Chapter 14

Bioinformatics Analysis of Functional Associations of PTMs

Pablo Minguez and Peer Bork

Abstract

Post-translational modifications (PTMs) are an important source of protein regulation; they fine-tune the function, localization, and interaction with other molecules of the majority of proteins and are partially responsible for their multifunctionality. Usually, proteins have several potential modification sites, and their patterns of occupancy are associated with certain functional states. These patterns imply cross talk among PTMs within and between proteins, the majority of which are still to be discovered. Several methods detect associations between PTMs; these have recently combined into a global resource, the PTMcode database, which contains already known and predicted functional associations between pairs of PTMs from more than 45,000 proteins in 19 eukaryotic species.

Key words Systems biology, Proteomics, Protein regulation, Post-translational modifications, Protein–protein interactions

1 Introduction

The cell is a very robust system where the final response to stimuli depends on many layers of regulation. In the last two decades, many new techniques have been developed to provide snapshots of genome and proteome regulation at different levels: transcription, posttranscription, translation, or posttranslation. These new types of experiments have changed partially the discovery workflow in science—now we may start to test a hypothesis with a lot of data and few assumptions, following a top-down strategy. Three main steps are required in this type of analysis: (1) filtering, mainly applying strong statistical controls in order to reduce the false positives rate; (2) annotation, where a biological meaning is superimposed on the statistics; and (3) integration, where results are merged with other levels of regulation to model cell behavior. Bioinformatics has come to help with the development of new algorithms, tools, and databases to address these three challenges.

Cathy H. Wu et al. (eds.), *Protein Bioinformatics: From Protein Modifications and Networks to Proteomics*, Methods in Molecular Biology, vol. 1558, DOI 10.1007/978-1-4939-6783-4_14,

For proteins, their role in a particular cell state is partially controlled by means of the addition of small moieties called post-translational modifications (PTMs). There are many PTM types described that are amino acid specific; their number, position, and combination present at a particular moment determine the final state of a protein. Mass spectrometry (MS) technology is able to explore the modification status of the whole proteome with a particular type of PTM at once. This technique requires an enrichment of these moieties in order to increase their detection threshold, which results in false positives as a side effect. An increasing number of experiments are now available that report mostly phosphorylation sites but also the occupancy of other PTM types such as acetylations, ubiquitinations, glycosylations, and others, of nearly the whole proteome under different conditions. Under this scenario there are two big challenges to address in order to translate the huge amount of data available into reliable information about the regulation of specific proteins. First, we should try to discriminate among all PTMs reported to focus on those with biological relevance. Several databases work to gather this type of information, measuring the conservation of the modified amino acids as a proxy for the prevalence of the PTM over evolution [1, 2], calculating the accessible surface area [3], or mapping the PTMs onto more or less complete maps of the regulatory elements of the proteins [1–5] and even onto their secondary [3] and tertiary structures [1, 3, 6]. The second challenge is to elucidate the possible cross-regulatory effects of PTM combinations under specific conditions. There are hundreds of described examples of PTM cross talk in the literature [7, 8] (*see* **Note 1**), and it is postulated that the function, localization, and interactions of most proteins depend partially on their modification pattern [9]. However, the search for their cross-regulatory effects is still an unexplored field. There are a few systematic efforts that go beyond the basic annotation of the experiment where the modifications were reported [4] (e.g., the cell cycle phase where they are present [2]) to more detailed information; for example, the downstream effects of the modifications [10] and their interaction with other PTMs [1] have been manually annotated, and text-mining tools have been used to report the functional processes in which the modifications are involved [11]. Although these approaches provide very accurate and valuable information, they can only be applied to low-throughput experiments (LTEs), while the majority of the PTMs reported come from high-throughput experiments (HTEs). The PTMcode database complements this information with several prediction methods to annotate pairs of PTMs as being functionally linked. To the best of our knowledge, PTMcode is the only resource that provides predictions of post-translational regulation of proteins based on the functional associations among their PTMs. In addition, PTMcode also provides predicted cross talk between

PTMs in interacting proteins that might regulate their binding, placing PTMcode as a bridge between protein-protein interactions (PPIs) and PTM databases.

Herein, we describe in detail how to explore functional associations between protein modifications using the PTMcode database.

2 Materials

Two types of users can profit from the data produced by the PTMcode database: those interested in the post-translational regulation of particular proteins or protein families [12, 13] and those interested in performing further high-throughput computational analyses on the PTMs, either using our curated dataset of PTMs in one or many species [14–16] or applying other algorithms to the predicted functionally associated PTMs in order to contribute to the deciphering of the global "PTMcode" [17, 18] (*see* **Note 2**). For both type of tasks, the PTMcode database has implemented ways to explore and download the data.

PTMcode is freely accessible at http://ptmcode.embl.de and requires no more than a modern browser installed in a standard computer. In order to access some PTMcode features, users should allow java applets to be run. Please check if your favorite browser is supported here (https://java.com/en/download/help/enable_browser.xml).

3 Methods

3.1 PTMs and PPIs Are the Main Sources of Protein Post-translational Regulation

Two interconnected processes regulate final protein function and localization: (1) the interaction with other molecules, mainly other proteins to form transient or stable complexes, and (2) the addition of PTMs [19] that may regulate the protein's binding activity. Many resources are dedicated to gather and curate these two types of events. The PTMcode database is not a substitute for PTM databases [2, 11, 20] nor the repositories that compile or predict PPIs [21–23]; instead it incorporates both types of regulation in a unique resource in order to provide a complete picture of the post-translational regulation of eukaryotic proteins. Thus, PTMcode compiles in its second release:

1. Post-translational modifications from six databases [2, 3, 11, 20, 22, 24] and nine proteome-wide experiments [25–33] summing up 316,546 experimentally verified PTMs of 69 different types. The PTMs are mapped onto sequences extracted from the eggNOG database [34], which chooses the largest transcript as representative of the protein. To avoid spurious

mapping due to different sequence or transcript versions, PTMcode checks that all PTMs for a particular protein coming from the same source modify the type of amino acid reported. If even one PTM is mapped onto the incorrect residue, we assume that the protein sequence from the source is not the same as the one in our database, and all PTMs coming from that source for that protein are discarded. This methodology permits us to build a consistent and accurate dataset of protein modifications for 45,361 proteins from 19 eukaryotes.

2. Protein-protein interactions extracted from the STRING database [35]. STRING compiles and scores known and predicted protein-protein associations based on several types of evidence. PTMcode collects high-confidence PPIs (score >0.700) that are based on experimental evidence of physical binding ("experiments" evidence type). In total, PTMcode includes 221,268 PPIs from the 19 eukaryotes.

3.2 Non-Experimentally Verified PTMs

In order to understand the modification landscape of the proteome, the study of individual proteins by means of LTEs, although very accurate, is clearly insufficient. We need HTEs to cover as many species, proteins, PTM types, and conditions as possible. Despite the efforts of the community, this goal is still far from being accomplished as only a few species have been subject to this type of screening. Thus, many different tools have been developed to predict modifications (http://www.cbs.dtu.dk/databases/PTMpredictions/).

The PTMcode database annotates a new category of PTMs that complements the information provided by experimentally verified modifications—the so-called "propagated PTMs." The principle behind them is that conservation of an amino acid over evolution is a proxy for the conservation of its function and so for its modifications [36–38].

Propagated PTMs are assigned using orthologous groups (OGs) from the eggNOG database. eggNOG builds OGs for proteins from thousands of species and organizes them in levels of inclusive taxa. For every protein, PTMcode selects the OG of the oldest eukaryotic level in which it is included and spreads the annotation of its experimentally verified PTMs across the conserved residues in a multiple sequence alignment (MSA) of all proteins in that group (Fig. 1). This naive exercise allows us to disseminate modifications from one species to others. We evaluated that 22.7 % of the experimentally verified human phosphoserines align with a known phosphorylation site in another species (15 % is the random expectation) [1], which is surely an underestimate as HTEs have been performed on only a few other species. Thus, the PTMcode database maps over 1,30,0000 non-verified PTMs in ~130,000 proteins and provides modification patterns for the proteomes of species that have not been subjects of HTEs.

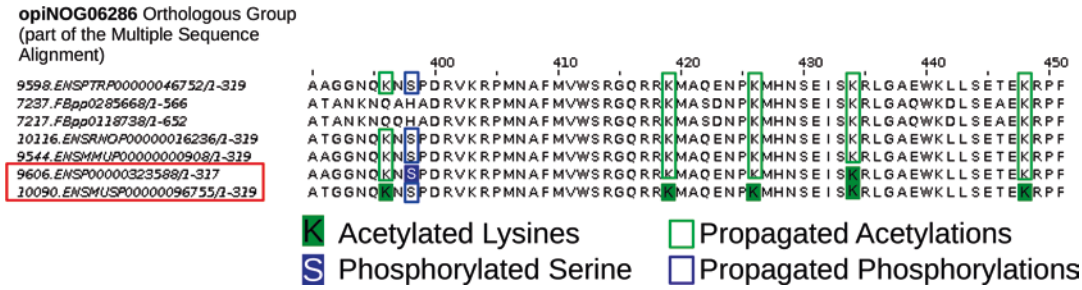

Fig. 1 Schema for PTM propagation. The SOX2 protein has two sources of PTMs coming from two HTEs performed in mouse and human (sequences highlighted with a *red rectangle*). The experimentally verified PTMs are mapped into the MSA of the OG, and the conserved amino acids in the columns with annotated PTMs are marked as "propagated PTMs"

To show the impact of this type of novel annotation, nine species had, as a result of our "PTM collection pipeline," less than 500 "real" modifications, but by including propagated PTMs, these numbers increased more than 250-fold to a level comparable with species with HTE data.

PTMcode predicts functional associations for propagated PTMs among themselves and with experimentally verified modifications. However, the propagated PTMs have to be considered potential PTMs and should be interpreted with more care than the ones found in experiments.

3.3 Channels for the Prediction of Functional Associations Between PTMs

There are many possible ways in which two particular PTMs might be functionally associated. For instance, they could be part of a molecular switch that controls protein function and/or localization [39], they could constitute a series of consecutive modifications [40], or they could contribute to the same final outcome of the protein even though they are added at different times and in different cell compartment (e.g., PTMs as a signal for protein transport).

In order to catch this wide variety of regulatory events, the PTMcode database implements five independent channels to predict the functional association between PTMs within the same protein; some of these are also applied to PTMs between interacting proteins. Below we describe these five channels in the context of PTMs within the same protein; for the association of PTMs between interacting proteins, see Subheading 3.7.

3.3.1 Coevolution Channel

The coevolution of two protein residues has been widely used as a proxy for their functional connection [41]. One of the most popular algorithms to address this concept is mutual information (MI) [42]. When applied to the MSA of a group of orthologous proteins, MI can estimate the coevolution of two residues (two columns in the MSA) by measuring the accuracy of predicting the amino acid

present in one position knowing the identity of the amino acid in the second position. To evaluate the functional association of two PTMs, PTMcode uses a slightly modified version of MI to penalize anticorrelation of residues (residues that have an opposite pattern of coevolution).

As discussed in [43], the signal coming from the MI evaluation must be compared to a background distribution of MI values to avoid spurious correlations due to phylogenetic influences (closely related species in the MSA) and small sample size (few species in the MSA). Some background distributions have been already proposed based on label randomization in the MSA [44] or on the set of MI values of all pairs of residues in the MSA [43]. PTMcode uses very strict criteria for its background distribution, using the MI values from non-modified residues in the MSA of the same type of amino acid and located in similar protein regions (ordered or disordered) as the two modified residues under evaluation. Pairs of PTMs with an MI value higher than 95 % of the background distribution are classified as coevolving. For residues lacking enough variability in the MSA column (very conserved or not conserved) to be able to compute MI values, we calculate the ratio of the conserved site in a MSA position to the total proteins and compare it to the distribution of the non-modified sites with the same limitations taken as background. Again pairs with a ratio above 95 % of the background distribution are selected as coevolving.

The coevolution channel is designed to extract a wide range of regulatory relationships as the underlying mechanisms might be very different and are not included in the definition of the algorithm.

3.3.2 Structural Distance Channel

The atomic proximity of two amino acids in the 3D structure of a protein is a widely accepted proof of the residues' functional association. Indeed protein contact maps representing matrices of all-against-all residue distances within a protein can be used to reconstruct its 3D structure [45]. The PTMcode database uses available protein 3D structures from the Protein Data Bank [46] to measure the distance between Cα-Cα atoms of all pairs of modified residues within the model. Although a threshold of 6–12 Å is usually accepted to determine contact between residues, we wanted to be more strict, so we calculated the threshold value based on known cases of PTMs that have a physical interaction. In total, we could measure 12 pairs of residues having this type of association and set as an optimal value for physical contact their average distance (4.69 Å).

The limitations of this channel are due to the availability of 3D protein structures and the mapping of the residues from sequence coordinates to positions in the structure.

3.3.3 Same Residue Channel

A very specific case of PTM cross talk is the direct competition of two types of modifications for the same protein residue [47]. There are two significant examples of this type of cross talk, the yin-yang molecular switches [48] where the same serines or threonines are modified with a phosphorylation or an O-linked glycosylation, co-regulating protein function and localization, and lysines that can be acetylated, SUMOylated, ubiquitinated, or methylated to produce different outcomes (e.g., in histone tails [49]).

PTMcode collects these events by checking the different PTM types that modify the same protein residue.

3.3.4 Manual Annotation Channel

In addition to predictions, PTMcode stores manually annotated PTM crosstalk events extracted from published papers. In these cases, a description of the interaction of both PTMs is provided.

3.3.5 PTM Hotspot Channel

From the analysis of the post-translational regulation of well-studied proteins such as the TP53 oncogene, we have learned that there are certain protein regions with an accumulation of PTMs [50] that act as regulatory centers (PTM hotspots). This concept was extended by Beltrao et al. [51] to many other eukaryotic proteins. PTMcode identifies PTM hotspots following the definition of Beltrao et al. and presents them within its complete framework of regulatory events. For each modified residues in a protein, we count the number of PTMs in a window of 31 amino acids (15 downstream and 15 upstream) and compare them using a Fisher exact test to the number of modifications in the whole protein. *P*-values are adjusted by false discovery rate and overlapping regions are collapsed.

3.4 The PTMcode Home Page

The users' entry page (Fig. 2) is divided into two panels:

1. The left panel (Fig. 2a) is dedicated to the exploration of particular combinations of PTM types. It is an option implemented for scientists that are interested in a particular type of cross talk, e.g., phosphorylations and O-linked glycosylations in yin-yang sites [52, 53] or phosphorylation linked to ubiquitination as a signal for degradation [7, 54]. From the browser wheel, users may select two types of PTMs and a table of all instances of known and predicted functional associations are displayed below. Two tabs divide the table into associations within the same protein and those regulating two interacting proteins.
2. The right panel (Fig. 2b) is designed to allow users to explore the predicted regulation of particular proteins. Again, two tabs separate the options to explore the regulation of a protein of interest or the regulation of its interactions with other proteins. The input is either a protein name or sequence (several examples are provided), and the search can be restricted to specific protein regions, residues, or PTM types of interest.

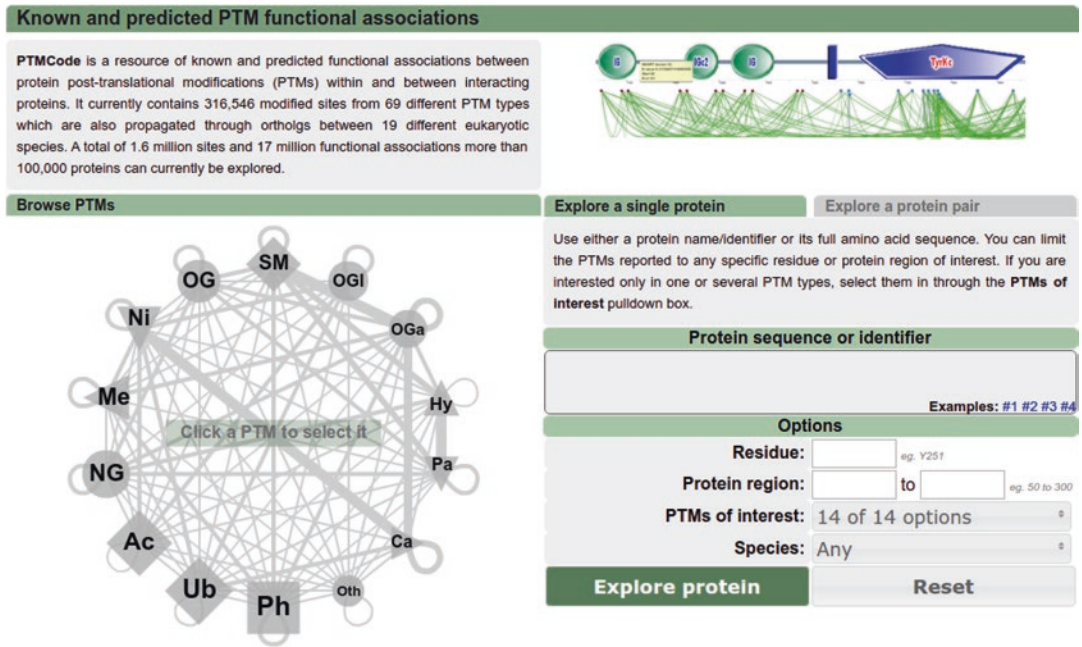

Fig. 2 The PTMcode home page

3.5 Exploring a Particular Protein

From the home page (http://ptmcode.embl.de), PTMcode provides several options to search for a protein of interest. Users may enter either the sequence (only exact matches are reported) or a protein identifier. In the case of protein IDs, PTMcode uses the protein ID dictionary from the STRING database, which has cross links between IDs from the major protein and gene resources. This dictionary does not report only synonymous IDs for the same protein but other cross links that help the users to identify the correct protein from the melting pot that represents the world of genes and protein names. If there is any source of conflict in the name provided, we redirect the user to a disambiguation page where the correct entry can be selected. If, in spite of the facilities implemented to find the desired protein, there is an unclear outcome or the protein is not found, we suggest searching for the protein ID in the Ensembl database (http://www.ensembl.org/) and, following the links to proteins and transcripts, getting either the ID or the sequence of the largest transcript associated with the input ID. Be aware that the protein sequence is our ultimate unique identification for a particular protein.

Users may also restrict the search to their favorite PTM types, known modified residues, or particular protein regions.

Once a protein is selected, PTMcode directs the user to its entry page that is divided into three panels (Fig. 3):

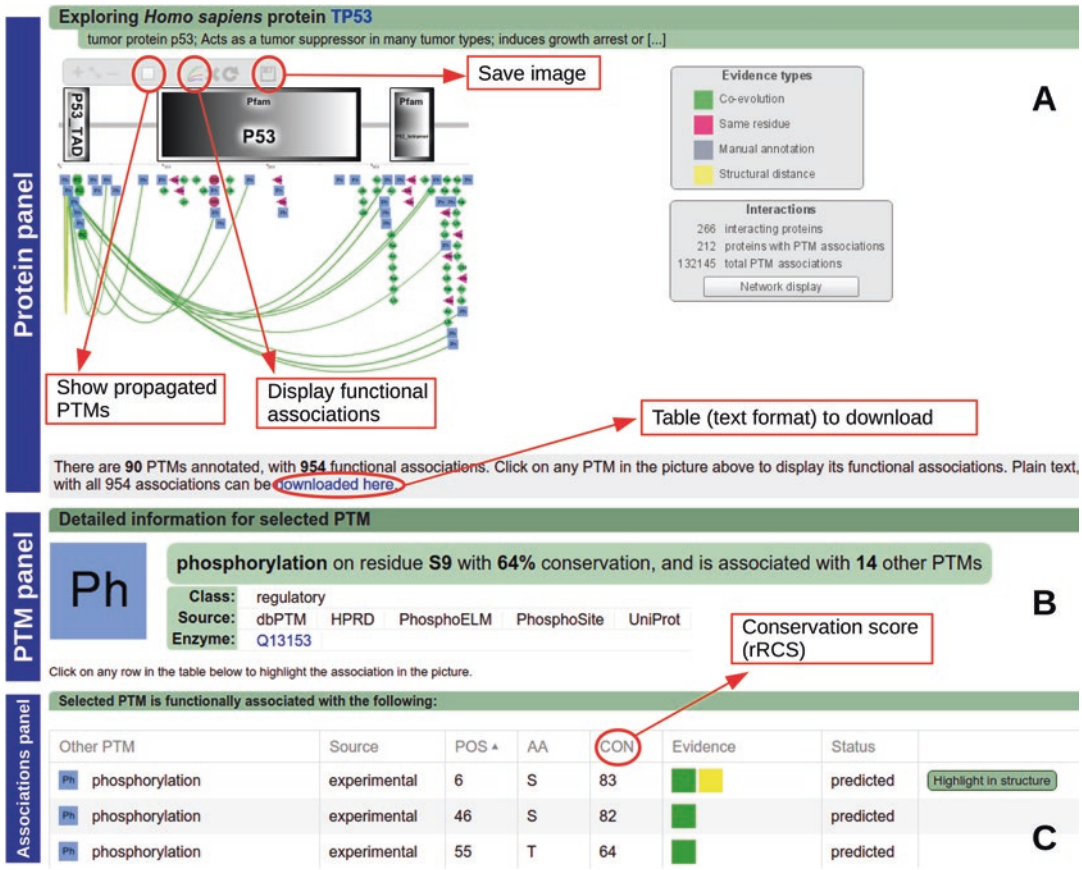

Fig. 3 PTMcode Results page. The three panels are shown for the TP53 human protein. In the "protein panel" (**a**), users may choose to show the "propagated" PTMs and display all functional associations or those related to one particular PTM or evidence type. They also can download the set of known and predicted associations either as an image or in text format. In the "PTM panel" (**b**) detailed information about the selected modification is shown. From the "associations panel" (**c**), the details of every PTM functionally associated with the selected modification are displayed including their rRCS (highlighted in *red*) and the evidence channels supporting the prediction

1. The protein panel. An interactive framework where an image of the protein is displayed (Fig. 3a) showing along the protein coordinates: (1) the protein globular domains annotated by the SMART database [55], (2) the PTMs, and (3) the hotspot regions. From this panel, users may explore the information about the domains (linked to SMART), PTMs, and their functional associations with other modifications. A zoom facility allows viewing of the details of particular protein regions and a checkbox permits inclusion of the "propagated PTMs" in the display. They are shown with a red border to distinguish them from the experimentally verified modifications. When the user clicks on their favorite PTM, the rest of the panels refresh in order to show the knowledge for that particular modification.

2. The PTM panel. Under the header "detailed information for selected PTM," a self-refreshing table appears each time a PTM is clicked in the "protein panel" (Fig. 3b). In here, some of the features of the selected PTM are displayed:
   (a) The type of modification.
   (b) A conservation score that reflects how the residue is conserved across orthologous proteins.
   (c) The number of functional associations predicted.
   (d) A general classification of the possible function of the PTM based on the PTM type: categories include "regulatory" (involved in regulation of protein function), "stabilizing" (required for conformational purposes), and "uncharacterized" (with unknown or unclear function).
   (e) The source from which we obtained the modification.
   (f) The enzyme that performs the modification (if annotated in the sources at the time of downloading the data).

   The information provided by this panel is complemented by a pop-up box that appears when the mouse is over a PTM in the "protein panel."
3. The PTM associations panel. A user-sortable table that lists the modifications that were predicted to be functionally associated with the selected PTM (Fig. 3c). The table provides (a) the modification type, (b) the source (in release 2 all are experimental), (c) the amino acid and position in the sequence, (d) the conservation score, (e) the set of evidence (channels) that supports this association, and (f) the status, either predicted or known.

The different questions that can be answered from these three panels are discussed in detail within the following subheadings.

3.6 How to Assess if a PTM Is Biologically Relevant

One of the hot topics in the field of protein PTMs is the discrimination between functional and nonfunctional modifications. This information is especially relevant for phosphorylation sites as from the early days of MS proteome-wide experiments; it has been postulated that some phosphorylations might result from promiscuous kinase activity [56]. In addition, it is possible that in some cases the conservation of phosphorylation events might be at the kinase-substrate level and not at the site level [38], especially in disordered regions. Still, phosphorylation sites have been found to be generally more conserved than both their flanking regions [57] and non-modified serines, threonines, and tyrosines in ordered and disordered regions [9]. Thus, although one cannot discard the possibility that non-conserved phosphorylations are functional, many resources use the conservation level of the site [58–60] and even of the protein [26, 61] to assess the functionality of PTMs.

The fact that phosphorylation is by far the most explored modification type also contributes to the special attention that it receives; in the future it is expected that other PTMs will be recognized as important sources of protein regulation [62], and so they will be also subject of this type of questioning. On the top of this, the conservation level of a modification site is valuable information in itself, as it reflects the evolutionary constraints of the function (if any) performed by the moiety.

The PTMcode database uses its own conservation algorithm, the *relative residue conservation score* (rRCS), to guide users in assessing the evolutionary constraints of a modification and its biological relevance. The rRCS of a PTM reports the conservation of the modified residue over orthologous proteins and is calculated as follow:

1. The protein in which the modification has been found is assigned to the oldest eukaryotic group of orthologous proteins provided by the eggNOG database [34].
2. Using the multiple alignment of the OG, the residue conservation score (RCS) is calculated for the residue. The RCS is the result of multiplying two components, the residue conservation ratio (RCR) that is the ratio of conserved sites and non-conserved sites, and the maximum branch length (MBL) of any two species containing the same residue as the PTM site from a species tree generated out of marker genes.
3. The modified residue is assigned to either an ordered or disordered region on the protein using DisEMBL [63].
4. The RCS is calculated for all residues in the OG of the same type of amino acid as the modified residue that are also in the same type of protein region (ordered or disordered). The set of scores generated here represents the background distribution used to calculate the rRCS.
5. The rRCS of the modified residue is calculated as the percentile of its RCS value in the background distribution. An rRCS >95 means that the modified residue is more conserved than the 95 % of the same amino acids within the same type of protein region.

For full details on rRCS algorithm and performance see [9]. Other people use rRCS for the same purpose [64, 65], and other resources have other types of conservation measurements based on their own algorithms [20] or on the visual inspection of orthologous protein alignments [2, 3].

Other types of data that could help to determine the biological relevance of a PTM, in the absence of specific annotation, are the number of publications where a particular PTM has been reported, the number of coevolving residues or the number of databases from which the PTM has been extracted. Indeed, in [1] we already

showed a significant positive correlation between the number of papers reporting a PTM, the normalized number of coevolving residues, and its conservation. This and other relevant information can be explored within the pop-up box that appears when the mouse is over the modification in the "protein panel." Other clues provided by PTMcode that can be used for the assessment of the biological importance of a PTM include (1) whether there is annotation concerning the enzyme(s) that perform the modification, (2) whether the modification is inside a hotspot region, or (3) whether the modification is in a globular domain as these domains are ordered regions and so are under more evolutionary constraints.

3.7 Exploring the Functional Associations Between PTMs

The known and predicted functional associations between PTMs within a particular protein can be explored in detail from the "protein panel" (Fig. 4). From the top menu of the frame, users can display all the associations for all the PTMs. A set of connecting lines will be displayed to illustrate the predicted cross talks. Propagated PTMs (if shown) can participate in functional associations among themselves and with experimentally verified PTMs. When a single PTM is clicked, the associations shown will be restricted to those involving that modification (Fig. 4a). The lines connecting PTMs are color coded according to their association type (*see* the "evidence type" menu for details (Fig. 4b)). Clicking an "evidence" square will display only the associations with the selected evidence type; a second click deactivates the selection.

3.7.1 Exploring Cross-Talk Events by Evidence Type

The evidence for cross talk between two PTMs is displayed in the "PTM associations panel." Users may click on each of the colored boxes in the evidence column, and a pop-up box will be displayed with further information, as detailed below:

1. Coevolution channel (green box). The conservation pattern of two modified amino acids within the MSA of the OG can be explored clicking on the "display in Jalview' button (Fig. 4c) in the pop-up box. In addition, we provide a list of species where the two amino acids are conserved.
2. Structural distance channel (yellow box). The pop-up box shows the two PTMs and the atomic distance between them (Fig. 4d). From the table row, the "highlight in structure" button will open a Jmol plug-in where the protein 3D structure with the two modified residues highlighted may be explored using full Jmol features (Fig. 4e).
3. Same residue channel (pink box). The pop-up box shows the two modified amino acids and a general annotation of their cross talk based on the types of PTMs involved. For instance, sites with a phosphorylation and an O-linked glycosylation reported are annotated as "competition" and sites with an O-linked glycosylation and a hydroxylation are annotated as "cooperation" [66].

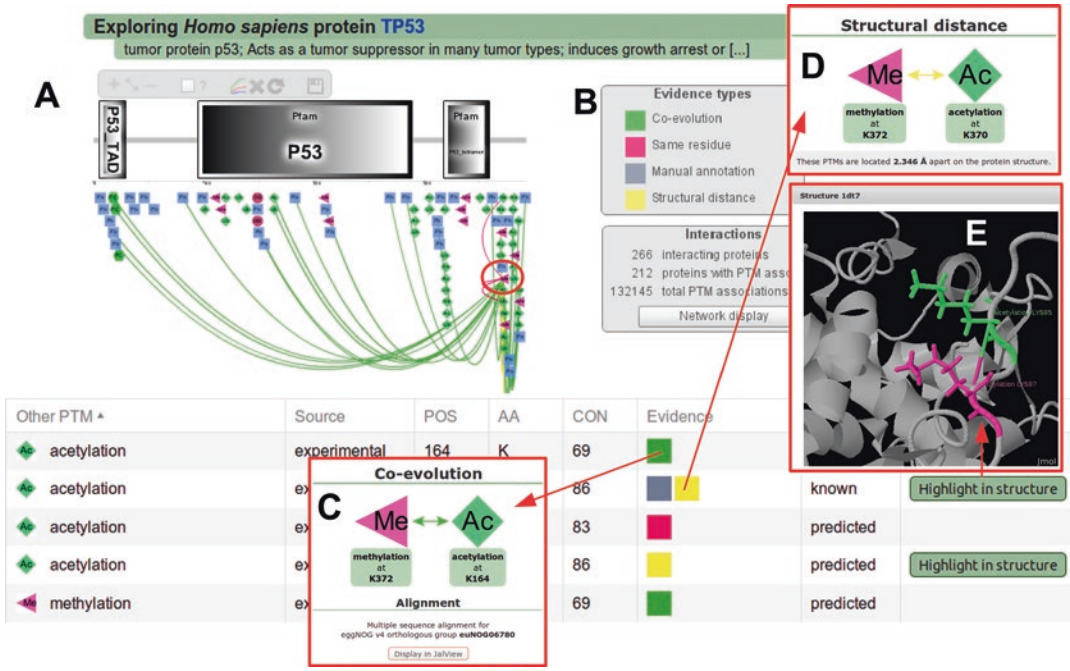

Fig. 4 Evidence channels for PTM functional associations. The human protein TP53 is methylated at position K372 (**a**). This methylation shows several functional associations with other PTMs supported by different "evidence channels" (**b**). For instance, it is found to be coevolving with the K164 residue that might be acetylated (**c**) and is found to be in contact with the K370 residue (**e**), also acetylated. This cross talk has been already reported in a scientific paper as indicated by the *gray square* (manual annotation) in the corresponding row of the table of functional associations

4. Manual annotation channel (grey box). The pop-up box shows the two modified residues, a link to the paper where the association has been reported and a single-sentence annotation that summarizes the effect of the cross talk.

3.7.2 Exploring PTM Hotspots

Inside the "protein panel," hotspots—regions with a significant concentration of PTMs—are indicated by a blue line (Fig. 5). Clicking on the blue line will highlight the region and display the included PTMs. A click on any of the PTMs will highlight the corresponding entry in the "protein panel." Hotspots that consist of "propagated" PTMs are only displayed if the "display propagated PTMs" checkbox is activated. To view an example of a PTM hotspot, users may check human cyclin-dependent kinase 12 (CRKRS). Interestingly, the homologous mouse protein has a hotspot in almost the same region, while in the rat protein, this region only appears as a hotspot if the "propagated" PTMs are displayed (Fig. 5). This observation highlights the need for annotation of "propagated" PTMs in species with no HTEs and supports the methodology that we use to calculate them.

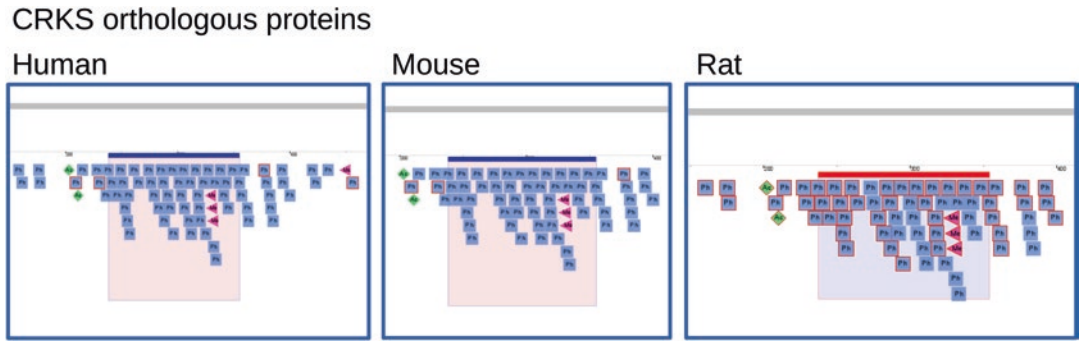

Fig. 5 PTM Hotspot representation. The hotspots found in three orthologous proteins are shown. Hotspots supported by experimentally verified PTMs are shown in *blue*; hotspots supported by "propagated" PTMs are shown in *red*

3.7.3 Export of Functional Association Data

The "protein panel" display can be downloaded using the top menu in png and jpg formats with a range of resolutions. The functional associations in text format can also be downloaded following the link below the "protein panel."

3.8 Functional Associations of PTMs in Interacting Proteins

As discussed earlier in this Chapter, PTMs and PPIs constitute an intricate layer of protein regulation that only recently has been connected using large-scale approaches. Several computational studies have shown enrichments in PTM clusters in protein complexes [18], a higher number of interaction partners in modified proteins compared to non-modified [67] and a higher degree of coevolution between PTMs located in interacting proteins compared to non-interacting [1]. On top of these simple associations, the "PTMcode" also plays a role in the regulation of protein interactions. For example, proteins with particular coevolving PTM types form bigger protein-protein networks than proteins with same type of PTMs that are not coevolving [9].

In order to predict functional associations between PTMs in interacting proteins, some of the channels described in Subheading 3.3 were adapted to address the particularities of this task. For the "coevolution channel," the two interacting proteins are mapped to their respective OGs, and their MSAs are pruned to keep only proteins of species in common to both. Then, the MI algorithm described in Subheading 3.3.1 is applied. For the "structural distance channel," we measure the distance of all pairs of modified residues in protein interfaces mapped in the structure of the protein complex (if available). The "same residue channel" and "hotspot" evidence sources are not applicable here, and manual annotation was not performed for this type of association. Be aware that we apply these predictions to all the possible pairs of PTMs between the two proteins, not only those located in the binding interface, so especially for the coevolution channel, the associations may encompass a wide variety of mechanisms.

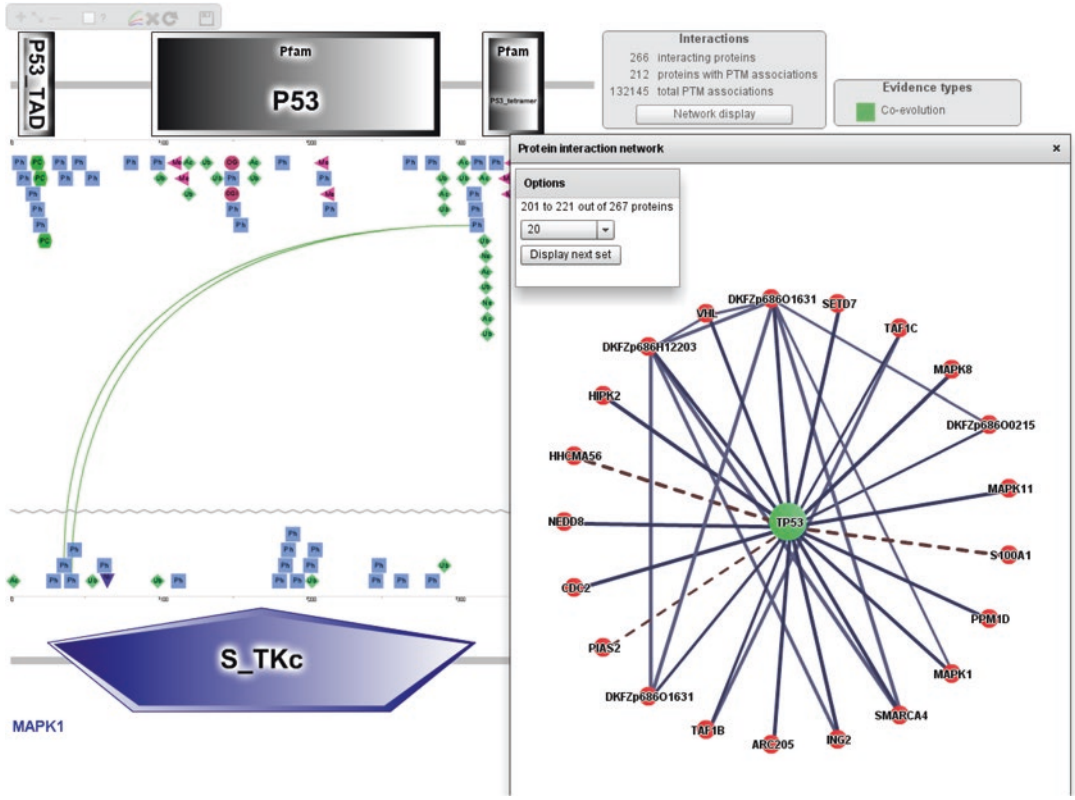

Fig. 6 Functional associations of PTMs in interacting proteins. The human protein TP53 has many reported interactions. The interaction with MAPK1 is shown within the "protein panel"; from here predicted functional associations between their PTMs on MAPK1 and TP53 can be explored in detail

The functional associations of PTMs in interacting proteins can be explored from two entry points:

1. From the home page (*see* Subheading 3.4), the tab "explore a protein pair" allows the user to display the list of interacting partners of his/her favorite protein. From that list, a particular interacting protein can be selected, and the two proteins are displayed in parallel within the "protein panel."
2. If a single protein is being explored, as described in previous Subheadings, the "interactions menu" provides information about PPIs (*see* our definition of PPI in Subheading 3.1) and their predicted functional associations. The "network display" button will open a new panel with a representation of the PPI network (Fig. 6). From here, any PPI with a continuous edge can be explored (dashed edges represent PPIs with no PTMs functionally linked).

Once a PPI is selected, the "protein panel" shows now the two proteins (Fig. 6), their PTMs, and their functional associations. Clicking on any of them displays evidence and features as described for single proteins in previous Subheadings.

4 Notes

1. For a general review on specific cross talk between different PTM types, we suggest reading the supplementary material in [9].
2. To download the whole dataset of PTMs and functional associations, users may visit the "data" tab at the top of any PTMcode page. A direct download link is provided for associations regulating proteins and PPIs.

References

1. Minguez P, Letunic I, Parca L et al (2015) PTMcode v2: a resource for functional associations of post-translational modifications within and between proteins. Nucleic Acids Res 43:D494–D502. doi:10.1093/nar/gku1081
2. Gnad F, Gunawardena J, Mann M (2010) PHOSIDA 2011: the posttranslational modification database. Nucleic Acids Res 39:D253–D260. doi:10.1093/nar/gkq1159
3. Lu C-T, Huang K-Y, Su M-G et al (2013) DbPTM 3.0: an informative resource for investigating substrate site specificity and functional association of protein post-translational modifications. Nucleic Acids Res 41:D295–D305. doi:10.1093/nar/gks1229
4. Sadowski I, Breitkreutz B-J, Stark C et al. (2013) The PhosphoGRID *Saccharomyces cerevisiae* protein phosphorylation site database: version 2.0 update. Database (Oxford) 2013: bat026. doi:10.1093/database/bat026
5. Naegle KM, Gymrek M, Joughin BA et al (2010) PTMScout, a web resource for analysis of high throughput post-translational proteomics studies. Mol Cell Proteomics 9:2558–2570. doi:10.1074/mcp.M110.001206
6. Craveur P, Rebehmed J, de Brevern AG (2014) PTM-SD: a database of structurally resolved and annotated posttranslational modifications in proteins. Database (Oxford) 2014:bau041. doi:10.1093/database/bau041
7. Hunter T (2007) The age of crosstalk: phosphorylation, ubiquitination, and beyond. Mol Cell 28:730–738. doi:10.1016/j.molcel.2007.11.019
8. Beltrao P, Trinidad JC, Fiedler D et al (2009) Evolution of phosphoregulation: comparison of phosphorylation patterns across yeast species. PLoS Biol 7:e1000134. doi:10.1371/journal.pbio.1000134
9. Minguez P, Parca L, Diella F et al (2012) Deciphering a global network of functionally associated post-translational modifications. Mol Syst Biol 8:599. doi:10.1038/msb.2012.31
10. The UniProt Consortium (2014) UniProt: a hub for protein information. Nucleic Acids Res 43:D204–D212. doi:10.1093/nar/gku989
11. Hornbeck PV, Zhang B, Murray B et al (2015) PhosphoSitePlus, 2014: mutations, PTMs and recalibrations. Nucleic Acids Res 43:D512–D520. doi:10.1093/nar/gku1267
12. Filtz TM, Vogel WK, Leid M (2014) Regulation of transcription factor activity by interconnected post-translational modifications. Trends Pharmacol Sci 35:76–85. doi:10.1016/j.tips.2013.11.005
13. Sun B, Zhang M, Cui P et al (2015) Nonsynonymous single-nucleotide variations on some posttranslational modifications of human proteins and the association with diseases. Comput Math Methods Med 2015:124630. doi:10.1155/2015/124630
14. Duan G, Walther D (2015) The roles of post-translational modifications in the context of protein interaction networks. PLoS Comput Biol 11:e1004049. doi:10.1371/journal.pcbi.1004049
15. Park CY, Krishnan A, Zhu Q et al (2014) Tissue-aware data integration approach for the inference of pathway interactions in metazoan organisms. Bioinformatics 31:1093–1101. doi:10.1093/bioinformatics/btu786
16. von Appen A, Kosinski J, Sparks L et al (2015) In situ structural analysis of the human nuclear pore complex. Nature 526:140–143. doi:10.1038/nature15381
17. Huang Y, Xu B, Zhou X et al (2015) Systematic characterization and prediction of post-translational modification cross-talk. Mol Cell Proteomics 14:761–770. doi:10.1074/mcp.M114.037994
18. Woodsmith J, Kamburov A, Stelzl U (2013) Dual coordination of post translational modifications in human protein networks.

PLoS Comput Biol 9:e1002933. doi:10.1371/journal.pcbi.1002933

19. Creixell P, Linding R (2012) Cells, shared memory and breaking the PTM code. Mol Syst Biol 8:598
20. Dinkel H, Chica C, Via A et al (2011) Phospho.ELM: a database of phosphorylation sites—update 2011. Nucleic Acids Res 39:D261–D267. doi:10.1093/nar/gkq1104
21. Orchard S, Ammari M, Aranda B et al (2014) The MIntAct project—IntAct as a common curation platform for 11 molecular interaction databases. Nucleic Acids Res 42:D358–D363. doi:10.1093/nar/gkt1115
22. Keshava Prasad TS, Goel R, Kandasamy K et al (2009) Human protein reference database—2009 update. Nucleic Acids Res 37:D767–D772. doi:10.1093/nar/gkn892
23. Szklarczyk D, Franceschini A, Wyder S et al (2015) STRING v10: protein-protein interaction networks, integrated over the tree of life. Nucleic Acids Res 43:D447–D452. doi:10.1093/nar/gku1003
24. The UniProt Consortium (2014) Activities at the Universal Protein Resource (UniProt). Nucleic Acids Res 42:D191–D198. doi:10.1093/nar/gkt1140
25. Danielsen JMR, Sylvestersen KB, Bekker-Jensen S et al (2011) Mass spectrometric analysis of lysine ubiquitylation reveals promiscuity at site level. Mol Cell Proteomics 10:M110.003590. doi:10.1074/mcp.M110.003590
26. Choudhary C, Kumar C, Gnad F et al (2009) Lysine acetylation targets protein complexes and co-regulates major cellular functions. Science 325:834–840. doi:10.1126/science.1175371
27. Henriksen P, Wagner SA, Weinert BT et al (2012) Proteome-wide analysis of lysine acetylation suggests its broad regulatory scope in *Saccharomyces cerevisiae*. Mol Cell Proteomics 11:1510–1522. doi:10.1074/mcp.M112.017251
28. Lundby A, Secher A, Lage K et al (2012) Quantitative maps of protein phosphorylation sites across 14 different rat organs and tissues. Nat Commun 3:876. doi:10.1038/ncomms1871
29. Matic I, Schimmel J, Hendriks IA et al (2010) Site-specific identification of SUMO-2 targets in cells reveals an inverted SUMOylation motif and a hydrophobic cluster SUMOylation motif. Mol Cell 39:641–652. doi:10.1016/j.molcel.2010.07.026
30. Murray CI, Kane LA, Uhrigshardt H et al (2011) Site-mapping of in vitro S-nitrosation in cardiac mitochondria: implications for cardioprotection. Mol Cell Proteomics 10:M110.004721. doi:10.1074/mcp.M110.004721
31. Weinert BT, Wagner SA, Horn H et al (2011) Proteome-wide mapping of the Drosophila acetylome demonstrates a high degree of conservation of lysine acetylation. Sci Signal 4:ra48. doi:10.1126/scisignal.2001902
32. Zielinska DF, Gnad F, Schropp K et al (2012) Mapping N-glycosylation sites across seven evolutionarily distant species reveals a divergent substrate proteome despite a common core machinery. Mol Cell 46:542–548. doi:10.1016/j.molcel.2012.04.031
33. Wagner SA, Beli P, Weinert BT et al (2011) A proteome-wide, quantitative survey of in vivo ubiquitylation sites reveals widespread regulatory roles. Mol Cell Proteomics 10:M111.013284. doi:10.1074/mcp.M111.013284
34. Powell S, Forslund K, Szklarczyk D et al (2014) eggNOG v4.0: nested orthology inference across 3686 organisms. Nucleic Acids Res 42:D231–D239. doi:10.1093/nar/gkt1253
35. Franceschini A, Szklarczyk D, Frankild S et al (2013) STRING v9.1: protein-protein interaction networks, with increased coverage and integration. Nucleic Acids Res 41:D808–D815. doi:10.1093/nar/gks1094
36. Boekhorst J, van Breukelen B, Heck A, Snel B (2008) Comparative phosphoproteomics reveals evolutionary and functional conservation of phosphorylation across eukaryotes. Genome Biol 9:R144. doi:10.1186/gb-2008-9-10-r144
37. Chen SC-C, Chen F-C, Li W-H (2010) Phosphorylated and nonphosphorylated serine and threonine residues evolve at different rates in mammals. Mol Biol Evol 27:2548–2554. doi:10.1093/molbev/msq142
38. Tan CSH, Bodenmiller B, Pasculescu A et al (2009) Comparative analysis reveals conserved protein phosphorylation networks implicated in multiple diseases. Sci Signal 2:ra39. doi:10.1126/scisignal.2000316
39. Humphrey SJ, James DE, Mann M (2015) Protein phosphorylation: a major switch mechanism for metabolic regulation. Trends Endocrinol Metab 26:676–687. doi:10.1016/j.tem.2015.09.013
40. Byeon I-JL, Li H, Song H et al (2005) Sequential phosphorylation and multisite interactions characterize specific target recognition by the FHA domain of Ki67. Nat Struct Mol Biol 12:987–993. doi:10.1038/nsmb1008
41. de Juan D, Pazos F, Valencia A (2013) Emerging methods in protein co-evolution. Nat Rev Genet 14:249–261. doi:10.1038/nrg3414

42. Cover TM, Thomas JA (1991) Elements of information theory. John Wiley & Sons, New York
43. Martin LC, Gloor GB, Dunn SD, Wahl LM (2005) Using information theory to search for co-evolving residues in proteins. Bioinformatics 21:4116–4124. doi:10.1093/bioinformatics/bti671
44. Skerker JM, Perchuk BS, Siryaporn A et al (2008) Rewiring the specificity of two-component signal transduction systems. Cell 133:1043–1054. doi:10.1016/j.cell.2008.04.040
45. Pietal MJ, Bujnicki JM, Kozlowski LP (2015) GDFuzz3D: a method for protein 3D structure reconstruction from contact maps, based on a non-Euclidean distance function. Bioinformatics 31:3499–3505. doi:10.1093/bioinformatics/btv390
46. Berman HM, Kleywegt GJ, Nakamura H, Markley JL (2012) The future of the protein data bank. Biopolymers. doi:10.1002/bip.22132
47. Seet BT, Dikic I, Zhou M-M, Pawson T (2006) Reading protein modifications with interaction domains. Nat Rev Mol Cell Biol 7:473–483. doi:10.1038/nrm1960
48. Hart GW, Greis KD, Dong LY et al (1995) O-linked N-acetylglucosamine: the "yin-yang" of Ser/Thr phosphorylation? Nuclear and cytoplasmic glycosylation. Adv Exp Med Biol 376:115–123
49. Latham JA, Dent SYR (2007) Cross-regulation of histone modifications. Nat Struct Mol Biol 14:1017–1024. doi:10.1038/nsmb1307
50. Brooks CL, Gu W (2003) Ubiquitination, phosphorylation and acetylation: the molecular basis for p53 regulation. Curr Opin Cell Biol 15:164–171. doi:10.1016/S0955-0674(03)00003-6
51. Beltrao P, Albanèse V, Kenner LR et al (2012) Systematic functional prioritization of protein posttranslational modifications. Cell 150:413–425. doi:10.1016/j.cell.2012.05.036
52. Zeidan Q, Hart GW (2010) The intersections between O-GlcNAcylation and phosphorylation: implications for multiple signaling pathways. J Cell Sci 123:13–22. doi:10.1242/jcs.053678
53. Butt AM, Khan IB, Hussain M et al (2011) Role of post translational modifications and novel crosstalk between phosphorylation and O-beta-GlcNAc modifications in human claudin-1, −3 and −4. Mol Biol Rep 39:1359–1369. doi:10.1007/s11033-011-0870-7
54. Vodermaier HC (2004) APC/C and SCF: controlling each other and the cell cycle. Curr Biol 14:R787–R796. doi:10.1016/j.cub.2004.09.020
55. Letunic I, Doerks T, Bork P (2015) SMART: recent updates, new developments and status in 2015. Nucleic Acids Res 43:D257–D260. doi:10.1093/nar/gku949
56. Lienhard GE (2008) Non-functional phosphorylations? Trends Biochem Sci 33:351–352. doi:10.1016/j.tibs.2008.05.004
57. Wang Z, Ding G, Geistlinger L et al (2011) Evolution of protein phosphorylation for distinct functional modules in vertebrate genomes. Mol Biol Evol 28:1131–1140. doi:10.1093/molbev/msq268
58. Gnad F, Ren S, Cox J et al (2007) PHOSIDA (phosphorylation site database): management, structural and evolutionary investigation, and prediction of phosphosites. Genome Biol 8:R250. doi:10.1186/gb-2007-8-11-r250
59. Holt LJ, Tuch BB, Villén J et al (2009) Global analysis of Cdk1 substrate phosphorylation sites provides insights into evolution. Science 325:1682–1686. doi:10.1126/science.1172867
60. Tan CSH, Bader GD (2012) Phosphorylation sites of higher stoichiometry are more conserved. Nat Methods 9:317. doi:10.1038/nmeth.1941
61. Zielinska DF, Gnad F, Wiśniewski JR, Mann M (2010) Precision mapping of an in vivo N-glycoproteome reveals rigid topological and sequence constraints. Cell 141:897–907. doi:10.1016/j.cell.2010.04.012
62. Martínez-Ruiz A, Lamas S (2004) S-nitrosylation: a potential new paradigm in signal transduction. Cardiovasc Res 62:43–52. doi:10.1016/j.cardiores.2004.01.013
63. Linding R, Jensen LJ, Diella F et al (2003) Protein disorder prediction: implications for structural proteomics. Structure 11:1453–1459
64. Ullah S, Lin S, Xu Y et al (2016) dbPAF: an integrative database of protein phosphorylation in animals and fungi. Sci Rep 6:23534. doi:10.1038/srep23534
65. Pan Z, Liu Z, Cheng H et al (2014) Systematic analysis of the in situ crosstalk of tyrosine modifications reveals no additional natural selection on multiply modified residues. Sci Rep 4:7331. doi:10.1038/srep07331
66. Wang ZA, Singh D, van der Wel H, West CM (2011) Prolyl hydroxylation- and glycosylation-dependent functions of Skp1 in O2-regulated development of Dictyostelium. Dev Biol 349:283–295. doi:10.1016/j.ydbio.2010.10.013
67. Yachie N, Saito R, Sugiyama N et al (2011) Integrative features of the yeast phosphoproteome and protein–protein interaction map. PLoS Comput Biol 7:e1001064. doi:10.1371/journal.pcbi.1001064