites.[7] While phosphorylation has been observed naturally on many residues including aspartic acid,[8] arginine,[9] cysteine,[10] histidine,[11,12] the most commonly observed phosphorylation sites in proteomics experiments are serine (S), threonine (T), and tyrosine (Y).

Given the physiological significance of PTMs, it is important to reliably determine which proteins within a sample are modified. While PTMs change a protein's overall mass and charge, allowing for analysis with techniques like gel electrophoresis,[13] liquid chromatography (LC) coupled tandem mass spectrometry (MS/MS) is the preferred technique for monitoring most PTMs.[14]. There are two general approaches to analyzing PTMs with LC-MS/MS: top-down, which measures PTMs on intact proteins, and bottom-up (sometimes referred to as shotgun proteomics), which uses enzymatic digestion to process proteins into peptides before measurement. Top-down proteomics can identify the number of occupied phosphosites on a protein molecule, but it can sometimes be a challenge to determine the exact sites of phosphorylation without additional experiments.[15] In contrast, bottom-up proteomics simplifies identifying sites of phosphorylation by analyzing shorter peptide sequences at the cost of context in multiply-phosphorylated proteins.[16] In some experiments[17] these techniques are used in conjunction to take advantage of both methods.

Bottom-up proteomics methods for measuring phosphopeptides:

Figure 1 outlines a basic workflow of bottom-up proteomics for PTM analysis. As proteomics samples are processed, proteins are typically denatured and digested with an enzyme such as trypsin, which cuts each protein at regular sites. Due to the relatively low abundance of naturally occurring phosphopeptides, samples must first be enriched using techniques such as metal-affinity chromatography,[18,19] titanium dioxide,[20] strong cation exchange,[21] or antibodies.[22,23] to produce quantifiable amounts of phosphopeptides from the sample. Phosphopeptides are then separated using an LC, ionized using electrospray, and taken into the MS where both precursor and sequence-specific fragment ions are measured. From those measurements, data analysis helps identify peptides and assign specific sites of modification.

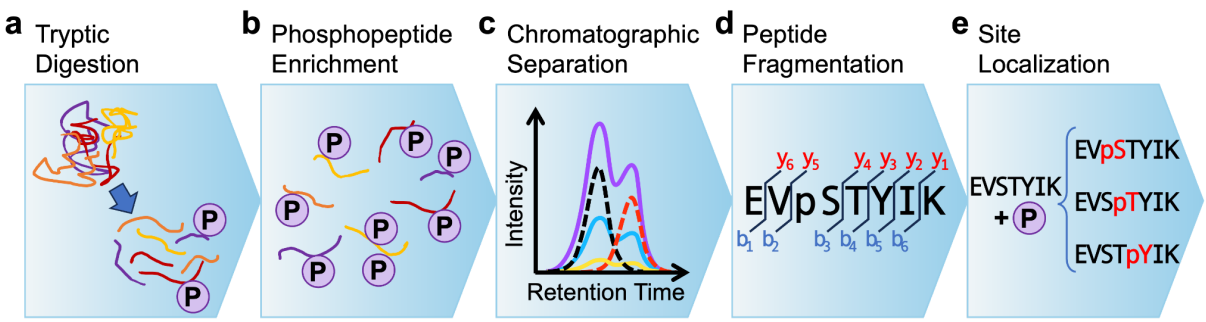

Figure 1: Outline of a standard bottom-up LC-MS/MS phosphoproteomics workflow. **(a)** Peptide samples are prepared by lysis, reduction, alkylation, and enzymatic digestion. **(b)** After digestion, phosphopeptides are enriched and prepared for LC-MS/MS measurement. **(c)** Peptides are chromatographically separated to reduce interference from competing signals and improve mass spectrum interpretation. **(d)** Eluting peptides are ionized and fragmented to produce sequence-level data. **(e)** Pattern recognition algorithms are used to identify phosphopeptides and fragmentation data is interpreted to identify the most likely site of phosphorylation in each peptide.

There are several different ways that a precursor ion can be ionized and fragmented, which have additional considerations when analyzing PTMs. While most current experiments use positive ion mode for ionizing phosphopeptides, negative ion mode has historically been a powerful technique for identifying phosphopeptides[24] and the combination of both positive and negative ion modes can improve phosphopeptide detection confidence.[25] Phosphopeptides are commonly analyzed using resonance collision-induced dissociation (CID)[26] or beam-type CID[27] (commonly referred to on Thermo instruments by the branded name, HCD[28]). CID methods produce b- and y-type ions, where beam-type CID undergoes multiple fragmentation events and typically produces long runs of y-ions while resonance CID results in the most energetically favored fragmentation pathways producing both b- and y-type ions with similar efficiency. For PTMs, both approaches of ion generation have advantages and disadvantages for assigning phosphopeptides.[29] Additionally, CID methods impart energy into peptides directly through collision,[30,31] which can cause the loss of PTMs during fragmentation as neutral losses,[32] and complicate their interpretation. An alternative fragmentation approach uses electron transfer dissociation (ETD)[33,34] or electron capture dissociation (ECD).[35] ETD and ECD use free radicals to make peptides unstable, fragmenting them along the backbone while typically leaving PTMs intact.[36] Tradeoffs between all fragmentation techniques suggest utilizing a combinational approach to fragmentation, which can help to better assign phosphopeptides.[37,38]

Once ions are fragmented, they can be analyzed in a variety of ways. Data-dependent acquisition (DDA)[39] uses precursor MS1 measurements to trigger selected (data-dependent) MS2s preferentially on the most abundant peptide signals. Instruments configured to collect DDA measurements typically use a top-N configuration, where the top-N most abundant precursor ions are selected for MS2 acquisition from the most recently collected MS1 spectrum. In general, this cycle is repeated every one to five seconds so that a new MS1 spectrum can indicate peptides that are newly eluting from the HPLC column into the mass spectrometer. In most DDA configurations, peptides elute over many cycles and a process called dynamic exclusion[40,41] is employed where recently measured precursor m/zs are put on an exclusion list for a specified duration such that precursors are not repeatedly re-measured in each cycle. As such, most peptides are identified by only a single MS2, allowing for more data acquisition time spent measuring low-abundance peptides. No matter what fragmentation technique is used, database search engines[42–47] such as Mascot[48] assign numerical scores to each potential peptide-spectrum match (PSM).[49] Search engines can be configured to search for variable PTMs,[50] which add the mass of modifications to user-specified residues.

Data-independent acquisition (DIA)[51,52] is an alternate acquisition method where MS2s are collected in a systematic manner regardless of precursor intensities. Systematic measurements from extracted fragment ion chromatograms can improve overall confidence in detecting specific phosphopeptides because each ion is measured multiple times over the elution profile of the peptide. However, precursor masses can be critical to assigning PTMs and peptide-centric DIA search engines[53] have to infer the specific precursor mass of a peptide from a wider precursor isolation window. As a result, care must be taken to either associate those fragment ions with precursors using tools like DIA-Umpire[54] or search for ions that are specific to the modified peptide. Targeted measurements, such as Parallel reaction monitoring (PRM),[55] can be thought of as a subtype of DIA where specific precursors are selected for fragmentation in targeted retention time windows. This approach marries the benefits of DIA with specific precursor information at the cost of measuring only a limited number of peptides per acquisition.

Challenges in interpreting phosphopeptide positional isomers:

The sheer number of kinases[56] and prevalence of serines, threonines, and tyrosines means that many proteins can be phosphorylated in multiple locations, where those sites are frequently at neighboring residues.[57] Modified peptides that differ only in which acceptor residue is phosphorylated are isomers because they have the same molecular formula, yet have different structures. As such, knowing the specific site of phosphorylation can be extremely informative, but these positional isomers can be very difficult to differentiate. For example, **Figure 2** shows major phosphosites of Human CDK1 (cyclin-dependent kinase 1), an important regulator of the cell cycle by controlling cell division.[58] The three most commonly-cited phosphosites (T14, Y15, and Y19) all fall within the same tryptic peptide: IGEGTYGVVYK, which has exact conservation across human, mouse, rat, chicken, and fruit flies, with nearly identical homology in yeast (***V***GEGTYGVVYK). Each of these residues serves the same inhibitory function but differs in what upstream kinases phosphorylate them. T14 and Y15 can both be phosphorylated by MYT1,[59] while Y15 can also be phosphorylated by PKC[60] and WEE1.[61] Little is understood about kinases that target Y19, but it has been observed that the SWE1 ortholog of WEE1 phosphorylates Y19 in *S. cerevisiae*.[62] Phosphorylation at each site produces the same CDK1 inhibition, yet observing each site indicates different upstream biology.

Due to the prevalence of positional isomers within the proteome, the use of standard search engine scores is often not enough to confidently identify isoforms.[63] That said, simply

considering the difference between the top two highest-scoring peptides can be an effective approach to localization. This concept was first considered in 1995 when Yates et al[50] used deltaCn, or the normalized difference between the best and second-best cross-correlation scores to identify phosphopeptides. The Mascot delta score (MD-score) approach[64] similarly considers the top two highest scores of a Mascot search, where the difference in Ions Scores relates to the separation between the quality of ions assigned to the first and second best peptides considered for that PSM. Since positional isomers have the same precursor mass and many of the same fragment ions, generally both the first and second-best peptides considered for a PSM are different positional isomers of the same peptide. As such, the MD-score frequently becomes a measure of how much better the top positional isomer scores above the other isomeric forms. Similarly, the SLIP score[65] for Protein Prospector[66,67] computes the difference between expectation values, analogous to probability values adjusted for the number of precursors considered within a given mass tolerance.

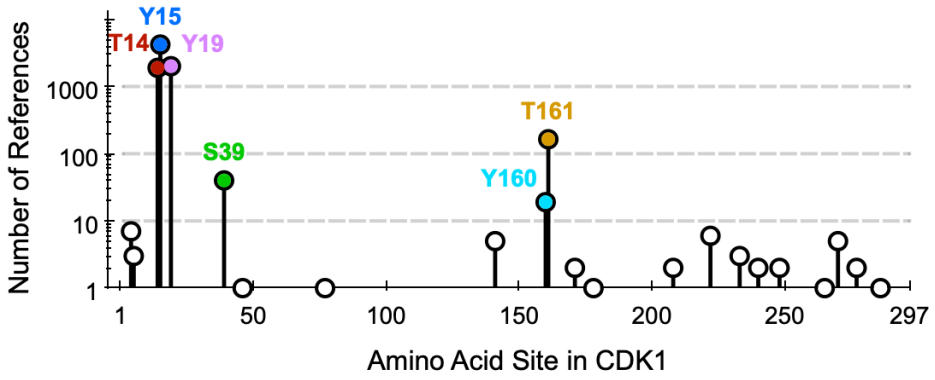

Figure 2: CDK phosphosites with at least 10 literature references. Highlighted phosphosites represent well-documented positional isomers. These positional isomers may serve similar functions and be located adjacent to one another, such as inhibition (T14, Y14, Y19), but can differ in what kinases can phosphorylate them. Phosphosite reference counts were derived from PhosphoSitePlus.[68]

While search engine scores consider the presence or absence of every potential sequence-specific fragment ion, only some ions change between positional isomers. **Figure 3** shows a spectrum from the Phosphopedia database[69] associated with each singly-phosphorylated form of the IGEGTYGVVYK peptide from CDK1 discussed previously in **Figure 2**. Notably, each positional isomer has the same precursor mass, but different fragmentation patterns. For example, y2 to y5 (as well as b6 to b10) fragment ions are mass shifted between the pY15 and pY19 isomers. These peaks are called site-determining ions and can be used to differentiate the isoforms of a specific peptide. In some cases, only a few site-determining ions are produced, such as only y6 and b5 differentiating between pT14 and pY15. Notably, the overall relative intensities of the fragmentation patterns,[30] as well as the LC retention times,[69] can differ between serine/threonine and tyrosine phosphorylated peptides, which can be used as an additional source of localization information. However, serine and threonine phosphorylated peptides generally have more similar relative fragmentation patterns and retention times due to the closeness of the chemistries of those amino acids.

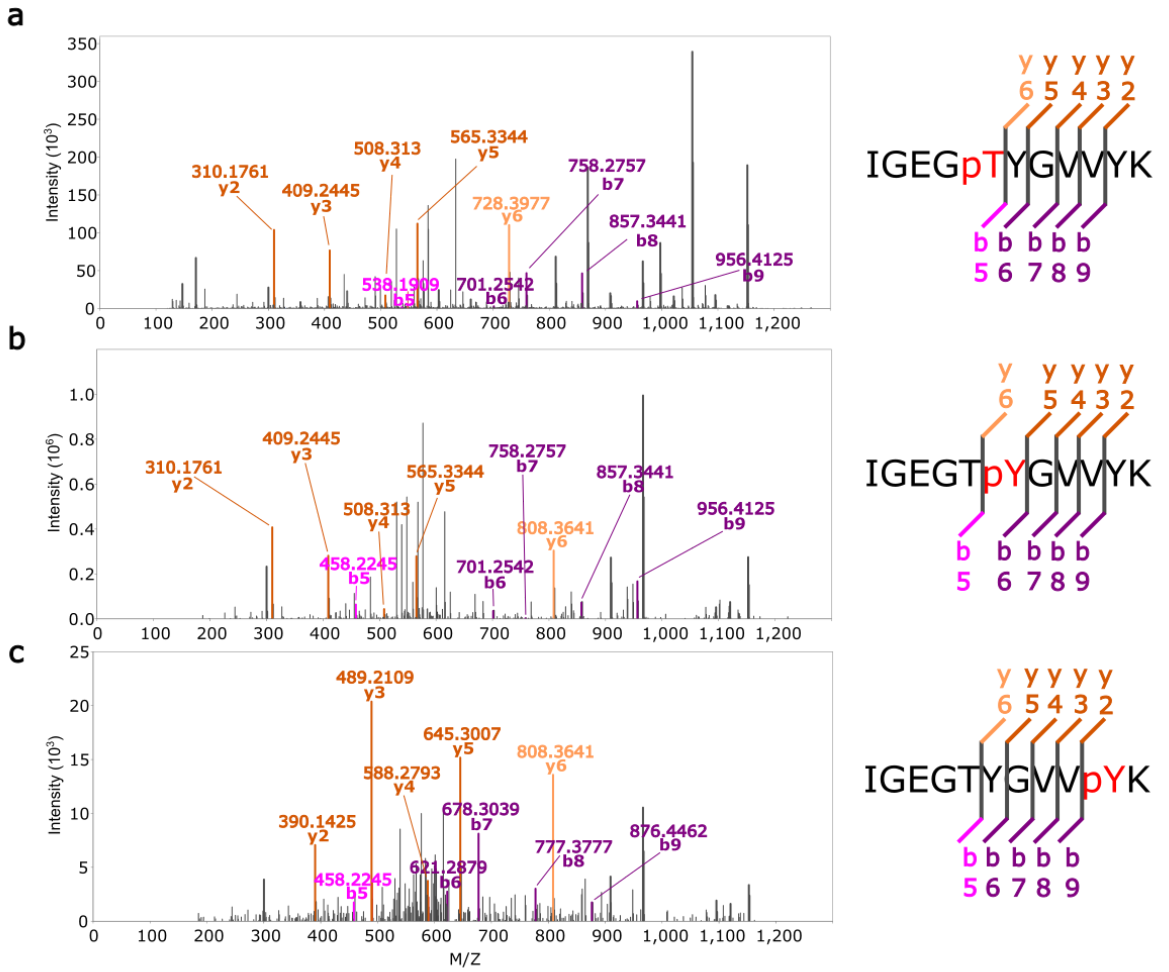

Figure 3: Resonance CID MS2 spectra from the CDK1 peptide IGEGTYGVVYK from the Phosphopedia library.[69] Each spectrum corresponds to one of three positional isomers. Peaks are highlighted according to their type (b-ions are colored purple while y-ions are colored orange) and what isomer they can be used to identify (lighter peaks identify pT14 while darker peaks identify pY19. **(a)** The pT14 isomer; b5 shows a -80 m/z downshift while y6 shows a +80 m/z upshift. (**b)** The pY15 isomer; note that this particular species does possess any unique site-determining ions and can only be identified in comparison to pT14 or pY19. **(c)** The pY19 isomer; b6 through b9 show -80 m/z downshifts while y2 through y6 show +80 m/z upshifts.

When a phosphopeptide has only two acceptor sites then each positional isomer has an equal number of unique ions that could distinguish that isomer. In peptides with more than two acceptor sites, the internal acceptor sites will not have any unique site-determining ions that can be used to definitively identify it. For instance, in the peptide IGEGTYGVVYK, the pT14 isomer can be distinguished from both pY15 and pY19 via the unique b5 and y6 fragment ions. Similarly, pY19 can be distinguished from both pT14 and pY15 by the isomer-specific b6-b9 and y2-y5 fragment ions. However, the pY15 isomer shares all of its ions with both pT14 and pY19, making it impossible to identify pY15 using site-determining ions alone. As such, most approaches to identifying the site of phosphorylation consider the delta between the best and second-best matches. In this case, the pY15 isomer can be localized compared to pT14 or

pY19, but not both simultaneously. If the delta between the top and second-best match is sufficiently large, that result would imply that any further result (third-best species and so forth) would not score better than the second-best match, allowing them to be discounted from consideration.

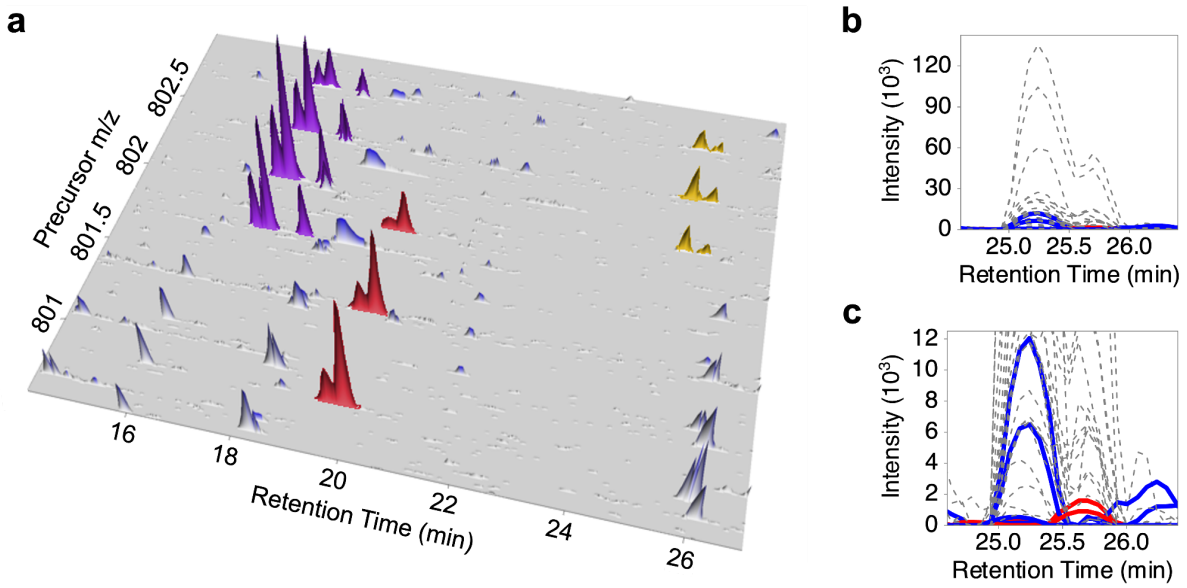

Figure 4: Positional isomers observed in a human phosphopeptide-enriched sample. **(a)** A three-dimensional precursor intensity landscape, where highlighted precursors are suggestive of positional isomers producing multiple peaks with the same m/z value and isotopic distribution, but at different retention times. The yellow peak corresponds to the singly phosphorylated IRS1 peptide SRTESITATSPASMVGGKPGSFR while the red and purple peaks are unassigned. If the retention time between these peaks is not sufficiently different enough, dynamic exclusion in DDA may cause only one isoform to be analyzed. **(b)** Extracted chromatograms of the peptide SRTESITAPASMVGGKPGSFR, where site-determining ions specific to the ambiguous SRTE(pSIT)ATSPASMVGGKPGSFR (e.g., either pS307 or pT309) are highlighted in red and site-determining ions belonging to SRTESITA(TpS)PASMVGGKPGSFR (e.g., either pT311 or pS312) are highlighted in blue. Ions belonging to both peptides are dashed gray lines. **(c)** A zoomed-in view of **(b)** showing lower-intensity peak shapes in greater detail.

The variety and scope of positional isomers in phosphopeptide-enriched datasets are most easily visualized by considering precursor-extracted ion chromatograms across retention time. **Figure 4a** presents a three-dimensional landscape of precursor intensities in a human phosphopeptide-enriched sample downloaded from the MassIVE repository (ProteomeXchange dataset PXD042974). By focusing on a narrow 800.5 to 803 m/z range from 15 to 26 minutes, this plot shows the isotopic structure of peptides in this range, indicating charge, as well as peptides with the same precursor mass and isotopic pattern, suggesting the same molecular formula. Three peptides have precursor mass patterns suggesting positional isomers, where the yellow peptide was identified as singly phosphorylated SRTESITATSPASMVGGKPGSFR from IRS1. Based on the precursor data, this peptide has two positional isomers. In **Figure 4b** and the corresponding blowup **4c**, the ambiguous isomer (pS307|pT309) is identified by the blue ions while the ambiguous isomer (pT311|pS312) is identified by the red ions, but pS307 and pT309 as well as pT311 and pS312 cannot be uniquely distinguished from each other due to

missing ions. In situations with localization ambiguity, existing literature can be a powerful tool[70] where pS307 and pS312 have been measured in 64 and 87 studies, respectively, from data in PhosphositePlus. Meanwhile, the other acceptor sites in SRTESITATSPASMVGGKPGSFR have only been observed in a handful of high-throughput studies. This result suggests the presence of IRS1 phosphorylation at S307 and S312, which are both phosphorylated by p70-S6 kinase.[71] The purple peptide (m/z=801.733, z=3) and red peptide (m/z=800.867, z=2) are unassigned but show similar precursor patterns.

While dynamic exclusion is a key tool for measuring low-abundance peptides with DDA, the approach can create complications when analyzing phosphopeptide positional isomers. Because positional isomers have the same molecular formula, they also have the same precursor mass. As such, if those positional isomers elute in close proximity to each other, the precursor mass from analyzing the first eluting positional isomer may still be on the instrument exclusion list during the time the second isomer elutes. In this case, the second positional isomer may never trigger an MS2 spectrum, even if its overall signal is higher than the first isomer.[72] Similarly, inherent stochasticity caused by the data-dependent nature of DDA may cause the first isomer to be missed if it is low in abundance, leaving room to collect an MS2 spectrum of the second isomer. Left unchecked, these characteristics of DDA cause a dramatic drop in the reproducibility of DDA-based phosphoproteomics experiments relative to DDA experiments of unmodified peptides.[73] Furthermore, positional isomers with the same precursors and nearby elution times can complicate precursor-based quantification and quantification approaches that use “match-between-runs”[74] where MS2-based detections in one injection can be transferred to other injections based on retention time and precursor mass matching alone. These factors underline the need for using isomer-specific fragment ions for detecting positional isomers.

Computational approaches to localize phosphates to specific sites in phosphopeptides:

Ascore[63] is the earliest method that utilizes site-determining ions to localize phosphate groups. The Ascore algorithm consists of three major steps. First, a set of fragment ions are generated for all known positional isomers from a given sequence. Second, these peaks are allocated into a series of 100 m/z wide bins and a binomial probability scorer is used to generate a Peptide Score for each isoform based on the number of matching peaks. The peak depth, or the ‘n’ most intense peaks within each 100 m/z window where n is an integer from 1 to 10 is determined by the largest delta Peptide Score for a given value of n. The last step is to recalculate the Peptide Score for the two highest scoring isomers, but considering only site-determining ions. In this scorer, the probability *P(x)* for PSM *x* is calculated as:

$$P(x) = \sum_{k=n}^{N} \binom{N}{k} p^{k} (1 - p)^{N-k} \qquad \text{(Eq. 1)}$$

where *N* is the number of possible ions that could differentiate between the two isomers and *n* is the number of those ions present in the spectrum. Assuming unit resolution, *p* is set to the peak depth divided by 100 (as the windows are 100 m/z wide). This formulation can be interpreted as the integration of the binomial distribution for the probability that the same or higher number of ions observed could be observed by random chance. Finally, the Peptide Score is set to $-10*\log_{10}(P(x))$ to ensure increasing numbers for more significant matches. The Ascore is the difference between these two site-specific Peptide Scores. This approach for scoring only site-determining ions for localization is demonstrably better than calculating the delta score between first and best-scoring peptides when considering all possible fragment ions, as demonstrated by Ascore’s precision and sensitivity improvements over both the MD-score and deltaCn.[63]

The fundamental Ascore algorithm using the binomial distribution for calculating localization probabilities has been implemented in several other contexts and tools.[75] For example, the Phosphate Localization score (PLscore)[76] simplifies the algorithm by utilizing a single peak density and extending the principles of Ascore to InsPecT[77] results. PhosphoRS[78] expands the idea of peak depth to allow for variable amounts of peaks within each window, thus accounting for an unequal distribution of peaks across m/z windows in a single spectrum. SLoMo (site localization of modifications)[79] is designed to allow for ETD/ECD data to be analyzed, while the Cscore (complementary score)[37] allows for both ETD/ECD and CID data to be utilized concurrently, increasing the confidence of peptide site localization. PhosCalc[80] can be used to analyze MS3 spectra[81] in addition to MS2 spectra, which can help assist in CID-based experiments where backbone fragmentation may be poor. Another method for utilizing MS3 in conjunction with MS2 spectra is PTM score used in MSQuant,[82] where localization is performed using the same approach as Ascore but limiting peak depth to four.

Other algorithms have been developed to score site-determining ions without assuming peaks are measured at random in a binomial distribution. For example, LuciPHOr[83] considers the likelihood odds ratio that a fragment peak would be assigned correctly or randomly and computes the cumulative sum of the log odds to assess each considered positional isomer, where the confidence of the reported isomer is the delta of the top two permutations. PTMProphet[84] uses a Bayesian approach to assess the confidence at each possible site using mixture models. This approach allows for ambiguous results to be returned when the exact site cannot be determined, but it can be limited to specific amino acids. PTMiner[85] is an approach for open-searching, where a database search engine is allowed to detect unanticipated modifications considering large mass tolerances. This strategy uses an empirical Bayesian-based approach to iteratively learn the prior probabilities for observing each type of modification on any given amino acid. Similar to the open-searching approach, pSite[86] uses a support vector machine to consider the confidence of each individual amino acid in *de novo*-derived sequences that have been assigned to proteins using sequence alignment.[87,88]

While Ascore and related algorithms primarily utilize database searching, the use of spectral libraries allows for incorporating sequence-specific features such as intensity values, neutral losses, and unusual fragments into a localization analysis process.[89] Recent methods allowed for the development of computationally predicted spectral libraries, where non-phosphorylated spectra are shifted by +80 Da to simulate phosphorylation.[90,91] First developed for Orbitraps, this method was expanded upon by Suni et al,[89,92] using a two-step process of enzymatic dephosphorylation followed by *in silico* rephosphorylation and has been extended to ToF platforms.[93] Dephosphorylated peptides share many of the same peaks as phosphorylated peptides, but are dramatically easier to detect. As such, library searching for “re-phosphorylated” forms of dephosphorylated peptides greatly improves phosphopeptide detection using SpectraST.[94] In this approach, localization is performed by recalculating the original deltaDot (delta score of the dot product) for the first and second-best matches that correspond to the same peptide sequence. A summary of these algorithms can be found in **Table 1**.

Phosphoproteomics analysis methods for data-independent acquisition:

While DDA typically attempts to fragment only a single peptide in each MS2 spectrum, DIA actively generates highly multiplexed MS2 spectra where multiple peptides are co-fragmented at the same time and extracted computationally.[95] Specter[96] is one approach to deconvolve this data using a non-negative least squares approach to find the simplest linear combination of library spectra that interpret a given window. Given complete data, this approach is able to distinguish highly similar library spectra including phosphopeptide positional isomers.

However, this approach uses intensity alone to identify positional isomers, with no consideration for site-determining ions. Time alignment is another method of grouping fragment ion signals. For example, PIQED[97] leverages the DIA-Umpire[54] tool to match time-extracted fragment ions with specific precursor ions in MS1 spectra to create pseudo-DDA spectra that can be searched using standard search engines. PIQED creates a pipeline that processes pseudo-DDA spectra through search engines, such as Comet,[98] using statistical analysis with the TPP,[99] including PTMProphet.[84]

Other approaches to analyzing phosphopeptides measured with DIA are built into DIA-specific processing tools. Inference of peptidoforms (IPF)[100] is a Bayesian hierarchical model integrated into the OpenSWATH[101] workflow that determines the most likely positional isomer (peptidoform) at a given retention time based on the elution pattern of site-determining ions. Thesaurus[72] is built into EncyclopeDIA[102] and uses an Ascore-like approach to model the probability of observing site-determining ions by chance. DIA MS2s are typically far more complex than DDA MS2s, and thus it is incorrect to assume the probability of observing an ion by chance is related to the number of peaks considered per m/z window. As a result, Thesaurus constructs a background distribution for each m/z window based on the frequency that any given m/z value is observed. Finally, site localization has been built into Spectronaut through directDIA[103,104] using a deconvolution approach similar to DIA-Umpire. A summary of these algorithms can be found in **Table 2**.

Estimating false localization rates across datasets:

Several experimental factors, such as data acquisition parameters, chromatography, instrument resolution, peak selection, and even the localization approach used, can cause results between algorithms to vary greatly.[105] For normal proteomics experiments, false discovery rate (FDR) estimation can control for false positives within a set of PSMs.[106,107] The target-decoy approach uses the score distribution of decoy peptides to model a null distribution from which to evaluate target peptide matches.[108] This principle works because the decoy-based null distribution closely approximates incorrect target matches. A similar concept for analyzing the accuracy of site localization algorithms is the false localization rate (FLR). However, unlike with FDR, global decoy measurements do not accurately reflect incorrect localizations.[109] One key difference is that incorrect localizations still score similarly to correct peptides with only a few site-determining ions mismatched. As a result, FLR estimations lack a standardized metric, resulting in arbitrary cutoffs being used to evaluate the performance of localization algorithms.[105]

Chalkley et al.[109] assert that the true FLR rate of a dataset can only be truly measured with prior knowledge of correct localizations, which may be modeled with synthetic peptides or singly-phosphorylated peptides. Most current methods of FLR estimation utilize biologically improbable residues to represent decoys, such as with PhosphoFLR,[110] which considers decoy localizations to alanine and leucine. Another approach, DeepFLR,[111] shuffles the phosphate group randomly within the peptide sequence and predicts spectral libraries for both the target and decoy datasets, bypassing the cost required to generate synthetic phosphopeptides. Other methods for FLR estimation are integrated into localization algorithms themselves. SLIP scoring[65] takes advantage of amino acid distribution and constructs decoys from glutamine and proline localizations, which are residues commonly observed near actual phosphosites. LuciPHOr[83] uses a Bayesian approach similar to PeptideProphet[112] to consider score distributions of S/T/Y localizations to other amino acids. PTMProphet[84] computes probabilities for each potential site of phosphorylation and calculates the localization mean best probability as a reasonable estimate for FLR.

Benchmarking studies can also be helpful in validating FLR estimates between algorithms. Locard-Paulet et al.[113] exhaustively evaluated the performance of 22 different proteomic pipelines using synthetic phosphopeptide data with known phosphosites. These pipelines were composed of search engines, localization algorithms, and validation strategies embedded within different phosphoproteomics software tools. The direct comparison of FLR used in this study serves as a useful resource for adapting the score thresholds between pipelines, helping to mitigate the variation in FLR between algorithms at the same threshold.

Open questions when analyzing phosphoproteomics datasets:

Regardless of the data acquisition method used, several key challenges still persist within the site localization process. The use of resonance or beam-type CID-derived fragmentation techniques produces b- and y-ions for site identification,[105] however, this technique creates additional dataset-complicating ions. These new ions include neutral losses and gas-phase re-arrangements, which may result in failure to detect or correctly localize the phosphosite.[114] Efforts to incorporate and adjust for these complications within scoring techniques likewise face similar challenges, particularly with regard to creating a standardized method of reporting false localization rates (FLRs).[109]

In CID-based mass spectrometry, the phosphate group of a phosphopeptide can often be lost in the form of phosphoric acid (H_3PO_4) in positive ion mode or metaphosphoric acid (HPO_3) in negative ion mode.[115,116] Serine and threonine residues undergo these neutral losses through beta elimination at a rapid rate due to the labile nature of the phosphate group,[117] while tyrosine is typically unable to undergo neutral losses. However, isobaric tag labeling with tandem mass tags (TMT) has been observed to phosphotyrosine neutral losses.[118] For example, in **Figure 3a**, the CDK1-pT14 isomer has several neutral loss peaks (y7-98 = 811.435 m/z, y8-98 = 868.456 m/z, y9-98 = 997.499 m/z, and y10-98 = 1054.520 m/z) that are absent in the pY15 and pY19 spectra. These losses can undercut localization by shifting the mass of modified residues since the beta elimination neutral loss of phosphoric acid from phosphorylated serine and phosphorylated threonine produces the same fragment mass as beta elimination of water from an unmodified serine or threonine.[105] Some site-localization algorithms still attempt to use phosphate neutral loss ions as site-determining ions, but this has been demonstrated to increase localization error rates.[89] Library searching may provide an avenue to consider phosphoric acid neutral losses by leveraging differences in the observed intensity for each ion type.

During the fragmentation process of CID-based experiments, the same gas phase beta-elimination reactions that induce neutral losses can result in the translocation of the phosphate group to a nearby acceptor site.[114,119] These gas-phase rearrangements result in peptides or ions that appear as alternate positional isomers but at the elution time of the original isomer. While these gas-phase products are typically low in abundance, it has been estimated that as many as 37% of positional isomers that appear to be identically co-eluting are actually side-products formed in the gas-phase,[72] matching estimates from Plaumbo and Reid[114] showing that 36% of tested phosphopeptides underwent some degree of site rearrangement. Plaumbo and Reid suggest that CID-based proteomic data should be re-evaluated before making conclusions regarding biological significance, especially if localization software considers the possibility of multiple positional isomers from the same MS2 spectrum. In these cases, positional isomers can be validated by considering known kinase motifs, comparisons to similar data produced using ETD or ECD fragmentation, or by comparison to CID-MS/MS spectra of synthetic phosphopeptide standards.[114] While gas-phase rearrangement is reproducibly common across resonance and beam-type CID instruments, Aguiar et al[120] argue that from a practical perspective, it may have a limited impact on phosphopeptide localization.

This study considered two combinatorial libraries of singly phosphorylated peptides with two acceptor sites and found that rearrangement products were rarely observed to decrease localization accuracy in +2H MS2 spectra. The authors further argue that using ETD, which does not produce rearrangement products, did not dramatically affect localization confidence in Ascore. While combinatorial libraries contain significantly more phosphopeptides than the pool of phosphopeptides studied by Palumbo and Reid, the limited combinations of chemistry retrain the practical conclusions that can be drawn from this result, leaving the real effects of gas-phase rearrangement on proteomics studies still somewhat unresolved.

Conclusion:

The field of phosphoproteomics has enjoyed rapid growth over the last twenty years as proteomics instrumentation and methodologies have improved. While mass spectrometry is a powerful method for detecting phosphopeptides, challenges remain regarding determining the exact sites of phosphorylation, which are needed to improve our understanding of the biological context surrounding these modifications. New approaches to making mass spectrometry measurements with data-independent acquisition are improving our ability to detect and quantify phosphosites. Despite these advances, neutral losses, gas-phase rearrangements, and lack of standard FLR metrics remain open challenges in both DDA-based and DIA-based proteomic workflows. Representing ambiguity in incomplete phosphosite localization, both in the literature as well as in databases and data standard formats such as mzIdentML,[121,122] will be critical to our ability to interpret and reuse results from phosphoproteomics studies.

Acknowledgments:
This research was supported by NIH R01GM133981 and NIH R21CA267394 to BCS.

Competing interests:
The authors declare the following competing interests: BCS is a founder and shareholder in Proteome Software, which operates in the field of proteomics.

Table 1: DDA-Based Localization Algorithms

Name	Category	Summary Description	Reference
Ascore	Binomial	Uses sequence information to predict fragmentation patterns; looks for the presence or absence of site-determining ions.	63
Cscore	Binomial	Implementation of Ascore that utilizes both ETD and CID data concurrently to increase localization confidence.	37
MD-Score	Delta Score	Difference between the first and second best results from the MASCOT search engine for positional isomers.	64
LuciPHOr	Probabilistic	Computes the likelihood odds that a given peak is matched at random before calculating a cumulative sum of the log odds to represent the entire peptide. The score itself is the delta of the two most likely isoforms.	83
PhosCalc	Binomial	Implementation of Ascore that can be applied to both MS2 and MS3 spectra.	80
PhosphoRS	Binomial	Implementation of Ascore that uses variable peak densities per 100 m/z window to account for unequal peak distribution in a spectrum.	78
Phosphorylation Localization Score (PLS)	Binomial	Extends Ascore method to InsPecT; uses a single peak density to simplify the scoring process.	76
PTMiner	Binomial	Designed for high mass-tolerance open searching; uses an empirical Bayesian-based iterative learning strategy for site localization.	85
PTMProphet	Bayesian	Uses a Bayesian mixture model to evaluate confidence at each possible site.	84
PTM score (MSQuant)	Binomial	Implementation of Ascore that allows for the analysis of MS3 spectra along with MS2 spectra; only considers the top 4 most intense peaks per 100 m/z window.	82
pSite	Bayesian	Site-based approach that first evaluates individual site confidence using support vector machines for de novo derived sequences assigned through sequence alignment.	86
SLIP	Delta Score	Designed to work with Protein Prospector using expectation values. The SLIP score is the difference between the first and second-best expectation values.	65

SLoMo	Binomial	Implementation of Ascore that allows ETD and ECD data to be used.	79
Suni et al.	Library Matching	Constructs a spectral library by applying mass shifts to enzymatically dephosphorylated peptide spectra; localization scores are recalculated deltaDot scores from SpectraST using these libraries.	89

Table 2: DIA-Based Methods

Name	Category	Summary Description	Reference
Inference of Peptidoforms (IPF)	Bayesian	Applies a Bayesian hierarchical model to spectral library data to determine the most likely positional isomer or 'peptidoform' at a given retention time based on elution pattern.	100
PIQED	Workflow	Sequences DIA-Umpire to deconvolve DIA data through the creation of pseudo-DDA spectra that can be localized through TPP-compatible methods like PTMProphet.	97
Spectronaut	Closed Source	Allows for analysis of DIA without libraries through directDIA using deconvolution methods similar to DIA-Umpire.	104
Thesaurus	Probabilistic	Creates a background distribution using differential frequencies of peptides in a dataset and then calculates probabilities based on whether site-determining ions can be matched in that background distribution by random chance.	72

References:

(1) Mann, M.; Ong, S. E.; Grønborg, M.; Steen, H.; Jensen, O. N.; Pandey, A. Analysis of Protein Phosphorylation Using Mass Spectrometry: Deciphering the Phosphoproteome. *Trends Biotechnol.* **2002**, *20* (6), 261–268.

(2) Hart, G. W.; Housley, M. P.; Slawson, C. Cycling of O-Linked Beta-N-Acetylglucosamine on Nucleocytoplasmic Proteins. *Nature* **2007**, *446* (7139), 1017–1022.

(3) Rathert, P.; Dhayalan, A.; Ma, H.; Jeltsch, A. Specificity of Protein Lysine Methyltransferases and Methods for Detection of Lysine Methylation of Non-Histone Proteins. *Mol. Biosyst.* **2008**, *4* (12), 1186–1190.

(4) Zhao, S.; Xu, W.; Jiang, W.; Yu, W.; Lin, Y.; Zhang, T.; Yao, J.; Zhou, L.; Zeng, Y.; Li, H.; Li, Y.; Shi, J.; An, W.; Hancock, S. M.; He, F.; Qin, L.; Chin, J.; Yang, P.; Chen, X.; Lei, Q.; Xiong, Y.; Guan, K.-L. Regulation of Cellular Metabolism by Protein Lysine Acetylation. *Science* **2010**, *327* (5968), 1000–1004.

(5) Kirkpatrick, D. S.; Denison, C.; Gygi, S. P. Weighing in on Ubiquitin: The Expanding Role of Mass-Spectrometry-Based Proteomics. *Nat. Cell Biol.* **2005**, *7* (8), 750–757.

(6) Hendriks, I. A.; Vertegaal, A. C. O. A Comprehensive Compilation of SUMO Proteomics. *Nat. Rev. Mol. Cell Biol.* **2016**, *17* (9), 581–595.

(7) Cesaro, L.; Pinna, L. A. Prevalence and Significance of the Commonest Phosphorylated Motifs in the Human Proteome: A Global Analysis. *Cell. Mol. Life Sci.* **2020**, *77* (24), 5281–5298.

(8) Collet, J. F.; Stroobant, V.; Pirard, M.; Delpierre, G.; Van Schaftingen, E. A New Class of Phosphotransferases Phosphorylated on an Aspartate Residue in an Amino-Terminal DXDX(T/V) Motif. *J. Biol. Chem.* **1998**, *273* (23), 14107–14112.

(9) Fuhrmann, J.; Schmidt, A.; Spiess, S.; Lehner, A.; Turgay, K.; Mechtler, K.; Charpentier, E.; Clausen, T. McsB Is a Protein Arginine Kinase That Phosphorylates and Inhibits the Heat-Shock Regulator CtsR. *Science* **2009**, *324* (5932), 1323–1327.

(10) Sun, F.; Ding, Y.; Ji, Q.; Liang, Z.; Deng, X.; Wong, C. C. L.; Yi, C.; Zhang, L.; Xie, S.; Alvarez, S.; Hicks, L. M.; Luo, C.; Jiang, H.; Lan, L.; He, C. Protein Cysteine Phosphorylation of SarA/MgrA Family Transcriptional Regulators Mediates Bacterial Virulence and Antibiotic Resistance. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109* (38), 15461–15466.

(11) Fuhs, S. R.; Hunter, T. pHisphorylation: The Emergence of Histidine Phosphorylation as a Reversible Regulatory Modification. *Curr. Opin. Cell Biol.* **2017**, *45*, 8–16.

(12) Leijten, N. M.; Heck, A. J. R.; Lemeer, S. Histidine Phosphorylation in Human Cells; a Needle or Phantom in the Haystack? *Nat. Methods* **2022**, *19* (7), 827–828.

(13) DeKroon, R. M.; Robinette, J. B.; Osorio, C.; Jeong, J. S. Y.; Hamlett, E.; Mocanu, M.; Alzate, O. Analysis of Protein Posttranslational Modifications Using DIGE-Based Proteomics. *Methods Mol. Biol.* **2012**, *854*, 129–143.

(14) Witze, E. S.; Old, W. M.; Resing, K. A.; Ahn, N. G. Mapping Protein Post-Translational Modifications with Mass Spectrometry. *Nat. Methods* **2007**, *4* (10), 798–806.

(15) Ge, Y.; Rybakova, I. N.; Xu, Q.; Moss, R. L. Top-down High-Resolution Mass Spectrometry of Cardiac Myosin Binding Protein C Revealed That Truncation Alters Protein Phosphorylation State. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106* (31), 12658–12663.

(16) Siuti, N.; Kelleher, N. L. Decoding Protein Modifications Using Top-down Mass Spectrometry. *Nat. Methods* **2007**, *4* (10), 817–821.

(17) Borchers, C. H.; Thapar, R.; Petrotchenko, E. V.; Torres, M. P.; Speir, J. P.; Easterling, M.; Dominski, Z.; Marzluff, W. F. Combined Top-down and Bottom-up Proteomics Identifies a Phosphorylation Site in Stem-Loop-Binding Proteins That Contributes to High-Affinity RNA Binding. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103* (9), 3094–3099.

(18) Andersson, L.; Porath, J. Isolation of Phosphoproteins by Immobilized Metal (Fe3+) Affinity

Chromatography. *Anal. Biochem.* **1986**, *154* (1), 250–254.
(19) Stensballe, A.; Andersen, S.; Jensen, O. N. Characterization of Phosphoproteins from Electrophoretic Gels by Nanoscale Fe(III) Affinity Chromatography with off-Line Mass Spectrometry Analysis. *Proteomics* **2001**, *1* (2), 207–222.
(20) Pinkse, M. W. H.; Uitto, P. M.; Hilhorst, M. J.; Ooms, B.; Heck, A. J. R. Selective Isolation at the Femtomole Level of Phosphopeptides from Proteolytic Digests Using 2D-NanoLC-ESI-MS/MS and Titanium Oxide Precolumns. *Anal. Chem.* **2004**, *76* (14), 3935–3943.
(21) Beausoleil, S. A.; Jedrychowski, M.; Schwartz, D.; Elias, J. E.; Villén, J.; Li, J.; Cohn, M. A.; Cantley, L. C.; Gygi, S. P. Large-Scale Characterization of HeLa Cell Nuclear Phosphoproteins. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101* (33), 12130–12135.
(22) Rush, J.; Moritz, A.; Lee, K. A.; Guo, A.; Goss, V. L.; Spek, E. J.; Zhang, H.; Zha, X.-M.; Polakiewicz, R. D.; Comb, M. J. Immunoaffinity Profiling of Tyrosine Phosphorylation in Cancer Cells. *Nat. Biotechnol.* **2005**, *23* (1), 94–101.
(23) Stokes, M. P.; Farnsworth, C. L.; Moritz, A.; Silva, J. C.; Jia, X.; Lee, K. A.; Guo, A.; Polakiewicz, R. D.; Comb, M. J. PTMScan Direct: Identification and Quantification of Peptides from Critical Signaling Proteins by Immunoaffinity Enrichment Coupled with LC-MS/MS. *Mol. Cell. Proteomics* **2012**, *11* (5), 187–201.
(24) Shou, W.; Verma, R.; Annan, R. S.; Huddleston, M. J.; Chen, S. L.; Carr, S. A.; Deshaies, R. J. Mapping Phosphorylation Sites in Proteins by Mass Spectrometry. In *Methods in Enzymology*; Academic Press, 2002; Vol. 351, pp 279–296.
(25) Busman, M.; Schey, K. L.; Oatis, J. E.; Knapp, D. R. Identification of Phosphorylation Sites in Phosphopeptides by Positive and Negative Mode Electrospray Ionization-Tandem Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **1996**, *7* (3), 243–249.
(26) Louris, J. N.; Cooks, R. G.; Syka, J. E. P.; Kelley, P. E.; Stafford, G. C.; Todd, J. F. J. Instrumentation, Applications, and Energy Deposition in Quadrupole Ion-Trap Tandem Mass Spectrometry. *Anal. Chem.* **1987**, *59* (13), 1677–1685.
(27) Nemeth-Cawley, J. F.; Rouse, J. C. Identification and Sequencing Analysis of Intact Proteins via Collision-Induced Dissociation and Quadrupole Time-of-Flight Mass Spectrometry. *J. Mass Spectrom.* **2002**, *37* (3), 270–282.
(28) Olsen, J. V.; Macek, B.; Lange, O.; Makarov, A.; Horning, S.; Mann, M. Higher-Energy C-Trap Dissociation for Peptide Modification Analysis. *Nat. Methods* **2007**, *4* (9), 709–712.
(29) Jedrychowski, M. P.; Huttlin, E. L.; Haas, W.; Sowa, M. E.; Rad, R.; Gygi, S. P. Evaluation of HCD- and CID-Type Fragmentation within Their Respective Detection Platforms for Murine Phosphoproteomics. *Mol. Cell. Proteomics* **2011**, *10* (12), M111.009910.
(30) DeGnore, J. P.; Qin, J. Fragmentation of Phosphopeptides in an Ion Trap Mass Spectrometer. *J. Am. Soc. Mass Spectrom.* **1998**, *9* (11), 1175–1188.
(31) Sleno, L.; Volmer, D. A. Ion Activation Methods for Tandem Mass Spectrometry. *J. Mass Spectrom.* **2004**, *39* (10), 1091–1112.
(32) Lehmann, W. D.; Krüger, R.; Salek, M.; Hung, C.-W.; Wolschin, F.; Weckwerth, W. Neutral Loss-Based Phosphopeptide Recognition: A Collection of Caveats. *J. Proteome Res.* **2007**, *6* (7), 2866–2873.
(33) Syka, J. E. P.; Coon, J. J.; Schroeder, M. J.; Shabanowitz, J.; Hunt, D. F. Peptide and Protein Sequence Analysis by Electron Transfer Dissociation Mass Spectrometry. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101* (26), 9528–9533.
(34) Mechref, Y. Use of CID/ETD Mass Spectrometry to Analyze Glycopeptides. *Curr. Protoc. Protein Sci.* **2012**, *Chapter 12*, 12.11.1–12.11.11.
(35) Zubarev, R. A.; Kelleher, N. L.; McLafferty, F. W. Electron Capture Dissociation of Multiply Charged Protein Cations. A Nonergodic Process. *J. Am. Chem. Soc.* **1998**, *120* (13), 3265–3266.
(36) Hart-Smith, G. A Review of Electron-Capture and Electron-Transfer Dissociation Tandem

Mass Spectrometry in Polymer Chemistry. *Anal. Chim. Acta* **2014**, *808*, 44–55.
(37) Hansen, T. A.; Sylvester, M.; Jensen, O. N.; Kjeldsen, F. Automated and High Confidence Protein Phosphorylation Site Localization Using Complementary Collision-Activated Dissociation and Electron Transfer Dissociation Tandem Mass Spectrometry. *Anal. Chem.* **2012**, *84* (22), 9694–9699.
(38) Vandenbogaert, M.; Hourdel, V.; Jardin-Mathé, O.; Bigeard, J.; Bonhomme, L.; Legros, V.; Hirt, H.; Schwikowski, B.; Pflieger, D. Automated Phosphopeptide Identification Using Multiple MS/MS Fragmentation Modes. *J. Proteome Res.* **2012**, *11* (12), 5695–5703.
(39) Stahl, D. C.; Swiderek, K. M.; Davis, M. T.; Lee, T. D. Data-Controlled Automation of Liquid Chromatography/tandem Mass Spectrometry Analysis of Peptide Mixtures. *J. Am. Soc. Mass Spectrom.* **1996**, *7* (6), 532–540.
(40) Davis, M. T.; Spahr, C. S.; McGinley, M. D.; Robinson, J. H.; Bures, E. J.; Beierle, J.; Mort, J.; Yu, W.; Luethy, R.; Patterson, S. D. Towards Defining the Urinary Proteome Using Liquid Chromatography-Tandem Mass Spectrometry. II. Limitations of Complex Mixture Analyses. *Proteomics* **2001**, *1* (1), 108–117.
(41) Kohli, B. M.; Eng, J. K.; Nitsch, R. M.; Konietzko, U. An Alternative Sampling Algorithm for Use in Liquid Chromatography/tandem Mass Spectrometry Experiments. *Rapid Commun. Mass Spectrom.* **2005**, *19* (5), 589–596.
(42) Eng, J. K.; McCormack, A. L.; Yates, J. R. An Approach to Correlate Tandem Mass Spectral Data of Peptides with Amino Acid Sequences in a Protein Database. *J. Am. Soc. Mass Spectrom.* **1994**, *5* (11), 976–989.
(43) Craig, R.; Beavis, R. C. TANDEM: Matching Proteins with Tandem Mass Spectra. *Bioinformatics* **2004**, *20* (9), 1466–1467.
(44) Tabb, D. L.; Fernando, C. G.; Chambers, M. C. MyriMatch: Highly Accurate Tandem Mass Spectral Peptide Identification by Multivariate Hypergeometric Analysis. *J. Proteome Res.* **2007**, *6* (2), 654–661.
(45) Cox, J.; Neuhauser, N.; Michalski, A.; Scheltema, R. A.; Olsen, J. V.; Mann, M. Andromeda: A Peptide Search Engine Integrated into the MaxQuant Environment. *J. Proteome Res.* **2011**, *10* (4), 1794–1805.
(46) Bern, M.; Kil, Y. J.; Becker, C. Byonic: Advanced Peptide and Protein Identification Software. *Curr. Protoc. Bioinformatics* **2012**, *Chapter 13*, 13.20.1–13.20.14.
(47) Kim, S.; Pevzner, P. A. MS-GF+ Makes Progress towards a Universal Database Search Tool for Proteomics. *Nat. Commun.* **2014**, *5*, 5277.
(48) Perkins, D. N.; Pappin, D. J.; Creasy, D. M.; Cottrell, J. S. Probability-Based Protein Identification by Searching Sequence Databases Using Mass Spectrometry Data. *Electrophoresis* **1999**, *20* (18), 3551–3567.
(49) Eng, J. K.; Searle, B. C.; Clauser, K. R.; Tabb, D. L. A Face in the Crowd: Recognizing Peptides through Database Search. *Mol. Cell. Proteomics* **2011**, *10* (11), R111.009522.
(50) Yates, J. R., 3rd; Eng, J. K.; McCormack, A. L.; Schieltz, D. Method to Correlate Tandem Mass Spectra of Modified Peptides to Amino Acid Sequences in the Protein Database. *Anal. Chem.* **1995**, *67* (8), 1426–1436.
(51) Venable, J. D.; Dong, M.-Q.; Wohlschlegel, J.; Dillin, A.; Yates, J. R. Automated Approach for Quantitative Analysis of Complex Peptide Mixtures from Tandem Mass Spectra. *Nat. Methods* **2004**, *1* (1), 39–45.
(52) Gillet, L. C.; Navarro, P.; Tate, S.; Röst, H.; Selevsek, N.; Reiter, L.; Bonner, R.; Aebersold, R. 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* **2012**, *11* (6), O111.016717.
(53) Ting, Y. S.; Egertson, J. D.; Payne, S. H.; Kim, S.; MacLean, B.; Käll, L.; Aebersold, R.; Smith, R. D.; Noble, W. S.; MacCoss, M. J. Peptide-Centric Proteome Analysis: An Alternative Strategy for the Analysis of Tandem Mass Spectrometry Data. *Mol. Cell.*

Proteomics **2015**, *14* (9), 2301–2307.
(54) Tsou, C.-C.; Avtonomov, D.; Larsen, B.; Tucholska, M.; Choi, H.; Gingras, A.-C.; Nesvizhskii, A. I. DIA-Umpire: Comprehensive Computational Framework for Data-Independent Acquisition Proteomics. *Nat. Methods* **2015**, *12* (3), 258–264, 7 p following 264.
(55) Peterson, A. C.; Russell, J. D.; Bailey, D. J.; Westphall, M. S.; Coon, J. J. Parallel Reaction Monitoring for High Resolution and High Mass Accuracy Quantitative, Targeted Proteomics. *Mol. Cell. Proteomics* **2012**, *11* (11), 1475–1488.
(56) Manning, G.; Whyte, D. B.; Martinez, R.; Hunter, T.; Sudarsanam, S. The Protein Kinase Complement of the Human Genome. *Science* **2002**, *298* (5600), 1912–1934.
(57) Schweiger, R.; Linial, M. Cooperativity within Proximal Phosphorylation Sites Is Revealed from Large-Scale Proteomics Data. *Biol. Direct* **2010**, *5*, 6.
(58) Kalous, J.; Jansová, D.; Šušor, A. Role of Cyclin-Dependent Kinase 1 in Translational Regulation in the M-Phase. *Cells* **2020**, *9* (7). https://doi.org/10.3390/cells9071568.
(59) Mueller, P. R.; Coleman, T. R.; Kumagai, A.; Dunphy, W. G. Myt1: A Membrane-Associated Inhibitory Kinase That Phosphorylates Cdc2 on Both Threonine-14 and Tyrosine-15. *Science* **1995**, *270* (5233), 86–90.
(60) LaGory, E. L.; Sitailo, L. A.; Denning, M. F. The Protein Kinase Cdelta Catalytic Fragment Is Critical for Maintenance of the G2/M DNA Damage Checkpoint. *J. Biol. Chem.* **2010**, *285* (3), 1879–1887.
(61) Szmyd, R.; Niska-Blakie, J.; Diril, M. K.; Renck Nunes, P.; Tzelepis, K.; Lacroix, A.; van Hul, N.; Deng, L.-W.; Matos, J.; Dreesen, O.; Bisteau, X.; Kaldis, P. Premature Activation of Cdk1 Leads to Mitotic Events in S Phase and Embryonic Lethality. *Oncogene* **2019**, *38* (7), 998–1018.
(62) Palou, G.; Palou, R.; Zeng, F.; Vashisht, A. A.; Wohlschlegel, J. A.; Quintana, D. G. Three Different Pathways Prevent Chromosome Segregation in the Presence of DNA Damage or Replication Stress in Budding Yeast. *PLoS Genet.* **2015**, *11* (9), e1005468.
(63) Beausoleil, S. A.; Villén, J.; Gerber, S. A.; Rush, J.; Gygi, S. P. A Probability-Based Approach for High-Throughput Protein Phosphorylation Analysis and Site Localization. *Nat. Biotechnol.* **2006**, *24* (10), 1285–1292.
(64) Savitski, M. M.; Lemeer, S.; Boesche, M.; Lang, M.; Mathieson, T.; Bantscheff, M.; Kuster, B. Confident Phosphorylation Site Localization Using the Mascot Delta Score. *Mol. Cell. Proteomics* **2011**, *10* (2), M110.003830.
(65) Baker, P. R.; Trinidad, J. C.; Chalkley, R. J. Modification Site Localization Scoring Integrated into a Search Engine. *Mol. Cell. Proteomics* **2011**, *10* (7), M111.008078.
(66) Clauser, K. R.; Baker, P.; Burlingame, A. L. Role of Accurate Mass Measurement (+/- 10 Ppm) in Protein Identification Strategies Employing MS or MS/MS and Database Searching. *Anal. Chem.* **1999**, *71* (14), 2871–2882.
(67) Chalkley, R. J.; Baker, P. R.; Huang, L.; Hansen, K. C.; Allen, N. P.; Rexach, M.; Burlingame, A. L. Comprehensive Analysis of a Multidimensional Liquid Chromatography Mass Spectrometry Dataset Acquired on a Quadrupole Selecting, Quadrupole Collision Cell, Time-of-Flight Mass Spectrometer: II. New Developments in Protein Prospector Allow for Reliable and Comprehensive Automatic Analysis of Large Datasets. *Mol. Cell. Proteomics* **2005**, *4* (8), 1194–1204.
(68) Hornbeck, P. V.; Zhang, B.; Murray, B.; Kornhauser, J. M.; Latham, V.; Skrzypek, E. PhosphoSitePlus, 2014: Mutations, PTMs and Recalibrations. *Nucleic Acids Res.* **2015**, *43* (Database issue), D512–D520.
(69) Lawrence, R. T.; Searle, B. C.; Llovet, A.; Villén, J. Plug-and-Play Analysis of the Human Phosphoproteome by Targeted High-Resolution Mass Spectrometry. *Nat. Methods* **2016**, *13* (5), 431–434.
(70) Lee, D. C. H.; Jones, A. R.; Hubbard, S. J. Computational Phosphoproteomics: From

Identification to Localization. *Proteomics* **2015**, *15* (5-6), 950–963.
(71) Xu, X.; Sarikas, A.; Dias-Santagata, D. C.; Dolios, G.; Lafontant, P. J.; Tsai, S.-C.; Zhu, W.; Nakajima, H.; Nakajima, H. O.; Field, L. J.; Wang, R.; Pan, Z.-Q. The CUL7 E3 Ubiquitin Ligase Targets Insulin Receptor Substrate 1 for Ubiquitin-Dependent Degradation. *Mol. Cell* **2008**, *30* (4), 403–414.
(72) Searle, B. C.; Lawrence, R. T.; MacCoss, M. J.; Villén, J. Thesaurus: Quantifying Phosphopeptide Positional Isomers. *Nat. Methods* **2019**, *16* (8), 703–706.
(73) Wolf-Yadlin, A.; Hautaniemi, S.; Lauffenburger, D. A.; White, F. M. Multiple Reaction Monitoring for Robust Quantitative Proteomic Analysis of Cellular Signaling Networks. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104* (14), 5860–5865.
(74) Lim, M. Y.; Paulo, J. A.; Gygi, S. P. Evaluating False Transfer Rates from the Match-between-Runs Algorithm with a Two-Proteome Model. *J. Proteome Res.* **2019**, *18* (11), 4020–4026.
(75) Phanstiel, D. H.; Brumbaugh, J.; Wenger, C. D.; Tian, S.; Probasco, M. D.; Bailey, D. J.; Swaney, D. L.; Tervo, M. A.; Bolin, J. M.; Ruotti, V.; Stewart, R.; Thomson, J. A.; Coon, J. J. Proteomic and Phosphoproteomic Comparison of Human ES and iPS Cells. *Nat. Methods* **2011**, *8* (10), 821–827.
(76) Albuquerque, C. P.; Smolka, M. B.; Payne, S. H.; Bafna, V.; Eng, J.; Zhou, H. A Multidimensional Chromatography Technology for in-Depth Phosphoproteome Analysis. *Mol. Cell. Proteomics* **2008**, *7* (7), 1389–1396.
(77) Tanner, S.; Shu, H.; Frank, A.; Wang, L.-C.; Zandi, E.; Mumby, M.; Pevzner, P. A.; Bafna, V. InsPecT: Identification of Posttranslationally Modified Peptides from Tandem Mass Spectra. *Anal. Chem.* **2005**, *77* (14), 4626–4639.
(78) Taus, T.; Köcher, T.; Pichler, P.; Paschke, C.; Schmidt, A.; Henrich, C.; Mechtler, K. Universal and Confident Phosphorylation Site Localization Using phosphoRS. *J. Proteome Res.* **2011**, *10* (12), 5354–5362.
(79) Bailey, C. M.; Sweet, S. M. M.; Cunningham, D. L.; Zeller, M.; Heath, J. K.; Cooper, H. J. SLoMo: Automated Site Localization of Modifications from ETD/ECD Mass Spectra. *J. Proteome Res.* **2009**, *8* (4), 1965–1971.
(80) MacLean, D.; Burrell, M. A.; Studholme, D. J.; Jones, A. M. PhosCalc: A Tool for Evaluating the Sites of Peptide Phosphorylation from Mass Spectrometer Data. *BMC Res. Notes* **2008**, *1*, 30.
(81) Ulintz, P. J.; Yocum, A. K.; Bodenmiller, B.; Aebersold, R.; Andrews, P. C.; Nesvizhskii, A. I. Comparison of MS(2)-Only, MSA, and MS(2)/MS(3) Methodologies for Phosphopeptide Identification. *J. Proteome Res.* **2009**, *8* (2), 887–899.
(82) Olsen, J. V.; Blagoev, B.; Gnad, F.; Macek, B.; Kumar, C.; Mortensen, P.; Mann, M. Global, in Vivo, and Site-Specific Phosphorylation Dynamics in Signaling Networks. *Cell* **2006**, *127* (3), 635–648.
(83) Fermin, D.; Walmsley, S. J.; Gingras, A.-C.; Choi, H.; Nesvizhskii, A. I. LuciPHOr: Algorithm for Phosphorylation Site Localization with False Localization Rate Estimation Using Modified Target-Decoy Approach. *Mol. Cell. Proteomics* **2013**, *12* (11), 3409–3419.
(84) Shteynberg, D. D.; Deutsch, E. W.; Campbell, D. S.; Hoopmann, M. R.; Kusebauch, U.; Lee, D.; Mendoza, L.; Midha, M. K.; Sun, Z.; Whetton, A. D.; Moritz, R. L. PTMProphet: Fast and Accurate Mass Modification Localization for the Trans-Proteomic Pipeline. *J. Proteome Res.* **2019**, *18* (12), 4262–4272.
(85) An, Z.; Zhai, L.; Ying, W.; Qian, X.; Gong, F.; Tan, M.; Fu, Y. PTMiner: Localization and Quality Control of Protein Modifications Detected in an Open Search and Its Application to Comprehensive Post-Translational Modification Characterization in Human Proteome. *Mol. Cell. Proteomics* **2019**, *18* (2), 391–405.
(86) Yang, H.; Chi, H.; Zhou, W.-J.; Zeng, W.-F.; Liu, C.; Wang, R.-M.; Wang, Z.-W.; Niu, X.-N.; Chen, Z.-L.; He, S.-M. pSite: Amino Acid Confidence Evaluation for Quality Control of De

Novo Peptide Sequencing and Modification Site Localization. *J. Proteome Res.* **2018**, *17* (1), 119–128.
(87) Searle, B. C.; Dasari, S.; Turner, M.; Reddy, A. P.; Choi, D.; Wilmarth, P. A.; McCormack, A. L.; David, L. L.; Nagalla, S. R. High-Throughput Identification of Proteins and Unanticipated Sequence Modifications Using a Mass-Based Alignment Algorithm for MS/MS de Novo Sequencing Results. *Anal. Chem.* **2004**, *76* (8), 2220–2230.
(88) Han, Y.; Ma, B.; Zhang, K. SPIDER: Software for Protein Identification from Sequence Tags with de Novo Sequencing Error. *J. Bioinform. Comput. Biol.* **2005**, *3* (3), 697–716.
(89) Suni, V.; Imanishi, S. Y.; Maiolica, A.; Aebersold, R.; Corthals, G. L. Confident Site Localization Using a Simulated Phosphopeptide Spectral Library. *J. Proteome Res.* **2015**, *14* (5), 2348–2359.
(90) Hu, Y.; Li, Y.; Lam, H. A Semi-Empirical Approach for Predicting Unobserved Peptide MS/MS Spectra from Spectral Libraries. *Proteomics* **2011**, *11* (24), 4702–4711.
(91) Hu, Y.; Lam, H. Expanding Tandem Mass Spectral Libraries of Phosphorylated Peptides: Advances and Applications. *J. Proteome Res.* **2013**, *12* (12), 5971–5977.
(92) Suni, V.; Suomi, T.; Tsubosaka, T.; Imanishi, S. Y.; Elo, L. L.; Corthals, G. L. SimPhospho: A Software Tool Enabling Confident Phosphosite Assignment. *Bioinformatics* **2018**, *34* (15), 2690–2692.
(93) Takai, A.; Tsubosaka, T.; Hirano, Y.; Hayakawa, N.; Tani, F.; Haapaniemi, P.; Suni, V.; Imanishi, S. Y. Optimization of TripleTOF Spectral Simulation and Library Searching for Confident Localization of Phosphorylation Sites. *PLoS One* **2019**, *14* (12), e0225885.
(94) Lam, H.; Deutsch, E. W.; Eddes, J. S.; Eng, J. K.; King, N.; Stein, S. E.; Aebersold, R. Development and Validation of a Spectral Library Searching Method for Peptide Identification from MS/MS. *Proteomics* **2007**, *7* (5), 655–667.
(95) Pino, L. K.; Just, S. C.; MacCoss, M. J.; Searle, B. C. Acquiring and Analyzing Data Independent Acquisition Proteomics Experiments without Spectrum Libraries. *Mol. Cell. Proteomics* **2020**. https://doi.org/10.1074/mcp.P119.001913.
(96) Peckner, R.; Myers, S. A.; Jacome, A. S. V.; Egertson, J. D.; Abelin, J. G.; MacCoss, M. J.; Carr, S. A.; Jaffe, J. D. Specter: Linear Deconvolution for Targeted Analysis of Data-Independent Acquisition Mass Spectrometry Proteomics. *Nat. Methods* **2018**, *15* (5), 371–378.
(97) Meyer, J. G.; Mukkamalla, S.; Steen, H.; Nesvizhskii, A. I.; Gibson, B. W.; Schilling, B. PIQED: Automated Identification and Quantification of Protein Modifications from DIA-MS Data. *Nat. Methods* **2017**, *14* (7), 646–647.
(98) Eng, J. K.; Jahan, T. A.; Hoopmann, M. R. Comet: An Open-Source MS/MS Sequence Database Search Tool. *Proteomics* **2013**, *13* (1), 22–24.
(99) Deutsch, E. W.; Mendoza, L.; Shteynberg, D.; Farrah, T.; Lam, H.; Tasman, N.; Sun, Z.; Nilsson, E.; Pratt, B.; Prazen, B.; Eng, J. K.; Martin, D. B.; Nesvizhskii, A. I.; Aebersold, R. A Guided Tour of the Trans-Proteomic Pipeline. *Proteomics* **2010**, *10* (6), 1150–1159.
(100) Rosenberger, G.; Liu, Y.; Röst, H. L.; Ludwig, C.; Buil, A.; Bensimon, A.; Soste, M.; Spector, T. D.; Dermitzakis, E. T.; Collins, B. C.; Malmström, L.; Aebersold, R. Inference and Quantification of Peptidoforms in Large Sample Cohorts by SWATH-MS. *Nat. Biotechnol.* **2017**, *35* (8), 781–788.
(101) Röst, H. L.; Rosenberger, G.; Navarro, P.; Gillet, L.; Miladinović, S. M.; Schubert, O. T.; Wolski, W.; Collins, B. C.; Malmström, J.; Malmström, L.; Aebersold, R. OpenSWATH Enables Automated, Targeted Analysis of Data-Independent Acquisition MS Data. *Nat. Biotechnol.* **2014**, *32* (3), 219–223.
(102) Searle, B. C.; Pino, L. K.; Egertson, J. D.; Ting, Y. S.; Lawrence, R. T.; MacLean, B. X.; Villén, J.; MacCoss, M. J. Chromatogram Libraries Improve Peptide Detection and Quantification by Data Independent Acquisition Mass Spectrometry. *Nat. Commun.* **2018**, *9* (1), 5128.

(103) Bekker-Jensen, D. B.; Bernhardt, O. M.; Hogrebe, A.; Martinez-Val, A.; Verbeke, L.; Gandhi, T.; Kelstrup, C. D.; Reiter, L.; Olsen, J. V. Rapid and Site-Specific Deep Phosphoproteome Profiling by Data-Independent Acquisition without the Need for Spectral Libraries. *Nat. Commun.* **2020**, *11* (1), 787.
(104) Martinez-Val, A.; Bekker-Jensen, D. B.; Hogrebe, A.; Olsen, J. V. Data Processing and Analysis for DIA-Based Phosphoproteomics Using Spectronaut. *Methods Mol. Biol.* **2021**, *2361*, 95–107.
(105) Potel, C. M.; Lemeer, S.; Heck, A. J. R. Phosphopeptide Fragmentation and Site Localization by Mass Spectrometry: An Update. *Anal. Chem.* **2019**, *91* (1), 126–141.
(106) Elias, J. E.; Gygi, S. P. Target-Decoy Search Strategy for Increased Confidence in Large-Scale Protein Identifications by Mass Spectrometry. *Nat. Methods* **2007**, *4* (3), 207–214.
(107) Käll, L.; Canterbury, J. D.; Weston, J.; Noble, W. S.; MacCoss, M. J. Semi-Supervised Learning for Peptide Identification from Shotgun Proteomics Datasets. *Nat. Methods* **2007**, *4* (11), 923–925.
(108) Tabb, D. L. What's Driving False Discovery Rates? *J. Proteome Res.* **2008**, *7* (1), 45–46.
(109) Chalkley, R. J.; Clauser, K. R. Modification Site Localization Scoring: Strategies and Performance. *Mol. Cell. Proteomics* **2012**, *11* (5), 3–14.
(110) Ramsbottom, K. A.; Prakash, A.; Riverol, Y. P.; Camacho, O. M.; Martin, M.-J.; Vizcaíno, J. A.; Deutsch, E. W.; Jones, A. R. Method for Independent Estimation of the False Localization Rate for Phosphoproteomics. *J. Proteome Res.* **2022**, *21* (7), 1603–1615.
(111) Zong, Y.; Wang, Y.; Yang, Y.; Zhao, D.; Wang, X.; Shen, C.; Qiao, L. DeepFLR Facilitates False Localization Rate Control in Phosphoproteomics. *Nat. Commun.* **2023**, *14* (1), 2269.
(112) Keller, A.; Nesvizhskii, A. I.; Kolker, E.; Aebersold, R. Empirical Statistical Model to Estimate the Accuracy of Peptide Identifications Made by MS/MS and Database Search. *Anal. Chem.* **2002**, *74* (20), 5383–5392.
(113) Locard-Paulet, M.; Bouyssié, D.; Froment, C.; Burlet-Schiltz, O.; Jensen, L. J. Comparing 22 Popular Phosphoproteomics Pipelines for Peptide Identification and Site Localization. *J. Proteome Res.* **2020**, *19* (3), 1338–1345.
(114) Palumbo, A. M.; Reid, G. E. Evaluation of Gas-Phase Rearrangement and Competing Fragmentation Reactions on Protein Phosphorylation Site Assignment Using Collision Induced Dissociation-MS/MS and MS3. *Anal. Chem.* **2008**, *80* (24), 9735–9747.
(115) Boersema, P. J.; Mohammed, S.; Heck, A. J. R. Phosphopeptide Fragmentation and Analysis by Mass Spectrometry. *J. Mass Spectrom.* **2009**, *44* (6), 861–878.
(116) Palumbo, A. M.; Smith, S. A.; Kalcic, C. L.; Dantus, M.; Stemmer, P. M.; Reid, G. E. Tandem Mass Spectrometry Strategies for Phosphoproteome Analysis. *Mass Spectrom. Rev.* **2011**, *30* (4), 600–625.
(117) Penkert, M.; Yates, L. M.; Schümann, M.; Perlman, D.; Fiedler, D.; Krause, E. Unambiguous Identification of Serine and Threonine Pyrophosphorylation Using Neutral-Loss-Triggered Electron-Transfer/Higher-Energy Collision Dissociation. *Anal. Chem.* **2017**, *89* (6), 3672–3680.
(118) Everley, R. A.; Huttlin, E. L.; Erickson, A. R.; Beausoleil, S. A.; Gygi, S. P. Neutral Loss Is a Very Common Occurrence in Phosphotyrosine-Containing Peptides Labeled with Isobaric Tags. *J. Proteome Res.* **2017**, *16* (2), 1069–1076.
(119) Palumbo, A. M.; Tepe, J. J.; Reid, G. E. Mechanistic Insights into the Multistage Gas-Phase Fragmentation Behavior of Phosphoserine- and Phosphothreonine-Containing Peptides. *J. Proteome Res.* **2008**, *7* (2), 771–779.
(120) Aguiar, M.; Haas, W.; Beausoleil, S. A.; Rush, J.; Gygi, S. P. Gas-Phase Rearrangements Do Not Affect Site Localization Reliability in Phosphoproteomics Data Sets. *J. Proteome Res.* **2010**, *9* (6), 3103–3107.
(121) Jones, A. R.; Eisenacher, M.; Mayer, G.; Kohlbacher, O.; Siepen, J.; Hubbard, S. J.;

Selley, J. N.; Searle, B. C.; Shofstahl, J.; Seymour, S. L.; Julian, R.; Binz, P.-A.; Deutsch, E. W.; Hermjakob, H.; Reisinger, F.; Griss, J.; Vizcaíno, J. A.; Chambers, M.; Pizarro, A.; Creasy, D. The mzIdentML Data Standard for Mass Spectrometry-Based Proteomics Results. *Mol. Cell. Proteomics* **2012**, *11* (7), M111.014381.

(122) Seymour, S. L.; Farrah, T.; Binz, P.-A.; Chalkley, R. J.; Cottrell, J. S.; Searle, B. C.; Tabb, D. L.; Vizcaíno, J. A.; Prieto, G.; Uszkoreit, J.; Eisenacher, M.; Martínez-Bartolomé, S.; Ghali, F.; Jones, A. R. A Standardized Framing for Reporting Protein Identifications in mzIdentML 1.2. *Proteomics* **2014**, *14* (21-22), 2389–2399.