

qPTM: an updated database for PTM dynamics in human, mouse, rat and yeast

Kai Yu^{1,†}, Ye Wang^{1,†}, Yongqiang Zheng^{1,†}, Zekun Liu^{1,†}, Qingfeng Zhang¹, Siyu Wang¹, Qi Zhao¹, Xiaolong Zhang¹, Xiaoxing Li², Rui-Hua Xu^{1,3,*} and Ze-Xian Liu^{1,*}

¹State Key Laboratory of Oncology in South China, Collaborative Innovation Center for Cancer Medicine, Sun Yat-sen University Cancer Center, Guangzhou 510060, China, ²Precision Medicine Institute, First Affiliated Hospital, Sun Yat-sen University, Guangzhou 510080, China and ³Research Unit of Precision Diagnosis and Treatment for Gastrointestinal Cancer, Chinese Academy of Medical Sciences, Guangzhou 510060, China

Received July 21, 2022; Revised August 26, 2022; Editorial Decision September 08, 2022; Accepted September 14, 2022

ABSTRACT

Post-translational modifications (PTMs) are critical molecular mechanisms that regulate protein functions temporally and spatially in various organisms. Since most PTMs are dynamically regulated, quantifying PTM events under different states is crucial for understanding biological processes and diseases. With the rapid development of high-throughput proteomics technologies, massive quantitative PTM proteome datasets have been generated. Thus, a comprehensive one-stop data resource for surfing big data will benefit the community. Here, we updated our previous phosphorylation dynamics database qPhos to the qPTM (<http://qptm.omicsbio.info>). In qPTM, 11 482 553 quantification events among six types of PTMs, including phosphorylation, acetylation, glycosylation, methylation, SUMOylation and ubiquitylation in four different organisms were collected and integrated, and the matched proteome datasets were included if available. The raw mass spectrometry based false discovery rate control and the recurrences of identifications among datasets were integrated into a scoring system to assess the reliability of the PTM sites. Browse and search functions were improved to facilitate users in swiftly and accurately acquiring specific information. The results page was revised with more abundant annotations, and time-course dynamics data were visualized in trend lines. We expected the qPTM database to be a much more powerful and comprehensive data repository for the PTM research community.

INTRODUCTION

Protein post-translational modification (PTM), a reversible covalent modification after protein translation, is one of the most important regulatory mechanisms in physiological processes and diseases/cancers (1–9). Various PTMs regulate protein structure and function dynamics by altering residue hydrophobicity, charge state and protein stability (1,2). Massive studies have explored PTMs since Swedish scientist Olof Hammarsten first discovered phosphorylation in 1883 (10). Over 680 types of PTMs have been discovered thus far (<https://www.uniprot.org/docs/ptmlist>), among which phosphorylation and ubiquitination have been the most studied due to their high abundance in cells (3,6). In 1992, Edmond H. Fischer and Edwin G. Krebs shared the Nobel Prize in Physiology or Medicine for their discovery of reversible protein phosphorylation as a biological regulatory mechanism (11), while Irwin Rose, Aaron Ciechanover and Avram Hershko were rewarded the prize in 2004 for identifying ubiquitination as a transferable signal for the degradation of proteins by the proteasome (12). Recently, lysine modifications, such as acetylation (13), crotonylation (14), succinylation (15) and lactylation (15) were discovered as ubiquitous PTMs. Various PTMs orchestrate biological processes (1–9). Thus, dissecting PTM dynamics is critical for understanding the cellular signaling network.

Recently, high-throughput proteomics techniques have greatly advanced and boosted the identification and quantification of PTM events in cells and organs (16–18). For example, quantitative phosphoproteome profiling was performed in proteogenomic characterization of patient cohorts with cancers, such as lung adenocarcinoma (19–21) and hepatocellular carcinoma (22,23), and dissection of the aberrances of phosphorylation signaling provided important clues of carcinogenesis, cancer development and treat-

*To whom correspondence should be addressed. Tel: +86 20 8734 2025; Fax: +86 20 8734 2522; Email: liuzx@sysucc.org.cn
Correspondence may also be addressed to Rui-Hua Xu. Tel: +86 20 8734 3228; Fax: +86 20 8734 3392; Email: xurh@sysucc.org.cn

†The authors wish it to be known that, in their opinion, the first four authors should be regarded as Joint First Authors.

ment. Quantitative *N*-glycoproteome analyses were performed in breast cancer and high-grade serous ovarian cancer studies (24,25). Sun *et al.* discovered that the ubiquitination of Ku80 was closely associated with the invasion and migration of HCC cells through crosstalk analysis between the proteome and ubiquitylome (26). Acetylome quantification by Krug *et al.* revealed the crucial roles of acetylation in regulating the DNA damage response and metabolism (27). Furthermore, Grimes *et al.* outlined the cell signaling pathway of lung cancer by coupling analyses of the phosphoproteome, methylproteome, and acetylproteome in lung cancer cell lines (28). Taken together, quantitative PTMomics data could facilitate the understanding of molecular biological mechanisms.

With the sharp increase in profiled PTM events, a series of distinguished studies have contributed to hosting the massive amount of PTM data. As the most comprehensive infrastructure of the protein knowledgebase, UniProt curated and hosted massive functional annotations, including PTMs (29). Specialized databases, including PhosphositePlus (30) and dbPTM (31), were constructed and maintained for over 15 years to curate PTM events, and databases, such as SysPTM (32), HPRD (33) and PHOSIDA (34), have also contributed significantly, while iPTMnet (35) and PTMcode (36) have provided integrated resources for the network regulation and functional associations of PTMs. Several databases have also been developed to collect PTM-specific data, such as dbPAF/EPSPD (37,38) for phosphorylation, hUbiquitome (39) for ubiquitination, CarbonylDB (40) for carbonylation, O-GlycBase/N-GlycositeAtlas/GlycoProtDB (41–43) for glycosylation, and CPLA/CPLM/PLMD (44–46) for lysine modifications. Furthermore, the ProteomeXchange (PX) consortium members (PRIDE, PeptideAtlas, MASSIVE, jPOST, iProX, Panorama Public) (47) hosted the mass spectrometry proteomics (including PTM proteomics) data, while ProteomeScout (48), ProteomicsDB (48,49) and piNET (50) were developed to store proteomics datasets and provide online analysis tools. However, these data repositories did not provide easy access to PTM dynamics among the different conditions/states.

Previously, we developed the qPhos (51) database to host the phosphorylation dynamics data. Here, we updated qPhos to qPTM (<http://qptm.omicsbio.info>), which contains six types of PTMs, including phosphorylation, acetylation, glycosylation, methylation, SUMOylation and ubiquitylation in four different organisms, including human, mouse, rat and yeast. In total, 11 482 553 quantification events for 660 030 sites on 40 728 proteins were collected and integrated into qPTM, and the matched proteome datasets were curated if available. With the limited available raw mass spectrometry (MS) data, 8 658 490 quantification events were confidently identified with 1% site-level false discovery rate (FDR) control. Furthermore, the newly designed browse and search pages enable users to quickly retrieve interested data from over tens of millions of quantitative events. In addition, we visualized time-course and treatment concentration gradient dynamic data to help users better understand the trends over time or concentration. Taken together, the qPTM database provides a comprehensive platform to access quantitative PTMomics data for the

community and is a reliable resource for further computational or experimental considerations.

DATA COLLECTION AND DATABASE CONSTRUCTION

By integrating quantitative PTMomics datasets curated from published literature and annotations from various resources, we established a comprehensive platform for host PTM dynamics data. The scheme for constructing the qPTM database is shown in Figure 1. The quantified events were collected from the literature published before January 2022 in PubMed for PTMs, including acetylation, glycosylation, methylation, phosphorylation, SUMOylation and ubiquitylation, which are the most studied PTMs. Text mining was conducted on the abstracts of the literature by searching PTM-related words, including ‘phosphoproteome’, ‘acetylproteome’, ‘*N*-glycoproteome’, ‘*O*-glycoproteome’, ‘lysine methylome’, ‘arginine methylome’, ‘ubiquitylome’ and ‘SUMO proteome’. To avoid missing data, additional keywords, such as ‘lysine acetylation’, ‘acetylome’ and ‘protein SUMOylation’, were also applied. To avoid data explosion, nomenclatures from high-throughput mass spectrometry experiments, including ‘quantitative’, ‘label-free’, ‘SILAC’, ‘enrichment’ and ‘mass spectrometer’, were employed as qualifier words. Quantitative PTMomics datasets were collected from ProteomeXchange (47) if not available in the literature.

Based on the coverage of PTMomics studies in model organisms, quantitative PTM datasets in *Homo sapiens*, *Mus musculus*, *Rattus norvegicus* and *Saccharomyces cerevisiae* were included in the qPTM database. To guarantee data quality, all matched literature was filtered manually through a stringent procedure as described previously (51). The details about PTM quantifications, including labeling methods, enrichment methods and mass spectrometry, were integrated into the database. In addition, quantitative proteome datasets simultaneously coupled with PTMomics were also collected if available. The modified peptides and residues were remapped to the reference proteome sequence downloaded from the UniProt database (Release 2021_01) (52). Consistent with qPhos, the unmapped raw peptides accounted for 4.19%, 5.87%, 6.16% and 5.93% in human, mouse, rat and yeast, respectively. Identifiers or names of PTMomics were uniformly mapped to UniProtKB accession.

To help users use quantitative PTMs and proteomics data more conveniently, we annotated the quantitative PTM sites with various external resources, including UniProt (52), ExPaSy (53), dbPAF (37), PLMD (44), PTMD (54) and DrugBank (Release 2021_05) (55). The addition of PLMD and dbPAF was a great supplement to UniProt for PTM site information, and PTMD provided PTM-disease association annotations (44). The kinase-substrate relationships were integrated as previously described (51,56,57), while the potential relationship between protein acetyltransferases and acetylation sites was predicted by Deep-PLA (58).

QUALITY CONTROL

To assess the reliability of the PTM sites, we developed a five-star scoring system by integrating the raw MS data-

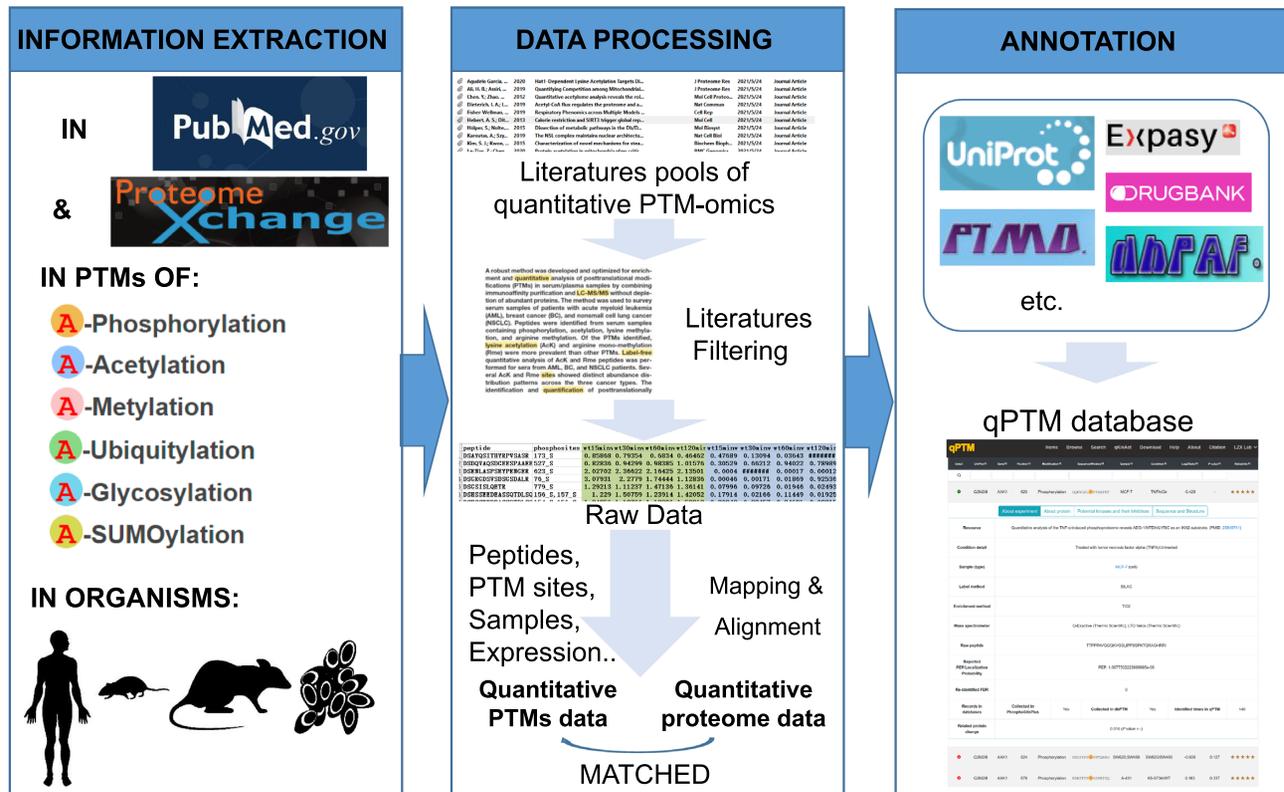


Figure 1. Workflow for the construction of the qPTM database.

based FDR control and the recurrences of identification among datasets. The raw mass spectrometry data and the associated sample information were available for 85% of the curated studies and retrieved from PRIDE (59), iProX (60), jPOSTdb (51) and MassIVE (61). The final raw data files with a total capacity exceeding 25T, consisting of 30 055 MS raw files corresponding to 536 publications, were enrolled in the subsequent MS search. We performed the MS search following previously published search strategies. For specific PTMs, all MS raw files were integrated to jointly search using MaxQuant 1.6.14 (MQ) (62) against the UniProt database (Release 2021.01) (52). Searches were performed with the following FASTA files of the corresponding species from UniProt: UP000005640.9606 (*Homo sapiens*), UP00000589_10090 (*Mus musculus*), UP000002494_10116 (*Rattus norvegicus*) and UP000002311_559292 (*Saccharomyces cerevisiae*). We used Mono 6.12.0.90 to enable MaxQuant to run on the Linux operating system (63). The default values were used for all parameters unless stated otherwise, including 1% PSM FDR and 1% site-level FDR. The minimum peptide length was set to seven amino acids, and peptides were allowed to have a maximum of two missed cleavages. Cysteine (C) carbamidomethylation was set as a fixed modification, while methionine (M) oxidation and protein N-terminal acetylation were used as variable modifications in all searches, as is the default in MaxQuant. For the phosphoproteomic data, phosphorylation of serine (S), threonine (T) and tyrosine (Y) were also set as variable modifications. For the acetylproteomic data, lysine (K) acetylation was set as an additional variable modification.

For glycosylation data, asparagine (N) and glutamine (Q) deamidations were set as additional variable modifications. For methylation data, methylation to lysine (K) or arginine (R), dimethylation to lysine (K) or arginine (R), and trimethylation to lysine (K) were set as additional variable modifications. For SUMO data, QQTGG to lysine (K) was set as an additional variable modification. For ubiquitylation data, GlyGly to lysine (K) was set as an additional variable modification. The identified PTMs and relevant FDRs were annotated to the curated data as re-identified FDR to enhance the reliability of qPTM.

For the datasets without raw MS data, the reliability of the PTM sites could be assessed through the recurrences of identification among datasets, which were characterized by three major standards in the five-star scoring system as follows. (i) The occurrences of the PTM site in the qPTM database. We counted the identified times of each modified site in all datasets. Because the sites were filtered through a cutoff of site-level FDR <1% according to the literature, the more it was identified, the more it was reliable. Thus, modified sites were given zero points when identified once, one point when identified twice, and so on. The identified score was no >4 for each site. (ii) The data reliability in the proteomics study. The posterior error probability (PEP) and localization probability data of each modified site were curated from the datasets if available. PEP, the probability that observed PSM is uncorrected, was under 1%, while sites with a localization probability higher than 75% were considered high-confidence sites. The PTM sites that met the criterion were given a point. (iii) The occurrences of

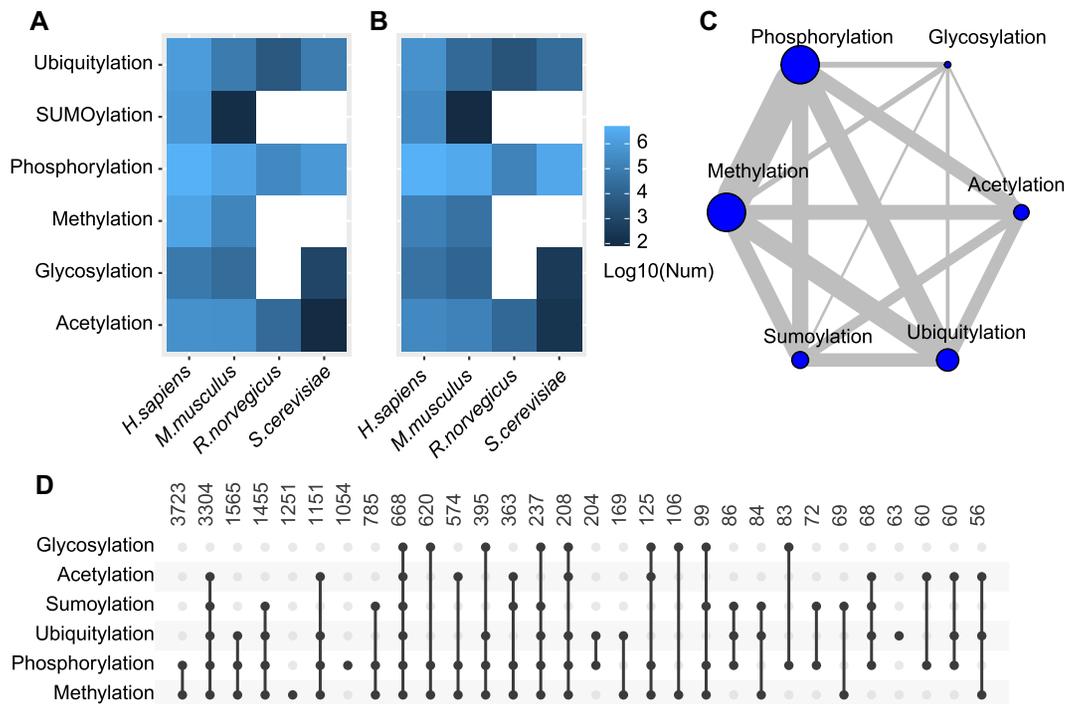


Figure 2. Heatmaps for the (A) modified site number and (B) protein number distribution of different PTMs and model organisms. (C) The proteins that were concurrently modified by the two types of PTMs. Node size and width of the line connecting PTM nodes represent the modified protein number (Log10) of the corresponding PTM and the number of shared modified proteins, respectively. (D) The concurrently modified proteins between no less than three PTMs in *H. sapiens*, while the detailed number is shown at the bottom. Detailed numbers of concurrences are provided in Supplementary Table S2. Crosstalks and concurrently modified proteins of *M. musculus*, *R. norvegicus* and *S. cerevisiae* are shown in Supplementary Figure S2.

the PTM site in other classical PTM databases. Modified sites collected in dbPTM or PhosphoSitePlus were also given a point. Taken together, the five-star scoring system assigned each PTM site with a reliability of 0–5 bright star(s) according to its recurrences of identification among datasets, and red bordered the bright star(s) if it had a site-level FDR < 0.01 according to the raw MS data-based re-identification.

Previously, Ochoa *et al.* (64) curated 112 different datasets of phospho-enriched proteins from 104 different human cells or tissues and compared the identified phosphoproteome against human phosphosites reported by MS in the PhosphoSitePlus database. Similarly, the frequency of high-confidence sites supported by MS/MS of all curated datasets in qPTM showed a similar pattern compared with PhosphoSitePlus and dbPTM (Supplementary Figure S1A and B). Additionally, we compared the phosphosites curated in PhosphoSitePlus, dbPTM and qPTM, as shown in Supplementary Figure S1C, 56.4% of phosphosites in qPTM overlapped with the other two databases. These results demonstrate similar patterns for the distribution of multiple integrated datasets and the reliability of data in qPTM. The occurrences of the PTM site in different datasets in the qPTM database could help assess its reliability. The distribution of reliability scores in the database and in different species is summarized in Supplementary Figure S1D. In addition, we found that the higher the star level was, the more sites that could be identified after re-identification, and the more sites with site-level FDR < 1% (Supplementary Figure S1E and F). These site-level FDR values were

cross-validated with the five-star scoring system, proving the reliability of our data.

DATABASE CONTENT

In the current release, qPTM contains 11 482 553 quantification events for 660 030 non-redundant PTM sites on 40 728 proteins under 2596 conditions in four different organisms collected from over 600 published studies. The detailed summaries for the datasets and literatures were provided in the ‘Summary of curated datasets’ and ‘Summary of curated literature’ sections of the ‘Help’ page on the qPTM website. The detailed summary for each organism is listed in Supplementary Table S1. Obviously, phosphorylation has the most substrates among PTMs (Figure 2A and B). Interestingly, methylation has a considerable number of substrates in a relatively small amount of quantitative data. Methylation was previously considered in nucleoproteins and transcription factors apart from histones. It was also found in many cytoplasmic proteins recently, and this greatly expanded its functional diversity (65). Currently, proteome-wide quantification of SUMOylation, methylation and glycosylation are insufficient in rats, while SUMOylation and methylation data are insufficient in yeast.

In eukaryotes, proteins undergo a variety of PTMs that are interrelated and coupled at different stages of biological processes (66,67). We analyzed PTMomics crosstalk in *Homo sapiens* by constructing a network using the iGragh R package (Figure 2C). The size of each node repre-



Figure 3. The BROWSE and SEARCH options of qPTM. (A) Summaries of the selected organisms among PTMs of three browse options. After selection of species and PTM, three browse approaches were provided, including (B) genes, (C) conditions and (D) samples, which are listed in alphabetical order below. (E) The advanced search function in the SEARCH page allows users to submit a combination of up to 10 terms for searching.

sents the number of modified proteins, and the width of the line represents the protein number of cooccurrences among different PTMs. Indeed, crosstalk between methylation and phosphorylation was most extensive, consistent with previous studies (68). Additionally, the interaction between methylation and acetylation might serve as a mechanism to control transcriptional activity (69,70). Abundant crosstalk between methylation, acetylation, ubiquitylation and SUMOylation was observed because they shared the modified lysine residues (Figure 2D). Crosstalks and concurrently modified proteins of *M. musculus*, *R. norvegicus* and *S. cerevisiae* are shown in Supplementary Figure S2. Although only a few substrates were identified for several PTMs, at least one crosstalk was observed. The intense PTM crosstalk suggested that different types of PTMs could competitively or dynamically regulate a considerable proportion of modified proteins.

NEW FEATURES

New browse and search function

With the tremendous growth in the quantity of PTMomics data in qPTM, it is challenging to locate and access a specific quantitative event from large amounts of data. We improved the browse and search functions to help users find their interested PTM dynamic data quickly.

On the BROWSE page, users can first select an organism of interest. Summaries of three browse options, including genes, conditions and samples of each PTM in the selected organism, are visualized in the bar plot below. As the

mouse hovers over the bar of each PTM, the quantity of gene, condition or sample is shown (Figure 3A). By clicking an item in the list of one browse option, the results of a certain gene/sample/condition will be shown on the result page (Figure 3B and C). This helps users quickly choose the data of interest from a bulk of items.

The advanced search function was provided on the SEARCH page, which provided keyword-based queries in UniProt accession, protein and gene names, protein functions and description of conditions and samples. In addition, selection of organisms and PTMs was also provided, which greatly narrowed down the result (Figure 3E). Users can submit up to ten search terms, which can be specified in different areas and combined with three operators of 'and', 'or' and 'not' to query PTM data accurately. Thus, wherever on the BROWSE or SEARCH page, users can always access interested data swiftly and accurately.

Enhanced result page

The PTM sites can be further filtered on the result page by conditions, sample names or modification types at the top of the page. The information for each quantitative event was organized by tabular format with UniProt accession, gene name, PTM position within the protein, modification type, sequence window, sample name, the abbreviation of condition, log₂-transformed ratio and *P* value. To distinguish between different PTM types, modification sites of the sequence window were labeled in different colors (Figure 4A). The details of each quantitative PTM event were

A

Result

Search content: Sample = arpe-19; Organism: All; Modification: All

There are 12128 entries with your keyword(s).

Condition: 5 selected | Sample: ARPE-19 | Modification: Phosphoryl... | Filter

Phosphorylation | Acetylation | Methylation | Ubiquitylation | Glycosylation | SUMOylation | Download

Label	Uniprot ID	Gene ID	Protein ID	Modification ID	Sample ID	Condition ID	Log2Ratio ID	FDR ID	Identified ID		
	P04792	AAK1	605	Phosphorylation	Q2F7YQ	PF2AVQC	ARPE-19	POS 15min/Ctr	0.04	0.645	★★★★★

B

P04792 | HSPB1 | 78 | Phosphorylation | PAYSRALSRQLSSGV | ARPE-19 | POS 15min/Ctr | -0.294 | 0.656 | ★★★★★

About experiment | About protein | Potential kinases and their inhibitors | Sequence and Structure

Resource: Quantitative phosphoproteomics reveals involvement of multiple signaling pathways in early phagocytosis by the retinal pigmented epithelium. (PMID: 29976645)

Condition detail: Photoreceptor outer segment (POS) stimulated 120 min/Unstimulated

Sample (type): ARPE-19 (cell)

Label method: SILAC

Enrichment method: ICK2

Mass spectrometer: Q Exactive Hybrid quadrupole-orbitrap mass spectrometer (Thermo Fisher Scientific)

Raw peptide: ALSRQLSSGVSEIR

Reported PEP/Localization Probability: PEP: 1.67527e-15

Re-identified FDR: 0

Records in databases: Collected in PhosphoSitePlus: Yes | Collected in dbPTM: Yes | Identified times in iPTM: 52

Related protein change: 0.217 (P value = 0.001)

D

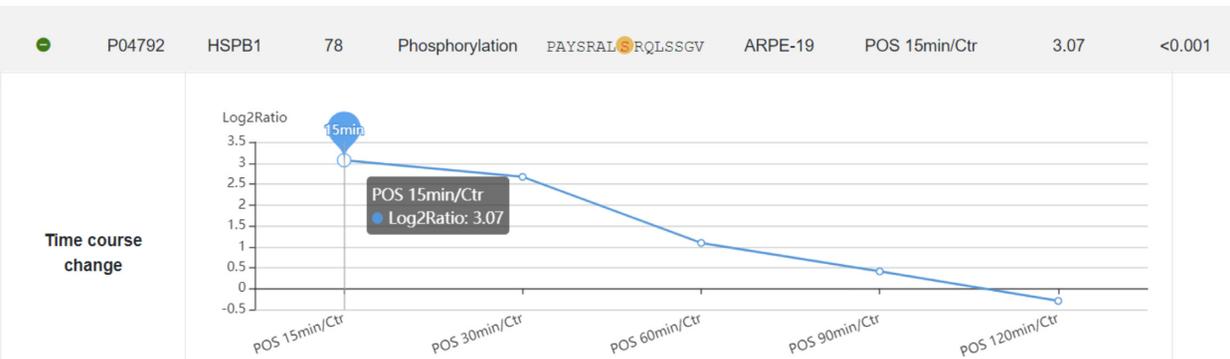
About experiment | About protein | Potential kinases and their inhibitors | Sequence and Structure

Exp.	Kinase	Inhibitor	Structure	Source
	MAPKAPK2	-	-	-
	MAPKAPK3	-	-	-

GPS	Kinase	Inhibitor	Structure	Source
	AGC	-	-	-
	CAMK	-	-	-

iGPS	Kinase	Inhibitor	Structure	Source
	CAMK/DCLK1	DB12010		DrugBank
	CAMK/PSKH2	-	-	-

F



C

About experiment | About protein | Potential kinases and their inhibitors | Sequence and Structure

Uniprot accession: HSPB1_HUMAN, P04792, B2R4N8, Q6F147, Q96C20, Q96E17, Q9UC31, Q9UC34, Q9UC35, Q9UC36

Entrez ID: 3315

Genbank protein ID: NP_001531.1

Genbank nucleotide ID: NM_001540.3

Protein name: Heat shock protein beta-1.28 kDa heat shock protein, Estrogen-regulated 24 kDa protein, Heat shock 27 kDa protein, Stress-responsive protein 27

Gene name: HSPB1, HSP27, HSP28

Organism: Homo sapiens (Human) NCBI Taxa ID=9606

Functional description: Small heat shock protein which functions as a molecular chaperone probably maintaining denatured proteins in a foldin...

PTMs for protein (Count: 63)	Position	Peptide	Modification	Source
	9	TERRVFFSLRGGPM	Phosphorylation	dbPAF
	15	FSLRGGPMDFRDM	Phosphorylation	UniProt

PTMDs for site (Count: 2)	Disease	Modification	Influence	PMID
	Breast cancer/tumor/carcinoma	Phosphorylation	Upregulation	17697330
	Pancreatic cancer/carcinoma/adenocarcinoma	Phosphorylation	Presence	22012255

Sequence (Length: 205)

10 20 30 40 50 60 70
MTERRVFFSL IRGFSNDPFR DWYFHSRLFD QAFQLRFLPE ENSQWLQGSF WFOYVRFLEF AAATESPAWLA

80 90 100 110 120 130 140
FAYSRLSRQ LSSVSEIRH TADMWVSLD VHSFAPDEL VYVQDQVEI TQVREERQDE HGVTSRCTFR

E



Figure 4. Examples of the enhanced result pages. (A) Overview of a returned result page. Detailed information will be shown by clicking the '+' button. Detailed information was sorted and restored in the (B) 'About experiment', (C) 'About protein', (D) 'Potential kinases and their inhibitors' and (E) 'Sequence and Structure' sections. (F) Example of the visualization of time-course quantitative events. The condition corresponding to the quantitative events is labeled in the trend line.

reorganized. Users can view detailed information by clicking the plus bottom, including 'About experiment', 'About protein', 'Potential kinases and their inhibitors' and 'Sequence and structure'. In the 'About experiment' section, source literature reference, detailed description of condition, sample and type, labeling methods, enrichment methods, mass spectrometer equipment, and raw peptide are shown. It is worth noting that \log_2 transformed fold change and the P value of the modified protein under the same circumstances were also shown, if available, which enables researchers to find connections between proteomics and PTMomics (Figure 4B). The protein information, such as database accessions, protein/gene name/alias, fundamental descriptions, PTMs, and sequences, from the UniProt database is shown in the 'About protein' column. Furthermore, to help researchers better understand the relationship between the PTM site and certain diseases, we annotated information from PTMD. 'Potential kinases and their inhibitors' showed the experimentally identified and predicted upstream kinases for the PTM sites (Figure 4C). Furthermore, the inhibitors annotated by DrugBank for the kinases are shown (Figure 4D). As mentioned previously, the sequence and structure properties of the protein were visualized in the 'Sequence and Structure' section (Figure 4E).

Visualization of time-course dynamics

We noticed that time-course and concentration gradient conditions accounted for a considerable proportion of 43.4% in all included conditions, while the corresponding quantitative PTM events accounted for 40.9%. Under these conditions, researchers focus more on dynamic changes in trends rather than individual time points or concentrations. Thus, we visualized quantitative events of time-course and concentration gradient conditions in trend lines, which were exhibited in the 'About Experiment' column. For example, as shown in the trend line (Figure 4F), the expression of HSPB1-S78 in ARPE-19 cells decreased dramatically over time. The ARPE-19 cell line was exposed to photoreceptor outer segments (POSS), and phosphorylated peptides were quantified at 15, 30, 60, 90 and 120 min. ARPE-19 is an immortalized retinal pigmented epithelium (RPE) whose principal function is the clearance of shed POS through a process resembling phagocytosis (71,72). Dysfunction of this process contributes to retinal degenerative disorders. HSPB1, a downstream substrate of the MAPK signaling pathway (73), functions as a molecular chaperone to maintain denatured proteins (74). The trend of HSPB1-S78 over time might reveal its role in early phagocytosis and thus affect retinal homeostasis.

DISCUSSION

As critical molecular mechanisms in biological processes and diseases, PTMs greatly expand proteome complexity and diversity (1,2). Abnormal PTM levels are frequently observed in response to stimuli, diseases and cancers (3–8), which has inspired scientists to study quantitative PTMomics. High-throughput technologies currently enable the detection of PTM quantitative changes under different conditions (17), thus leading to a deeper understanding

of various physiological processes and intractable diseases. With the exponential increase in quantitative PTMomics data, a comprehensive platform is expected to provide a resource to integrate PTM quantification event data. qPTM is the first and unique repository to curate and organize PTM dynamics data. In addition, various annotations, including PTM-corresponding quantitative proteomics data, protein information, PTM information, potential upstream kinases and inhibitors, and sequence and structure properties, were integrated. Furthermore, visualization of time-course or concentration gradient quantitative data was also provided for a better view of the trend.

With the development of high-throughput proteomics techniques, the amounts of PTMomics datasets increased rapidly. The quality control became more and more important especially for the integration of large-scale datasets, and it was shown that combining multiple datasets could lead to the aggregation of false-positive hits (75,76). In this study, we performed the re-identification and FDR control following the previously published searching strategies (64). Considering the raw MS data was inaccessible in a large-proportion of published literatures, the recurrences of identification among datasets were used as the reliability indicator alternatively. To balance between the unified quality control and the data abundance reported by the literatures, we did not perform site-level FDR filtering, but annotated the quality information to each site and marked significance by adding the 'red border' to the star in the five-star scoring system if the site-level FDR was $<1\%$. Although this is not a perfect solution for quality control, it is the practical one and we will keep up with the progresses in the community.

Meanwhile, there were also limitations in the database. The quantitative PTM data from the four organisms were mainly isolated, and it was difficult to perform homology analysis among species due to the unshared conditions. Some other essential model organisms or PTMs, such as *C. elegans* and *D. melanogaster*, and palmitoylation were not included because of insufficient quantitative data. Besides, the crosstalk of PTMs was not curated due to data limitations. The development of proximity labeling facilitated subcellular PTMomics and helped understand protein functions at the subcellular level (61,77). However, we only collected a few subcellular quantitative PTMomics data. Taken together, although improvement is still needed, the qPTM database could serve as a comprehensive platform for accessing PTM dynamics systematically and conveniently. The qPTM database will be regularly updated to track the progress of quantitative PTM dynamics.

DATA AVAILABILITY

qPTM is an open-access web server, which can be accessed at <http://qptm.omicsbio.info>.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

FUNDING

National Key R&D Program of China [2021YFA1302100]; National Natural Science Foundation of China [81972239,

91953123, 81930065, 82173128]; Science and Technology Program of Guangzhou [202206080011]; Program for Guangdong Introducing Innovative and Entrepreneurial Teams [2017ZT07S096]; Tip-Top Scientific and Technical Innovative Youth Talents of Guangdong Special Support Program [2019TQ05Y351]; Science and Technology Program of Guangdong [2019B020227002]; Natural Science Foundation of Guangdong Province [2019A1515010634]; CAMS Innovation Fund for Medical Sciences (CIFMS) [2019-I2M-5-036]. Funding for open access charge: National Key R&D Program of China [2021YFA1302100].

Conflict of interest statement. None declared.

REFERENCES

- Walsh, C.T., Garneau-Tsodikova, S. and Gatto, G.J. Jr (2005) Protein posttranslational modifications: the chemistry of proteome diversifications. *Angew. Chem. Int. Ed Engl.*, **44**, 7342–7372.
- Harmel, R. and Fiedler, D. (2018) Features and regulation of non-enzymatic post-translational modifications. *Nat. Chem. Biol.*, **14**, 244–252.
- Cohen, P. (2002) The origins of protein phosphorylation. *Nat. Cell Biol.*, **4**, E127–E130.
- Rodriguez-Paredes, M. and Lyko, F. (2019) The importance of non-histone protein methylation in cancer therapy. *Nat. Rev. Mol. Cell Biol.*, **20**, 569–570.
- Narita, T., Weinert, B.T. and Choudhary, C. (2019) Functions and mechanisms of non-histone protein acetylation. *Nat. Rev. Mol. Cell Biol.*, **20**, 156–174.
- Welchman, R.L., Gordon, C. and Mayer, R.J. (2005) Ubiquitin and ubiquitin-like proteins as multifunctional signals. *Nat. Rev. Mol. Cell Biol.*, **6**, 599–609.
- Yang, X. and Qian, K. (2017) Protein O-GlcNAcylation: emerging mechanisms and functions. *Nat. Rev. Mol. Cell Biol.*, **18**, 452–465.
- Hendriks, I.A. and Vertegaal, A.C. (2016) A comprehensive compilation of SUMO proteomics. *Nat. Rev. Mol. Cell Biol.*, **17**, 581–595.
- Huang, H., Sabari, B.R., Garcia, B.A., Allis, C.D. and Zhao, Y. (2014) SnapShot: histone modifications. *Cell*, **159**, 458–458.
- H.O. (1883) Zur frage ob casein ein einheitlicher stoff sei. *Hoppe-Seyler's Zeitsch. Physiol. Chem.*, **7**, 227–273.
- Anderson, C. (1992) Nobel prize given for work on protein phosphorylation. *Nature*, **359**, 570.
- Hershko, A. (2005) The ubiquitin system for protein degradation and some of its roles in the control of the cell-division cycle (Nobel lecture). *Angew. Chem. Int. Ed Engl.*, **44**, 5932–5943.
- Kim, S.C., Sprung, R., Chen, Y., Xu, Y., Ball, H., Pei, J., Cheng, T., Kho, Y., Xiao, H., Xiao, H. *et al.* (2006) Substrate and functional diversity of lysine acetylation revealed by a proteomics survey. *Mol. Cell*, **23**, 607–618.
- Tan, M., Luo, H., Lee, S., Jin, F., Yang, J.S., Montellier, E., Buchou, T., Cheng, Z., Rousseaux, S., Rajagopal, N. *et al.* (2011) Identification of 67 histone marks and histone lysine crotonylation as a new type of histone modification. *Cell*, **146**, 1016–1028.
- Zhang, Z., Tan, M., Xie, Z., Dai, L., Chen, Y. and Zhao, Y. (2011) Identification of lysine succinylation as a new post-translational modification. *Nat. Chem. Biol.*, **7**, 58–63.
- Witte, E.S., Old, W.M., Resing, K.A. and Ahn, N.G. (2007) Mapping protein post-translational modifications with mass spectrometry. *Nat. Methods*, **4**, 798–806.
- Choudhary, C. and Mann, M. (2010) Decoding signalling networks by mass spectrometry-based proteomics. *Nat. Rev. Mol. Cell Biol.*, **11**, 427–439.
- Olsen, J.V. and Mann, M. (2013) Status of large-scale analysis of post-translational modifications by mass spectrometry. *Mol. Cell. Proteomics*, **12**, 3444–3452.
- Chen, Y.J., Roumeliotis, T.I., Chang, Y.H., Chen, C.T., Han, C.L., Lin, M.H., Chen, H.W., Chang, G.C., Chang, Y.L., Wu, C.T. *et al.* (2020) Proteogenomics of Non-smoking lung cancer in east asia delineates molecular signatures of pathogenesis and progression. *Cell*, **182**, 226–244.
- Gillette, M.A., Satpathy, S., Cao, S., Dhanasekaran, S.M., Vasaiakar, S.V., Petralia, F., Li, Y., Liang, W.W., Reva, B. *et al.* (2020) Proteogenomic characterization reveals therapeutic vulnerabilities in lung adenocarcinoma. *Cell*, **182**, 200–225.
- Xu, J.Y., Zhang, C., Wang, X., Zhai, L., Ma, Y., Mao, Y., Qian, K., Sun, C., Liu, Z., Jiang, S. *et al.* (2020) Integrative proteomic characterization of human lung adenocarcinoma. *Cell*, **182**, 245–261.
- Gao, Q., Zhu, H., Dong, L., Shi, W., Chen, R., Song, Z., Huang, C., Li, J., Dong, X., Zhou, Y. *et al.* (2019) Integrated proteogenomic characterization of HBV-Related hepatocellular carcinoma. *Cell*, **179**, 561–577.
- Jiang, Y., Sun, A., Zhao, Y., Ying, W., Sun, H., Yang, X., Xing, B., Sun, W., Ren, L., Hu, B. *et al.* (2019) Proteomics identifies new therapeutic targets of early-stage hepatocellular carcinoma. *Nature*, **567**, 257–261.
- Wang, Z., Liu, H., Yan, Y., Yang, X., Zhang, Y. and Wu, L. (2020) Integrated proteomic and N-Glycoproteomic analyses of human breast cancer. *J. Proteome Res.*, **19**, 3499–3509.
- Sinha, A., Hussain, A., Ignatchenko, V., Ignatchenko, A., Tang, K.H., Ho, V.W.H., Neel, B.G., Clarke, B., Bernardini, M.Q., Ailles, L. *et al.* (2019) N-Glycoproteomics of patient-derived xenografts: a strategy to discover tumor-associated proteins in high-grade serous ovarian cancer. *Cell Syst.*, **8**, 345–351.
- Sun, Y., Zheng, X., Yuan, H., Chen, G., Ouyang, J., Liu, J., Liu, X., Xing, X. and Zhao, B. (2020) Proteomic analyses reveal divergent ubiquitylation patterns in hepatocellular carcinoma cell lines with different metastasis potential. *J. Proteomics*, **225**, 103834.
- Krug, K., Jaehnig, E.J., Satpathy, S., Blumenberg, L., Karpova, A., Anurag, M., Miles, G., Mertins, P., Geffen, Y., Tang, L.C. *et al.* (2020) Proteogenomic landscape of breast cancer tumorigenesis and targeted therapy. *Cell*, **183**, 1436–1456.
- Grimes, M., Hall, B., Foltz, L., Levy, T., Rikova, K., Gaiser, J., Cook, W., Smirnova, E., Wheeler, T., Clark, N.R. *et al.* (2018) Integration of protein phosphorylation, acetylation, and methylation data sets to outline lung cancer signaling networks. *Sci. Signal*, **11**, eaaq1087.
- Consortium, U. (2021) UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Res.*, **49**, D480–D489.
- Hornbeck, P.V., Kornhauser, J.M., Latham, V., Murray, B., Nandhikonda, V., Nord, A., Skrzypek, E., Wheeler, T., Zhang, B. and Gnäd, F. (2019) 15 years of PhosphoSitePlus®: integrating post-translationally modified sites, disease variants and isoforms. *Nucleic Acids Res.*, **47**, D433–D441.
- Huang, K.Y., Lee, T.Y., Kao, H.J., Ma, C.T., Lee, C.C., Lin, T.H., Chang, W.C. and Huang, H.D. (2019) dbPTM in 2019: exploring disease association and cross-talk of post-translational modifications. *Nucleic Acids Res.*, **47**, D298–D308.
- Li, J., Jia, J., Li, H., Yu, J., Sun, H., He, Y., Lv, D., Yang, X., Glocker, M.O., Ma, L. *et al.* (2014) SysPTM 2.0: an updated systematic resource for post-translational modification. *Database (Oxford)*, **2014**, bau025.
- Keshava Prasad, T.S., Goel, R., Kandasamy, K., Keerthikumar, S., Kumar, S., Mathivanan, S., Telikicherla, D., Raju, R., Shafreen, B., Venugopal, A. *et al.* (2009) Human protein reference database—2009 update. *Nucleic Acids Res.*, **37**, D767–D772.
- Gnäd, F., Gunawardena, J. and Mann, M. (2011) PHOSIDA 2011: the posttranslational modification database. *Nucleic Acids Res.*, **39**, D253–D260.
- 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. (2018) iPTMnet: an integrated resource for protein post-translational modification network discovery. *Nucleic Acids Res.*, **46**, D542–D550.
- Minguez, P., Letunic, I., Parca, L., Garcia-Alonso, L., Dopazo, J., Huerta-Cepas, J. and Bork, P. (2015) PTMcode v2: a resource for functional associations of post-translational modifications within and between proteins. *Nucleic Acids Res.*, **43**, D494–D502.
- Ullah, S., Lin, S., Xu, Y., Deng, W., Ma, L., Zhang, Y., Liu, Z. and Xue, Y. (2016) dbPAF: an integrative database of protein phosphorylation in animals and fungi. *Sci. Rep.*, **6**, 23534.
- Lin, S., Wang, C., Zhou, J., Shi, Y., Ruan, C., Tu, Y., Yao, L., Peng, D. and Xue, Y. (2021) EPSD: a well-annotated data resource of protein phosphorylation sites in eukaryotes. *Brief Bioinform.*, **22**, 298–307.
- Du, Y., Xu, N., Lu, M. and Li, T. (2011) hUbiquitome: a database of experimentally verified ubiquitination cascades in humans. *Database (Oxford)*, **2011**, bar055.

40. Rao, R.S.P., Zhang, N., Xu, D. and Møller, I.M. (2018) CarbonylDB: a curated data-resource of protein carbonylation sites. *Bioinformatics*, **34**, 2518–2520.
41. Gupta, R., Birch, H., Rapacki, K., Brunak, S. and Hansen, J.E. (1999) O-GLYCBASE version 4.0: a revised database of O-glycosylated proteins. *Nucleic Acids Res.*, **27**, 370–372.
42. Kaji, H., Shikanai, T., Sasaki-Sawa, A., Wen, H., Fujita, M., Suzuki, Y., Sugahara, D., Sawaki, H., Yamauchi, Y., Shinkawa, T. *et al.* (2012) Large-scale identification of N-glycosylated proteins of mouse tissues and construction of a glycoprotein database, GlycoProtDB. *J. Proteome Res.*, **11**, 4553–4566.
43. Sun, S., Hu, Y., Ao, M., Shah, P., Chen, J., Yang, W., Jia, X., Tian, Y., Thomas, S. and Zhang, H. (2019) N-GlycositeAtlas: a database resource for mass spectrometry-based human N-linked glycoprotein and glycosylation site mapping. *Clin. Proteomics*, **16**, 35.
44. Xu, H., Zhou, J., Lin, S., Deng, W., Zhang, Y. and Xue, Y. (2017) PLMD: an updated data resource of protein lysine modifications. *J. Genet. Genomics*, **44**, 243–250.
45. Liu, Z., Wang, Y., Gao, T., Pan, Z., Cheng, H., Yang, Q., Cheng, Z., Guo, A., Ren, J. and Xue, Y. (2014) CPLM: a database of protein lysine modifications. *Nucleic Acids Res.*, **42**, D531–D536.
46. Liu, Z., Cao, J., Gao, X., Zhou, Y., Wen, L., Yang, X., Yao, X., Ren, J. and Xue, Y. (2011) CPLA 1.0: an integrated database of protein lysine acetylation. *Nucleic Acids Res.*, **39**, D1029–D1034.
47. Deutsch, E.W., Bandeira, N., Sharma, V., Perez-Riverol, Y., Carver, J.J., Kundu, D.J., Garcia-Seisdedos, D., Jarnuczak, A.F., Hewapathirana, S., Pullman, B.S. *et al.* (2020) The proteomexchange consortium in 2020: enabling ‘big data’ approaches in proteomics. *Nucleic Acids Res.*, **48**, D1145–D1152.
48. Matlock, M.K., Holehouse, A.S. and Naegle, K.M. (2015) ProteomeScout: a repository and analysis resource for post-translational modifications and proteins. *Nucleic Acids Res.*, **43**, D521–D530.
49. Samaras, P., Schmidt, T., Frejno, M., Gessulat, S., Reinecke, M., Jarzab, A., Zecha, J., Mergner, J., Giansanti, P., Ehrlich, H.C. *et al.* (2020) ProteomicsDB: a multi-omics and multi-organism resource for life science research. *Nucleic Acids Res.*, **48**, D1153–D1163.
50. Shamsaei, B., Chojnacki, S., Pilarczyk, M., Najafabadi, M., Niu, W., Chen, C., Ross, K., Matlock, A., Muhlich, J., Chutipongtanate, S. *et al.* (2020) piNET: a versatile web platform for downstream analysis and visualization of proteomics data. *Nucleic Acids Res.*, **48**, W85–W93.
51. Moriya, Y., Kawano, S., Okuda, S., Watanabe, Y., Matsumoto, M., Takami, T., Kobayashi, D., Yamanouchi, Y., Araki, N. and Yoshizawa, A.C. (2019) The iPOST environment: an integrated proteomics data repository and database. *Nucleic Acids Res.*, **47**, D1218–D1224.
52. Farriol-Mathis, N., Garavelli, J.S., Boeckmann, B., Duvaud, S., Gasteiger, E., Gateau, A., Veuthey, A.L. and Bairoch, A. (2004) Annotation of post-translational modifications in the swiss-prot knowledge base. *Proteomics*, **4**, 1537–1550.
53. Artimo, P., Jonnalagedda, M., Arnold, K., Baratin, D., Csardi, G., de Castro, E., Duvaud, S., Flegel, V., Fortier, A., Gasteiger, E. *et al.* (2012) ExpASY: SIB bioinformatics resource portal. *Nucleic Acids Res.*, **40**, W597–W603.
54. Xu, H., Wang, Y., Lin, S., Deng, W., Peng, D., Cui, Q. and Xue, Y. (2018) PTMD: a database of human Disease-associated Post-translational modifications. *Genomics Proteomics Bioinformatics*, **16**, 244–251.
55. Wishart, D.S., Feunang, Y.D., Guo, A.C., Lo, E.J., Marcu, A., Grant, J.R., Sajed, T., Johnson, D., Li, C., Sayeeda, Z. *et al.* (2018) DrugBank 5.0: a major update to the drugbank database for 2018. *Nucleic Acids Res.*, **46**, D1074–D1082.
56. Wang, C., Xu, H., Lin, S., Deng, W., Zhou, J., Zhang, Y., Shi, Y., Peng, D. and Xue, Y. (2020) GPS 5.0: an update on the prediction of Kinase-specific phosphorylation sites in proteins. *Genomics Proteomics Bioinformatics*, **18**, 72–80.
57. Song, C., Ye, M., Liu, S., Cheng, H., Jiang, X., Han, G., Songyang, Z., Tan, Y., Wang, H., Ren, J. *et al.* (2012) Systematic analysis of protein phosphorylation networks from phosphoproteomic data. *Mol. Cell. Proteomics*, **11**, 1070–1083.
58. Yu, K., Zhang, Q., Liu, Z., Du, Y., Gao, X., Zhao, Q., Cheng, H., Li, X. and Liu, Z.X. (2020) Deep learning based prediction of reversible HAT/HDAC-specific lysine acetylation. *Brief Bioinform*, **21**, 1798–1805.
59. Perez-Riverol, Y., Csordas, A., Bai, J., Bernal-Llinares, M., Hewapathirana, S., Kundu, D.J., Inuganti, A., Griss, J., Mayer, G. and Eisenacher, M. (2019) The PRIDE database and related tools and resources in 2019: improving support for quantification data. *Nucleic Acids Res.*, **47**, D442–D450.
60. Ma, J., Chen, T., Wu, S., Yang, C., Bai, M., Shu, K., Li, K., Zhang, G., Jin, Z. and He, F. (2019) iProX: an integrated proteome resource. *Nucleic Acids Res.*, **47**, D1211–D1217.
61. Choi, M., Carver, J., Chiva, C., Tzouros, M., Huang, T., Tsai, T.-H., Pullman, B., Bernhardt, O.M., Hüttenhain, R. and Teo, G.C. (2020) MassIVE: quant: a community resource of quantitative mass spectrometry-based proteomics datasets. *Nat. Methods*, **17**, 981–984.
62. Cox, J. and Mann, M. (2008) MaxQuant enables high peptide identification rates, individualized ppb-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.*, **26**, 1367–1372.
63. Sinitcyn, P., Tiwary, S., Rudolph, J., Gutenbrunner, P., Wichmann, C., Yilmaz, S., Hamzei, H., Salinas, F. and Cox, J. (2018) MaxQuant goes linux. *Nat. Methods*, **15**, 401–401.
64. Ochoa, D., Jarnuczak, A.F., Viéitez, C., Gehre, M., Soucheray, M., Mateus, A., Kleefeldt, A.A., Hill, A., Garcia-Alonso, L. and Stein, F. (2020) The functional landscape of the human phosphoproteome. *Nat. Biotechnol.*, **38**, 365–373.
65. Biggar, K.K. and Li, S.S. (2015) Non-histone protein methylation as a regulator of cellular signalling and function. *Nat. Rev. Mol. Cell Biol.*, **16**, 5–17.
66. Hunter, T. (2007) The age of crosstalk: phosphorylation, ubiquitination, and beyond. *Mol. Cell*, **28**, 730–738.
67. Venne, A.S., Kollipara, L. and Zahedi, R.P. (2014) The next level of complexity: crosstalk of posttranslational modifications. *Proteomics*, **14**, 513–524.
68. Huang, Y., Xu, B., Zhou, X., Li, Y., Lu, M., Jiang, R. and Li, T. (2015) Systematic characterization and prediction of post-translational modification cross-talk. *Mol. Cell. Proteomics*, **14**, 761–770.
69. Ivanov, G.S., Ivanova, T., Kurash, J., Ivanov, A., Chuikov, S., Gizatullin, F., Herrera-Medina, E.M., Rauscher, F. 3rd, Reinberg, D. and Barlev, N.A. (2007) Methylation-acetylation interplay activates p53 in response to DNA damage. *Mol. Cell. Biol.*, **27**, 6756–6769.
70. Jiang, Y., Trescott, L., Holcomb, J., Zhang, X., Brunzelle, J., Sirinupong, N., Shi, X. and Yang, Z. (2014) Structural insights into estrogen receptor α methylation by histone methyltransferase SMYD2, a cellular event implicated in estrogen signaling regulation. *J. Mol. Biol.*, **426**, 3413–3425.
71. Ruggiero, L., Connor, M.P., Chen, J., Langen, R. and Finnemann, S.C. (2012) Diurnal, localized exposure of phosphatidylserine by rod outer segment tips in wild-type but not Itgb5^{-/-} or Mfge8^{-/-} mouse retina. *Proc. Natl. Acad. Sci. U.S.A.*, **109**, 8145–8148.
72. Chiang, C.K., Tworak, A., Kevany, B.M., Xu, B., Mayne, J., Ning, Z., Figeys, D. and Palczewski, K. (2017) Quantitative phosphoproteomics reveals involvement of multiple signaling pathways in early phagocytosis by the retinal pigmented epithelium. *J. Biol. Chem.*, **292**, 19826–19839.
73. Doshi, B.M., Hightower, L.E. and Lee, J. (2010) HSPB1, actin filament dynamics, and aging cells. *Ann. N.Y. Acad. Sci.*, **1197**, 76–84.
74. Kainuma, S., Tokuda, H., Yamamoto, N., Kuroyanagi, G., Fujita, K., Kawabata, T., Sakai, G., Matsushima-Nishiwaki, R., Kozawa, O. and Otsuka, T. (2017) Heat shock protein 27 (HSPB1) suppresses the PDGF-BB-induced migration of osteoblasts. *Int. J. Mol. Med.*, **40**, 1057–1066.
75. Ezkurdia, I., Vazquez, J., Valencia, A. and Tress, M. (2014) Analyzing the first drafts of the human proteome. *J. Proteome Res.*, **13**, 3854–3855.
76. Reiter, L., Claassen, M., Schrimpf, S.P., Jovanovic, M., Schmidt, A., Buhmann, J.M., Hengartner, M.O. and Aebersold, R. (2009) Protein identification false discovery rates for very large proteomics data sets generated by tandem mass spectrometry. *Mol. Cell. Proteomics*, **8**, 2405–2417.
77. Liu, Y., Zeng, R., Wang, R., Weng, Y., Wang, R., Zou, P. and Chen, P.R. (2021) Spatiotemporally resolved subcellular phosphoproteomics. *Proc. Natl. Acad. Sci. U.S.A.*, **118**, e2025299118.