

EPSD 2.0: An Updated Database of Protein Phosphorylation Sites Across Eukaryotic Species

Miaomiao Chen ^{1,#}, Yujie Gou ^{1,#}, Ming Lei ¹, Leming Xiao ¹, Miaoying Zhao ¹,
Xinhe Huang ¹, Dan Liu ¹, Zihao Feng ¹, Di Peng ^{1,*}, Yu Xue ^{1,2,*}

¹Department of Bioinformatics and Systems Biology, Key Laboratory of Molecular Biophysics of the Ministry of Education, Hubei Bioinformatics and Molecular Imaging Key Laboratory, College of Life Science and Technology, Huazhong University of Science and Technology, Wuhan 430074, China

²Nanjing University Institute of Artificial Intelligence Biomedicine, Nanjing 210031, China

*Corresponding authors: xueyu@hust.edu.cn (Xue Y), pengdi@hust.edu.cn (Peng D).

#Equal contribution.

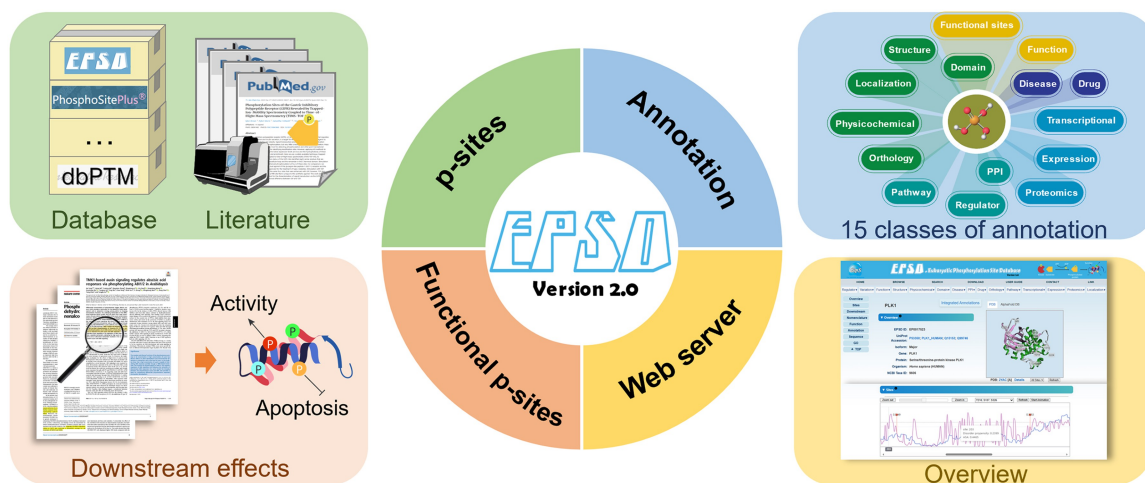
Handling Editor: Jiangning Song

Abstract

As one of the most crucial post-translational modifications, protein phosphorylation regulates a broad range of biological processes in eukaryotes. Biocuration, integration, and annotation of reported phosphorylation events will deliver a valuable resource for the community. Here, we present an updated database, the eukaryotic phosphorylation site database 2.0 (EPSD 2.0), which includes 2,769,163 experimentally identified phosphorylation sites (p-sites) in 362,707 phosphoproteins from 223 eukaryotes. From the literature, 873,718 new p-sites identified through high-throughput phosphoproteomic research were first collected, and 1,078,888 original phosphopeptides together with primary references were reserved. Then, this dataset was merged into EPSD 1.0, comprising 1,616,804 p-sites within 209,326 proteins across 68 eukaryotic organisms. We also integrated 362,190 additional known p-sites from 10 public databases. After redundancy clearance, we manually re-checked each p-site and annotated 88,074 functional events for 32,762 p-sites, covering 58 types of downstream effects on phosphoproteins, and regulatory impacts on 107 biological processes. In addition, phosphoproteins and p-sites in 8 model organisms were meticulously annotated utilizing information supplied by 100 external platforms encompassing 15 areas. These areas included kinase/phosphatase, transcription regulators, three-dimensional structures, physicochemical characteristics, genomic variations, functional descriptions, protein domains, molecular interactions, drug–target associations, disease-related data, orthologs, transcript expression levels, proteomics, subcellular localization, and regulatory pathways. We expect that EPSD 2.0 will become a useful database supporting comprehensive studies on phosphorylation in eukaryotes. The EPSD 2.0 database is freely accessible online at <https://epsd.biocuckoo.cn/>.

Key words: Post-translational modification; Phosphorylation; Functional phosphorylation site; Phosphoproteomics; PLK1.

Graphical abstract



Received: 6 November 2024; Revised: 16 January 2025; Accepted: 20 February 2025.

© The Author(s) 2025. Published by Oxford University Press and Science Press on behalf of the Beijing Institute of Genomics, Chinese Academy of Sciences / China National Center for Bioinformation and Genetics Society of China.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

Introduction

Protein phosphorylation serves as a crucial post-translational modification (PTM), orchestrating a wide range of biological functions [1–5]. In eukaryotes, phosphorylation primarily modifies specific amino acids, especially serine (S), threonine (T), or tyrosine (Y) residue on targeted substrates catalyzed by protein kinases [6]. Phosphorylation at different sites on a protein may exert distinct functions under various biological contexts [2]. For example, phosphorylation of the human PLK1 protein at S137 promotes cancer cell survival in response to mitochondrial dysfunction [7], while the phosphorylation at T210 is required for enhancing neuroprotective autophagy [8]. In particular, aberrant or inadequate phosphorylation has been implicated in multiple diseases, including neurodegenerative diseases [9], cancer [10], and cardiovascular disorders [11]. Consequently, pinpointing phosphorylation sites (p-sites) in eukaryotes is essential for understanding the regulatory patterns and underlying mechanisms of phosphorylation in diverse biological processes.

Tandem mass spectrometry (MS/MS) serves as a leading technique for identifying p-sites in proteins [12,13]. Besides Orbitrap mass spectrometers [14], recent progress has led to the development of MS/MS instruments with much higher speed, throughput, and resolution. For example, by incorporating ion mobility spectrometry as a new dimension, MS/MS technology has advanced to a four-dimensional (4D) stage [15,16]. Utilizing trapped ion mobility spectrometry coupled with a time-of-flight (timsTOF) mass spectrometer, Kramer et al. identified nearly 30,000 p-sites from bone marrow samples of 44 acute myeloid leukemia patients and 6 healthy controls [17]. Recently, the Asymmetric Track Lossless (Astral), a novel type of mass analyzer, achieved higher scanning speed, resolving power, sensitivity, and low-ppm mass accuracy [18,19]. Coupled with Orbitrap, the leading Orbitrap Astral mass spectrometer helped in identifying 81,120 unique p-sites within 12 h of measurement in a data-independent acquisition (DIA) manner [20]. These advancements in MS/MS have facilitated the continuous identification of numerous p-sites.

To manage and utilize these data, many public resources have been developed to curate p-sites, including Phospho.ELM [21,22], PhosphoSitePlus [23,24], dbPTM [25,26], SysPTM [27,28], iPTMnet [29,30], PHOSIDA (phosphorylation site database) [31], PhosphoPep [32], LymphOS [33,34], PlantsP [35], PhosPhAt 4.0 (The *Arabidopsis* Protein Phosphorylation Site Database) [36,37], P³DB (Plant Protein Phosphorylation Database) [38], MPPD (Medicago PhosphoProtein Database) [39], FPD (Fungi Phosphorylation Database) [40], UniProt [41,42], Plant PTM Viewer [43], HPRD (Human Protein Reference Database) [44], BioGRID [45], RegPhos (Regulatory Network in Protein Phosphorylation) [46,47], Pf-Phospho [48], Scop3P [49], HisPhosSite [50], and Nphos [51] (Table S1). Among these databases, PhosphoSitePlus, dbPTM, SysPTM, iPTMnet, Plant PTM Viewer, and UniProt provide data for multiple PTM types in addition to phosphorylation. PhosphoSitePlus and dbPTM also offer functional annotations on the downstream effects of p-sites. Additionally, most of the existing resources mainly curate O-phosphosites, whereas HisPhosSite and Nphos focus on the collection of N-phosphosites. Of note, Nphos contains 11,710 experimentally verified N-phosphosites covering pHis, pLys, and pArg residues, and provides an online service for predicting N-phosphosites, serving as a highly useful resource for N-phosphorylation [51]. With the development of artificial intelligence (AI) technologies,

more structured and comprehensive datasets are needed to support the training of large models. Therefore, there remains a need for the collection, curation, and combination for extensive constantly experimentally verified p-sites. Also, it is very desirable to obtain the specific functional mechanisms of phosphorylation events, so it is important and useful to annotate the downstream effects of p-sites.

In 2019, we released EPSD 1.0 that comprised 1,616,804 p-sites by re-curating data from the dbPPT [52] and dbPAF [53] databases, along with p-sites identified in high-throughput (HTP) studies and 13 publicly accessible phosphorylation databases. In this release of EPSD 2.0, we further re-curated the p-sites in EPSD 1.0, collected p-sites from HTP experiments and 10 additional public databases (Figure 1A; Table S2). This resulted in a dataset comprising 2,769,163 p-sites on 362,707 phosphoproteins across 223 species, encompassing 3,364,760 phosphopeptides. Furthermore, we annotated 88,074 functional events for 32,762 p-sites through extensive literature review. These annotations of downstream effects covered 58 types of functions on phosphoproteins, and 107 regulatory impacts on biological processes. Additionally, we meticulously annotated the dataset by incorporating the information from 100 extra resources. Ultimately, EPSD 2.0 (~36.2 GB) exhibits roughly a 2.5-fold increase in the data volume compared to EPSD 1.0 (~14.1 GB). Compared to other existing databases, EPSD 2.0 offers more experimentally validated p-sites across more eukaryotic species, alongside extensive functional annotations on downstream effects of p-sites. We believe that EPSD 2.0 will serve as a more comprehensive and valuable database for the scientific community.

Database construction and content

Data collection, curation, and integration

This update not only expanded the collection of new p-sites but also curated the downstream effects of these p-sites (Figure 1). First, we performed a PubMed search employing various keywords, such as “large-scale phosphorylation”, “phosphoproteome”, “phosphoproteomics”, “phosphoproteomic”, “mass spectrometry phosphorylation”, “MS phosphorylation”, “phospho-residue”, “phospho-site”, and “phosphosite”. This search aimed to get published studies containing HTP phosphoproteomic data. In these HTP studies, different MS/MS instruments were used for phosphoproteomic profiling, and different software packages were used for data processing. However, all these studies followed a standard procedure for data processing, in which the false discovery rates for the peptide–spectrum match, p-site, and protein decoy fraction were all set to < 1%, supporting the high quality of all studies. To ensure the consistency with the original studies, we directly obtained HTP p-sites from the supplementary data of each paper if available. Additionally, to capture articles containing p-sites with descriptions of their downstream effects, we conducted targeted searches on PubMed using combinations of keywords such as “phosphorylation”, “phosphorylated”, and “phospho-” along with “site”, “residue”, “serine”, “threonine”, “tyrosine”, “Ser”, “Thr”, and “Tyr”. Moreover, we reviewed the literature that served as the source of existing low-throughput (LTP) p-sites in the EPSD 2.0 database, aiming to extract descriptions of functional information related to p-sites. Here, we defined a functional event as a specific p-site coupled with one of its reported downstream effects. To ensure the data quality, the

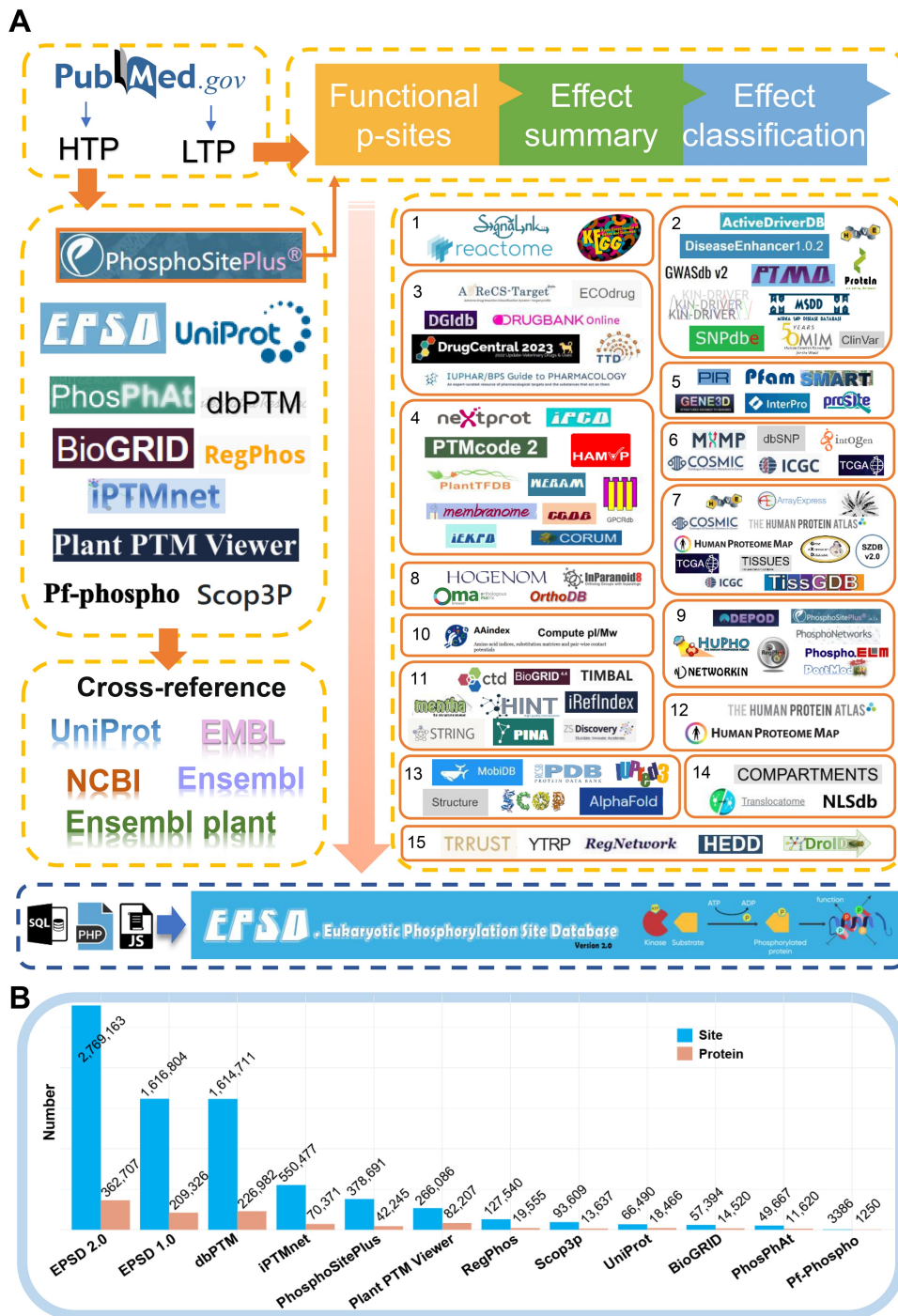


Figure 1 The procedure for development of EPSSD 2.0

A. First, we manually collected experimentally identified p-sites from PubMed. Then, known p-sites in 10 additional phosphorylation databases were integrated. In addition to basic annotations, we further annotated eight model organisms, including *Homo sapiens*, *Mus musculus*, *Rattus norvegicus*, *Drosophila melanogaster*, *Caenorhabditis elegans*, *Arabidopsis thaliana*, *Schizosaccharomyces pombe*, and *Saccharomyces cerevisiae*, by integrating the knowledge from 100 additional resources that covered 15 aspects, including kinase/phosphatase, transcription regulators, three-dimensional structures, physicochemical characteristics, genomic variations, functional descriptions, protein domains, molecular interactions, drug-target associations, disease-related data, orthologs, transcript expression levels, proteomics, subcellular localization, and regulatory pathways. **B.** The numbers of p-sites and phosphoproteins curated and integrated from EPSSD and other existing resources. p-site, phosphorylation site; EPSSD, the eukaryotic phosphorylation site database; HTP, high-throughput; LTP, low-throughput.

primary references were presented for both HTP and LTP p-sites.

To obtain the position information, all the modified residues with collected peptides were aligned with the reference sequences obtained from the UniProt database [42]. For the

annotation of fundamental information of modified proteins, we also retained relative data from UniProt [42], including protein names, gene names, gene synonyms, Ensembl IDs, functional information, keywords, and sequences. In total, we obtained 873,718 nonredundant new p-sites of 125,180

proteins from 575 HTP phosphoproteomic studies. By collecting functional p-sites with their summarized descriptions from PubMed and PhosphoSitePlus [24], we classified the functional descriptions into short categories, and finally retained 88,074 functional downstream events for 32,762 p-sites.

Next, these p-sites were merged with the existing dataset in EPSD 1.0 [1] and ten other public databases, including PhosphoSitePlus [24], dbPTM [26], UniProt [42], PhosphoAt [37], BioGRID [45], RegPhos [47], iPTMnet [30], Plant PTM Viewer [43], Pf-Phospho [48], and Scop3P [49] (Figure 1A). For each database, the obtained phosphopeptides were remapped to the benchmark sequences of UniProt, and exact positions of p-sites were pinpointed. To ensure the consistency and reliability of data, the origins of p-sites obtained from other databases were documented. Ultimately, we obtained 2,769,163 nonredundant p-sites on 362,707 phosphoproteins across 223 species with annotations of 88,074 functional events (Figure 1B; Tables S3 and S4). In comparison to EPSD 1.0 and other phosphorylation databases, EPSD 2.0 exhibited an approximate 71% enhancement in collection of p-sites and a nearly 2-fold increase in functional events.

Multilayer annotation for phosphoproteins and p-sites

Designed as a protein-centered database, EPSD 2.0 assigned each phosphoprotein with an automatically generated EPSD ID (EP-) as the main identifier, and utilized a UniProt accession as the alternative identifier, e.g., EP0017523 corresponds to the PLK1 protein in *Homo sapiens* with UniProt ID P53350. To provide detailed information for each phosphoprotein entry, some basic annotations were integrated from UniProt [42], for example, accession numbers of UniProt/RefSeq/Ensembl, National Center for Biotechnology Information (NCBI) taxa IDs, gene names and synonyms, NCBI gene IDs, protein names and synonyms, protein and nucleotide sequences, Gene Ontology (GO) terms, and keywords. Additionally, the representative three-dimensional (3D) structure of the protein is provided, if available, from the Protein Data Bank (PDB) database [38] or the AlphaFold database [54], with p-sites visualized on the structure. For each p-site, a flanking peptide of 15 amino acids centered around the middle-phosphorylated residue was provided with its reference information and PubMed IDs. For p-sites from HTP experiments, the initial peptides detected via MS and their cell or tissue origins were preserved when available. MS data analyzing software like MaxQuant can compute and assign the localization probability (LP) score for every candidate p-site [55]. The LP score quantifies the likelihood of correctly identifying p-sites within protein sequences based on MS data. The LP score varies between 0 and 1, with higher values indicating a greater probability that a site is an authentic p-site. Accordingly, we classified the p-sites with pre-calculated LP scores directly obtained from these studies into four classes: I (> 0.75), II ($0.50 < LP \leq 0.75$), III ($0.25 \leq LP \leq 0.50$) and IV (< 0.25) [55]. To ensure the consistency with the original papers, we retained all p-sites from HTP experiments without applying any additional filters, following the settings of EPSD 1.0.

In addition to the basic information, we reviewed each functional description from literature, summarized them into short phrases, and classified them into two types: “Effect on Protein”

and “Effect on Biological Process” (Figure 1A). For example, “enzymatic activity, induce” falls under “Effect on Protein”, while “cell growth, inhibit” is one category of “Effect on Biological Process”. Moreover, we meticulously annotated phosphoproteins in 8 model organisms, leveraging information from 100 external resources (Figure 1A). The annotation covered 15 different aspects, including kinase/phosphatase, transcription regulators, 3D structures, physicochemical characteristics, genomic variations, functional descriptions, protein domains, molecular interactions, drug–target associations, disease-related data, orthologs, transcript expression levels, proteomics, subcellular localization, and regulatory pathways. (Figure 1A; Table S2). These annotations were directly used, and no inconsistencies were found. File S1 provided in-depth details on how annotations from each resource were processed. The p-sites and annotation datasets in EPSD 2.0 are available at <https://epsd.biocuckoo.cn/Download.php>.

Data statistics in the EPSD 2.0

Compared to EPSD 1.0, the updated EPSD 2.0 has added 1,152,359 p-sites and 153,381 new phosphoproteins. Additionally, the number of eukaryotes has increased by over 3-fold to 223 species, including 95 animals, 20 protists, 61 plants, and 48 fungi. The heatmap in Figure 2A illustrates the number of phosphoproteins, p-sites, phospho-serine (pS) residues, phospho-threonine (pT) residues, and phospho-tyrosine (pY) residues across the top 80 organisms. The specific numbers for all organisms are presented in Table S3. The results revealed that *H. sapiens* and *Mus musculus* occupy approximately 57.99% of all the p-sites.

EPSD 1.0 did not provide descriptions of potential functional effects related to p-sites. In EPSD 2.0, we annotated 88,074 functional downstream events on 32,762 p-sites, including 58 categories related to protein function and 107 categories about biological processes (Figure 2B). Notably, the most common effect on proteins is “molecular association, regulation”, indicating that phosphorylation on proteins usually alters the interactions with other molecules. Additionally, when a site was phosphorylated, its downstream effects on biological processes were most frequently involved in the disease progression.

Among all the p-sites, there were 1,930,151 (69.70%) pS, 638,944 (23.07%) pT, and 200,020 (7.22%) pY residues, with the remaining p-sites involving other residues like histidine. The distribution of p-sites per phosphoprotein revealed that 260,683 phosphoproteins (71.87%) were phosphorylated with at least two p-sites, highlighting multisite phosphorylation as a dominant regulatory mechanism for substrate phosphoproteins (Figure 2C). Specifically, 7840 (2.16%) phosphoproteins exhibited 50 or more p-sites, demonstrating the highly complex phosphorylation regulatory patterns in these proteins (Figure 2C).

Usage

EPSD 2.0 features an intuitive and easy-to-navigate interface. Using the human PLK1 protein, here we demonstrate how to utilize the EPSD 2.0 website. Users can perform a one-click search on the home or search page by selecting “Protein Name” as the search option and entering the corresponding name “Serine/threonine-protein kinase PLK1” in the input box of “Substrate Search” field (Figure 3A). After clicking the “Submit” button or pressing “Enter”, a table will be

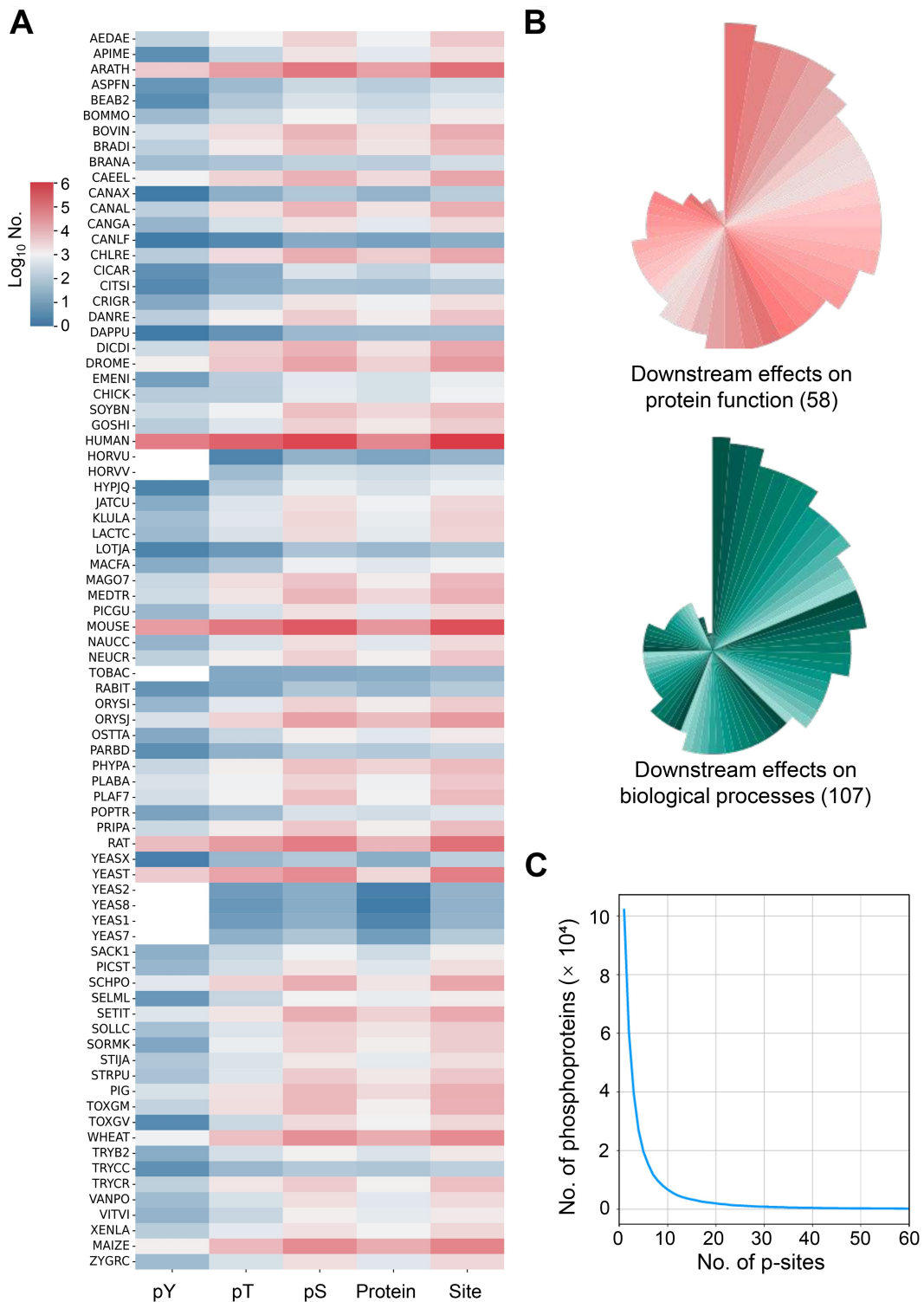


Figure 2 The data statistics in EPSD 2.0

A. Heatmap showing the distribution of p-sites, phosphoproteins, and pY/pT/pS residues for the top 80 species in EPSD 2.0. **B.** Downstream effects on protein function and biological processes. **C.** Distribution of p-sites in protein substrates. pS, phospho-serine; pT, phospho-threonine; pY, phospho-tyrosine; AEDAE, *Aedes aegypti*; APIME, *Apis mellifera*; ARATH, *Arabidopsis thaliana*; ASPFN, *Aspergillus flavus* (strain ATCC 200026 / FGSC A1120 / NRRL 3357 / JCM 12722 / SRRC 167); BEAB2, *Beauveria bassiana* (strain ARSEF 2860); BOMMO, *Bombyx mori*; BOVIN, *Bos taurus*; BRADI, *Brachypodium distachyon*; BRANA, *Brassica napus*; CAEEL, *Caenorhabditis elegans*; CANAX, *Candida albicans*; CANAL, *Candida albicans* (strain SC5314 / ATCC MYA-2876); CANGA, *Candida glabrata* (strain ATCC 2001 / CBS 138 / JCM 3761 / NBRC 0622 / NRRL Y-65); CANLF, *Canis lupus familiaris*; CHLRE, *Chlamydomonas reinhardtii*; CICAR, *Cicer arietinum*; CITSI, *Citrus sinensis*; CRIGR, *Cricetulus griseus*; DANRE, *Danio rerio*; DAPPU, *Daphnia pulex*; DICDI, *Dictyostelium discoideum*; DROME, *Drosophila melanogaster*; EMENI, *Emericella nidulans* (strain FGSC A4 / ATCC 38163 / CBS 112.46 / NRRL 194 / M139); CHICK, *Gallus gallus*; SOYBN, *Glycine max*; GOSHI, *Gossypium hirsutum*; HUMAN, *Homo sapiens*; HORVU, *Hordeum vulgare*; HORVV, *Hordeum vulgare* subsp. *vulgare*; HYPJQ, *Hypocrea jecorina* (strain QM6a); JATCU, *Jatropha curcas*; KLULA, *Kluyveromyces lactis* (strain ATCC 8585 / CBS 2359 / DSM 70799 / NBRC 1267 / NRRL Y-1140 / WM37); LACTC, *Lachancea thermotolerans* (strain ATCC 56472 / CBS 6340 / NRRL Y-8284); LOTJA, *Lotus japonicus*; MACFA, *Macaca fascicularis*; MAGO7, *Magnaporthe oryzae* (strain 70-15 / ATCC MYA-4617 / FGSC 8958); MEDTR, *Medicago truncatula*; PICGU, *Meyerozyma guilliermondii* (strain ATCC 6260 / CBS 566 / DSM 6381 / JCM 1539 / NBRC 10279 / NRRL Y-324); MOUSE, *Mus musculus*;

generated to display the search results, which includes the columns of EPSD ID, UniProt accession, gene name, protein name, and species (Figure 3A). The same result can also be obtained from the browse page by clicking the “*Homo sapiens*” in the “Browse by Species” section (Figure 3A). Then, a “View” page will be displayed to provide more detailed information by clicking on “EP0017523”, the EPSD ID corresponding to the human PLK1 protein.

The “View” page typically comprises ten sections. In the “Overview” section, we offer the basic information, including UniProt ID, protein/gene name, organism, and NCBI taxa ID. In addition, if the protein was mapped to the PDB database [56] or the AlphaFold database [54], its typical 3D structure would be shown by 3Dmol.js [57] with p-sites highlighted on the structure (Figure 3B). In the section of “Sites”, users can view the p-sites with corresponding disorder propensity and surface accessibility calculated by IUPred [58] and NetSurfP [59], respectively. Previous research has shown that p-sites are predominantly situated in intrinsically disordered regions [60,61], and the residues with higher surface accessibility are more likely to be phosphorylated [62,63]. In turn, phosphorylation can result in conformational changes of proteins, which may alter disorder regions and surface accessibility [61,64]. Therefore, we consider this information will assist biologists in identifying and prioritizing p-sites with potential functional significance. Additionally, all the p-sites are presented in a tabular list along with corresponding source information in the next “Experimentally identified p-sites” section. The reference information is categorized into three types: HTP, database, and LTP. A “View” button allows users to display further details such as position, phosphopeptides, LP scores, cell/tissue sources, data sources, and PMIDs (Figure 3C). For p-sites with information on downstream effects, users can view these annotations along with PMIDs in the section of “Downstream” (Figure 3C). The remaining sections provide annotations for functional descriptions, protein and gene sequences, GO terms, and keywords.

Users can access further detailed annotations for this protein in the “Annotation” section or by clicking the “Integrated Annotations” button at the top of the “Overview” section (Figure 3B). Subsequently, clicking on each category will redirect to the page that displays the relevant annotation content. For instance, users can click on “PhosphoSitePlus” under the “Regulator” title to access protein kinases that phosphorylate protein PLK1 curated from PhosphoSitePlus database [24] (Figure 3D). On the page of annotation, various types of annotation can be accessed through the left column (Figure 3D). For example, the physicochemical property can also be viewed by clicking the option “Physicochemical” on the left (Figure 3D).

All the detailed information for the 15 aspects of annotation can be viewed in Figure 4, using human PLK1 as an example.

In addition to the protein search, the search page also offers four other options, such as “Advanced Search”, “Peptide Search”, “Batch Search”, and “BLAST Search”. Each option is tailored to provide different functionalities for more convenient searching. The “Peptide Search” allows users to search the curated phosphopeptides such as “KK_pTLCGTPNYIAPEVLSK”, where the “p” indicates that the following amino acid is a p-site. The “Advanced Search” allows users to enter multiple conditions to perform more precise database queries. The “Batch Search” option permits multiple input keywords in a line-by-line format. Moreover, the “BLAST Search” option enables the submission of a single protein sequence in FASTA (Fast-All) format to identify identical or homologous proteins in the database.

Discussion

Protein phosphorylation serves as a crucial and well-studied PTM that modulates protein function, activity, localization, stability, and interactions with other molecules, thereby playing an essential role in virtually all cellular processes [2–7,63–66]. In recent years, MS/MS has continuously generated a number of p-sites across various species [12,13]. The collection and integration of these p-sites can offer valuable resources for elucidating phosphorylation’s regulatory mechanisms. Previously, we developed EPSD 1.0 to curate and annotate p-sites [1]. As the biological research delves deeper into more specific mechanism of biological questions, the functional annotation of p-sites has become increasingly important. In the latest version, EPSD 2.0 introduces significant improvements, particularly in the augmentation of p-sites and inclusion of their functional annotations. Totally, EPSD 2.0 contains 2,769,163 p-sites occurring on 362,707 unique proteins across 223 species, and annotates 88,074 downstream functional events of 32,762 p-sites, covering 58 types of downstream effects on phosphoproteins and 107 regulatory impacts on biological processes. With a total size of 36.2 GB, EPSD 2.0 shows a 2.5-fold expansion in data volume compared to EPSD 1.0 (Table S5).

The information in EPSD 2.0 also reveals some interesting insights about phosphorylation. For example, analyzing the downstream effects of p-sites on human PLK1 reveals that the phosphorylation of T210 can trigger lots of effects, such as inducing its own enzymatic activity, altering intracellular localization, and inducing autophagy. Surprisingly, T210 phosphorylation can both induce and inhibit carcinogenesis, which may seem counterintuitive at first glance. By reviewing

Figure 2 Continued

NAUCC, *Naumovozyma castellii* (strain ATCC 76901 / CBS 4309 / NBRC 1992 / NRRL Y-12630); NEUCR, *Neurospora crassa* (strain ATCC 24698 / 74-OR23-1A / CBS 708.71 / DSM 1257 / FGSC 987); TOBAC, *Nicotiana tabacum*; RABIT, *Oryctolagus cuniculus*; ORYSI, *Oryza sativa* subsp. *indica*; ORYSJ, *Oryza sativa* subsp. *japonica*; OSTTA, *Ostreococcus tauri*; PARBD, *Paracoccidioides brasiliensis* (strain Pb18); PHYP, *Physcomitrium patens*; PLABA, *Plasmodium berghei* (strain Anka); PLAF7, *Plasmodium falciparum* (isolate 3D7); POPTR, *Populus trichocarpa*; PRIPA, *Pristionchus pacificus*; RAT, *Rattus norvegicus*; YEASX, *Saccharomyces cerevisiae*; YEAST, *Saccharomyces cerevisiae* (strain ATCC 204508 / S288c); YEAS2, *Saccharomyces cerevisiae* (strain JAY291); YEAS8, *Saccharomyces cerevisiae* (strain Lalvin EC1118 / Prise de mousse); YEAS1, *Saccharomyces cerevisiae* (strain RM11-1a); YEAS7, *Saccharomyces cerevisiae* (strain YJM789); SACK1, *Saccharomyces kudriavzevii* (strain ATCC MYA-4449 / AS 2.2408 / CBS 8840 / NBRC 1802 / NCYC 2889); PICST, *Scheffersomyces stipitis* (strain ATCC 58785 / CBS 6054 / NBRC 10063 / NRRL Y-11545); SCHPO, *Schizosaccharomyces pombe* (strain 972 / ATCC 24843); SELML, *Selaginella moellendorffii*; SETIT, *Setaria italica*; SOLLC, *Solanum lycopersicum*; SORMK, *Sordaria macrospora* (strain ATCC MYA-333 / DSM 997 / K(L3346) / K-hell); STIJA, *Stichopus japonicus*; STRPU, *Strongylocentrotus purpuratus*; PIG, *Sus scrofa*; TOXGM, *Toxoplasma gondii* (strain ATCC 50611 / Me49); TOXGV, *Toxoplasma gondii* (strain ATCC 50861 / VEG); WHEAT, *Triticum aestivum*; TRYB2, *Trypanosoma brucei brucei* (strain 927/4 GUTat10.1); TRYCC, *Trypanosoma cruzi* (strain CL Brener); TRYCR, *Trypanosoma cruzi* Dm28c; VANPO, *Vanderwaltozyma polyspora* (strain ATCC 22028 / DSM 70294); VITVI, *Vitis vinifera*; XENLA, *Xenopus laevis*; MAIZE, *Zea mays*; ZYGRC, *Zygosaccharomyces rouxii* (strain ATCC 2623 / CBS 732 / NBRC 1130 / NCYC 568 / NRRL Y-229).

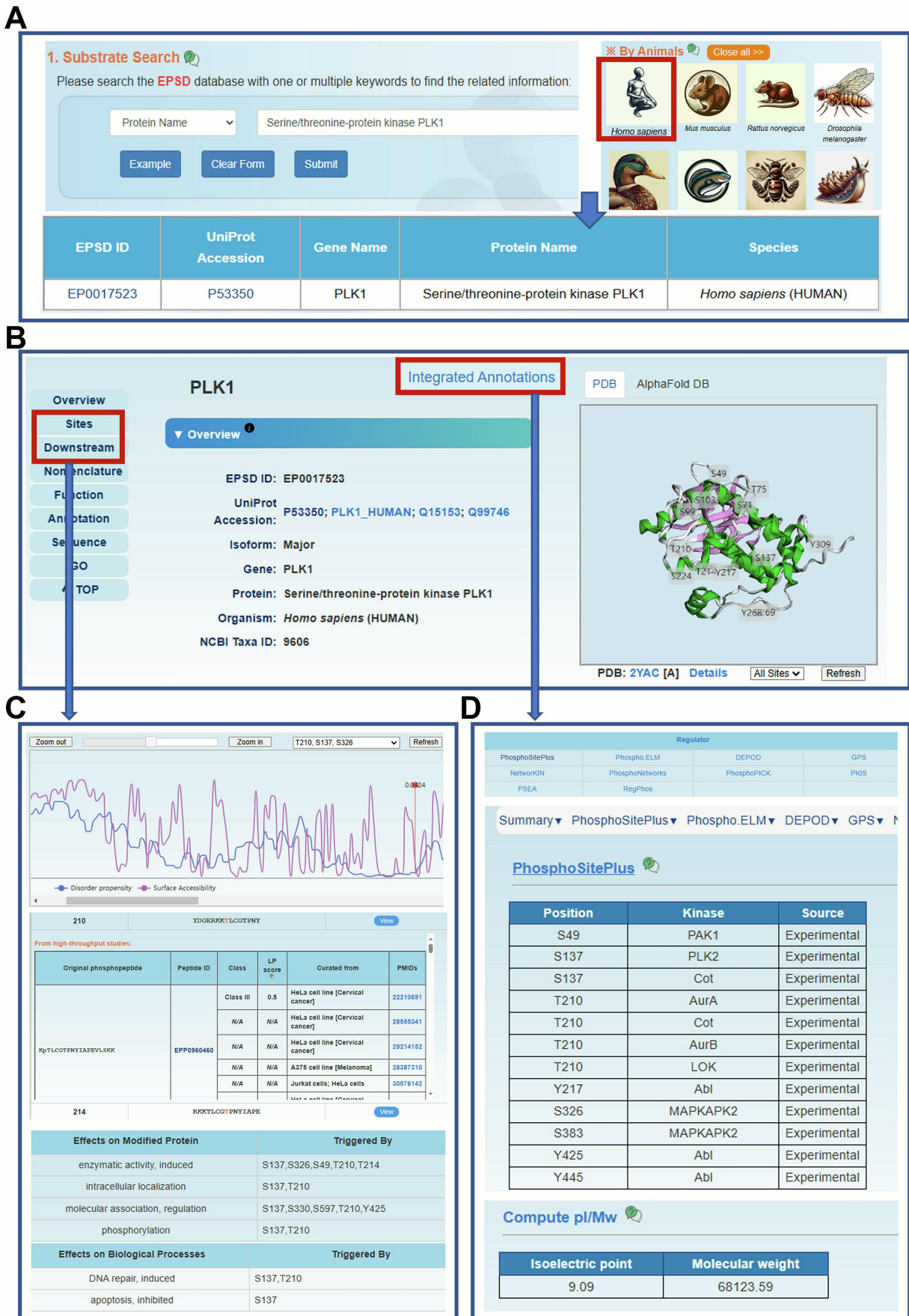


Figure 3 The browse and search options of EPSP 2.0

A. The “Substrate Search” and “Browse by Species” options. **B.** The overview of human PLK1. **C.** The basic information and details on p-sites. **D.** The annotation of PLK1. “Regulator” was presented as an example.

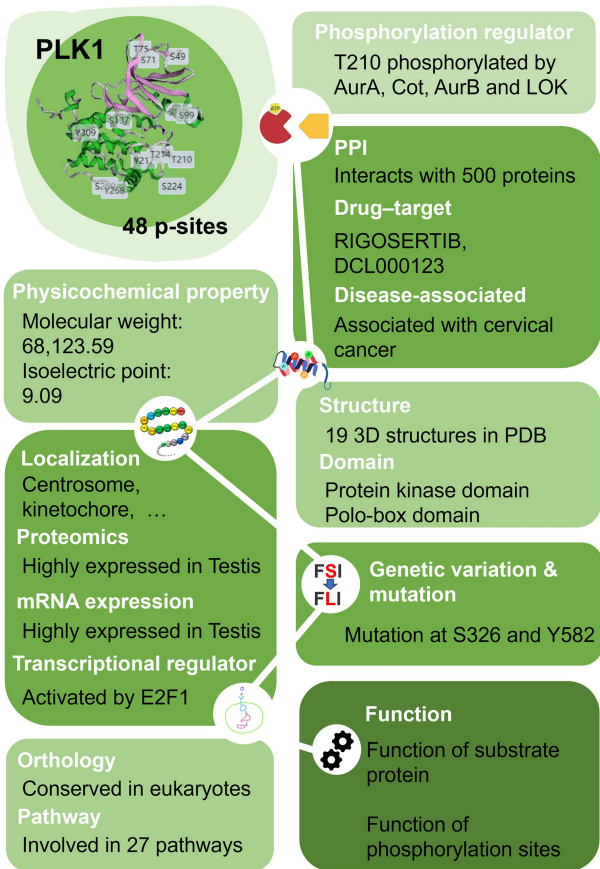


Figure 4 The overview example of integrated annotations

The overview of integrated annotations for human PLK1 protein. A brief summary of all the data resources used in this study is shown in Table S2. The details about the processing of each resource are shown in File S1. mRNA, messenger RNA; PPI, protein-protein interaction.

the original literature, this paradox can be explained. PLK1 is often regarded as a key oncogenic protein because of its critical function in promoting tumor cell division [67]. For example, Zhu et al. discovered that VRK2 kinase phosphorylated PLK1 at T210 to prevent its ubiquitin-dependent proteasomal degradation [68]. This led to the overexpression of PLK1, which drives cancer progression [69]. However, some scientists found that PLK1 overexpression could induce chromosomal instability and suppress tumor development [70]. Yu et al. demonstrated that PLK1 phosphorylated at T210 could act as a tumor suppressor by maintaining the integrity of the spindle assembly checkpoint, preventing chromosomal instability, and inhibiting tumor development in the presence of functional DAB2IP [71]. Understanding these complex regulatory mechanisms is important for the development of effective anti-cancer strategies and for gaining insight into cell cycle regulation. This type of data is important, but its collection is still insufficient.

Although EPSD 2.0 maintains 2,769,163 p-sites, only 32,762 p-sites (1.18%) have been annotated with downstream effects. Thus, the functional effects of most of the p-sites remain to be studied. For example, the S187 residue of human partitioning defective 3 homolog (PARD3; EPSD ID: EP0065790) has been identified as a p-site from multiple phosphoproteomic studies. This site is evolutionarily conserved, and its orthologous counterpart in *Drosophila melanogaster* is S201 of Bazooka (EPSD ID: EP0024205), which

was also annotated as an HTP p-site. Based on these annotations, Loyer et al. further uncovered that CDK1 phosphorylates the S201 site of Bazooka to orchestrate neuroblast polarization and sensory organ formation [72] (Figure S1A). In *Arabidopsis thaliana*, Fu et al. identified phosphorylation of protein MALE DISCOVERER 1 (MDIS1; EPSD ID: EP0151139) at S377, an HTP p-site annotated in EPSD, which regulates the auto-phosphorylation of its co-receptor, MDIS1-interacting receptor like kinase 1 (MIK1) [73] (Figure S1B). In addition, Hansen et al. used all annotated phosphoproteins in *M. musculus* from EPSD, and revealed that phosphorylation changes of placental proteins occurred much earlier than messenger RNA (mRNA) expression changes in response to pulmonary maternal inflammation [74] (Figure S1C). Besides providing useful clues for further experimental design, EPSD also contributes a helpful dataset toward the development of highly accurate predictors. For example, using the data from EPSD 1.0, we previously developed an on-line service of Group-based Prediction System (GPS) 6.0, which can hierarchically predict kinase-specific p-sites for up to 44,046 protein kinases in 185 eukaryotic species [75]. We believe that more useful predictors will be developed based on the data of EPSD 2.0.

EPSD 2.0 will be regularly maintained and continuously updated to include newly reported p-sites in the literature. Additionally, more downstream functional events will be collected after reviewing more articles due to the importance of the downstream effects of p-sites and the shortage of their curation. Also, further annotations from external public resources will be included to enhance the comprehensiveness and connectivity of the database.

Data availability

EPSD 2.0 is accessible at <https://epsd.biocuckoo.cn/>. EPSD 2.0 has been submitted to Database Commons [76] at the National Genomics Data Center (NGDC), China National Center for Bioinformation (CNCB), which is publicly accessible at <https://ngdc.cncb.ac.cn/databasecommons/database/id/6945>.

CRedit author statement

Miaomiao Chen: Methodology, Software, Investigation, Data curation, Visualization, Writing – original draft, Writing – review & editing. **Yujie Gou:** Investigation, Data curation, Visualization, Writing – review & editing. **Ming Lei:** Data curation. **Leming Xiao:** Data curation. **Miaoying Zhao:** Data curation. **Xinhe Huang:** Data curation. **Dan Liu:** Data curation. **Zihao Feng:** Data curation. **Di Peng:** Writing – original draft, Writing – review & editing. **Yu Xue:** Writing – original draft, Writing – review & editing, Supervision, Project administration, Funding acquisition. All authors have read and approved the final manuscript.

Competing interests

The authors have declared no competing interests.

Supplementary material

Supplementary material is available at *Genomics, Proteomics & Bioinformatics* online (<https://doi.org/10.1093/gpbjnl/qzaf057>).

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Grant Nos. 32341020 and 32341021), the National Key R&D Program of China (Grant Nos. 2022YFC2704304 and 2021YFF0702000), the Interdisciplinary Research Program of HUST (Grant Nos. 2023JCYJ010 and 2024JCYJ013), the Hubei Province Postdoctoral Outstanding Talent Tracking Support Program, and the Research Core Facilities for Life Science (HUST), China.

ORCID

0000-0003-2458-1260 (Miaomiao Chen)

0000-0002-5399-9757 (Yujie Gou)

0009-0007-5860-0534 (Ming Lei)

0000-0002-6152-492X (Leming Xiao)

0009-0003-2890-0975 (Miaoying Zhao)

0009-0009-2138-1037 (Xinhe Huang)

0000-0002-8376-7287 (Dan Liu)

0000-0001-5667-3819 (Zihao Feng)

0000-0002-1249-3741 (Di Peng)

0000-0002-9403-6869 (Yu Xue)

References

- Lin S, Wang C, Zhou J, Shi Y, Ruan C, Tu Y, et al. EPSD: a well-annotated data resource of protein phosphorylation sites in eukaryotes. *Brief Bioinform* 2021;22:298–307.
- Bilbrough T, Piemontese E, Seitz O. Dissecting the role of protein phosphorylation: a chemical biology toolbox. *Chem Soc Rev* 2022;51:5691–730.
- Shindo S, Kakizaki S, Sakaki T, Kawasaki Y, Sakuma T, Negishi M, et al. Phosphorylation of nuclear receptors: novelty and therapeutic implications. *Pharmacol Ther* 2023;248:108477.
- Zhang WJ, Zhou Y, Zhang Y, Su YH, Xu T. Protein phosphorylation: a molecular switch in plant signaling. *Cell Rep* 2023;42:112729.
- Narasimamurthy R, Virshup DM. The phosphorylation switch that regulates ticking of the circadian clock. *Mol Cell* 2021;81:1133–46.
- Bauman AL, Scott JD. Kinase- and phosphatase-anchoring proteins: harnessing the dynamic duo. *Nat Cell Biol* 2002;4:E203–6.
- Matsumoto T, Wang PY, Ma W, Sung HJ, Matoba S, Hwang PM. Polo-like kinases mediate cell survival in mitochondrial dysfunction. *Proc Natl Acad Sci U S A* 2009;106:14542–6.
- Chen LL, Wang YB, Song JX, Deng WK, Lu JH, Ma LL, et al. Phosphoproteome-based kinase activity profiling reveals the critical role of MAP2K2 and PLK1 in neuronal autophagy. *Autophagy* 2017;13:1969–80.
- Lee JH, Paull TT. Cellular functions of the protein kinase ATM and their relevance to human disease. *Nat Rev Mol Cell Biol* 2021;22:796–814.
- Dong Y, Hu H, Zhang X, Zhang Y, Sun X, Wang H, et al. Phosphorylation of PHF2 by AMPK releases the repressive H3K9me2 and inhibits cancer metastasis. *Signal Transduct Target Ther* 2023;8:95.
- Zhang TT, Lei QQ, He J, Guan X, Zhang X, Huang Y, et al. Bestrophin3 deficiency in vascular smooth muscle cells activates MEK2/3–MAPK signaling to trigger spontaneous aortic dissection. *Circulation* 2023;148:589–606.
- Safabakhsh S, Panwar P, Barichello S, Sangha SS, Hanson PJ, Petegem FV, et al. The role of phosphorylation in atrial fibrillation: a focus on mass spectrometry approaches. *Cardiovasc Res* 2022;118:1205–17.
- Chen Y, Weckwerth W. Mass spectrometry untangles plant membrane protein signaling networks. *Trends Plant Sci* 2020;25:930–44.
- Yates JR, Ruse CI, Nakorchevsky A. Proteomics by mass spectrometry: approaches, advances, and applications. *Annu Rev Biomed Eng* 2009;11:49–79.
- Lerner R, Baker D, Schwitter C, Neuhaus S, Hauptmann T, Post JM, et al. Four-dimensional trapped ion mobility spectrometry lipidomics for high throughput clinical profiling of human blood samples. *Nat Commun* 2023;14:937.
- Brown KA, Morris RK, Eckhardt SJ, Ge Y, Gellman SH. Phosphorylation sites of the gastric inhibitory polypeptide receptor (GIPR) revealed by trapped-ion-mobility spectrometry coupled to time-of-flight mass spectrometry (TIMS-TOF MS). *J Am Chem Soc* 2023;145:28030–7.
- Kramer MH, Zhang Q, Sprung R, Day RB, Erdmann-Gilmore P, Li Y, et al. Proteomic and phosphoproteomic landscapes of acute myeloid leukemia. *Blood* 2022;140:1533–48.
- Stewart H, Grinfeld D, Wagner A, Kholomeev A, Biel M, Giannakopoulos A, et al. A conjoined rectilinear collision cell and pulsed extraction ion trap with auxiliary DC electrodes. *J Am Soc Mass Spectrom* 2024;35:74–81.
- Guzman UH, Martinez-Val A, Ye Z, Damoc E, Arrey TN, Pashkova A, et al. Ultra-fast label-free quantification and comprehensive proteome coverage with narrow-window data-independent acquisition. *Nat Biotechnol* 2024;42:1855–66.
- Lancaster NM, Sinitcyn P, Fornoy P, Peters-Clarke TM, Fecher C, Smith AJ, et al. Fast and deep phosphoproteome analysis with the Orbitrap Astral mass spectrometer. *Nat Commun* 2024;15:7016.
- Diella F, Cameron S, Gemund C, Linding R, Via A, Kuster B, et al. Phospho.ELM: a database of experimentally verified phosphorylation sites in eukaryotic proteins. *BMC Bioinformatics* 2004;5:79.
- Dinkel H, Chica C, Via A, Gould CM, Jensen LJ, Gibson TJ, et al. Phospho.ELM: a database of phosphorylation sites—update 2011. *Nucleic Acids Res* 2011;39:D261–7.
- Hornbeck PV, Zhang B, Murray B, Kornhauser JM, Latham V, Skrzypek E. PhosphoSitePlus, 2014: mutations, PTMs and recalibrations. *Nucleic Acids Res* 2015;43:D512–20.
- Hornbeck PV, Kornhauser JM, Latham V, Murray B, Nandhikonda V, Nord A, et al. 15 years of PhosphoSitePlus®: integrating post-translationally modified sites, disease variants and isoforms. *Nucleic Acids Res* 2019;47:D433–41.
- Huang KY, Lee TY, Kao HJ, Ma CT, Lee CC, Lin TH, et al. dbPTM in 2019: exploring disease association and cross-talk of post-translational modifications. *Nucleic Acids Res* 2019;47:D298–308.
- Li Z, Li S, Luo M, Jhong JH, Li W, Yao L, et al. dbPTM in 2022: an updated database for exploring regulatory networks and functional associations of protein post-translational modifications. *Nucleic Acids Res* 2022;50:D471–9.
- Li H, Xing X, Ding G, Li Q, Wang C, Xie L, et al. SysPTM: a systematic resource for proteomic research on post-translational modifications. *Mol Cell Proteomics* 2009;8:1839–49.
- Li J, Jia J, Li H, Yu J, Sun H, He Y, et al. SysPTM 2.0: an updated systematic resource for post-translational modification. *Database (Oxford)* 2014;2014:bau025.
- Ross KE, Huang H, Ren J, Arighi CN, Li G, Tudor CO, et al. iPTMnet: integrative bioinformatics for studying PTM networks. *Methods Mol Biol* 2017;1558:333–53.
- Huang H, Arighi CN, Ross KE, Ren J, Li G, Chen SC, et al. iPTMnet: an integrated resource for protein post-translational modification network discovery. *Nucleic Acids Res* 2018;46:D542–50.
- Gnad F, Gunawardena J, Mann M. PHOSIDA 2011: the post-translational modification database. *Nucleic Acids Res* 2011;39:D253–60.
- Bodenmiller B, Aebersold R. Phosphoproteome resource for systems biology research. *Methods Mol Biol* 2011;694:307–22.

- [33] Ovelheiro D, Carrascal M, Casas V, Abian J. LymPHOS: design of a phosphosite database of primary human T cells. *Proteomics* 2009;9:3741–51.
- [34] Nguyen TD, Vidal-Cortes O, Gallardo O, Abian J, Carrascal M. LymPHOS 2.0: an update of a phosphosite database of primary human T cells. *Database (Oxford)* 2015;2015:bav115.
- [35] Nuhse TS, Stensballe A, Jensen ON, Peck SC. Phosphoproteomics of the *Arabidopsis* plasma membrane and a new phosphorylation site database. *Plant Cell* 2004;16:2394–405.
- [36] Heazlewood JL, Durek P, Hummel J, Selbig J, Weckwerth W, Walther D, et al. PhosphAt: a database of phosphorylation sites in *Arabidopsis thaliana* and a plant-specific phosphorylation site predictor. *Nucleic Acids Res* 2008;36:D1015–21.
- [37] Zulawski M, Braginets R, Schulze WX. PhosphAt goes kinases—searchable protein kinase target information in the plant phosphorylation site database PhosphAt. *Nucleic Acids Res* 2013;41:D1176–84.
- [38] Yao Q, Ge H, Wu S, Zhang N, Chen W, Xu C, et al. P³DB 3.0: from plant phosphorylation sites to protein networks. *Nucleic Acids Res* 2014;42:D1206–13.
- [39] Rose CM, Venkateswaran M, Grimsrud PA, Westphall MS, Sussman MR, Coon JJ, et al. Medicago PhosphoProtein Database: a repository for *Medicago truncatula* phosphoprotein data. *Front Plant Sci* 2012;3:122.
- [40] Bai Y, Chen B, Li M, Zhou Y, Ren S, Xu Q, et al. FPD: a comprehensive phosphorylation database in fungi. *Fungal Biol* 2017;121:869–75.
- [41] UniProt Consortium. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res* 2019;47:D506–15.
- [42] UniProt Consortium. UniProt: the Universal Protein Knowledgebase in 2023. *Nucleic Acids Res* 2023;51:D523–31.
- [43] Willems P, Horne A, Van Parys T, Goormachtig S, De Smet I, Botzki A, et al. The Plant PTM Viewer, a central resource for exploring plant protein modifications. *Plant J* 2019;99:752–62.
- [44] Goel R, Harsha HC, Pandey A, Prasad TS. Human Protein Reference Database and Human Proteinpedia as resources for phosphoproteome analysis. *Mol Biosyst* 2012;8:453–63.
- [45] Oughtred R, Stark C, Breitkreutz BJ, Rust J, Boucher L, Chang C, et al. The BioGRID interaction database: 2019 update. *Nucleic Acids Res* 2019;47:D529–41.
- [46] Lee TY, Bo-Kai Hsu J, Chang WC, Huang HD. RegPhos: a system to explore the protein kinase–substrate phosphorylation network in humans. *Nucleic Acids Res* 2011;39:D777–87.
- [47] Huang KY, Wu HY, Chen YJ, Lu CT, Su MG, Hsieh YC, et al. RegPhos 2.0: an updated resource to explore protein kinase–substrate phosphorylation networks in mammals. *Database (Oxford)* 2014;2014:bau034.
- [48] Gupta P, Venkadesan S, Mohanty D. Pf-Phospho: a machine learning-based phosphorylation sites prediction tool for *Plasmodium* proteins. *Brief Bioinform* 2022;23:bbac249.
- [49] Ramasamy P, Turan D, Tichshenko N, Hulstaert N, Vandermarliere E, Vranken W, et al. Scop3P: a comprehensive resource of human phosphosites within their full context. *J Proteome Res* 2020;19:3478–86.
- [50] Zhao J, Zou L, Li Y, Liu X, Zeng C, Xu C, et al. HisPhosSite: a comprehensive database of histidine phosphorylated proteins and sites. *J Proteomics* 2021;243:104262.
- [51] Zhao MX, Ding RF, Chen Q, Meng J, Li F, Fu S, et al. Nphos: database and predictor of protein N-phosphorylation. *Genomics Proteomics Bioinformatics* 2024;22:qzae032.
- [52] Cheng H, Deng W, Wang Y, Ren J, Liu Z, Xue Y. dbPPT: a comprehensive database of protein phosphorylation in plants. *Database (Oxford)* 2014;2014:bau121.
- [53] Ullah S, Lin S, Xu Y, Deng W, Ma L, Zhang Y, et al. dbPAF: an integrative database of protein phosphorylation in animals and fungi. *Sci Rep* 2016;6:23534.
- [54] Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. *Nature* 2021;596:583–9.
- [55] Humphrey SJ, Yang G, Yang P, Fazakerley DJ, Stockli J, Yang JY, et al. Dynamic adipocyte phosphoproteome reveals that Akt directly regulates mTORC2. *Cell Metab* 2013;17:1009–20.
- [56] Burley SK, Bhikadiya C, Bi C, Bittrich S, Chao H, Chen L, et al. RCSB Protein Data Bank (RCSB.org): delivery of experimentally-determined PDB structures alongside one million computationally-determined PDB structures alongside one million computed structure models of proteins from artificial intelligence/machine learning. *Nucleic Acids Res* 2023;51:D488–508.
- [57] Rego N, Koes D. 3Dmol.js: molecular visualization with WebGL. *Bioinformatics* 2015;31:1322–4.
- [58] Erdos G, Pajkos M, Dosztanyi Z. IUPred3: prediction of protein disorder enhanced with unambiguous experimental annotation and visualization of evolutionary conservation. *Nucleic Acids Res* 2021;49:W297–303.
- [59] Hoie MH, Kiehl EN, Petersen B, Nielsen M, Winther O, Nielsen H, et al. NetSurfP-3.0: accurate and fast prediction of protein structural features by protein language models and deep learning. *Nucleic Acids Res* 2022;50:W510–5.
- [60] Iakoucheva LM, Radivojac P, Brown CJ, O'Connor TR, Sikes JG, Obradovic Z, et al. The importance of intrinsic disorder for protein phosphorylation. *Nucleic Acids Res* 2004;32:1037–49.
- [61] Wright PE, Dyson HJ. Intrinsically disordered proteins in cellular signalling and regulation. *Nat Rev Mol Cell Biol* 2015;16:18–29.
- [62] Pearlman SM, Serber Z, Ferrell JE Jr. A mechanism for the evolution of phosphorylation sites. *Cell* 2011;147:934–46.
- [63] Nishi H, Hashimoto K, Panchenko AR. Phosphorylation in protein–protein binding: effect on stability and function. *Structure* 2011;19:1807–15.
- [64] Edreira MM, Li S, Hochbaum D, Wong S, Gorfe AA, Ribeiro-Neto F, et al. Phosphorylation-induced conformational changes in Rap1b: allosteric effects on switch domains and effector loop. *J Biol Chem* 2009;284:27480–6.
- [65] Zhao MX, Chen Q, Li F, Fu S, Huang B, Zhao Y. Protein phosphorylation database and prediction tools. *Brief Bioinform* 2023;24:bbad090.
- [66] Noujaim J, Payne LS, Judson I, Jones RL, Huang PH. Phosphoproteomics in translational research: a sarcoma perspective. *Ann Oncol* 2016;27:787–94.
- [67] Strebhardt K, Ullrich A. Targeting polo-like kinase 1 for cancer therapy. *Nat Rev Cancer* 2006;6:321–30.
- [68] Zhu H, Li Q, Zhao Y, Peng H, Guo L, Zhu J, et al. Vaccinia-related kinase 2 drives pancreatic cancer progression by protecting Plk1 from Chfr-mediated degradation. *Oncogene* 2021;40:4663–74.
- [69] Smith MR, Wilson ML, Hamanaka R, Chase D, Kung H, Longo DL, et al. Malignant transformation of mammalian cells initiated by constitutive expression of the polo-like kinase. *Biochem Biophys Res Commun* 1997;234:397–405.
- [70] de Carcer G, Venkateswaran SV, Salgueiro L, El Bakkali A, Somogyi K, Rowald K, et al. Plk1 overexpression induces chromosomal instability and suppresses tumor development. *Nat Commun* 2018;9:3012.
- [71] Yu L, Lang Y, Hsu CC, Chen WM, Chiang JC, Hsieh JT, et al. Mitotic phosphorylation of tumor suppressor DAB2IP maintains spindle assembly checkpoint and chromosomal stability through activating PLK1–Mps1 signal pathway and stabilizing mitotic checkpoint complex. *Oncogene* 2022;41:489–501.
- [72] Loyer N, Hogg EKJ, Shaw HG, Pasztor A, Murray DH, Findlay GM, et al. A CDK1 phosphorylation site on *Drosophila* PAR-3 regulates neuroblast polarisation and sensory organ formation. *Elife* 2024;13:e97902.
- [73] Fu Q, Liu Q, Zhang R, Chen J, Guo H, Ming Z, et al. Large-scale analysis of the N-terminal regulatory elements of the kinase

- domain in plant receptor-like kinase family. *BMC Plant Biol* 2024;24:174.
- [74] Hansen SSK, Krautz R, Rago D, Havelund J, Stigliani A, Faergeman NJ, et al. Pulmonary maternal immune activation does not cross the placenta but leads to fetal metabolic adaptation. *Nat Commun* 2024;15:4711.
- [75] Chen M, Zhang W, Gou Y, Xu D, Wei Y, Liu D, et al. GPS 6.0: an updated server for prediction of kinase-specific phosphorylation sites in proteins. *Nucleic Acids Res* 2023;51:W243–50.
- [76] Ma L, Zou D, Liu L, Shireen H, Abbasi AA, Bateman A, et al. Database Commons: a catalog of worldwide biological databases. *Genomics Proteomics Bioinformatics* 2023;21:1054–8.