the plant journal

SEB Society for Experimental Biology

The Plant Journal (2019) doi: 10.1111/tpj.14372

RESOURCE

Functional analysis tools for post-translational modification: a post-translational modification database for analysis of proteins and metabolic pathways

Edward R. Cruz[1], Hung Nguyen[2], Tin Nguyen[2] and Ian S. Wallace[1,3,*]

[1]*Department of Biochemistry and Molecular Biology, University of Nevada, Reno, Reno, NV 89557, USA,*
[2]*Department of Computer Science and Engineering, University of Nevada, Reno, Reno, NV 89557, USA, and*
[3]*Department of Chemistry, University of Nevada, Reno, Reno, NV 89557, USA*

Received 1 March 2019; revised 3 April 2019; accepted 17 April 2019.
*For correspondence (e-mail iwallace@unr.edu).

SUMMARY

Post-translational modifications (PTMs) are critical regulators of protein function, and nearly 200 different types of PTM have been identified. Advances in high-resolution mass spectrometry have led to the identification of an unprecedented number of PTM sites in numerous organisms, potentially facilitating a more complete understanding of how PTMs regulate cellular behavior. While databases have been created to house the resulting data, most of these resources focus on individual types of PTM, do not consider quantitative PTM analyses or do not provide tools for the visualization and analysis of PTM data. Here, we describe the Functional Analysis Tools for Post-Translational Modifications (FAT-PTM) database (https://bioinformatics.cse.unr.edu/fat-ptm/), which currently supports eight different types of PTM and over 49 000 PTM sites identified in large-scale proteomic surveys of the model organism *Arabidopsis thaliana*. The FAT-PTM database currently supports tools to visualize protein-centric PTM networks, quantitative phosphorylation site data from over 10 different quantitative phosphoproteomic studies, PTM information displayed in protein-centric metabolic pathways and groups of proteins that are co-modified by multiple PTMs. Overall, the FAT-PTM database provides users with a robust platform to share and visualize experimentally supported PTM data, develop hypotheses related to target proteins or identify emergent patterns in PTM data for signaling and metabolic pathways.

Keywords: post-translational modifications, proteomic database, metabolic regulation, quantitative proteomics, mass spectrometry.

INTRODUCTION

Post-translational modifications (PTMs) can control virtually every aspect of protein behavior, including enzymatic activity, subcellular localization, protein–protein interaction networks and protein stability. Nearly 200 PTMs have been identified (Garavelli, 2004), and it is estimated that over half of all eukaryotic proteins are post-translationally modified at some point in their lifetime. Additionally, co-modification of proteins with more than one PTM has been increasingly observed (Beltrao *et al.*, 2013), suggesting that PTMs functionally interact to integrate multiple signals affecting a target protein. While phosphorylation on serine, threonine and tyrosine residues is one of the most often studied and common PTMs in eukaryotes, other PTMs, such as lysine acetylation, ubiquitinylation, SUMOylation and methylation on lysine residues as well as *O*- and *N*-type glycosylation, have become increasingly appreciated as important regulators of protein function (Kühn, 2016; Strasser, 2016; Turnbull and Hemsley, 2017; Withers and Dong, 2017; Augustine and Vierstra, 2018).

Mass spectrometry has supplanted other techniques as the primary method for identifying PTMs in protein samples. Various technical advances have led to an explosion of PTMs being identified in a number of model organisms, and these advances include the widespread implementation of very high mass resolution analysis devices coupled with electron transfer dissociation (ETD) as a fragmentation method (Coon *et al.*, 2005), the development of

chemical and antibody-based enrichment strategies for PTMs (Larsen *et al.*, 2005; Xu *et al.*, 2010, 2017; Kim *et al.*, 2011; Majeran *et al.*, 2018; Zeng *et al.*, 2018) and advanced data analysis techniques that identify PTMs of unknown function (Tsur *et al.*, 2005; Huang *et al.*, 2013). Furthermore, quantitative mass spectrometry has become more common through the implementation of isotopically encoded labels, isobaric tags and advances in spectral counting methodology. Therefore, quantitative analyses of PTM dynamics can be utilized to study regulation under changing cellular conditions.

Numerous databases have been created to house PTM-related information resulting from mass spectrometry experiments. For example, PhosphoGRID (Breitkreutz *et al.*, 2013), PHOSIDA (Gnad *et al.*, 2010), PhosphoELM (Via *et al.*, 2010), iPTMNet (Huang *et al.*, 2017) and dbPTM (Kao *et al.*, 2015) have all been developed to store PTM data for various model organisms. Additionally, PhosPhat (Zulawski *et al.*, 2012) and P^3DB (Yao *et al.*, 2013) have become critically important repositories for the examination of phosphorylation events in plant model and non-model organisms. Despite the large number of databases devoted to the function of PTMs, many of these databases contain a limited range of species, do not include quantitative information associated with PTMs, do not consider multiple PTM types or do not provide tools for analyzing PTM data in the context of biological pathways. Therefore, there is an important need to develop computational and bioinformatic infrastructure to meet these demands.

Here, we describe Functional Analysis Tools for Post-Translational Modifications (FAT-PTM), an integrated database to examine the function and co-occurrence of PTMs in *Arabidopsis thaliana*. The FAT-PTM database provides various types of PTM data (phosphorylation, ubiquitinylation, acetylation, etc.) that can be mapped onto individual proteins or displayed in the context of larger metabolic pathways or protein complexes. Additionally, FAT-PTM allows users to visualize quantitative phosphoproteomic information from a variety of publicly available experiments. The resulting network visualizations and all underlying PTM data are easily exported and accessible for further analysis or publication. The FAT-PTM platform can also be easily modified in the future to include additional model organisms or functional analysis tools.

RESULTS

Database construction and interface

Multiple sources of PTM data were compiled to construct the foundational PTM information for FAT-PTM (Figure 1). Phosphorylation events in PhosPhAt 4.0 (Zulawski *et al.*, 2012) were initially utilized because this database is a well-established Arabidopsis resource for phosphorylation data and also provides the number of spectral observations supporting each phosphorylation event. We queried PhosPhAt 4.0 for every protein in the Arabidopsis proteome and retrieved 8585 proteins with 190 680 spectra supporting 32 023 mapped phosphorylation sites from varying tissues and stress conditions (Table 1). Additionally, we obtained data supporting ubiquitinylation, SUMOylation, *N*-glycosylation, *O*-GlcNAcylation, acylation, *S*-nitrosylation and lysine acetylation events at 16 094 sites in 7238 proteins identified in large-scale proteomic studies from the literature (Miller *et al.*, 2010, 2013; Hemsley *et al.*, 2013; Kim *et al.*, 2013; Svozil *et al.*, 2014; Hu *et al.*, 2015; Johnson and Vert, 2016; Walton *et al.*, 2016; Xu *et al.*, 2016, 2017; Koskela *et al.*, 2018; Liu *et al.*, 2018; Majeran *et al.*, 2018; Rytz *et al.*, 2018; Zeng *et al.*, 2018) (Table 1). We also collected data supporting 19 043 mapped phosphorylation sites in 5783 proteins resulting from quantitative

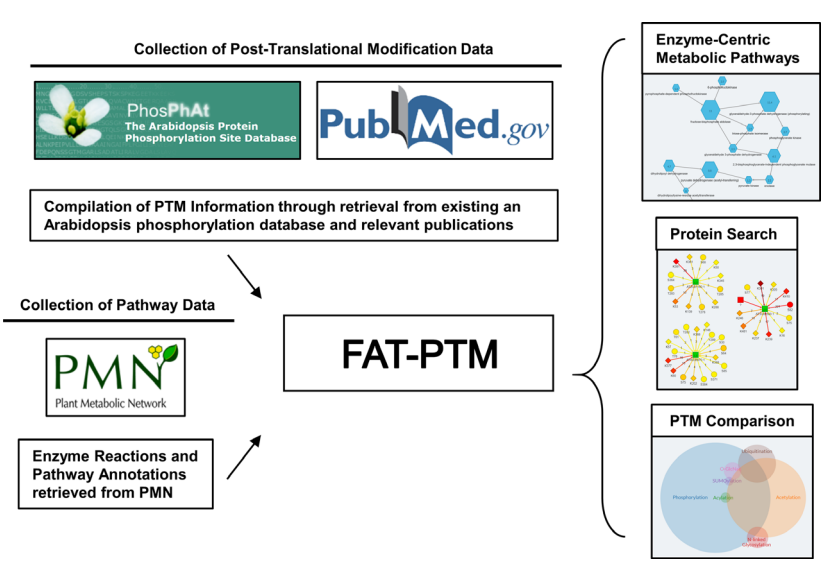

Figure 1. Construction of the Functional Analysis Tools for Post-Translational Modification (FAT-PTM) database.
The FAT-PTM database combines the phosphorylation site data from the existing PhosPhAt database as well as proteomic data supporting seven other types of post-translational modification and quantitative phosphorylation site studies from the literature. Additionally, metabolic pathway information from the Plant Metabolic Network was used to construct enzyme-centric metabolic networks. The FAT-PTM database currently contains three modules to functionally analyze post-translational modifications in individual proteins, metabolic pathways or protein complexes, and groups of proteins that are co-modified by multiple user-defined post-translational modifications.

Table 1 Statistics on post-translational modification data for the FAT-PTM database

Modification	Spectra	Sites	Proteins
Phosphorylation	190 680	33 023	8585
Lysine acetylation	172 834	10 784	4524
Quantitative phosphorylation	90 215	19 043	5783
Ubiquitination	3389	3350	3222[a]
S-nitrosylation	2513	1035	897
O-GlcNAcylation	2492	362	212[a]
N-glycosylation	2257	716	478
SUMOylation	78	68	444[a]
Acylation	72	72	637[a]

[a]These values include proteins that are associated with a particular modification, but the modification sites were not precisely mapped.

phosphoproteomics studies (Benschop *et al.*, 2007; Nühse *et al.*, 2007; Mithoe *et al.*, 2012; Zhang *et al.*, 2013, 2018; Chen and Hoehenwarter, 2015; Minkoff *et al.*, 2015; Roitinger *et al.*, 2015; Cho *et al.*, 2016; Bhaskara *et al.*, 2017; Wang *et al.*, 2018) (Table 1). To our knowledge, this is the most complete repository of Arabidopsis PTMs to date.

The resulting PTM-associated data were organized into a series of modules that can be utilized to query and output data from FAT-PTM. The 'Protein' module is a search tool for single or multiple Arabidopsis Gene Identifier (AGI) codes or gene symbols and allows for the output of an experimentally supported PTM network for each protein with associated quantitative or non-quantitative data. The 'Pathway' module visualizes PTM data associated with aggregated protein families superimposed on various metabolic pathways harvested from the Plant Metabolic Network (Schläpfer *et al.*, 2017) to investigate how post-translational phosphorylation events may regulate metabolism. The 'Co-PTM' tool allows users to identify protein groups that are modified with one or more PTMs and output a list of the resulting proteins and PTM data associated with these proteins. Each of these modules will be discussed in more detail below.

Protein-based PTM networks

The FAT-PTM protein module is a simple protein search tool which allows users to query the FAT-PTM database with an AGI number or gene symbol. The output of this search is a network representation of the queried protein as a central node, experimentally supported PTMs as peripheral nodes and connecting edges that reflect PTM attributes. The edge attributes can display either the number of spectra supporting each PTM site or can display quantitative fold-change information for phosphorylation sites from a variety of quantitative phosphoproteomic studies that utilize stable isotopic or isobaric tag labeling followed by mass spectrometry. Each protein network can also be displayed as a lollipop plot to show where PTMs reside in the context of a given protein sequence.

A series of example outputs for individual proteins is shown in Figure 2(a). Each protein is represented as a central square-shaped node, and PTMs associated with that protein are represented as different shaped nodes surrounding the central node with the modified amino acid and residue index indicated. In the 'Post-translational modifications' view, edges connecting the central protein node to the PTM nodes indicate the number of spectra supporting each PTM site assignment. The edges and associated PTM nodes are colored using a yellow (low spectral support) to red (high spectral support) color scale. In cases such as ubiquitinylation events, where a PTM was identified associated with a protein but the corresponding PTM site was not mapped due to experimental design (Miller *et al.*, 2010, 2013; Hemsley *et al.*, 2013; Kim *et al.*, 2013; Svozil *et al.*, 2014; Johnson and Vert, 2016; Xu *et al.*, 2017), the PTM node is indicated with a '?' symbol for the residue index, and the edge-symbol pair is not colored, to indicate the lack of reliable spectral observation information. When a protein node is selected, the gene view sidebar encloses links to other commonly used databases, such as TAIR and Araport (Rosen *et al.*, 2014; Berardini *et al.*, 2015), along with protein sequence, common annotations and FAT-PTM's phosphorylation scoring method, PhoScore. Furthermore, individual sites contain experimental information, including the phosphorylated peptide fragment, the experimental conditions, the tissue information used to prepare the sample and the PubMed ID of the publication where the site was identified. The generated graph can be exported and saved to an individual user's account with modifications to the layout, naming and color scale if they wish to change it. Alternatively, the resulting data can also be represented as a lollipop plot that illustrates where PTMs occur in the context of a queried protein sequence. In this view, the protein is represented as a bar, with individual protein domains collected from Pfam (Bateman *et al.*, 2018) indicated along the protein sequence representation. Each PTM site is represented as a vertical line with a height that increases based on the number of spectral observations that support the PTM site. In this way, a user can visualize all the experimentally supported PTM sites in a queried protein using a preferred visualization output and easily export the resulting data for publication.

The 'Protein' module can also be used to visualize quantitative changes in phosphopeptide abundance across a variety of conditions by using the 'quantitative phosphorylation' option. In this option, quantitative changes in phosphopeptide abundance are displayed along the edges of protein-centric networks, and multiple graphs can be generated for proteins across a variety of conditions. For example, Figure 2(b) displays protein-centric networks of Arabidopsis H^+-ATPase 1 (AHA1) across a time course (3, 7

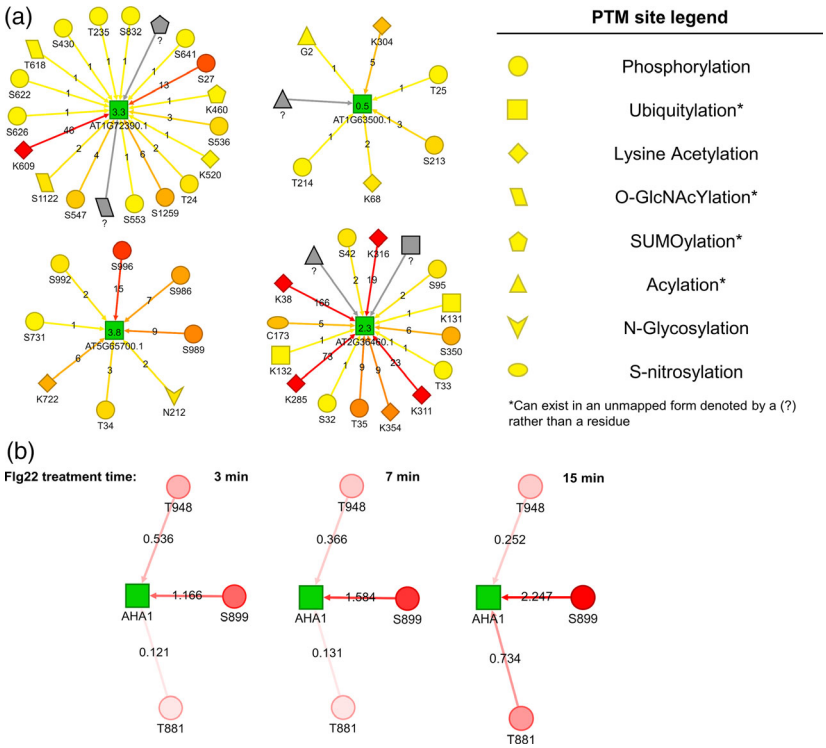

Figure 2. Protein visualizations based on spectral observations and quantitative information.
The Functional Analysis Tools for Post-Translational Modification (FAT-PTM) database was used to create post-translational modification (PTM) network visualizations of various representative proteins.
(a) Representative outputs of various proteins from the 'post-translational modifications' setting in the 'Protein' module are shown. Each protein is represented as a central green node with its Arabidopsis Gene Identifier number indicated. Experimentally supported PTM sites are shown as nodes with different shaped symbols as indicated in the legend. Edges connecting each PTM node to the central protein node as well as the PTM node itself are colored on a yellow to red scale, where warmer colors represent a higher number of spectral observations supporting each site. Grey sites with no residue index indicate PTM sites that are experimentally supported in the protein but have not been mapped to a discrete residue.
(b) Similar networks displaying quantitative abundance changes can be generated using the 'Quantitative phosphorylation' setting in the 'Protein' module. In the presented example, quantitative phosphorylation changes for three experimentally supported phosphorylation sites (T881, S899 and T948) in the AHA1 H^+-ATPase are shown over a time course of flg22 treatment. The fold change of each phosphorylation site compared with control is represented numerically along the edge.

and 15 min) of flg22 treatment. In this example, each phosphorylation site displays different abundance change kinetics, with S899 increasing rapidly over the entire time course, T881 increasing in abundance only after 15 min and T948 decreasing in abundance during the time course. Therefore, this quantitative visualization scheme can be used to examine complex PTM site dynamics under changing experimental conditions. Similar to the 'Post-translational modifications' option, quantitative phosphorylation site changes can be viewed in a lollipop plot format. The PTM information is presented in a similar manner as described above, but the height of the site representation reflects the $\log_2$ fold-change for a particular PTM site, and the direction of the line indicates whether the site increased or decreased in abundance under the selected quantitative conditions.

Overall, the FAT-PTM protein module allows users to quickly assess whether any of eight different PTMs are experimentally supported within any queried Arabidopsis protein, identify PTM sites that are more or less experimentally supported based on spectral observations supporting a particular site and visualize quantitative abundance change data for phosphorylation site PTMs if they are available. These tools will allow users to quickly evaluate basic information regarding PTMs associated with their proteins of interest and rapidly display or aggregate the resulting data.

The FAT-PTM 'Pathway' tool

To extend the utility of FAT-PTM beyond individual proteins or protein families, we reasoned that it would be valuable to superimpose PTM data onto metabolic pathways and protein complexes, since many of the observed PTMs could serve to regulate enzymatic activity or protein interactions within this context. For these metabolic pathways, a scoring method was devised that accounts for the total number of phosphorylation sites and total spectral observations supporting these sites in a given protein to

comparatively analyze different protein groups. This value is referred to as the PhoScore, which is calculated as described in Experimental Procedures. The PhoScore value was designed to increase in proportion to both the number of phosphorylation sites in a target protein and the number of spectra supporting each site. For example, Figure 3(a) shows hypothetical enzyme classes with node sizes scaled by the PhoScore value and the corresponding individual PTM networks for each corresponding enzyme class. A hypothetical protein that contains only a single phosphorylation site that is weakly experimentally supported will generate a low PhoScore value, and similarly, proteins that have a relatively large number of phosphorylation sites each with low experimental support will have a similar low PhoScore value. Only proteins that have a relatively large number of phosphorylation sites that are all supported by many spectra will have high PhoScore values. Figure 3(b) shows a three-dimensional plot displaying PhoScore values for all proteins in the FAT-PTM database compared with their corresponding number of phosphorylation sites and number of total spectra supporting phosphorylation sites within that protein. These results demonstrate that proteins containing a relatively large number of phosphorylation sites or spectra supporting sites can achieve similarly large PhoScore values. To scale metabolic pathways based on the extent of phosphorylation, PhoScore values were standardized to account for multiple isoforms of the same enzyme class. For protein families, the total PhoScore value further accounts for the number of contributing protein family members as described in Experimental Procedures.

To display PTM data in the context of metabolic pathways, pathway information and reactions were parsed from the Plant Metabolic Network (Schläpfer *et al.*, 2017) and used to form enzyme-centric metabolic pathways, where each node represents an enzyme class in a pathway that is scaled according to the PhoScore value. The FAT-PTM database currently contains 615 unique pathways for Arabidopsis. Metabolic pathways are searchable in FAT-PTM, and the resulting graph of the metabolic pathway formed represents enzyme classes as hexagons that can be expanded to display the genes comprising each enzyme class (Figure 4). When enzyme classes are selected, the enzyme class sidebar viewer displays gene information, PTM data, individual PhoScore values for each protein and information regarding additional pathways in which a target protein group participates. PhoScore scaling of metabolic pathways allows for easier identification of enzyme classes that are subject to larger numbers of phosphorylation events, which could potentially reflect critical metabolic regulatory points.

As an example of the 'Pathway' module function, the superpathway of lysine, threonine and methionine biosynthesis is shown in Figure 4 as an enzyme-centric metabolic network with sequential reaction steps indicated. As described above, the size of each enzyme class node is scaled according to the PhoScore value, and these nodes can be expanded to view the individual genes that contribute to each enzyme class. Additionally, the individual protein nodes can be further expanded to view the experimentally supported PTM sites associated with each protein in a particular enzyme class. Based on the premise that proteins which are subject to a larger number of PTMs may be potential regulatory points in a metabolic pathway, this network predicts that the aspartate kinase (AK),

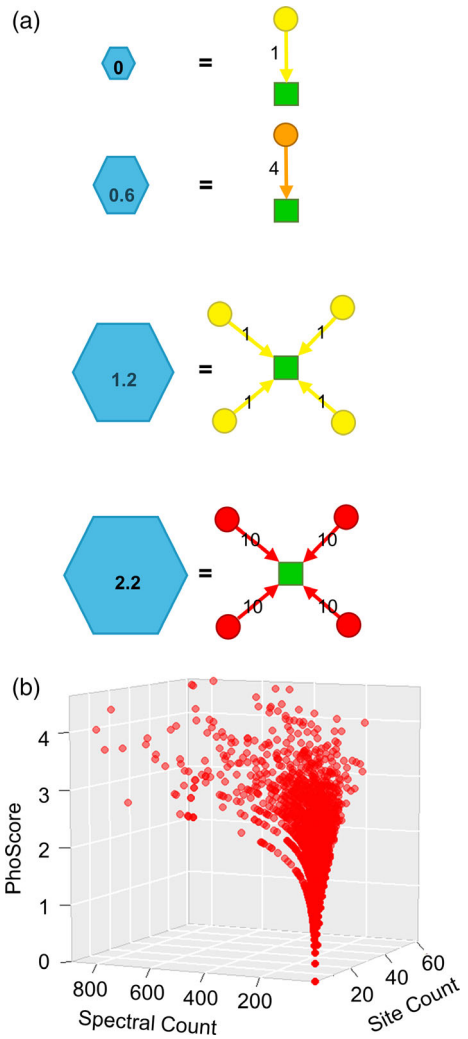

Figure 3. PhoScore representation of proteins in pathways.
To simplify network representations of multiple proteins in pathways, the PhoScore value was utilized.
(a) Hypothetical enzyme classes scaled according to their PhoScore value are displayed alongside the individual post-translational modification networks that compose the enzyme class.
(b) PhoScore values for all Arabidopsis proteins with experimentally supported phosphorylation sites are shown plotted against the number of phosphorylation sites and total number of spectral observations for each protein.

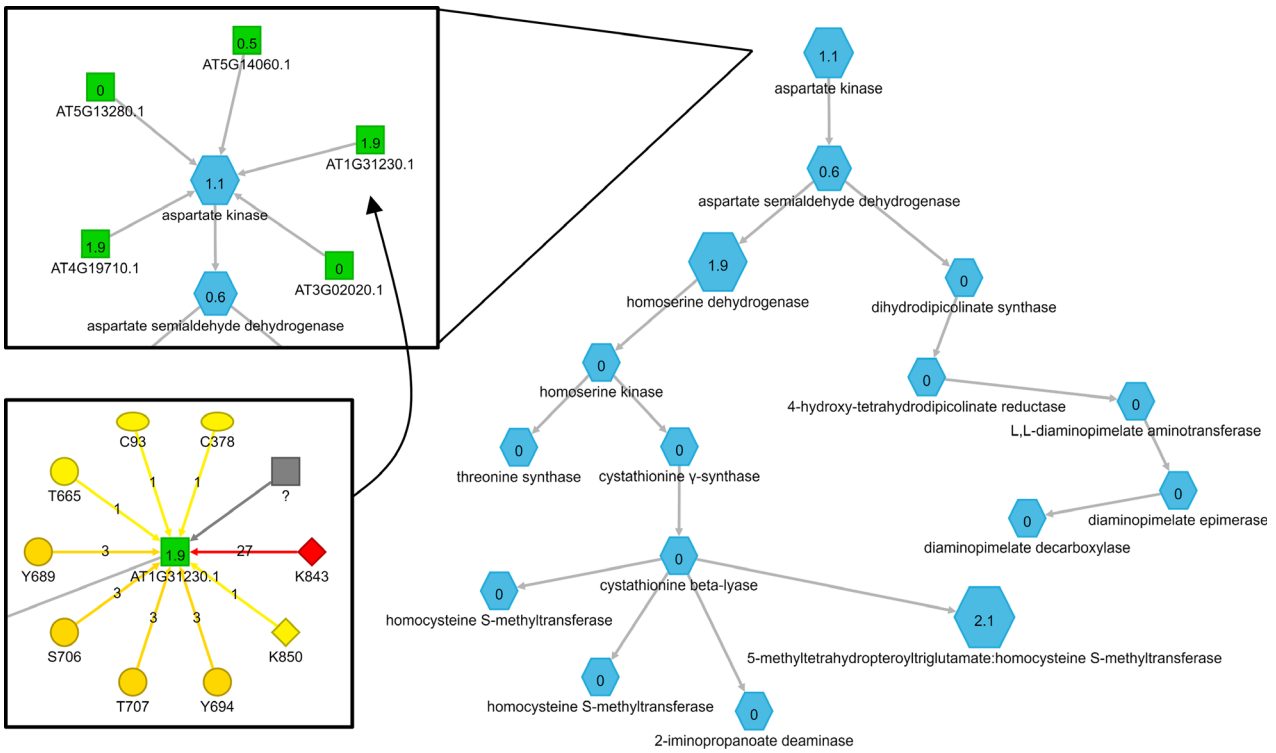

Figure 4. The Functional Analysis Tools for Post-Translational Modification (FAT-PTM) pathway analysis tool.
The FAT-PTM 'Pathway' module uses the PhoScore representation to visualize experimentally supported phosphorylation sites in the context of metabolic pathways. Each step of the enzymatic reaction pathway is represented by a blue hexagon of variable size that is scaled based on the PhoScore value for that protein family. PhoScore values are numerically reported in the center of each node. Pathway steps are connected by a directed network of edges that indicate the path of metabolite flux. The superpathway of lysine, threonine and methionine biosynthesis is shown as an example. Each node in the protein-centric network can be expanded (inset 1) to visualize protein isoforms (represented as green squares) comprising that node, and each of these protein nodes can be expanded (inset 2) to visualize post-translational modifications associated with that protein.

homoserine dehydrogenase (HSDH) and 5-methyltetrahydropteroyltriglutamate:homocysteine *S*-methyltransferase (methionine synthase) steps are likely to be regulatory due to their relatively high PhoScore values. Since AK is the first entry point into the biosynthesis of lysine, threonine and methionine, regulation of this enzyme could control the allocation of aspartate into the biosynthetic pathways for these amino acids. Additionally, HSDH represents the committed step that diverts aspartate semialdehyde to threonine and methionine biosynthesis. Therefore, this regulatory step may be subject to increased post-translational regulation in order to sense the cellular metabolic needs of these amino acids over lysine (Clark and Lu, 2015).

The FAT-PTM 'Pathway' tool can also be used to visualize PTMs in the context of protein complexes. For example, the 'Pathway' tool was used to create representations of the cellulose synthase complex (Figure 5a) and the 26S proteasome (Figure 5b). Visualized protein complexes allow for the identification of potential regulation points in these protein complexes and plausible associated functions. For example, the relatively large PhoScore value of the Rpn3 subunit in the 19S proteasome regulatory particle might suggest that this protein subunit is a point of regulation for proteasome function, and indeed it was recently demonstrated that the Rpn3 subunit of mouse proteasomes is subject to phosphorylation as the proteasome ages, leading to increased proteasome turnover (Tomita *et al.*, 2019). These representations were created in the FAT-PTM pathway drawing function that can be accessed following registration for an account in FAT-PTM. This tool can be utilized to draw representations of metabolic pathways, protein complexes, signaling pathways or any other desired subset of proteins and visualize each node in the pathway as user-defined protein families. In addition, these representations can be submitted through an account to improve annotations of these pathways as new data become available. Overall, the FAT-PTM 'Pathway' tool allows users to display PTM events visually in the context of enzyme-centric metabolic networks or create user-defined networks of protein complexes or signaling pathways, which may facilitate the identification of critical signaling regulatory points and associated PTM sites that could be manipulated to control plant metabolism, growth or development.

The FAT-PTM 'Co-modification' tool

Due to improvements in enrichment strategies for different types of PTMs (Larsen *et al.*, 2005; Xu *et al.*, 2010, 2017; Kim *et al.*, 2011; Majeran *et al.*, 2018; Zeng *et al.*, 2018), it has become increasingly clear that many proteins are modified by multiple PTMs and that the combination of these PTMs often dictates different protein regulatory states (Beltrao *et al.*, 2013). We reasoned that it would be beneficial to develop a tool that quickly identifies proteins that contain multiple user-defined PTMs and output data related to these proteins. Therefore, we developed the 'Co-modification' module (Figure 6), which allows users to select proteins in the FAT-PTM dataset that are modified with multiple PTMs. The output of each search is a Venn diagram (Figure 6a), and users can scroll over the union of the Venn diagram to identify all proteins that are modified by both selected PTMs as well as output the list of co-modified proteins and the sites that are modified. Currently, the 'Co-modification' tool supports eight different PTMs (Table 1), and an analysis performed with the 'Co-modification' tool indicates that 2728, 1280, 340, 49 and 4 proteins in the current FAT-PTM database are subject to two, three, four, five or six different types of unique PTM based on experimentally supported datasets (Figure 6b).

To demonstrate the utility of the 'Co-modification' module, this tool was used to identify sites in which a single amino acid in a protein was modified by multiple PTMs. The lysine modifications ubiquitinylation, SUMOylation and acetylation were first examined, and we found 234 proteins containing 289 precisely mapped sites that have shared lysine acetylation and ubiquitinylation sites, suggesting that competition between lysine acetylation and ubiquitinylation sites is widespread in Arabidopsis. Furthermore, two proteins (At2g17410.1 and At5g62390.1) contain two precisely mapped sites (K447 and K212 respectively) that have shared lysine acetylation and SUMOylation sites. While this result does indicate that these two modifications can compete with one another, this number of shared sites is quite low compared with other organisms and may reflect the currently low number of SUMOylation events with precise sequence localization in the FAT-PTM database. We further examined the PTM site overlap between serine/threonine phosphorylation and *O*-GlcNAc modification, since site-specific competitive modification of proteins with these PTMs is known to differentially regulate protein function (van der Laarse *et al.*, 2018). In this analysis, we identified 29 proteins and 37 serine or threonine sites that contained experimentally supported phosphorylation and *O*-GlcNAcylation modifications. Figure 6(c) illustrates two examples of these modification sites in At1 g79090 and At3 g13990. These results suggest that the competitive regulation between phosphorylation and *O*-GlcNAcylation modifications is relatively widespread in Arabidopsis. Overall, the Co-PTM module can be used to quickly identify groups of proteins that are modified by multiple user-defined PTMs and examine functional relationships among these co-modified proteins.

DISCUSSION

Post-translational modifications can regulate virtually every aspect of protein behavior in cellular contexts, and advances in protein mass spectrometry in recent years have facilitated the identification and quantification of unprecedented numbers of PTM events. Various databases have been created to house this information, but many of these databases are devoted to individual types of PTM, do

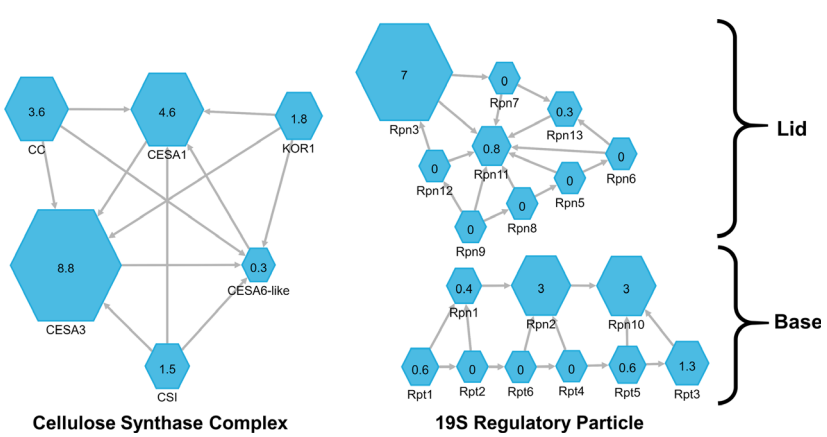

Figure 5. PhoScore representations of protein complexes using the Functional Analysis Tools for Post-Translational Modification (FAT-PTM) pathway drawing tool.

The FAT-PTM database contains a pathway drawing tool that allows users to construct defined signaling or metabolic pathways as well as protein complexes. The subunits of (a) the cellulose synthase complex and (b) the 19S proteasome regulatory particle are shown as examples. The pathway drawing tool allows users to define individual proteins that comprise protein families, which are represented as nodes. Additionally, users can define connections between nodes in a user-defined manner.

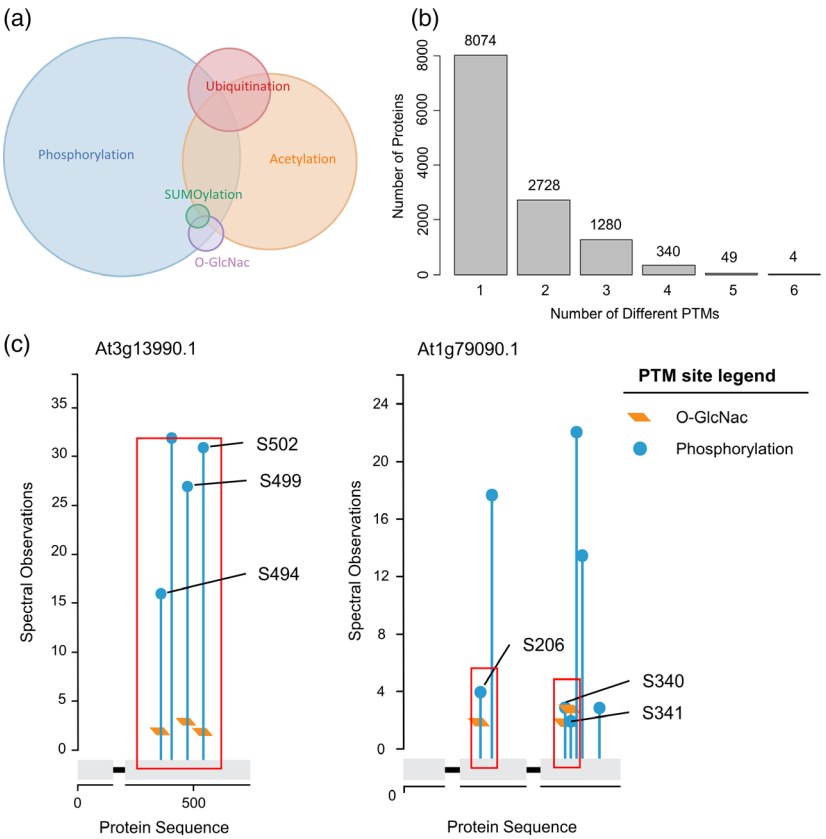

Figure 6. The Functional Analysis Tools for Post-Translational Modification (FAT-PTM) 'Co-modification' tool.
The FAT-PTM database contains a 'Co-modification' module that allows users to select multiple protein modifications and output a list of proteins that are modified by these multiple modifications.
(a) The initial output of these results is a Venn diagram, where the junction of the Venn diagram illustrates proteins that are co-modified.
(b) The number of proteins that contain multiple modifications in FAT-PTM.
(c) Lollipop plot examples of proteins that are modified at the same site by multiple modifications. Phosphorylation sites are indicated with blue circles and *O*-GlcNAc sites are indicated with green parallelograms. The number of spectra experimentally supporting each site is indicated on the *y*-axis, and the residue indices of co-modified sites are indicated.

not include quantitative PTM information or do not provide a broader context for PTM information outside of the individually modified protein. In this study, we have developed the FAT-PTM database, which seeks to circumvent some of these limitations and facilitate the identification of emergent patterns in PTM data. The FAT-PTM database allows the visualization of multiple PTM types in individual proteins or protein networks based on experimentally supported PTM sites (Figure 2). Additionally, FAT-PTM can display quantitative phosphorylation site dynamics based on multiple studies that use relative quantification of PTM dynamics under a variety of developmental or stress conditions (Figure 2b). In addition, FAT-PTM can display PTM networks in the context of larger metabolic pathways or protein complexes in order to gain a clearer understanding of how PTMs may regulate these groups of proteins (Figures 4 and 5), and the database can quickly identify groups of proteins that are co-modified by more than one PTM (Figure 6). Overall, these unique tools may lead to a more comprehensive understanding of PTM regulatory processes in Arabidopsis.

The large majority of plant proteomic studies focused on PTMs have been performed in Arabidopsis, and PTM dynamics in this organism are likely to be important for understanding how PTM dynamics influence growth and development in agronomically important plant species. However, an increasing number of proteomic studies are focusing on these important crop species, and it will therefore be an important future goal to integrate other plant species into the FAT-PTM database, including rice, soybean, corn and others. Additionally, establishing phosphoproteomic visualization resources for an evolutionary continuum of plant species will be critically important for understanding the evolution of PTM dynamics across a wide array of plant species. This information may also facilitate the identification of ancient and conserved PTM sites that critically influence protein function in metabolic or signaling pathways.

A unique aspect of FAT-PTM compared with other plant-focused PTM databases is the ability to quickly visualize and export quantitative phosphoproteomic data from a variety of experimental conditions. These tools may allow for the quick assessment of which phosphorylation sites in a target protein are regulated by a particular developmental or environmental condition and then develop functional analysis strategies accordingly. For example, this tool may suggest experimental conditions that could be implemented to evaluate phosphorylation site mutants that may not have a physiological effect under normal growth conditions but may exhibit phenotypic defects under specific stress conditions. Based on the rapid rate of adoption of quantitative mass spectrometry-based strategies for the

evaluation of PTM dynamics, we anticipate that numerous additional conditions may be added to FAT-PTM in the future. Additionally, these quantitative methods are increasingly being applied to the study of non-phosphorylation modifications, such as lysine acetylation (Liu *et al.*, 2018), so quantitative information regarding these modifications will be added in the future. This inclusion of multiple quantitative datasets may allow for the investigation of the relationship and co-variance of multiple PTMs under identical growth conditions.

Increasing evidence also suggests that proteins are subject to multiple PTMs, and that these PTM patterns can be used to integrate multiple stimuli for a combined protein output (Beltrao *et al.*, 2013; van der Laarse *et al.*, 2018). Indeed, our current database indicates that 35.3% of the Arabidopsis proteins in the FAT-PTM database are modified by two or more distinct PTMs. The 'Co-modification' tool that we have developed as well as the visualization tools contained within FAT-PTM to examine multiple PTMs will be useful for understanding how these regulatory modifications influence the global regulation of protein behavior and activity in the context of signaling and metabolic pathways. Through the use of these tools, more informed functional analysis experiments can be designed to examine the downstream effect of these modifications.

The rate-limiting step in the evaluation of PTM function is developing a functional understanding of the role that each PTM plays in cellular and organismal growth and development. While mass spectrometry has led to the identification and quantification of thousands of PTM sites in a single experiment, functional evaluation of such sites is still evaluated by laborious and non-ideal methods. The FAT-PTM database is designed to serve as a hypothesis-generating platform to identify individual PTM sites that are worthy of functional analysis, based on their degree of experimental support or quantitative abundance changes under particular conditions. However, it is exciting to note that new technological developments may facilitate more rapid investigations of PTM function in target proteins. For example, orthogonal translation systems that allow for the site-specific incorporation of unnatural amino acids, such as acetyl-lysine (Guo *et al.*, 2008) and phosphoserine (Pirman *et al.*, 2015), could be used to evaluate how individual PTMs identified in FAT-PTM functionally impact target proteins in a relatively high-throughput fashion (Barber *et al.*, 2018). Additionally, the development of high-precision genome-editing strategies using methods based on CRISPR–Cas9 (Komor *et al.*, 2016; Gaudelli *et al.*, 2017; Roy *et al.*, 2018) will conceivably allow the high-throughput targeted mutation and functional evaluation of PTM sites in plant genomes, suggesting that PTM sites in plant species may be functionally analyzed more rapidly.

EXPERIMENTAL PROCEDURES

Collection of experimental phosphorylation data and pathway annotations

All genes in the Arabidopsis genome were retrieved from Araport, including annotations and protein sequences for each AGI identification number. All AGI identification numbers were individually queried in the PhosPhAt 4.0 application (Zulawski *et al.*, 2012) through a web crawler that parsed all available experimental data including modified peptide sequence, PubMed ID, treatment and tissue. For metabolic networks, Arabidopsis-specific metabolic pathway information was obtained from AraCyc (Mueller *et al.*, 2003), part of the Plant Metabolic Network, and the reaction and pathway information was mapped from the PathwayTools flat files by creating a reaction × gene matrix and a reaction × compound/class matrix. The reactions were merged with the reaction identification numbers in each pathway, and the enzyme-centric metabolic networks formed were manually curated and reviewed.

Data processing and website structure

When forming the experimental data for use in FAT-PTM from PhosPhAt 4.0, phosphorylation site predictions were omitted due to a lack of spectral support by only accounting for modified peptide sequences that had precisely mapped phosphorylation events indicated by (pS), (pT) and (pY). The additional data encompassed in FAT-PTM, including modification type, localization site and total number of spectral observations supporting each PTM site, were collected from published studies in the literature (Miller *et al.*, 2010, 2013; Hemsley *et al.*, 2013; Kim *et al.*, 2013; Svozil *et al.*, 2014; Hu *et al.*, 2015; Johnson and Vert, 2016; Walton *et al.*, 2016; Xu *et al.*, 2016, 2017; Koskela *et al.*, 2018; Liu *et al.*, 2018; Majeran *et al.*, 2018; Rytz *et al.*, 2018; Zeng *et al.*, 2018). For quantitative phosphoproteomics studies, abundance fold-change information for phosphopeptides based on ratiometric quantification of isobaric tag or stable isotope reporters was also collected from supplemental files in the following references (Benschop *et al.*, 2007; Nühse *et al.*, 2007; Mithoe *et al.*, 2012; Zhang *et al.*, 2013, 2018; Chen and Hoehenwarter, 2015; Minkoff *et al.*, 2015; Roitinger *et al.*, 2015; Cho *et al.*, 2016; Bhaskara *et al.*, 2017; Wang *et al.*, 2018). The resulting information retrieved from PhosPhAt and other sources was stored in a MongoDB database management system. The website was developed using NodeJS and deployed on a Linux server.

Calculation of PhoScore

For an individual phosphorylated protein, PhoScore is calculated by the following equation:

$$\text{PhoScore} = \log(\text{Spectral Count} \times \text{Site Count})$$

where Spectral Count is the total number of phosphorylation site spectra observed for a protein and Site Count is the total number of experimentally supported phosphorylation sites a protein contains.

A Combined PhoScore for a protein group containing *n* family members is calculated by the following equation:

$$\text{Combined PhoScore} = \frac{\sum_{i=1}^{n} \text{PS}_i}{n}$$

where PS is the PhoScore of an individual protein in the enzyme class and *n* is the total number of proteins in an enzyme class with experimentally supported phosphorylation spectra.

ACKNOWLEDGEMENTS

This work was supported in part by a National Science Foundation CAREER Award (MCB 1750359) to ISW, a University of Nevada, Reno Office of the Vice President for Research and Innovation micro grant to ISW and TN, the Ronald E. McNair Post-Baccalaureate Achievement Program to ERC and an ASPB 2019 Summer Undergraduate Research Fellowship to ERC. We would also like to thank Dr Jeffrey F. Harper, Won C. Yim and Won-Gyu Choi for helpful commentary during the writing of this manuscript.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

ERC and ISW conceived the original research. ISW and TN supervised the experiments. ERC and HN performed the experiments. ERC, HN, TN and ISW wrote the manuscript.

DATA AVAILABILITY

We collected the data used to construct the FAT-PTM database from PhosPhat4.0 (http://phosphat.uni-hohenheim.de/) and studies in the literature, which are cited in the body of the manuscript. All data associated with various protein modifications (AGI number, type of modification, number of spectral observations, quantitative fold changes and PubMed IDs of original publications producing the data) can be exported from the FAT-PTM website (http://bioinformatics.cse.unr.edu/fat-ptm).

REFERENCES

Augustine, R.C. and Vierstra, R.D. (2018) SUMOylation: re-wiring the plant nucleus during stress and development. *Curr. Opin. Plant Biol.* **45**, 143–154.

Barber, K.W., Muir, P., Szeligowski, R.V., Rogulina, S., Gerstein, M., Sampson, J.R., Isaacs, F.J. and Rinehart, J. (2018) Encoding human serine phosphopeptides in bacteria for proteome-wide identification of phosphorylation-dependent interactions. *Nat. Biotechnol.* **36**, 638–644.

Bateman, A., Smart, A., Luciani, A. *et al.* (2018) The Pfam protein families database in 2019. *Nucleic Acids Res.* **47**, D427–D432.

Beltrao, P., Bork, P., Krogan, N.J. and van Noort, V. (2013) Evolution and functional cross-talk of protein post-translational modifications. *Mol. Syst. Biol.* **9**, 714.

Benschop, J.J., Mohammed, S., O'Flaherty, M., Heck, A.J.R., Slijper, M. and Menke, F.L.H. (2007) Quantitative phosphoproteomics of early elicitor signaling in Arabidopsis. *Mol. Cell Proteomics*, **6**, 1198–1214.

Berardini, T.Z., Reiser, L., Li, D., Mezheritsky, Y., Muller, R., Strait, E. and Huala, E. (2015) The Arabidopsis information resource: making and mining the "gold standard" annotated reference plant genome. *Genesis*, **53**, 474–485.

Bhaskara, G.B., Wen, T.-N., Nguyen, T.T. and Verslues, P.E. (2017) Protein phosphatase 2Cs and microtubule-associated stress protein 1 control microtubule stability, plant growth, and drought response. *Plant Cell*, **29**, 169–191.

Breitkreutz, B.-J., Stark, C., Dolinski, K. *et al.* (2013) The PhosphoGRID Saccharomyces cerevisiae protein phosphorylation site database: version 2.0 update. *Database*, **2013**, bat026.

Chen, Y. and Hoehenwarter, W. (2015) Changes in the phosphoproteome and metabolome link early signaling events to rearrangement of photosynthesis and central metabolism in salinity and oxidative stress response in Arabidopsis. *Plant Physiol.* **169**, 3021–3033.

Cho, H.-Y., Shih, M.-C., Wang, Y.-T. and Wen, T.-N. (2016) Quantitative phosphoproteomics of protein kinase SnRK1 regulated protein phosphorylation in Arabidopsis under submergence. *J. Exp. Bot.* **67**, 2745–2760.

Clark, T.J. and Lu, Y. (2015) Analysis of loss-of-function mutants in aspartate kinase and homoserine dehydrogenase genes points to complexity in the regulation of aspartate-derived amino acid contents. *Plant Physiol.* **168**, 1512–1526.

Coon, J.J., Ueberheide, B., Syka, J.E.P., Dryhurst, D.D., Ausio, J., Shabanowitz, J. and Hunt, D.F. (2005) Protein identification using sequential ion/ion reactions and tandem mass spectrometry. *Proc. Natl Acad. Sci. USA*, **93**, 9463–9468.

Garavelli, J.S. (2004) The RESID database of protein modifications as a resource and annotation tool. *Proteomics*, **4**, 1527–1533.

Gaudelli, N.M., Komor, A.C., Rees, H.A., Packer, M.S., Badran, A.H., Bryson, D.I. and Liu, D.R. (2017) Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage. *Nature*, **551**, 464–471.

Gnad, F., Gunawardena, J. and Mann, M. (2010) PHOSIDA 2011: the post-translational modification database. *Nucleic Acids Res.* **39**, D253–D260.

Guo, J., Wang, J., Lee, J.S. and Schultz, P.G. (2008) Site-specific incorporation of methyl- and acetyl-lysine analogues into recombinant proteins. *Angew. Chem. Int. Ed. Engl.* **47**, 6399–6401.

Hemsley, P.A., Weimar, T., Lilley, K.S., Dupree, P. and Grierson, C.S. (2013) A proteomic approach identifies many novel palmitoylated proteins in Arabidopsis. *New Phytol.* **197**, 805–814.

Hu, J., Huang, X., Chen, L., Sun, X., Lu, C., Zhang, L., Wang, Y. and Zuo, J. (2015) Site-specific nitrosoproteomic identification of endogenously S-nitrosylated proteins in Arabidopsis. *Plant Physiol.* **167**, 1731–1746.

Huang, X., Huang, L., Peng, H. *et al.* (2013) ISPTM: an iterative search algorithm for systematic identification of post-translational modifications from complex proteome mixtures. *J. Proteome Res.* **12**, 3831–3842.

Huang, H., Arighi, C.N., Ross, K.E., Ren, J., Li, G., Chen, S.-C., Wang, Q., Cowart, J., Vijay-Shanker, K. and Wu, C.H. (2017) iPTMnet: an integrated resource for protein post-translational modification network discovery. *Nucleic Acids Res.* **46**, D542–D550.

Johnson, A. and Vert, G. (2016) Unraveling K63 polyubiquitination networks by sensor-based proteomics. *Plant Physiol.* **171**, 1808–1820.

Kao, H.-J., Jhong, J.-H., Huang, K.-Y., Cheng, K.-H., Su, M.-G., Hsieh, Y.-C., Lee, T.-Y. and Huang, H.-D. (2015) dbPTM 2016: 10-year anniversary of a resource for post-translational modification of proteins. *Nucleic Acids Res.* **44**, D435–D446.

Kim, W., Bennett, E.J., Huttlin, E.L. *et al.* (2011) Systematic and quantitative assessment of the ubiquitin-modified proteome. *Mol. Cell* **44**, 325–340.

Kim, D.-Y., Scalf, M., Smith, L.M. and Vierstra, R.D. (2013) Advanced proteomic analyses yield a deep catalog of ubiquitylation targets in Arabidopsis. *Plant Cell*, **25**, 1523–1540.

Komor, A.C., Kim, Y.B., Packer, M.S., Zuris, J.A. and Liu, D.R. (2016) Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature*, **533**, 420–424.

Koskela, M.M., Brünje, A., Ivanauskaite, A. *et al.* (2018) Chloroplast acetyltransferase NSI is required for state transitions in *Arabidopsis thaliana*. *Plant Cell*, **30**, 1695–1709.

Kühn, C. (2016) Review: post-translational cross-talk between brassinosteroid and sucrose signaling. *Plant Sci.* **248**, 75–81.

van der Laarse, S.A.M., Leney, A.C. and Heck, A.J.R. (2018) Crosstalk between phosphorylation and O-GlcNAcylation: friend or foe. *FEBS J.* **285**, 3152–3167.

Larsen, M.R., Thingholm, T.E., Jensen, O.N., Roepstorff, P. and Jørgensen, T.J.D. (2005) Highly selective enrichment of phosphorylated peptides from peptide mixtures using titanium dioxide microcolumns. *Mol. Cell Proteomics*, **4**, 873–886.

Liu, S., Yu, F., Yang, Z., Wang, T., Xiong, H., Chang, C., Yu, W. and Li, N. (2018) Establishment of dimethyl labeling-based quantitative acetylproteomics in Arabidopsis. *Mol. Cell Proteomics*, **17**, 1010–1027.

Majeran, W., Le Caer, J.-P., Ponnala, L., Meinnel, T. and Giglione, C. (2018) Targeted profiling of *Arabidopsis thaliana* subproteomes illuminates co- and posttranslationally N-terminal myristoylated proteins. *Plant Cell*, **30**, 543–562.

Miller, M.J., Barrett-Wilt, G.A., Hua, Z. and Vierstra, R.D. (2010) Proteomic analyses identify a diverse array of nuclear processes affected by small ubiquitin-like modifier conjugation in Arabidopsis. *Proc. Natl Acad. Sci. USA*, **107**, 16512–16517.

Miller, M.J., Scalf, M., Rytz, T.C., Hubler, S.L., Smith, L.M. and Vierstra, R.D. (2013) Quantitative proteomics reveals factors regulating RNA biology as

dynamic targets of stress-induced SUMOylation in Arabidopsis. *Mol. Cell Proteomics*, **12**, 449–463.

Minkoff, B.B., Stecker, K.E. and Sussman, M.R. (2015) Rapid phosphoproteomic effects of abscisic acid (ABA) on wild-type and ABA receptor-deficient *A. thaliana* mutants. *Mol. Cell Proteomics*, **14**, 1169–1182.

Mithoe, S.C., Boersema, P.J., Berke, L., Snel, B., Heck, A.J.R. and Menke, F.L.H. (2012) Targeted Quantitative Phosphoproteomics Approach for the Detection of Phospho-tyrosine Signaling in Plants. *J. Proteome Res.* **11**, 438–448.

Mueller, L.A., Zhang, P. and Rhee, S.Y. (2003) AraCyc: a biochemical pathway database for Arabidopsis. *Plant Physiol.* **132**, 453–460.

Nühse, T.S., Bottrill, A.R., Jones, A.M.E. and Peck, S.C. (2007) Quantitative phosphoproteomic analysis of plasma membrane proteins reveals regulatory mechanisms of plant innate immune responses. *Plant J.* **51**, 931–940.

Pirman, N.L., Barber, K.W., Aerni, H.R., Ma, N.J., Haimovich, A.D., Rogulina, S., Isaacs, F.J. and Rinehart, J. (2015) A flexible codon in genomically recoded *Escherichia coli* permits programmable protein phosphorylation. *Nat. Commun.* **6**, 8130.

Roitinger, E., Hofer, M., Köcher, T., Pichler, P., Novatchkova, M., Yang, J., Schlögelhofer, P. and Mechtler, K. (2015) Quantitative phosphoproteomics of the ataxia telangiectasia-mutated (ATM) and ataxia telangiectasia-mutated and Rad3-related (ATR) dependent DNA damage response in *Arabidopsis thaliana*. *Mol. Cell Proteomics*, **14**, 556–571.

Rosen, B.D., Cheng, C.-Y., Town, C.D. *et al.* (2014) Araport: the Arabidopsis information portal. *Nucleic Acids Res.* **43**, D1003–D1009.

Roy, K.R., Smith, J.D., Vonesch, S.C. *et al.* (2018) Multiplexed precision genome editing with trackable genomic barcodes in yeast. *Nat. Biotechnol.* **36**, 512–520.

Rytz, T.C., Miller, M.J., McLoughlin, F., Augustine, R.C., Marshall, R.S., Juan, Y., Charng, Y., Scalf, M., Smith, L.M. and Vierstra, R.D. (2018) SUMOylome profiling reveals a diverse array of nuclear targets modified by the SUMO ligase SIZ1 during heat stress. *Plant Cell*, **30**, 1077–1099.

Schläpfer, P., Zhang, P., Wang, C. *et al.* (2017) Genome-wide prediction of metabolic enzymes, pathways, and gene clusters in plants. *Plant Physiol.* **173**, 2041–2059.

Strasser, R. (2016) Plant protein glycosylation. *Glycobiology*, **26**, 926–939.

Svozil, J., Hirsch-Hoffmann, M., Dudler, R., Gruissem, W. and Baerenfaller, K. (2014) Protein abundance changes and ubiquitylation targets identified after inhibition of the proteasome with syringolin A. *Mol. Cell Proteomics*, **13**, 1523–1536.

Tomita, T., Hirayama, S., Sakurai, Y., Ohte, Y., Yoshihara, H., Saeki, Y., Hamazaki, J. and Murata, S. (2019) Specific modification of aged proteasomes revealed by tag-exchangeable knock-in mice. *Mol. Cell. Biol.* **39**, e00426–18.

Tsur, D., Tanner, S., Zandi, E., Bafna, V. and Pevzner, P.A. (2005) Identification of post-translational modifications by blind search of mass spectra. *Nat. Biotechnol.* **23**, 1562–1567.

Turnbull, D. and Hemsley, P.A. (2017) Fats and function: protein lipid modifications in plant cell signalling. *Curr. Opin. Plant Biol.* **40**, 63–70.

Via, A., Gould, C.M., Chica, C., Dinkel, H., Jensen, L.J., Diella, F. and Gibson, T.J. (2010) Phospho.ELM: a database of phosphorylation sites—update 2011. *Nucleic Acids Res.* **39**, D261–D267.

Walton, A., Stes, E., Cybulski, N. *et al.* (2016) It's time for some "site"-seeing: novel tools to monitor the ubiquitin landscape in *Arabidopsis thaliana*. *Plant Cell*, **28**, 6–16.

Wang, K., Yang, Z., Qing, D. *et al.* (2018) Quantitative and functional post-translational modification proteomics reveals that TREPH1 plays a role in plant touch-delayed bolting. *Proc. Natl Acad. Sci. USA*, **115**, E10265–E10274.

Withers, J. and Dong, X. (2017) Post-translational regulation of plant immunity. *Curr. Opin. Plant Biol.* **38**, 124–132.

Xu, G., Paige, J.S. and Jaffrey, S.R. (2010) Global analysis of lysine ubiquitination by ubiquitin remnant immunoaffinity profiling. *Nat. Biotechnol.* **28**, 868–873.

Xu, S.-L., Medzihradszky, K.F., Wang, Z.-Y., Burlingame, A.L. and Chalkley, R.J. (2016) N-glycopeptide profiling in arabidopsis inflorescence. *Mol. Cell Proteomics*, **15**, 2048–2054.

Xu, S.-L., Chalkley, R.J., Maynard, J.C. *et al.* (2017) Proteomic analysis reveals O-GlcNAc modification on proteins with key regulatory functions in Arabidopsis. *Proc. Natl Acad. Sci. USA*, **114**, E1536–E1543.

Yao, Q., Ge, H., Wu, S., Zhang, N., Chen, W., Xu, C., Gao, J., Thelen, J.J. and Xu, D. (2013) P3DB 3.0: from plant phosphorylation sites to protein networks. *Nucleic Acids Res.* **42**, D1206–D1213.

Zeng, W., Ford, K.L., Bacic, A. and Heazlewood, J.L. (2018) N-linked glycan micro-heterogeneity in glycoproteins of arabidopsis. *Mol. Cell Proteomics*, **17**, 413–421.

Zhang, H., Zhou, H., Berke, L., Heck, A.J.R., Mohammed, S., Scheres, B. and Menke, F.L.H. (2013) Quantitative phosphoproteomics after auxin-stimulated lateral root induction identifies an SNX1 protein phosphorylation site required for growth. *Mol. Cell Proteomics*, **12**, 1158–1169.

Zhang, C., Guo, X., Xie, H. *et al.* (2018) Quantitative phosphoproteomics of lectin receptor-like kinase VI.4 dependent abscisic acid response in *Arabidopsis thaliana*. *Physiol. Plant.* **165**, 728–745.

Zulawski, M., Braginets, R. and Schulze, W.X. (2012) PhosPhAt goes kinases—searchable protein kinase target information in the plant phosphorylation site database PhosPhAt. *Nucleic Acids Res.* **41**, D1176–D1184.

