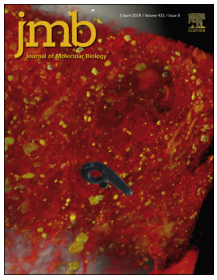
Review



Visualization and Analysis of Epiproteome Dynamics☆

Sandeep Kaur[1,2], Benedetta Baldi[2,3], Jenny Vuong[2,3] and Seán I. O'Donoghue[1,2,3]

*1 - **University of New South Wales (UNSW)**, Kensington, NSW 2052, Australia*
*2 - **Garvan Institute of Medical Research**, Darlinghurst, NSW 2010, Australia*
*3 - **Data 61, CSIRO**, Eveleigh, NSW 2015, Australia*

***Correspondence to Seán I. O'Donoghue:** Garvan Institute of Medical Research, Darlinghurst, NSW 2010, Australia. sandeep.kaur@unsw.edu.au, b.baldi@garvan.org.au, jenny.vuong@data61.csiro.au, sean@odonoghuelab.org.*
https://doi.org/10.1016/j.jmb.2019.01.044
Edited by Jodie Jenkinson

Abstract

The epiproteome describes the set of all post-translational modifications (PTMs) made to the proteins comprising a cell or organism. The extent of the epiproteome is still largely unknown; however, advances in experimental techniques are beginning to produce a deluge of data, tracking dynamic changes to the epiproteome in response to cellular stimuli. These data have potential to revolutionize our understanding of biology and disease. This review covers a range of recent visualization methods and tools developed specifically for dynamic epiproteome data sets. These methods have been designed primarily for data sets on phosphorylation, as this the most studied PTM; however, most of these methods are also applicable to other types of PTMs. Unfortunately, the currently available methods are often inadequate for existing data sets; thus, realizing the potential buried in epiproteome data sets will require new, tailored bioinformatics methods that will help researchers analyze, visualize, and interactively explore these complex data sets.

☆ This article is part of the special issue on "Biovisualization" edited by Donald E Ingber, Janet H Iwasa, Jodie Jenkinson, Gaël McGill, Seán O'Donoghue, Charles Reilly, Alain Viel.

Introduction

The "epiproteome," as envisioned by Margueron *et al*. [1], describes the set of all post-translational modifications (PTMs) made to the proteins comprising a cell or organism. Almost all proteins are believed to contain PTMs [2], typically resulting in small but significant changes in protein structure [3] as well as functional changes [4], such as gain or loss of enzymatic activity, alteration of protein interactions, and changes in subcellular localization. These changes play key roles in cellular regulation and greatly contribute to the complexity of living organisms [5].

Currently, over 200 different types of PTMs are characterized in UniProt [6].† Furthermore, some of these different types of PTMs can work in conjunction with each other [7,8], thus further increasing the combinatorial complexity with which PTMs regulate cellular systems. Some types of PTMs are irreversible (e.g., deamidation) and are often associated with processes such as aging and tissue injury responses [9,10]. In contrast, reversible PTMs (e.g., phosphorylation) are often implicated in fast, tightly regulated processes [11] (e.g., cell signaling pathways). Dysregulation and dysfunction of PTMs are associated with numerous diseases [12–17].

The most commonly occurring PTMs are phosphorylation, acetylation, glycosylation, methylation, and ubiquitination [18], of which phosphorylation is the most studied [19] (see the Phosphoproteomics section below for more details). Lysine acetylation and deacetylation were first discovered in histones [20], but since then, they have been reported in many other proteins—such as transcription factors, nuclear receptors, cytoskeletal proteins, and chaperones [21]—and are now also known to play

multiple roles in cellular regulation [22,23]. Glycosylation plays essential roles in protein folding, especially for membrane-bound and secreted proteins, and in immune regulation [24]. Methylation and demethylation are common on histone proteins, as well as on proteins regulating RNA metabolism, DNA repair, cell cycle progression, apoptosis, and signal transduction [25]. Polyubiquitination marks a protein for degradation, while monoubiquitination can regulate cellular processes, such as receptor transport, DNA repair, and viral budding [26–29].

While the overall importance of PTMs is clear [30], much is still unknown. In fact, much of our current understanding of biochemistry and molecular biology has been derived from experimental methods insensitive to PTMs—primarily due to difficulties in tracking such relatively small chemical changes. However, this is changing rapidly, driven by recent technical advances—making epiproteomics an emerging frontier in science [31].

Currently, mass spectrometry (MS) is the leading method for quantitative studies of the epiproteome [32]. For example, MS methods have recently been used for omics-scale studies of ubiquitylation [33] and SUMOylation [34] (i.e., the addition of "Small Ubiquitin-like MOdifier" proteins, observed to be upregulated in various cancers [35]). However, most epiproteome studies to date have focused on the phosphoproteome, mapping dynamic changes in response to specific cellular stimuli [36–41].

These emerging epiproteome data sets present several challenges when visualized using tools and methods developed for conventional "whole-protein" proteomics data [42–44]. First, the visual representation of each protein needs to be expanded to account for the fact that many proteins contain multiple PTM sites [45,46]—with some individual proteins having as many as 200 distinct PTM sites [47]. Second, the dynamic epiproteome can be surprisingly large and complex. For example, tens of thousands of distinct PTM sites and events can be involved in the cellular response to a single stimulus, such as the insulin hormone [47]. Finally, modern MS methods now allow tracking of events across multiple time points, as well as across multiple experimental conditions, thus adding dimensions of complexity to the analysis and visualization. Overlaying such data (temporal and concentrations) onto visualizations of biological pathways and networks has long been a vexing challenge that is still largely unmet.

For these reasons, the visualization tools and strategies developed for traditional proteomics analysis have, in many cases, proven to be inadequate to address dynamic epiproteome data sets. This calls for innovative, new visualization strategies to help scientists effectively explore these large and multidimensional data sets, and to communicate their findings [48,49]. This review covers a range of recent visualization methods and tools developed specifically for dynamic epiproteome data sets. As most of the currently available high-throughput data sets focus on phosphorylation, we will use this PTM as the primary example for illustration. However, many of the visualization methods and tools discussed can also be used with other types of PTM.

In the sub-section below (Phosphoproteomics section), we provide details specific to dynamic phosphoproteome data sets. In subsequent sections, we then discuss how the experimentally derived data can be visually explored (Visualizing Experimental Data section), and how this exploration can be enhanced by integration with prior knowledge (Visualizing Integrated Data section) and with predictive models (Visualizing Predictive Models section).

Phosphoproteomics

Phosphorylation (carried out by kinases) and dephosphorylation (carried out by phosphatases) often occur across large numbers of proteins during cellular signaling [47,50]. Each protein involved typically has multiple phosphosites that in turn regulate and modulate the protein's function in different ways [51,52]—for example, by inducing conformational changes that can expose or obstruct binding sites, affect protein folding and stability, or result in subcellular translocation [53,54]. Phosphorylation has been reported most often in serine residues (~ 87%), followed by threonine (~12%) and tyrosine (~2%) [55]. Other amino acids such as histidine and aspartate can also be phosphorylated; however, they are less likely to be detected using routine phosphoamino acid analysis [56,57].

A range of MS techniques have been described recently that can track changes in phosphorylation across a cellular proteome [36], generating data sets that contain three crucial pieces of information [58]:

1. The individual phosphorylation site (often with the surrounding peptide sequence)
2. The proteins on which these sites reside
3. The phosphosite ratios or fold changes ($\log_2$ratio) between the basal and the measured time points

Unfortunately, these MS techniques do not directly reveal which kinases and phosphatases have caused the observed phosphorylation events. For this reason, these studies are often augmented with knock-down or knock-out experiments targeting putative kinases and phosphatases. Such MS-quantified data sets are frequently provided to the community as supplementary data sets, and raw data may additionally be deposited in public repositories, such as the PRIDE Archive[‡] [59].

Until fairly recently, few of the available tools for analyzing proteomics data were readily applicable to phosphoproteome data or, more generally, to

epiproteome data. However, since around 2013, as the number of such data sets has increased, a range of visualization strategies have been developed, tailored for epiproteome data sets. These developments were partly aided by the 2013 DREAM 8 (Dialogue for Reverse Engineering Assessments and Methods) competition, which featured a visualization sub-challenge focused on time-series phosphoproteome data [60].

The documentation for each of these tools typically provides example visualizations with data sets selected to clearly demonstrate the strengths of each method; in this work, to help compare these tools, we have defined a benchmark phosphoproteome data set, derived from an experiment by Humphrey *et al*. [47] on insulin response within mouse adipose-like 3T3-L1 cells. In this experiment, Humphrey *et al*. were able to detect >37,000 distinct phosphorylation events involving nearly 6,000 distinct proteins, measured across nine time points (including basal). From these, we defined our benchmark data set by selecting a subset of 103 phosphorylation site profiles that were well understood, based on previous literature [61]. This subset —hereafter referred to as the "*Benchmark* data set"—was then used with each of the tools considered in this review to generate the figures presented below. In each case, the figures were modified using Adobe Illustrator and Adobe Photoshop to improve clarity for publication. One of the reviewed tools (DynaPho [62]) allowed only human data; in this case, we converted the mouse proteins and amino acid positions to the corresponding equivalents in human proteins. This was done by running a BLAST search [63] of each mouse protein against all human proteins in UniProt [6], then selecting the best matching human sequence. All data and input files used to generate the figures are provided in a supplemental Data in Brief article that accompanies this review [64]. A summary of the various tools reviewed in this paper is provided in Table 1.

Visualizing experimental data

Visual exploration of experimental epiproteome data is often a critical first step, helping to identify anomalies and to assess data validity. In addition, a visual overview of these data sets can help detect patterns that provide insight into the cascade of signaling events occurring in the experiment. In the sections below, we discuss two available tools for visualizing experimental data directly. These tools primarily use two visualization methods: profile plots (also known as parallel coordinate plots) and heat maps [65].

DynaPho (Data Summary)

DynaPho[§] [62] is a comprehensive web-based tool that supports a range of multiple analysis and visualization modules, tailored for time-course phosphoproteomics data sets. The "Data Summary" module of DynaPho allows direct visualization for selected phosphosites of interest (Fig. 1).

Methods

In the "Data Summary" module (Fig. 1), single or multiple time profiles can be selected from a spreadsheet view of a complete data set. Selected profiles are then shown on a single profile plot.

Features

This combined use of a spreadsheet and profile plot provides a simple and intuitive technique for exploring a data set and for revealing changes in the pattern of phosphorylation across specific sites.

Limitations

The straightforward approach used in this module can scale to a large number of profiles—however,

Table 1. An overview of tools and methods used to perform analysis and create visualizations in this review

Name	Platform	Analysis methods	Input	Interactive	Free	Image export
DynaPho	Web	Data summary, clustering, kinase and phosphatase assignments, protein interaction assignments	Quantified ratios or $\log_2$ of ratios of phosphorylation change at various time points	Yes	Yes	Yes
DiBS	Web	Data summary	Quantified ratios or $\log_2$ of ratios of phosphorylation change at various time points (1 MB file upload limit)	Yes	Yes, for one project	No (only for paid version)
PHOXTRACK	Web	Kinase enrichment	Phosphosite sequences (13aa max) and phosphorylation ratio (one time point)	Yes	Yes	Yes
PhosphoPath	Cytoscape plugin (desktop only)	Kinase assignments	Separate input for: proteins and their sites, and time series data	Yes	Yes	Yes
Kappa	Web and desktop	Modeling and simulation framework	Mathematical model in the Kappa-language format	Yes	Yes	Yes
Minardo	–	Kinase and phosphatase assignments	–	–	–	–

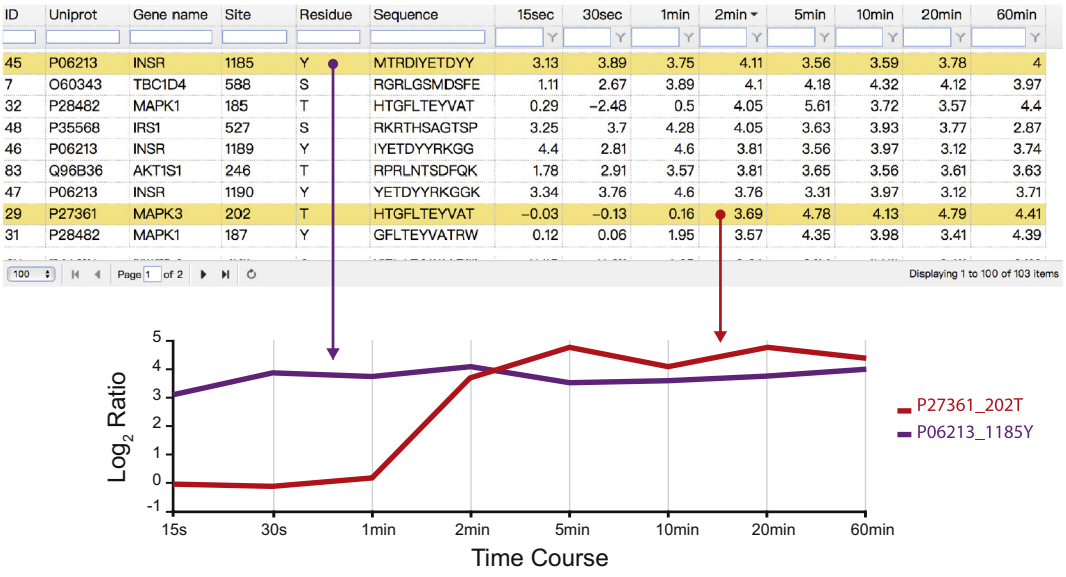


Fig. 1. DynaPho's "Data Summary" module showing part of the Benchmark data set. DynaPho provides an interactive spreadsheet (top), from which individual profiles can be selected by mouse clicking; these are then displayed in a profile plot (below). Top and bottom panels were created in DynaPho and modified in Illustrator.

only when these profiles are very similar. For profiles that are highly divergent, the plots rapidly become visually cluttered as the number of profiles increases, thus impeding the detection of trends. A second limitation is that, currently, DynaPho's advanced features are only available when using data on human proteins. Finally, the "Data Summary" module has relatively limited functionality and provides rather generic plotting facilities, similar to Excel or Google Sheets; a much richer set of phosphoproteomic-specific features are available in other DynaPho modules, covered below.

DynaPho (Profile Clustering)

The "Profile Clustering" module of DynaPho [62] overcomes the limitations mentioned above, enabling very similar time profiles to be identified and grouped together (Fig. 2).

Methods

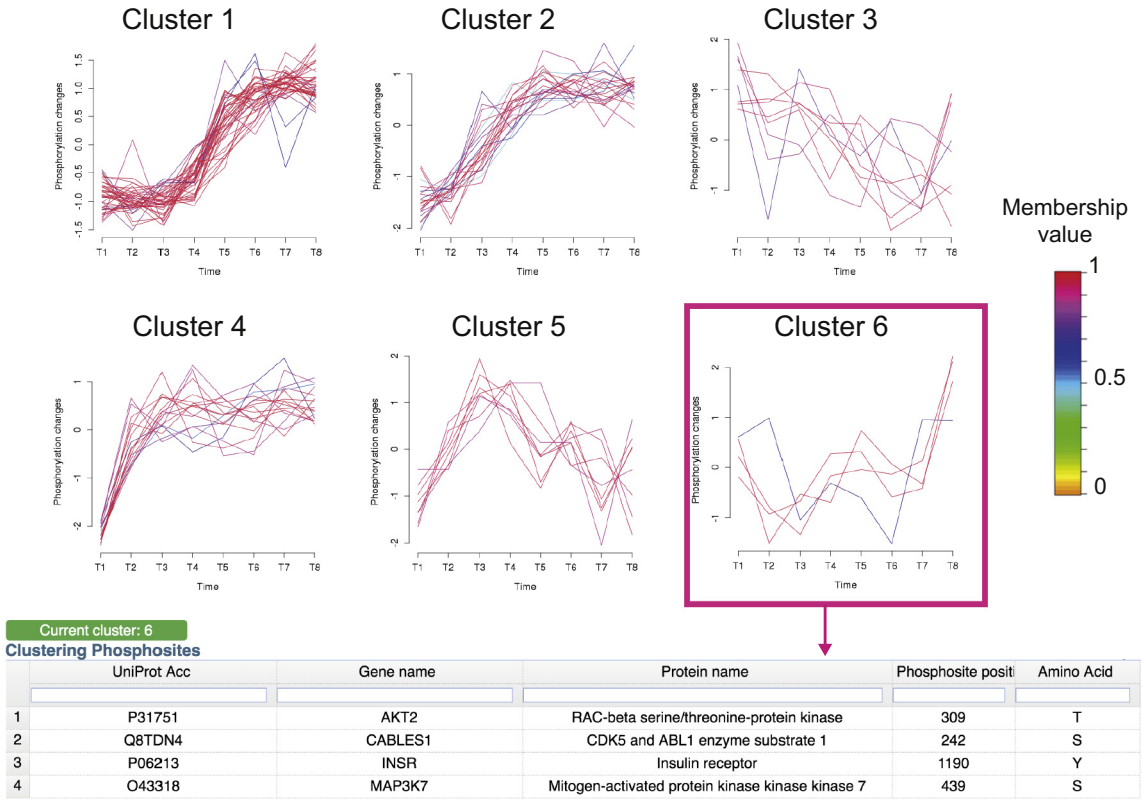
In the "Profile Clustering" module (Fig. 2), quantified ratios for phosphorylation changes are normalized, standardized, then clustered using the fuzzy c-means (FCM) algorithm [66]. This algorithm—also implemented in the Mfuzz R package [66]—has often been used for phosphoproteomics data sets [67,68]. FCM is a "soft clustering" technique, in which each individual profile is assigned a membership score (between 0 and 1, low to high, respectively) for each cluster, indicating the certainty of a profile's assignment to that cluster. Profiles in the same cluster are generally hypothesized to be regulated by the same kinases and phosphatases [69,70], and thus to have associated functions [71]. For epiproteome data, fuzzy clustering methods are considered more appropriate than "hard clustering" methods, in which an individual profile can only be assigned to one cluster. This is because (1) individual phosphorylation sites can be regulated by multiple enzymes, and (2) most epiproteome data sets to date have high variability arising from experimental uncertainties, noise, and intrinsic variations in underlying biological processes.

Features

Once clusters are calculated, a matrix of profile plots is then generated, one plot per cluster, each plot showing all profiles assigned to that cluster (Fig. 2). Individual time profiles are colored to show their membership score, as assigned by the FCM algorithm. Clicking on a particular cluster opens a table showing details of profiles assigned to that cluster.

Limitations

A key limitation of this DynaPho module is that the FCM membership score can only be shown via the rainbow color map—a poor visual encoding that can introduce visual artifacts and obscure true data patterns [72,73]. A more effective and scalable encoding could be achieved via a monochromatic color map. A second limitation is the absence of an overall trendline within each cluster (e.g., both of these suggestions are implemented in Fig. 7c from O'Donoghue *et al.* [49]).



	UniProt Acc	Gene name	Protein name	Phosphosite positi	Amino Acid
1	P31751	AKT2	RAC-beta serine/threonine-protein kinase	309	T
2	Q8TDN4	CABLES1	CDK5 and ABL1 enzyme substrate 1	242	S
3	P06213	INSR	Insulin receptor	1190	Y
4	O43318	MAP3K7	Mitogen-activated protein kinase kinase kinase 7	439	S

Fig. 2. DynaPho's "Profile Clustering" module showing all profiles in the Benchmark data set. For the Benchmark data set, six fuzzy c-means clusters (top) were automatically estimated and generated using the default parameters. When a cluster is selected, shown with a pink box around the profile plot, an interactive spreadsheet is displayed (below), containing further information for the profiles contained within. Top and bottom panels were created in DynaPho and modified in Illustrator.

DiBS

DiBS[¶] is a web-based tool designed to enable visual exploration of phosphoproteomic data using heat maps with a circular layout (Fig. 3).

Methods

In DiBS (Fig. 3), quantified ratios for phosphorylation changes are first normalized then visualized using a divergent color map (red = high levels of phosphorylation, green/blue = low). These colors are then used to construct a heat map [65] with a circular layout, where rows are aligned radially, showing time points used in the experiment, and columns are aligned along the circumference, sorted by the normalized phosphorylation changes in a selected row. Color hue is also used on the outermost row to encode the names of genes. Compared with profile plots (above), heat maps are often able to encode a greater density of data with reduced visual clutter; for example, compare Figs. 2 and 3, which both show the same 824 distinct values from the 103-site Benchmark data set.

Features

The web-based heat maps created by DiBs can be explored interactively, thus providing greatly enhanced functionality over a static image. A key interactive feature is the ability to reveal an informative pop-up showing details (gene name, time point, phosphorylation value) associated with each cell in the heat map upon mouse hover (Fig. 3). Additional features include the ability to hide or reveal selected data-values (e.g., hiding specific time points or sets of genes), and to define which information will be displayed in the pop-ups.

Limitations

Using color hue to encode gene names makes it difficult to identify more than 10–15 genes. Additional confusion can arise as color is also used to encode a second, unrelated, property (phosphorylation values). In the online version, these issues are partly mitigated via pop-ups; for static, published figures, gene names could be displayed explicitly around the circumference rather than in a legend off

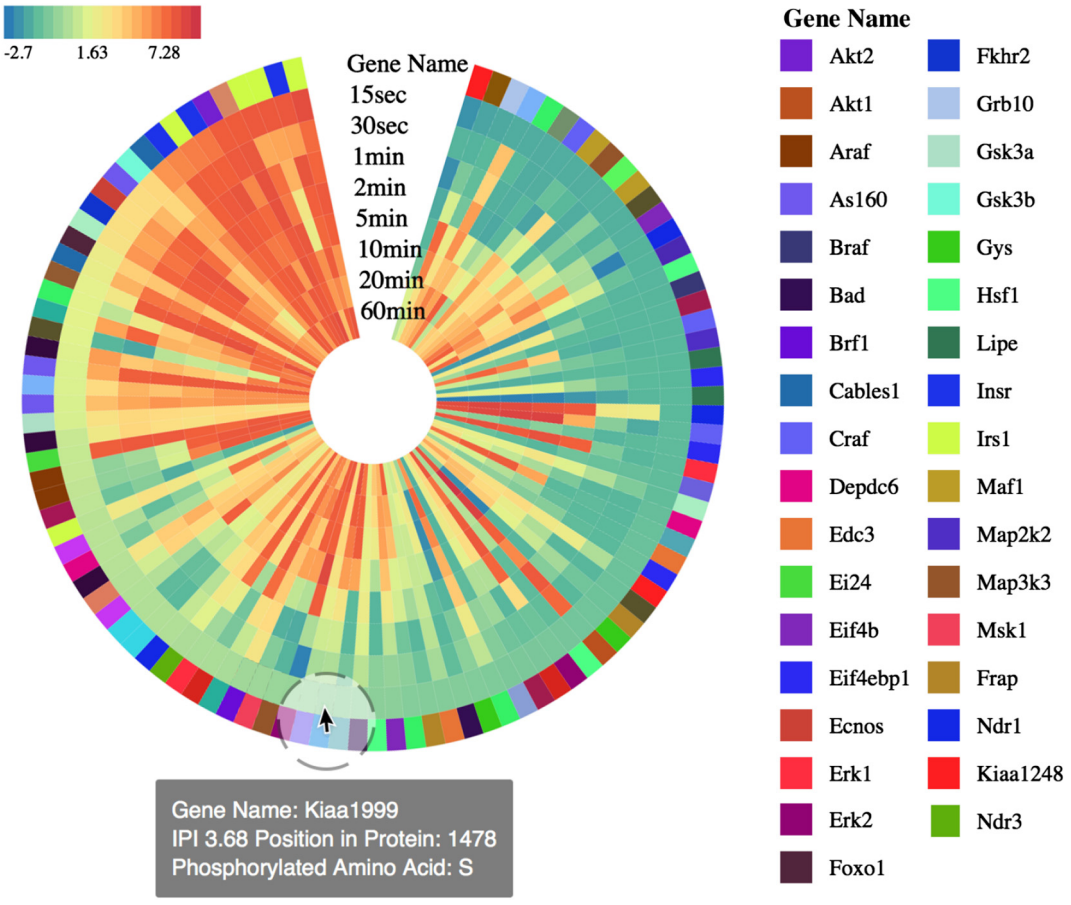


Fig. 3. DiBS tool showing the entire Benchmark data set as a circular heat map. The outermost row of the heat map shows colors that are associated with gene names (right hand side legend). All other rows in the heap map show phosphorylation abundance values using a diverging color map, with red and green/blue for high and low values, respectively (top legend). Columns in the heat map have been sorted by values at the 15-s time point (i.e., the row following gene names). Hovering over a cell displays a pop-up showing the gene name, site number (in an international protein identifier (IPI) protein sequence), and amino acid type. The figure was created in DiBS and modified in Illustrator.

to the side. An additional issue with color is that heat maps, like rainbow color maps, can introduce strong perceptual illusions, resulting in visual artifacts and obscuring true patterns [74]. Other limitations arise due to the circular layout: first, cells on the outermost row have greater area, and hence greater visual prominence, compared with inner cells, introducing mild perceptual biases. Second, this layout greatly limits the number of rows (time points) and columns (genes) that can be included; thus, DiBS can typically only display a small subset of a high-throughput data sets. Finally, some current technical limitations include the following: (1) restricted ability to define sort order in the heat map; (2) data sets file size must be $\leq$1 MB—for comparison, the complete Humphrey data set is ~4 MBs; and (3) some DiBS features cannot be used without paying a fee.

Visualizing Integrated Data

To understand and interpret a new epiproteome data set, it is usually necessary to integrate these data together with prior knowledge. For example, with a phosphoproteomics data set, prior knowledge is needed to infer which kinases and phosphatases have caused the observed phosphorylation changes. For this purpose, one key resource is PhosphoSitePlus [75], a compendium of kinases and their known substrate sites—it also includes similar information for other PTMs. Another key resource is DEPOD [76], a compendium of phosphatases and known substrate sites.

In addition, it is often very helpful to infer the subcellular location of epiproteome events (e.g., via a resource such as COMPARTMENTS [77]), to view relevant protein–protein interactions (e.g., via

BioGRID [78]), and to group proteins in the database based on functional categories (e.g., via resources using Gene Ontology (GO) [79]). Prior knowledge such as this is available from a wealth of online resources (Pathguide[††] [80] currently lists 702 pathway resources). However, such resources need to be used with caution since, overwhelmingly, they are built using data derived from experiments insensitive to PTMs.

In the sections below, we discuss several tools that are used to integrate epiproteomics data with prior and inferred knowledge. All these tools use graphs as the primary strategy for visualization—except for one (PHOXTRACK), which primarily uses bar charts.

Generic graph tools

Graphs have long been a core visual paradigm for organizing data on cellular systems [81]. As a result, a suite of powerful, generic tools and methods are available [42,43] that can be adapted to combine epiproteomics data with prior knowledge. Some commonly used generic tools include: the Cytoscape core distribution [82], Gephi [83], and Adobe Illustrator.

Methods

Within networks or pathway graphs showing epiproteomics data, a single protein is generally represented as a node connected to subnodes that represent the protein's PTM sites (Fig. 4). PTM reactions are typically shown with arrows connecting an enzyme to its substrates, for example, a kinase and a phosphosite. Reactions that inhibit or suppress are typically shown with hammerheads, while generic protein–protein interactions (e.g., co-occurrence in a protein complex [84]) are shown with undirected edges (Fig. 4). For each PTM site, coloring is often used to show experimentally measured PTM abundance values. To show changes in PTM abundances over time—or over a series of concentrations—this coloring can be changed dynamically [85] or, alternatively, a series of "small multiple" versions [86] of the network can be calculated [87].

Features

Graphics generated for epiproteomics data sets are almost always large and complex; a key feature that helps in managing these data is to hide textual information associated with nodes or edges, and selectively reveal this information upon mouse hover. Fortunately, this is widely supported, especially in web-based tools, which further allow linking to other resources with more detailed information. Here, graphic tools specifically designed for biological networks (e.g., Cytoscape [82]) often are particularly useful, for example, by automatically recognizing proteins identifiers in the imported data set and matching them to standard database entries, thereby enabling more useful information in the pop-ups, and facilitating a range of bioinformatics analysis methods. Another useful feature available in many network tools is the ability to interactively expand or collapse subnetworks: this can be especially useful for data sets in which proteins have a large number of PTM sites.

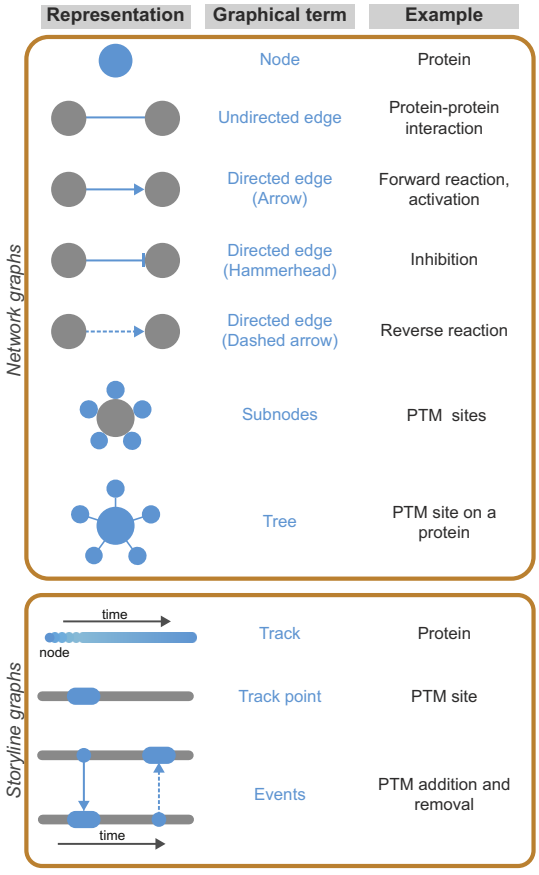


Fig. 4. Graphical terms used to describe networks and storylines. In each case, graphical elements are shown with blue coloring, followed by their denotation, and an example interpretation of the graphical element. A key complexity with epiproteome data is that each protein can have many PTM sites: this structure can be represented using a simple tree graph or, more concisely, with subnodes attached to a node.

Limitations

Using generic graph tools with epiproteome data can be challenging. A core limitation is that most epiproteome experiments to date have focused on tracking PTM changes across time, or in response to

varying stimuli. While such data can be overlaid onto generic graphs, the results are often cluttered and difficult to read [43,44,88]. Usually, when dynamic epiproteome data are overlaid onto a graph, some information is lost; one strategy to overcome this limitation would be to use profile plots (Figs. 1 and 2) connected to a graph via "brushing and linking" [89]. To overcome these limitations, a small but increasing set of tools is available, specifically tailored for integrating epiproteome data with prior knowledge—these are covered in the subsections below.

PhosphoPath

PhosphoPath [90] is a plugin for Cytoscape [82] that uses prior knowledge on kinases and protein interactions to help interpret dynamic phosphoproteomics data sets (Fig. 5).

Methods

In PhosphoPath (Fig. 5a), each data set is first shown as a network graph (Fig. 4), calculated using the so-called "force-directed layout," in which nodes are arranged so that all edges have—as nearly as possible—the same length [91]. In Fig. 5b, the layout is then modified, such that each protein is shown as a node with attached subnodes, indicating phosphorylation sites; arrows are used to connect these sites to kinase proteins, inferred via PhosphoSitePlus [75], while other physical interactions between proteins in the data set are shown using undirected edges, inferred from BioGRID [78]. Changes in PTM abundance are shown using a small, single-row, heat map placed close to selected phosphosites (Fig. 5b).

Features

PhosphoPath is able to take advantage of a range of Cytoscape features that support effective, interactive exploration of large graphs. This includes (1) a graph that facilitates navigation via zoom and pan, (2) automatic hiding or revealing of details as the user zooms in or out, (3) selective reveal of detail via mouse hover, and (4) ability to manually edit the layout (Fig. 5). In addition, a range of analysis methods are supported; this includes automated detection of pathways (obtained from Wikipathways [92]) enriched with proteins in the data set.

Limitations

PhosphoPath currently has several key limitations. First, many manual steps are required to upload data and to create an effective visualization. Second, the method used to assign kinases is often incomplete and does not account for the many other factors that influence phosphorylation abundance during cell signaling, such as protein degradation or phosphatases [93]. Third, no consideration is given to the subcellular location of phosphorylation events, although this can be critical for understanding cell

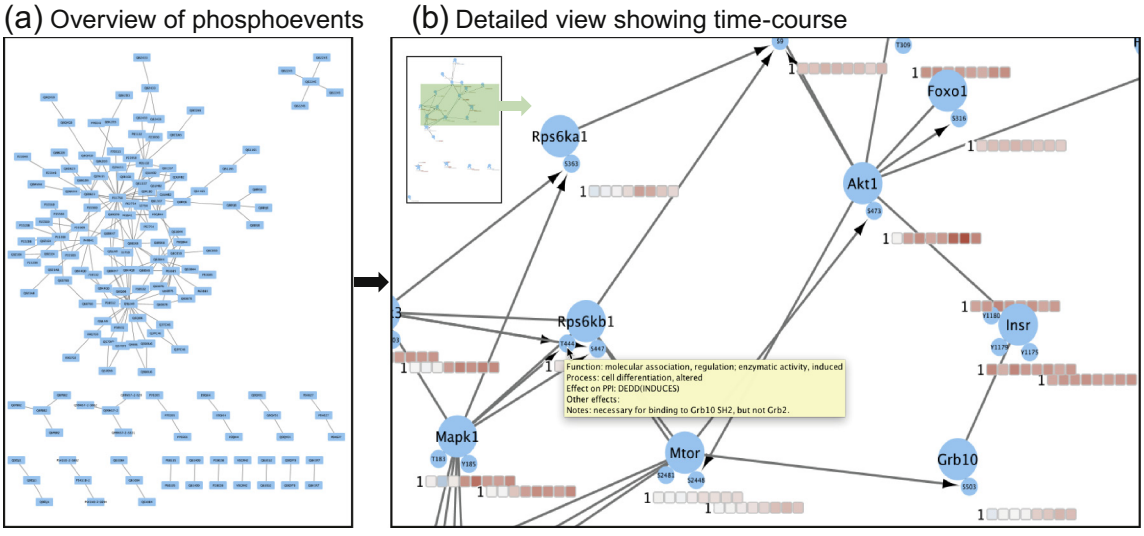


Fig. 5. PhosphoPath network for the entire Benchmark data set. Panel a shows the complete network generated by PhosphoPath via a force-directed layout. Here, proteins and PTM sites are both shown with the same node shape and are hence not easily distinguishable. Panel b shows part of the insulin pathway, which PhosphoPath detects as the most enriched pathway in the Benchmark data set. PTM sites are shown as subnodes located on nodes that represent the corresponding protein. Phosphorylation abundance values for each site are visualized as a single row heat map placed next to the site subnode. Arrows originate from kinases and terminate at phosphosites. Undirected edges show protein–protein interactions. The figure was created using the PhosphoPath plugin (version 3.2) in the Cytoscape tool (version 3.6.1), and panel b was further modified in Illustrator.

signaling [94]. Fourth, PhosphoPath creates a single view, in which the phosphorylation abundance data are split up and scattered across a large graph, making it difficult to see patterns. This could be improved by gathering the abundance data into a smaller number of heat maps (e.g., Fig. 3) or profile plots (e.g., Fig. 2). Finally, due to high interconnectivity, the graph can become unreadably cluttered for large data sets, e.g., with thousands of nodes.

DynaPho (Dynamic Networks)

The web-based tool DynaPho [62] also has a module called "Dynamic Networks" that uses prior knowledge on protein–protein interactions to help assign kinases and phosphatases in phosphoproteomics data sets (Fig. 6).

Methods

The "Dynamic Network" module (Fig. 6) generates a network graph (Fig. 4), using nodes of different shape to distinguish four types of proteins: kinases (shown as octagons), phosphatases (squares), transcription factors (triangles), and all other proteins in the data set (circles). Undirected edges between these nodes show protein–protein interactions derived from multiple sources, including BioGRID [78], HPRD [95], and IntAct [96]. In addition, small circular nodes are used to show phosphosites, each colored to show phosphorylation abundance values, and are joined to their corresponding proteins. Each network can also be switched to a mode where the phosphosites are hidden, and where each protein node is colored to indicate an average phosphorylation

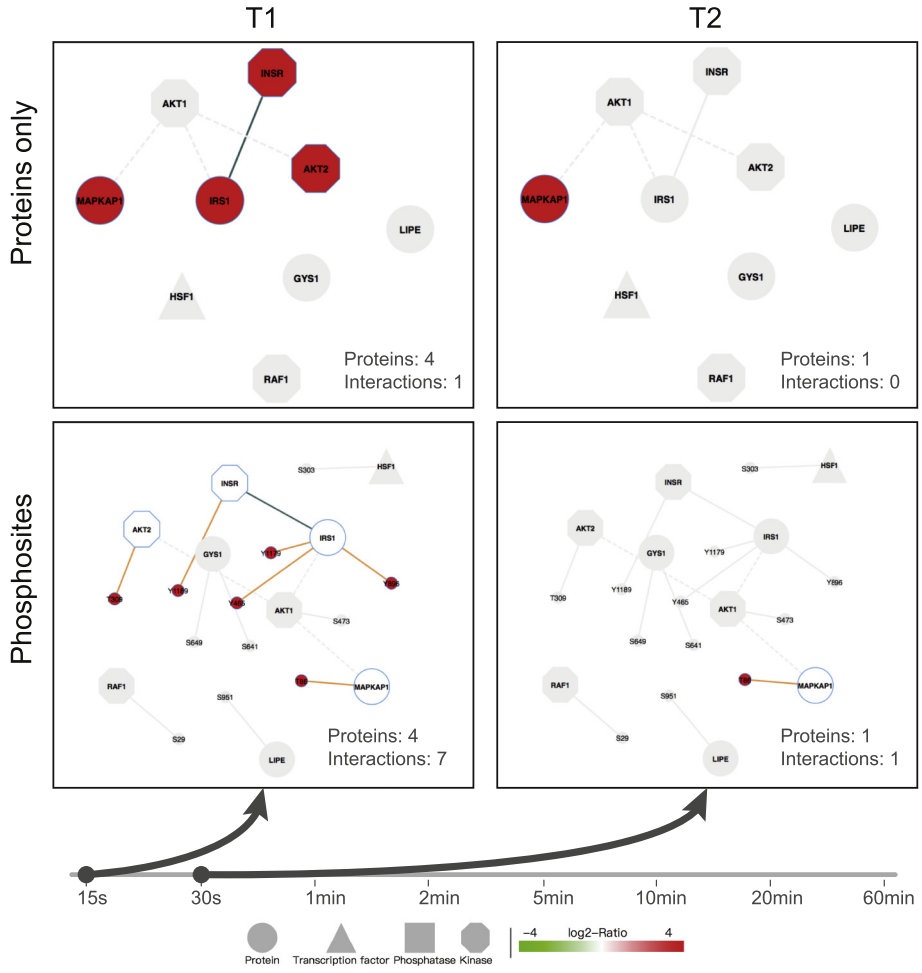


Fig. 6. DynaPho's "Dynamic Networks" module showing networks generated for the entire Benchmark data set. All four panels depict protein–protein interactions. This module provides two views, "Proteins only" (top two panes) and "Phosphosites" (bottom two panes) through a select bar. In the protein-only view, only the protein nodes are shown, colored according to the average abundance ratios of all its sites quantified in the data set. In the phosphosite view, the individual site nodes are also shown, attached to the protein nodes as subnodes, and colored according to their abundance ratios. A slider, seen at the bottom, allows visualizing the time specific network. The left column depicts the network at 15 s, and the right column depicts it at 30 s. This figure contains five screenshots (four panels and a slider), generated using DynaPho, with default parameters, brought together and modified in Illustrator.

abundance value for the corresponding phosphosites (Fig. 6, top). This can significantly reduce clutter, thus, helping the user manage larger networks.

Features

The tool provides a slider with values corresponding to each time point in the data set; this enables a user to explore dynamic changes in phosphorylation abundance values. The user can also explore the network with standard operations, such as zoom, pan, as well as basic editing (e.g., moving individual nodes). By presenting an integrated view that includes known interactions between kinases, phosphatases, and substrate proteins, this tool can help users manually assign which enzymes cause observed phosphorylation and dephosphorylation events.

Limitations

A key limitation of this tool is that it can only show one state of the network at a time, making it difficult to gain an overview of dynamic changes across the whole data set. This can be partly addressed by manually creating small multiple versions of the network—as done in Fig. 6. A second limitation is that the edges joining a protein to its phosphosites are quite variable in length, and often greater than edges between proteins; this contributes to visual clutter, making larger graphs harder to read. This could be improved via shorter edges or subnodes (Fig. 4). A third limitation is the lack of a network overview (e.g., Fig. 5b), which again makes it difficult to use with larger data sets. Finally, the graph can become unreadably cluttered for large data sets (e.g., with thousands of nodes).

DynaPho (Correlation Analysis)

DynaPho [62] also has a module called "Correlation Analysis" that uses prior knowledge together with correlation analysis to assign kinases and phosphatases in phosphoproteomics data sets (Fig. 7).

Methods

The "Correlation Analysis" module (Fig. 7) uses multiple resources (e.g., UniProt [6]) to first identify kinase or phosphatase proteins in the data set. A network graph (Fig. 4) is then generated in which

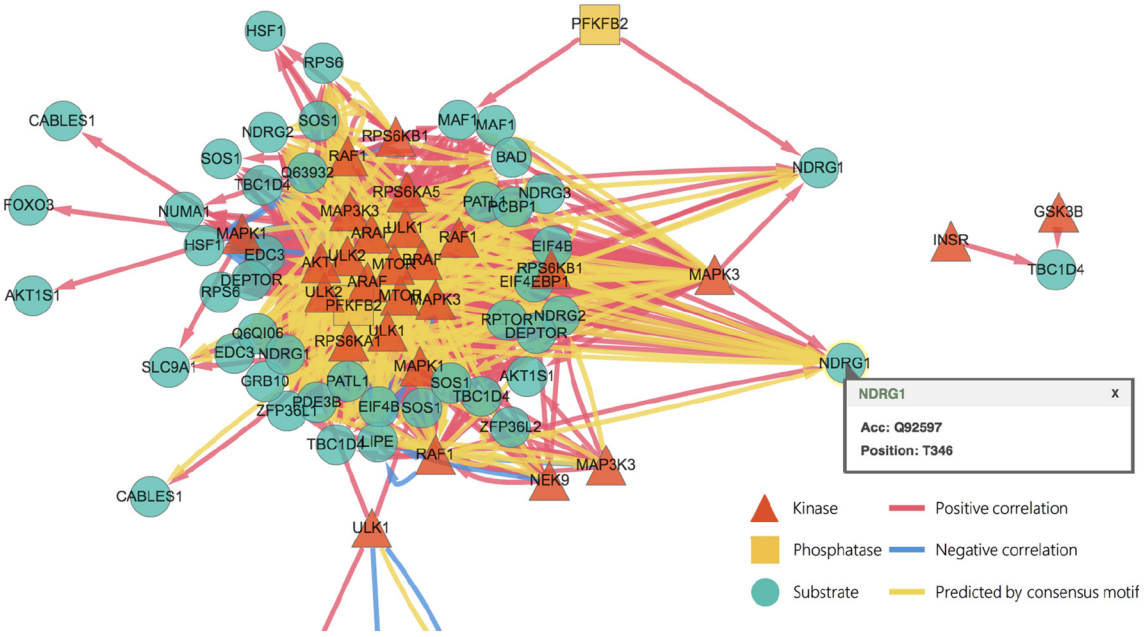


Fig. 7. DynaPho's "Correlation Analysis" network showing almost all of the Benchmark data set. DynaPho uses statistical correlation, gene ontology, and various kinase- and phosphatase-to-substrate databases, to infer potential kinases and phosphatases for the events detected at each phosphosite. The results are shown as a network in which each node represents a phosphosite and edge coloring distinguishes positive and negative correlations (red and blue, respectively). Additional evidence for kinase and phosphatase assignments based on sequence analysis is shown via yellow edges. Details about each site are revealed in a pop-up that appears upon mouse hover. Note that in this network, there are no nodes or edges to represent the protein that each site belongs to. Overall, the correlation network presented in this tool can visually organize phosphorylation events into groups, potentially with similar time profiles; this helps identify the kinases and phosphatases causing the phosphorylation and dephosphorylation events observed. The figure was created using DynaPho with default parameters and has been minimally modified in Illustrator.

each phosphosite is represented as a node, with different shapes distinguishing sites on kinases (triangles), on phosphatases (squares), or on other proteins in the data set (circles). Edges in the network indicate pairs of sites where phosphorylation abundance is correlated either positively or negatively, indicated with red and blue arrows, respectively. Details on each edge are also listed in a table below the network in the webtool.

Features

The tool provides a slider that enables a user to explore the effect of varying cutoffs for the correlation score. The user can also explore the network using standard operations, such as zoom, pan, and basic editing (e.g., moving individual nodes). Clicking on a node reveals a pop-up showing the protein identifier and amino acid position of that phosphosite. Overall, the correlation network presented in this tool can visually organize phosphorylation events into groups, potentially with similar time profiles; this helps identify the kinases and phosphatases causing the phosphorylation and dephosphorylation events observed.

Limitations

A key limitation is that the timing of events is not explicitly encoded visually; although timing is used implicitly to generate the graph, in the resulting layout it is difficult for a user to see and reason about the ordering of events. Another key limitation is that sites belonging to the same protein can be difficult to find as they are not interconnected with edges to the protein and are often widely dispersed throughout the graph. A further limitation is that node labels show only protein names; residue numbering identifying the site that each node represents is only shown when the user manually opens the pop-up for that node. Together, these limitations make it difficult for the user to reason about how multiple phosphorylation events coordinate to influence function in individual proteins, and to orchestrate cellular changes involving multiple proteins. Finally, as with other methods based around networks, the graph can become unreadably cluttered for large data sets.

Minardo

Minardo [61,97] is a novel layout strategy that uses storyline graphs (Fig. 4) to integrate phosphoproteomics data sets with prior knowledge on kinases, phosphatases, and protein subcellular location (Fig. 8). Minardo was designed to address some of the key limitations mentioned above in methods based on network graphs. By depicting events on a circular cellular topology, this layout is well suited for describing the transient cellular responses typically studied in epiproteomics experiments, in which the state of a cell is initially perturbed, then eventually returns to a baseline level. An automated web-based tool using this strategy is currently being developed by the authors; currently, two published examples using Minardo [97,98] are available online at http://minardo.org.

Methods

The Minardo layout (Fig. 8) produces a graph in which spatial position is used to explicitly encode the timing and subcellular location of phosphorylation events. These events are sorted by the time at which phosphorylation abundance is estimated to cross the 50% threshold between the minimum and maximum observed values. Proteins with only one phosphorylation event are shown as nodes, while proteins (or protein complexes) with multiple phosphorylation events are represented as a track—which can be thought of as a node stretched out in time (Fig. 4). Each phosphorylation event is indicated via an arrow connecting either track points or nodes (Fig. 4), with the arrow direction distinguishing the enzyme and substrate.

Features

In the online Minardo examples [97,98], hovering over each node reveals a pop-up showing details about that specific phosphorylation event; in addition, hovering highlights all related events, thus facilitating interactive exploration of the graph. Overall, by including time and subcellular localization explicitly in the layout, Minardo can help users clearly see and reason about the causal relationships between phosphorylation events and how they interact to orchestrate cellular functions.

Limitations

A key limitation of Minardo is that it is currently only a method, not a tool, making it difficult to apply to new data sets. A second limitation is that the use of storyline graphs for dynamics cellular systems is quite novel, thus potentially presenting a steep learning curve for some users. A related limitation is that some of the visual cues used in Minardo to simplify the layout may be too subtle to be easily noticed: for example, tracks showing either proteins or protein complexes are distinguished only via the coloring on the track labels (blue and black for proteins and complexes, respectively). Finally, while Minardo generally results in less visual clutter than network graphs (compare Figs. 7 and 8), the layout may still become unreadably cluttered for larger data sets.

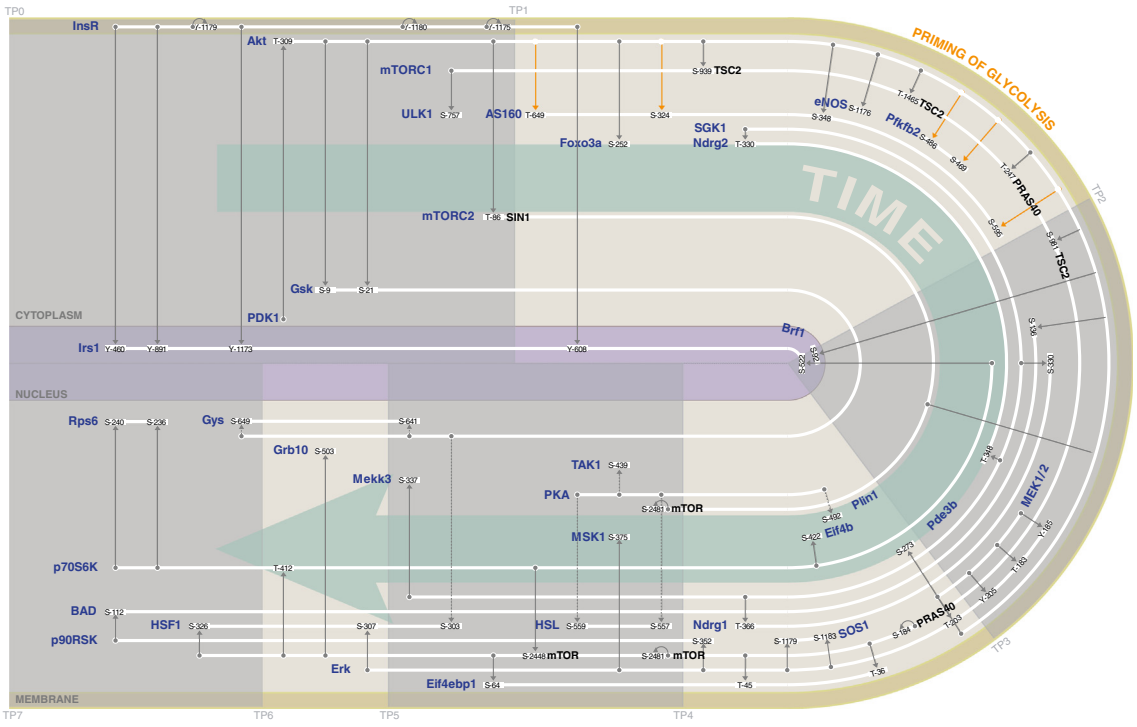


Fig. 8. Minardo layout showing the entire Benchmark data set. The Minardo layout arranges cellular events by time and subcellular location. Time flows in a clockwise fashion and is divided into differently shaded regions, corresponding to time intervals in the Benchmark data set. Tracks represent proteins or protein complexes, identified by text labels. For protein complexes, black labels show the names of protein components involved in phosphoevents. Phosphorylation and dephosphorylation events are shown via solid and dashed arrows, respectively. At the end point of each arrow, the sequence position of the phosphosite is shown. The image was created using the Minardo tool, currently being developed by the authors, with minor modifications made in Illustrator.

PHOXTRACK

The web-based tool PHOXTRACK (PHOsphosite-X-TRacing Analysis of Causal Kinases) [99] aims to identify the kinases responsible for the events observed in a particular phosphoproteomics data set. In contrast to most of the tools discussed above, the results generated by PHOXTRACK are visualized using only bar charts and heat maps (Fig. 9).

Methods

PHOXTRACK compares phosphopeptides observed in a specific phosphoproteomics data set against all phosphopeptides previously assigned to human and mouse kinases; these assignments are drawn from many resources, including PhosphoSitePlus [75], SwissProt [100], and Phospho.ELM [101]. For each time point in the data, PHOXTRACK generates a one-column heat map with rows showing the kinases likely to have either activated (red) or inhibited (blue) phosphoproteomic events in the data set (Fig. 9, left, “kinase regulation”). Another one-column heat map is also generated, with rows showing abundance values for all phosphorylation sites in the data set; this heat map can be used, together with a bar chart, to explore the phosphorylation sites assigned to each kinase (Fig. 9, right, “substrate phosphorylation”).

Features

PHOXTRACK provides several options for modifying its enrichment analysis; for example, the user can select which database is used for kinase assignments. Each resulting analysis can then be explored via a dropdown menu, which enables the user to select one of the enriched kinases, and see all phosphorylation sites assigned to that kinase (Fig. 9, right). Overall, the analyses performed by PHOXTRACK appear to be quite thorough, and results are well presented, making it easy to use and thus helpful for assigning kinases in phosphoproteomics data sets.

Limitations

A key limitation of PHOXTRACK is that it can only show data associated with one time point at a time,

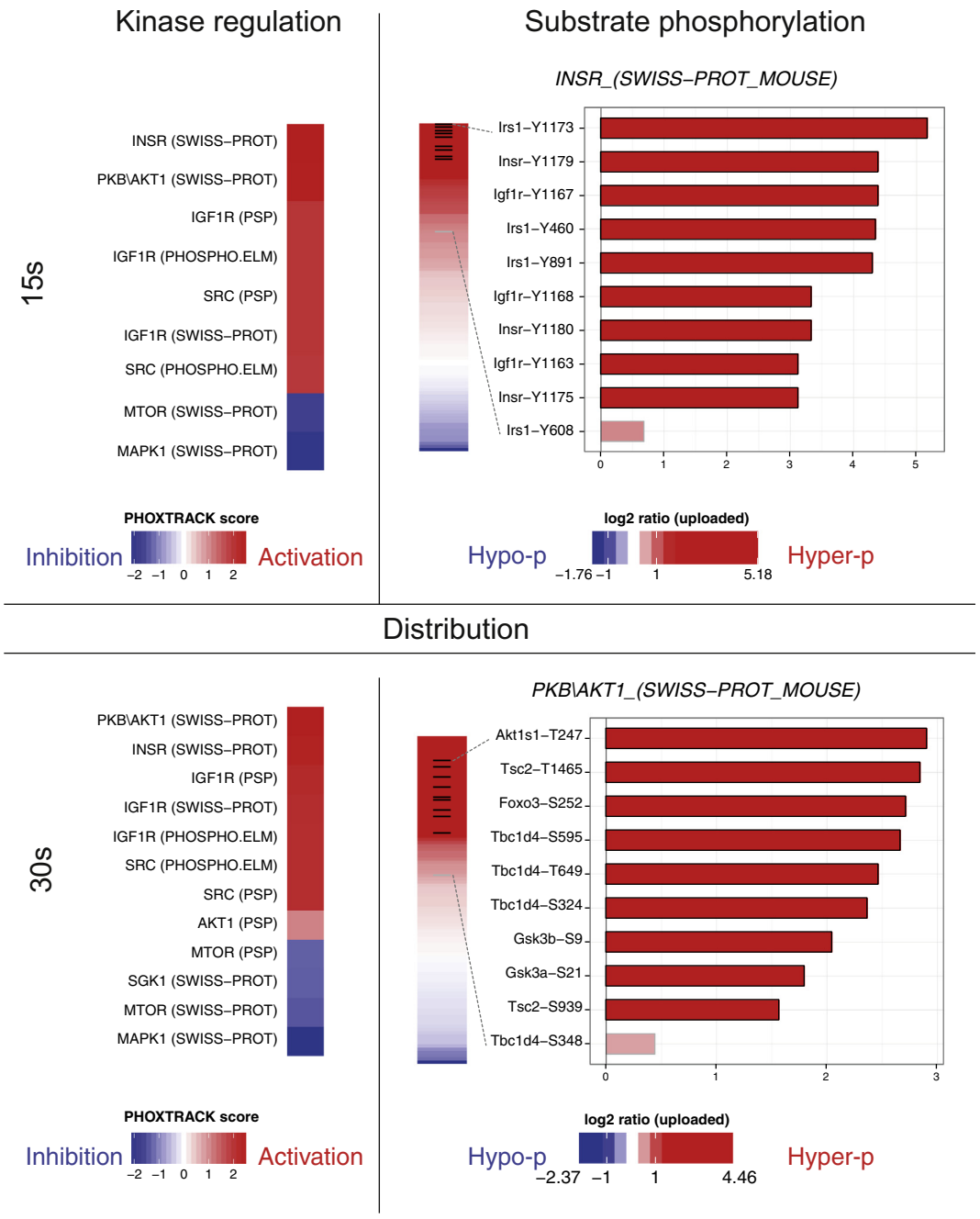


Fig. 9. PHOXTRACK kinase enrichment results for the entire Benchmark data set. PHOXTRACK presents two core pieces of information at any particular time point: “Kinase regulation” and “Substrate phosphorylation”. Kinase regulation displays kinases and their activity states, ordered by most active to most inactive at that time point. Substrate phosphorylation displays substrate sites for the first kinase in the Kinase regulation window, which have contributed the most towards that kinase's enrichment. Results for a different kinase can be selected from a select bar in the tool. On the left side of this window is a single column heat map containing the phosphorylation values at that time point of all the sites in the data set. This heat map has been sorted by most phosphorylated to most dephosphorylated. The black and gray lines in the heat map correspond to the bars in the bar chart. Black lines and the black bordered bars in the bar chart are the sites that have higher contribution to the kinase's enrichment, and the gray bordered bars represent sites that have contributed less. The four panels were created separately in PHOXTRACK using default parameters, then assembled together and modified in Illustrator.

making it difficult to see overall trends across multiple time points; this is partly mitigated by features that make it easy to save images for offline use. A minor limitation is the potential confusion that may arise from using red-blue diverging color-maps [102] for two very different properties, namely, PHOXTRACK enrichment scores and phosphorylation abundance values. A final limitation of PHOXTRACK and other methods using prior knowledge is the effect of biases in underlying databases [103]; for example, one of the known biases is that PHOXTRACK currently only uses kinase assignments for human and mouse proteins; however additional unknown biases may arise for proteins or processes that are either well-studied or understudied [103].

Visualizing Predictive Models

One way to test our understanding of a biological system is to create a model of the system, run a simulation, then check if it reproduces what we observe experimentally [104,105]. For this purpose, a wide range of modeling frameworks have been developed [106] using mathematical formalisms that range from highly detailed (e.g., differential equations) to very minimal (e.g., boolean networks). Most of these frameworks use standardized notations [107,108], which facilitates model sharing, reuse, and storage in online repositories such as BioModels [109], a large compendium of models (both manually curated and automatically generated). Many of these models can be visualized using standard tools (e.g., Cytoscape [82]), and can also be used to predict how the modeled system changes over time; this allows for comparison with experimental data sets and facilitates computational reasoning about biological processes [110].

To date, most systems biology models have been built manually, with a focus on validating them against experimental studies. For example, Di Camillo *et al*. [111] described a comprehensive, small-scale, highly predictive model of insulin signaling. Recently, however, driven by high-throughput experimental techniques, algorithms are emerging that allow predictive models to be automatically generated from large data sets. For example, Terfve *et al*. [112] described an algorithm named PHONEMeS that generates boolean models from high-throughput phosphorylation perturbational data sets. In the next section, we discuss Kappa [113,114], a formalism particularly well suited for modeling dynamic changes to the epiproteome.

Kappa

Kappa[‡‡] [115] is a modeling framework originally designed for simulating phosphoproteomic events, although it can also model many other types of biomolecular events. Accessing all of Kappa's capabilities requires using a suite of desktop tools; here, however, we focus on a limited set of novel and practical visual methods available online via the Kappa website[§§] (Fig. 10).

Methods

In Kappa, proteins and other macromolecules in a model can be visualized using a circular "contact map" in which each macromolecule is represented as an arc segment placed on the circumference of the circle, while specific interaction sites (e.g., PTM sites or domains) are represented as dots on the inner side of each arc segment (Fig. 10a). Interactions between sites are shown as chords. Once the model is defined, the user can run a simulation (using Gillespie's kinetics [116]), generating time profiles of each site (Fig. 10b). Also generated is an "influence map" (Fig. 10c), a somewhat complex graph in which each node represents a reaction (e.g., a PTM event), and each edge indicates the influence that one reaction has on another (green indicates positive influence; red indicates negative influence). The influence map is updated at each time point in the simulation.

Features

Hovering over a site in the contact map highlights its interactions with other components in the system (Fig. 10a); this makes the map usable and insightful, even with models containing many interactions. The simulated profile plots (Fig. 10b) provide a quick and simple method to visually inspect the simulation; the simulated data can be exported to other tools (e.g., DynaPho, Fig. 1) to enable detailed comparison against experimental results. The influence map (Fig. 10c) has buttons (start, next, and back) that enable navigation to versions of the map, covering all time points in the simulation. The influence map can be searched using Kappa's query language, thus facilitating highly detailed exploration of the simulation dynamics.

Limitations

A key limitation of Kappa is that the user needs to manually specify each reaction in the system. A second limitation is that the influence map is often cluttered and difficult to read (even for the relatively small Benchmark data set) and cannot be manually edited—however, this can be partly addressed using a standalone tool called DIN-viz [117]. A final limitation is that using Kappa effectively can require detailed knowledge of its modeling and query language.

Perspectives

The analysis and visualization tools discussed in this review have been developed predominantly for

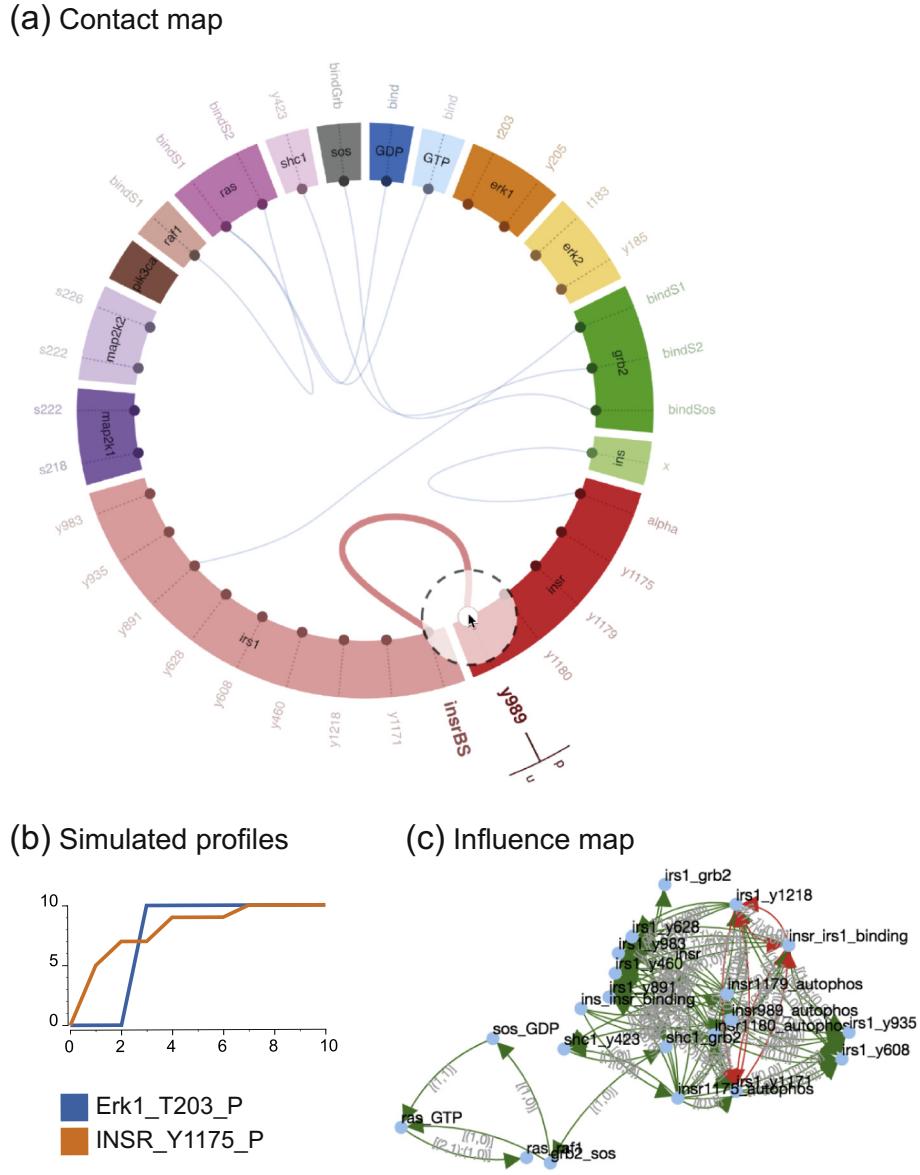


Fig. 10. Three visualizations for a small insulin signaling model from Kappa language's online tool. The model was created by surveying the literature for a small set of proteins, which started from the interaction with insulin and ended at ERK1 and ERK2 proteins, with the focus being the sites that were present in the Humphrey data set. Panel a depicts all the binding interactions represented in the model. It shows a contact map, where arc segments are used to depict proteins and dots in the inner side of the arcs represent the various protein components, such as domains and phosphorylation sites. Chords connect the interacting components. Hovering on a particular dot of an arc segment highlights the connecting cord, and reveals the possible states of the component (e.g. phosphorylated (p) and unphosphorylated (u) states of a phosphorylation site). Panels b and c are simulation outputs. Panel b shows simulated profiles that can be compared to experimental data (Fig. 1). Panel c shows an advanced queryable influence map. The network is a snapshot in simulation time, where the nodes represent the modeled reactions (rules) and the directed edges represent the influence that a reaction in the source node has on the reaction in the target node at that time point. The figure was generated using various features of Kappa's online tool, and panels a and b were modified in Illustrator.

phosphorylation—currently the most studied PTM. Many of these methods can also be used with epiproteome data on other types of PTMs (e.g., acetylation, glycosylation, or methylation). However, we can expect that analyzing specific PTMs may require some tailoring of visual analysis methods used. For example, many PTMs (e.g., prenylation [118] and myristoylation [119]) cause direct changes in subcellular localization; interpretation of such data sets is likely to require visual layout strategies

that explicitly depict cellular compartments and membranes—for example, in a manner similar to Minardo [120].

Over the next decade, we can expect a deluge of epiproteome data sets; these will also include multiple different types of PTMs studied in the same experiment. These experiments can be thought of as "hypothesis-generating engines" [121]; each experiment reveals new insights but leaves many parts of the puzzle missing, which, in turn, helps direct future research, including both experimental and modeling studies [112].

However, the process of gaining insight from epiproteomics techniques is becoming increasingly challenging due to the rapid growth in data volume and complexity. Nonetheless, from the methods and tools considered in this review, we can identify a general strategy for addressing these challenges; this strategy, in turn, suggests specific areas of focus where future bioinformatic developments could help to unravel the insights buried in epiproteome data sets.

(1) *Overview first*. A promising first step in analyzing an epiproteome data set is exemplified in PHOXTRACK [99], which creates a concise visual summary of key outcomes derived from the entire data set (Fig. 9). We anticipate that further innovations using this general approach will be very helpful, and are likely to also include graph-based methods.

(2) *Show Uncertainty and Missing Data*. Epiproteomics data sets often have high variability and missing data; these arise from limitations in experimental techniques, as well as from inherent features of the epiproteome. Additional uncertainties occur in the assignment of kinases and cellular functions, and in other data derived from prior knowledge. The currently available tools have only limited capabilities for visualizing uncertainties: for example, DynaPho's "Correlation Analysis" module [62] allows users to interactively vary the correlation score cutoff (Fig. 7). Effective visualization of uncertainty and incomplete data is an important but often vexing problem, ripe for future innovation [122,123].

(3) *Cluster and Assign Enzymes*. A core task is to assign the enzymes (kinases, methylases, ubiquitin-conjugating enzymes, etc.) responsible for the PTM events in a data set. This can be done by drawing upon prior knowledge (e.g., using PhosphoPath [90] or DynaPho's "Dynamic Networks" module [62]); it can also be done by identifying correlated profiles in the data set (e.g., DynaPho's "Correlation Analysis" module [62]). We expect that future methods will combine both approaches. An interrelated task is to cluster PTM profiles (e.g., Fig. 2); to date, this has mostly been done using generic clustering methods (e.g., FCM [66]); in the future, we expect to see more methods combining clustering with enzyme assignment (e.g., ClueR [70]).

(4) *Assign Function*. Another task is to identify general cellular functions or compartments implicated in a data set; as an example, several events in the Humphrey data set are involved in priming the cell cytoplasm for glycolysis (Fig. 8). Such functions can be found via a wide range of enrichment methods (e.g., listed on the GO home page [79][¶¶]). Many enrichment methods have been developed to analyze gene expression data; some are specific for proteomics data [124]; in the future, we can expect to see enrichment methods tailored specifically for epiproteomics data.

(5) *Integrate and Explore Details*. Information from the above steps needs to be integrated with experimental data to allow effective exploration of details. This is increasingly challenging as data sets grow rapidly in size and complexity. Currently, this integration is typically done using networks with force-directed layout [91], overlaid with additional data—for example, showing PTM abundance via color, and functional categories via node shapes (Figs. 5–7). Networks overlaid in such ways can encode multiple variables: but can be difficult to read and visually cluttered [88]. To improve this, we can expect the development of alternative graph layouts specifically tailored for epiproteome data; some examples already developed include Minardo [61,97] (Fig. 8), CellNetVis [125], and CerebralWeb [126], each of which uses position to encode key variables, such as time, causal flow, or subcellular location. In addition, tools will need to adopt strategies that facilitate interactive exploration of massive data sets: this includes filtering, zoom, overview maps (e.g., Fig. 5b), edge bundling [127], small multiples [86], revealing details upon mouse hover, and collapsing and expanding subgraphs. Especially useful will be alternative views (e.g., a graph and a profile plot) connected via "brushing and linking" [89].

One of the core goals when analyzing an epiproteome data set is to gain insight into how multiple PTM events are orchestrated to regulate overall changes in cell state. This goal is showcased in two online resources created by the authors using the Minardo layout [61,97] (Fig. 8): one resource shows how insulin stimulation leads to

phosphoevents that cause fat cells to switch from releasing energy to storing energy [97][†††]; the other resource shows how phosphoevents regulate key processes in mitosis [98][‡‡‡]. Each resource was published as a Cell SnapShot, one of the few publication formats that currently allow for inclusion of an online version with interactive features.

Unfortunately, due to many underlying complexities, it is currently quite difficult to create visualizations that clearly express the multidimensional data stories within an epiproteome data set, as highlighted in the two SnapShots. While part of the process can be automated, many manual steps are still needed; for comparison, the exemplary Roche metabolic pathway—showing thousands of reactions in one comprehensive view—results from ongoing work begun in 1965 [128][§§§]. However, there is hope that new approaches (e.g., machine learning) will improve the computer-aided design of graphs that visualize complex data sets [129,130].

In summary, advances in experimental techniques are set to produce a deluge of epiproteome data sets that reveal an unprecedented depth of information about the molecular processes in living cells. This calls for the development of new, tailored bioinformatics methods that will help researchers analyze, visualize, and interactively explore these complex data sets, with the goal of enabling new insights into the epiproteome to be effectively integrated with prior knowledge. These insights have potential to revolutionize our understanding of biology and disease, and to lead to more precise diagnosis and therapeutic intervention.

Acknowledgments

We gratefully acknowledge David James and Sean Humphrey for useful suggestions, and Renecia Lowe for help with proofreading. This work was partly supported by CSIRO's OCE Science Leader programme and Computational and Simulation Sciences platform, and partly by the Australian Research Council under Linkage Project LP140100574.

Received 11 October 2018;
Received in revised form 29 January 2019;
Accepted 29 January 2019
Available online 13 February 2019

Keywords:
phosphoproteome;
mass-spectrometry proteomics;
systems biology;
data visualization;
regulatory networks

†https://www.UniProt.org/help/post-translational_modification. Last visited January 22, 2019.

‡PRIDE Archive database website: https://www.ebi.ac.uk/pride/archive/, Last visited May 31, 2018.

§Website for DynaPho: http://140.112.52.89/dynapho/, Last visited June 12, 2018.

¶Website for DiBS: http://www.dibsvis.com/. Last visited June 11, 2018.

††Pathguide website: http://pathguide.org. Last visited September 24, 2018.

‡‡Kappa Language webpage: https://kappalanguage.org/. Last visited June 1, 2018.

§§Online interactive simulator for Kappa: https://tools.kappalanguage.org/try/. Last visited June 1, 2018.

¶¶A website for GO enrichment, http://www.geneontology.org/page/go-enrichment-analysis, additional tools are listed at the bottom of the page. Last visited June 12, 2018.

†††https://cell.com/cell/enhanced/odonoghue

‡‡‡https://cell.com/cell/enhanced/odonoghue2

§§§http://biochemical-pathways.com/

Abbreviations used:
PTM, post-translational modification; MS, mass spectrometry; FCM, fuzzy c-means; GO, Gene Ontology.

References

[1] R. Margueron, P. Trojer, D. Reinberg, The key to development: interpreting the histone code? Curr. Opin. Genet. Dev. 15 (2005) 163–176, https://doi.org/10.1016/j.gde.2005.01.005.

[2] A.P. Lothrop, M.P. Torres, S.M. Fuchs, Deciphering post-translational modification codes, FEBS Lett. 587 (2013) 1247–1257, https://doi.org/10.1016/j.febslet.2013.01.047.

[3] F. Xin, P. Radivojac, Post-translational modifications induce significant yet not extreme changes to protein structure, Bioinformatics 28 (2012) 2905–2913, https://doi.org/10.1093/bioinformatics/bts541.

[4] H. Nishi, K. Hashimoto, A.R. Panchenko, Phosphorylation in protein–protein binding: effect on stability and function, Structure 19 (2011) 1807–1815, https://doi.org/10.1016/j.str.2011.09.021.

[5] A.S. Gajadhar, F.M. White, System level dynamics of post-translational modifications, Curr. Opin. Biotechnol. 28 (2014) 83–87, https://doi.org/10.1016/j.copbio.2013.12.009.

[6] R. Apweiler, A. Bairoch, C.H. Wu, W.C. Barker, B. Boeckmann, S. Ferro, et al., UniProt: the Universal Protein knowledgebase, Nucleic Acids Res. 32 (2004) D115–D119, https://doi.org/10.1093/nar/gkh131.

[7] A.S. Venne, L. Kollipara, R.P. Zahedi, The next level of complexity: crosstalk of posttranslational modifications, Proteomics 14 (2014) 513–524, https://doi.org/10.1002/pmic.201300344.
[8] P. Minguez, I. Letunic, L. Parca, P. Bork, PTMcode: a database of known and predicted functional associations between post-translational modifications in proteins, Nucleic Acids Res. 41 (2013) D306–D311, https://doi.org/10.1093/nar/gks1230.
[9] K.A. Liddy, M.Y. White, S.J. Cordwell, Functional decorations: post-translational modifications and heart disease delineated by targeted proteomics, Genome Med. 5 (2013) 20, https://doi.org/10.1186/gm424.
[10] P.G. Hains, R.J.W. Truscott, Age-dependent deamidation of lifelong proteins in the human lens, Invest. Ophthalmol. Vis. Sci. 51 (2010) 3107–3114, https://doi.org/10.1167/iovs.09-4308.
[11] I. Scott, Regulation of cellular homoeostasis by reversible lysine acetylation, Essays Biochem. 52 (2012) 13–22, https://doi.org/10.1042/bse0520013.
[12] H. Abou-Abbass, H. Abou-El-Hassan, H. Bahmad, K. Zibara, A. Zebian, R. Youssef, et al., Glycosylation and other PTMs alterations in neurodegenerative diseases: current status and future role in neurotrauma, Electrophoresis 37 (2016) 1549–1561, https://doi.org/10.1002/elps.201500585.
[13] A.R. Wende, Post-translational modifications of the cardiac proteome in diabetes and heart failure, Proteomics Clin. Appl. 10 (2016) 25–38, https://doi.org/10.1002/prca.201500052.
[14] P.R. Gajjala, D. Fliser, T. Speer, V. Jankowski, J. Jankowski, Emerging role of post-translational modifications in chronic kidney disease and cardiovascular disease, Nephrol. Dial. Transplant. 30 (2015) 1814–1824, https://doi.org/10.1093/ndt/gfv048.
[15] A.L. Santos, A.B. Lindner, Protein posttranslational modifications: roles in aging and age-related disease, Oxid. Med. Cell. Longev. 2017 (2017) 5716409, https://doi.org/10.1155/2017/5716409.
[16] H. Xu, Y. Wang, S. Lin, W. Deng, D. Peng, Q. Cui, et al., PTMD: a database of human disease-associated post-translational modifications, Genomics Proteomics Bioinformatics 16 (2018) 244–251, https://doi.org/10.1016/j.gpb.2018.06.004.
[17] O. Pagel, S. Loroch, A. Sickmann, R.P. Zahedi, Current strategies and findings in clinically relevant post-translational modification-specific proteomics, Expert Rev. Proteomics 12 (2015) 235–253, https://doi.org/10.1586/14789450.2015.1042867.
[18] C.T. Walsh, S. Garneau-Tsodikova, G.J. Gatto, Protein posttranslational modifications: the chemistry of proteome diversifications, Angew. Chem. Int. Ed. Engl. 44 (2005) 7342–7372, https://doi.org/10.1002/anie.200501023.
[19] G.A. Khoury, R.C. Baliban, C.A. Floudas, Proteome-wide post-translational modification statistics: frequency analysis and curation of the swiss-prot database, Sci. Rep. 1 (2011) https://doi.org/10.1038/srep00090.
[20] V.G. Allfrey, A.E. Mirsky, Structural modifications of histones and their possible role in the regulation of RNA synthesis, Science 144 (1964) 559, https://doi.org/10.1126/science.144.3618.559.
[21] M.A. Glozak, N. Sengupta, X. Zhang, E. Seto, Acetylation and deacetylation of non-histone proteins, Gene 363 (2005) 15–23, https://doi.org/10.1016/j.gene.2005.09.010.
[22] S. Zhao, W. Xu, W. Jiang, W. Yu, Y. Lin, T. Zhang, et al., Regulation of cellular metabolism by protein lysine acetylation, Science 327 (2010) 1000–1004, https://doi.org/10.1126/science.1179689.
[23] T. Kouzarides, Acetylation: a regulatory modification to rival phosphorylation? EMBO J. 19 (2000) 1176–1179, https://doi.org/10.1093/emboj/19.6.1176.
[24] M. Dalziel, M. Crispin, C.N. Scanlan, N. Zitzmann, R.A. Dwek, Emerging principles for the therapeutic exploitation of glycosylation, Science 343 (2014) 1235681, https://doi.org/10.1126/science.1235681.
[25] K.K. Biggar, S.S.-C. Li, Non-histone protein methylation as a regulator of cellular signalling and function, Nat. Rev. Mol. Cell Biol. 16 (2015) 5–17, https://doi.org/10.1038/nrm3915.
[26] F. Ikeda, I. Dikic, Atypical ubiquitin chains: new molecular signals. "Protein Modifications: Beyond the Usual Suspects" review series, EMBO Rep. 9 (2008) 536–542, https://doi.org/10.1038/embor.2008.93.
[27] L. Hicke, Protein regulation by monoubiquitin, Nat. Rev. Mol. Cell Biol. 2 (2001) 195–201, https://doi.org/10.1038/35056583.
[28] P.P. Di Fiore, S. Polo, K. Hofmann, When ubiquitin meets ubiquitin receptors: a signalling connection, Nat. Rev. Mol. Cell Biol. 4 (2003) 491–497, https://doi.org/10.1038/nrm1124.
[29] K. Haglund, S. Sigismund, S. Polo, I. Szymkiewicz, P.P. Di Fiore, I. Dikic, Multiple monoubiquitination of RTKs is sufficient for their endocytosis and degradation, Nat. Cell Biol. 5 (2003) 461–466, https://doi.org/10.1038/ncb983.
[30] T.M. Karve, A.K. Cheema, Small changes huge impact: the role of protein posttranslational modifications in cellular homeostasis and disease, J. Amino Acids 2011 (2011) 207691, https://doi.org/10.4061/2011/207691.
[31] E.J. Needham, B.L. Parker, T. Burykin, D.E. James, S.J. Humphrey, Illuminating the dark phosphoproteome, Sci. Signal. 12 (2019)https://doi.org/10.1126/scisignal.aau8645.
[32] J.V. Olsen, M. Mann, Status of large-scale analysis of post-translational modifications by mass spectrometry, Mol. Cell. Proteomics 12 (2013) 3444–3452, https://doi.org/10.1074/mcp.O113.034181.
[33] L.K. Povlsen, P. Beli, S.A. Wagner, S.L. Poulsen, K.B. Sylvestersen, J.W. Poulsen, et al., Systems-wide analysis of ubiquitylation dynamics reveals a key role for PAF15 ubiquitylation in DNA-damage bypass, Nat. Cell Biol. 14 (2012) 1089–1098, https://doi.org/10.1038/ncb2579.
[34] M.J. Miller, M. Scalf, T.C. Rytz, S.L. Hubler, L.M. Smith, R.D. Vierstra, Quantitative proteomics reveals factors regulating RNA biology as dynamic targets of stress-induced SUMOylation in Arabidopsis, Mol. Cell. Proteomics 12 (2013) 449–463, https://doi.org/10.1074/mcp.M112.025056.
[35] J.-S. Seeler, A. Dejean, SUMO and the robustness of cancer, Nat. Rev. Cancer 17 (2017) 184–197, https://doi.org/10.1038/nrc.2016.143.
[36] N. Dephoure, K.L. Gould, S.P. Gygi, D.R. Kellogg, Mapping and analysis of phosphorylation sites: a quick guide for cell biologists, Mol. Biol. Cell 24 (2013) 535–542, https://doi.org/10.1091/mbc.E12-09-0677.
[37] K.B. Emdal, A.-K. Pedersen, D.B. Bekker-Jensen, K.P. Tsafou, H. Horn, S. Lindner, et al., Temporal proteomics of NGF-TrkA signaling identifies an inhibitory role for the E3 ligase Cbl-b in neuroblastoma cell differentiation, Sci. Signal. 8 (2015) ra40, https://doi.org/10.1126/scisignal.2005769.
[38] B. Zhu, Q. He, J. Xiang, F. Qi, H. Cai, J. Mao, et al., Quantitative phosphoproteomic analysis reveals key mechanisms of cellular proliferation in liver cancer cells,

Sci. Rep. 7 (2017) 10908, https://doi.org/10.1038/s41598-017-10716-0.
[39] J.V. Olsen, M. Vermeulen, A. Santamaria, C. Kumar, M.L. Miller, L.J. Jensen, et al., Quantitative phosphoproteomics reveals widespread full phosphorylation site occupancy during mitosis, Sci. Signal. 3 (2010) ra3, https://doi.org/10.1126/scisignal.2000475.
[40] R.K. Kandasamy, G.I. Vladimer, B. Snijder, A.C. Müller, M. Rebsamen, J.W. Bigenzahn, et al., A time-resolved molecular map of the macrophage response to VSV infection, Npj Syst. Biol. Appl. 2 (2016) 16027, https://doi.org/10.1038/npjsba.2016.27.
[41] N.L. Young, M.D. Plazas-Mayorca, B.A. Garcia, Systems-wide proteomic characterization of combinatorial post-translational modification patterns, Expert Rev. Proteomics 7 (2010) 79–92.
[42] G.A. Pavlopoulos, D. Malliarakis, N. Papanikolaou, T. Theodosiou, A.J. Enright, I. Iliopoulos, Visualizing genome and systems biology: technologies, tools, implementation techniques and trends, past, present and future, Gigascience 4 (2015) 38, https://doi.org/10.1186/s13742-015-0077-2.
[43] N. Gehlenborg, S.I. O'Donoghue, N.S. Baliga, A. Goesmann, M.A. Hibbs, H. Kitano, et al., Visualization of omics data for systems biology, Nat. Methods 7 (2010) S56–S68, https://doi.org/10.1038/nmeth.1436.
[44] M. Secrier, R. Schneider, Visualizing time-related data in biology, a review, Brief. Bioinform. 15 (2014) 771–782, https://doi.org/10.1093/bib/bbt021.
[45] R.A. Haeusler, T.E. McGraw, D. Accili, Biochemical and cellular properties of insulin receptor signalling, Nat. Rev. Mol. Cell Biol. 19 (2018) 31–44, https://doi.org/10.1038/nrm.2017.89.
[46] K. Sharma, R.C.J. D'Souza, S. Tyanova, C. Schaab, J.R. Wiśniewski, J. Cox, et al., Ultradeep human phosphoproteome reveals a distinct regulatory nature of Tyr and Ser/Thr-based signaling, Cell Rep. 8 (2014) 1583–1594, https://doi.org/10.1016/j.celrep.2014.07.036.
[47] S.J. Humphrey, G. Yang, P. Yang, D.J. Fazakerley, J. Stöckli, J. Y. Yang, et al., Dynamic adipocyte phosphoproteome reveals that Akt directly regulates mTORC2, Cell Metab. 17 (2013) 1009–1020, https://doi.org/10.1016/j.cmet.2013.04.010.
[48] J.D. Fekete, J.J. Van Wijk, J.T. Stasko, C. North, The value of information visualization, Inf. Vis. (2008) 1–18.
[49] S.I. O'Donoghue, B.F. Baldi, S.J. Clark, A.E. Darling, J.M. Hogan, S. Kaur, et al., Visualization of Biomedical Data. Visualization of Biomedical Data, 2018.
[50] R.A. McCloy, B.L. Parker, S. Rogers, R. Chaudhuri, V. Gayevskiy, N.J. Hoffman, et al., Global phosphoproteomic mapping of early mitotic exit in human cells identifies novel substrate dephosphorylation motifs, Mol. Cell. Proteomics 14 (2015) 2194–2212, https://doi.org/10.1074/mcp.M114.046938.
[51] M. Thomson, J. Gunawardena, Unlimited multistability in multisite phosphorylation systems, Nature 460 (2009) 274–277, https://doi.org/10.1038/nature08102.
[52] P. Cohen, The regulation of protein function by multisite phosphorylation—a 25 year update, Trends Biochem. Sci. 25 (2000) 596–601, https://doi.org/10.1016/S0968-0004(00)01712-6.
[53] H.C. Reinhardt, M.B. Yaffe, Phospho-Ser/Thr-binding domains: navigating the cell cycle and DNA damage response, Nat. Rev. Mol. Cell Biol. 14 (2013) 563–580, https://doi.org/10.1038/nrm3640.
[54] M.F. White, S.E. Shoelson, H. Keutmann, C.R. Kahn, A cascade of tyrosine autophosphorylation in the beta-subunit activates the phosphotransferase of the insulin receptor, J. Biol. Chem. 263 (1988) 2969–2980.
[55] J.V. Olsen, B. Blagoev, F. Gnad, B. Macek, C. Kumar, P. Mortensen, et al., Global, in vivo, and site-specific phosphorylation dynamics in signaling networks, Cell 127 (2006) 635–648, https://doi.org/10.1016/j.cell.2006.09.026.
[56] P.V. Attwood, M.J. Piggott, X.L. Zu, P.G. Besant, Focus on phosphohistidine, Amino Acids 32 (2007) 145–156, https://doi.org/10.1007/s00726-006-0443-6.
[57] P.G. Besant, P.V. Attwood, Detection and analysis of protein histidine phosphorylation, Mol. Cell. Biochem. 329 (2009) 93–106, https://doi.org/10.1007/s11010-009-0117-2.
[58] S. Munk, J.C. Refsgaard, J.V. Olsen, Systems analysis for interpretation of phosphoproteomics data, Phospho-Proteomics (2016) 341–360.
[59] J.A. Vizcaíno, A. Csordas, N. del-Toro, J.A. Dianes, J. Griss, I. Lavidas, et al., 2016 Update of the PRIDE database and its related tools, Nucleic Acids Res. 44 (2016) D447–D456, https://doi.org/10.1093/nar/gkv1145.
[60] S.M. Hill, L.M. Heiser, T. Cokelaer, M. Unger, N.K. Nesser, D. E. Carlin, et al., Inferring causal molecular networks: empirical assessment through a community-based effort, Nat. Methods 13 (2016) 310–318, https://doi.org/10.1038/nmeth.3773.
[61] D.K. Ma, C. Stolte, S. Kaur, M. Bain, S.I. O'Donoghue, Visual analytics of signalling pathways using time profiles, Signal and Image Analysis for Biomedical and Life Sciences 2015, pp. 3–22.
[62] C.-L. Hsu, J.-K. Wang, P.-C. Lu, H.-C. Huang, H.-F. Juan, DynaPho: a web platform for inferring the dynamics of time-series phosphoproteomics, Bioinformatics (2017), https://doi.org/10.1093/bioinformatics/btx443.
[63] S.F. Altschul, W. Gish, W. Miller, E.W. Myers, D.J. Lipman, Basic local alignment search tool, J. Mol. Biol. 215 (1990) 403–410, https://doi.org/10.1016/S0022-2836(05)80360-2.
[64] S. Kaur, B. Baldi, J. Vuong, S.I. O'Donoghue, A benchmark dataset for analyzing and visualizing the dynamic epiproteome. Data In Brief.
[65] N. Gehlenborg, B. Wong, Points of view: heat maps, Nat. Methods 9 (2012) 213, https://doi.org/10.1038/nmeth.1902.
[66] L. Kumar, M.E. Futschik, Mfuzz: a software package for soft clustering of microarray data, Bioinformation 2 (2007) 5–7.
[67] J.C. Bezdek, R. Ehrlich, W. Full, FCM: the fuzzy c-means clustering algorithm, Comput. Geosci. 10 (1984) 191–203.
[68] S. Chuai-Aree, C. Lursinsap, P. Sophasathit, S. Siripant, Fuzzy C-mean: a statistical feature classification of text and image segmentation method, Int. J. Uncertainty Fuzziness Knowledge Based Syst. 9 (2001) 661–671.
[69] W. Domanova, J. Krycer, R. Chaudhuri, P. Yang, F. Vafaee, D. Fazakerley, et al., Unraveling kinase activation dynamics using kinase-substrate relationships from temporal large-scale phosphoproteomics studies, PLoS One 11 (2016), e0157763. https://doi.org/10.1371/journal.pone.0157763.
[70] P. Yang, X. Zheng, V. Jayaswal, G. Hu, J.Y.H. Yang, R. Jothi, Knowledge-based analysis for detecting key signaling events from time-series phosphoproteomics data, PLoS Comput. Biol. 11 (2015), e1004403. https://doi.org/10.1371/journal.pcbi.1004403.
[71] Y. Li, X. Zhou, Z. Zhai, T. Li, Co-occurring protein phosphorylation are functionally associated, PLoS Comput. Biol. 13 (2017), e1005502. https://doi.org/10.1371/journal.pcbi.1005502.

[72] Y. Liu, J. Heer, Somewhere over the rainbow: an empirical assessment of quantitative colormaps, Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems 2018, p. 598.
[73] D. Borland, M.R. Taylor, Rainbow color map (still) considered harmful, IEEE Comput. Graph. Appl. 27 (2007) 14–17, https://doi.org/10.1109/MCG.2007.323435.
[74] B. Wong, Points of view: points of review (part 2), Nat. Methods 8 (2011) 189, https://doi.org/10.1038/nmeth0311-189.
[75] P.V. Hornbeck, J.M. Kornhauser, S. Tkachev, B. Zhang, E. Skrzypek, B. Murray, et al., PhosphoSitePlus: a comprehensive resource for investigating the structure and function of experimentally determined post-translational modifications in man and mouse, Nucleic Acids Res. 40 (2012) D261–D270, https://doi.org/10.1093/nar/gkr1122.
[76] G. Duan, X. Li, M. Köhn, The human DEPhOsphorylation database DEPOD: a 2015 update, Nucleic Acids Res. 43 (2015) D531–D535, https://doi.org/10.1093/nar/gku1009.
[77] J.X. Binder, S. Pletscher-Frankild, K. Tsafou, C. Stolte, S.I. O'Donoghue, R. Schneider, et al., COMPARTMENTS: unification and visualization of protein subcellular localization evidence, Database (Oxford) 2014 (2014) bau012, https://doi.org/10.1093/database/bau012.
[78] A. Chatr-Aryamontri, R. Oughtred, L. Boucher, J. Rust, C. Chang, N.K. Kolas, et al., The BioGRID interaction database: 2017 update, Nucleic Acids Res. 45 (2017) D369–D379, https://doi.org/10.1093/nar/gkw1102.
[79] M. Ashburner, C.A. Ball, J.A. Blake, D. Botstein, H. Butler, J.M. Cherry, et al., Gene Ontology: tool for the unification of biology, Nat. Genet. 25 (2000).
[80] G.D. Bader, M.P. Cary, C. Sander, Pathguide: a pathway resource list, Nucleic Acids Res. 34 (2006) D504–D506, https://doi.org/10.1093/nar/gkj126.
[81] G.A. Pavlopoulos, M. Secrier, C.N. Moschopoulos, T.G. Soldatos, S. Kossida, J. Aerts, et al., Using graph theory to analyze biological networks, BioData Min. 4 (2011) 10, https://doi.org/10.1186/1756-0381-4-10.
[82] P. Shannon, A. Markiel, O. Ozier, N.S. Baliga, J.T. Wang, D. Ramage, et al., Cytoscape: a software environment for integrated models of biomolecular interaction networks, Genome Res. 13 (2003) 2498–2504, https://doi.org/10.1101/gr.1239303.
[83] M. Bastian, S. Heymann, M. Jacomy, Gephi: an open source software for exploring and manipulating networks, Icwsm, 8, 2009, pp. 361–362.
[84] S. Fields, R. Sternglanz, The two-hybrid system: an assay for protein–protein interactions, Trends Genet. 10 (1994) 286–292, https://doi.org/10.1016/0168-9525(90)90012-U.
[85] C.H. You, L.B. Holder, D.J. Cook, Temporal and structural analysis of biological networks in combination with microarray data, Computational Intelligence in Bioinformatics and Computational Biology, 2008 CIBCB'08 IEEE Symposium On 2008, p. 62.
[86] E.R. Tufte, N.H. Goeler, R. Benson, Envisioning Information, vol. 126, Graphics press, Cheshire CT, 1990.
[87] A. Barsky, T. Munzner, J. Gardy, R. Kincaid, Cerebral: visualizing multiple experimental conditions on a graph with biological context, IEEE Trans. Vis. Comput. Graph. 14 (2008) 1253–1260.
[88] D. Merico, D. Gfeller, G.D. Bader, How to visually interpret biological data using networks, Nat. Biotechnol. 27 (2009) 921–924, https://doi.org/10.1038/nbt.1567.
[89] A. Buja, J.A. McDonald, J. Michalak, W. Stuetzle, Interactive data visualization using focusing and linking, Proceeding Visualization '91, IEEE Comput. Soc. Presstime 1991, pp. 156–163, https://doi.org/10.1109/VISUAL.1991.175794.
[90] L.M. Raaijmakers, P. Giansanti, P.A. Possik, J. Mueller, D. S. Peeper, A.J.R. Heck, et al., PhosphoPath: visualization of phosphosite-centric dynamics in temporal molecular networks, J. Proteome Res. 14 (2015) 4332–4341, https://doi.org/10.1021/acs.jproteome.5b00529.
[91] S.G. Kobourov, Force-directed drawing algorithms, in: R. Tamassia (Ed.), Handbook of Graph Drawing and Visualization, Citeseer, 2004.
[92] M. Kutmon, A. Riutta, N. Nunes, K. Hanspers, E.L. Willighagen, A. Bohler, et al., WikiPathways: capturing the full diversity of pathway knowledge, Nucleic Acids Res. 44 (2016) D488–D494, https://doi.org/10.1093/nar/gkv1024.
[93] L. Gelens, J. Qian, M. Bollen, A.T. Saurin, The importance of kinase-phosphatase integration: lessons from mitosis, Trends Cell Biol. 28 (2018) 6–21, https://doi.org/10.1016/j.tcb.2017.09.005.
[94] E.U. Azeloglu, R. Iyengar, Signaling networks: information flow, computation, and decision making, Cold Spring Harb. Perspect. Biol. 7 (2015) a005934, https://doi.org/10.1101/cshperspect.a005934.
[95] L. Baolin, H. Bo, HPRD: a high performance RDF database, in: K. Li, C. Jesshope, H. Jin, J.-L. Gaudiot (Eds.), Network and Parallel Computing, 4672, Springer Berlin Heidelberg, Berlin, Heidelberg 2007, pp. 364–374, https://doi.org/10.1007/978-3-540-74784-0_37.
[96] H. Hermjakob, L. Montecchi-Palazzi, C. Lewington, S. Mudali, S. Kerrien, S. Orchard, et al., IntAct: an open source molecular interaction database, Nucleic Acids Res. 32 (2004) D452–D455, https://doi.org/10.1093/nar/gkh052.
[97] D.K.G. Ma, C. Stolte, J.R. Krycer, D.E. James, S.I. O'Donoghue, SnapShot: Insulin/IGF1 Signaling, Cell 161 (2015), https://doi.org/10.1016/j.cell.2015.04.041 (948–948.e1).
[98] A. Burgess, J. Vuong, S. Rogers, M. Malumbres, S.I. O'Donoghue, Snapshot: phosphoregulation of mitosis, Cell 169 (2017), https://doi.org/10.1016/j.cell.2017.06.003 (1358–1358.e1).
[99] C. Weidner, C. Fischer, S. Sauer, PHOXTRACK-a tool for interpreting comprehensive data sets of post-translational modifications of proteins, Bioinformatics 30 (2014) 3410–3411, https://doi.org/10.1093/bioinformatics/btu572.
[100] B. Boeckmann, A. Bairoch, R. Apweiler, M.-C. Blatter, A. Estreicher, E. Gasteiger, et al., The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003, Nucleic Acids Res. 31 (2003) 365–370, https://doi.org/10.1093/nar/gkg095.
[101] F. Diella, S. Cameron, C. Gemünd, R. Linding, A. Via, B. Kuster, et al., Phospho.ELM: a database of experimentally verified phosphorylation sites in eukaryotic proteins, BMC Bioinf. 5 (2004) 79, https://doi.org/10.1186/1471-2105-5-79.
[102] K. Moreland, Diverging color maps for scientific visualization, International Symposium on Visual Computing, 92, 2009.
[103] W.A. Haynes, A. Tomczak, P. Khatri, Gene annotation bias impedes biomedical research, Sci. Rep. 8 (2018) 1362, https://doi.org/10.1038/s41598-018-19333-x.
[104] E. Bartocci, P. Lió, Computational modeling, formal analysis, and tools for systems biology, PLoS Comput. Biol. 12 (2016), e1004591. https://doi.org/10.1371/journal.pcbi.1004591.
[105] H. Kitano, Computational systems biology, Nature 420 (2002) 206–210, https://doi.org/10.1038/nature01254.

[106] D. Machado, R.S. Costa, M. Rocha, E.C. Ferreira, B. Tidor, I. Rocha, Modeling formalisms in systems biology, AMB Express 1 (2011) 45, https://doi.org/10.1186/2191-0855-1-45.
[107] M. Hucka, A. Finney, H.M. Sauro, H. Bolouri, J.C. Doyle, H. Kitano, et al., The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models, Bioinformatics 19 (2003) 524–531, https://doi.org/10.1093/bioinformatics/btg015.
[108] N. Le Novère, M. Hucka, H. Mi, S. Moodie, F. Schreiber, A. Sorokin, et al., The systems biology graphical notation, Nat. Biotechnol. 27 (2009) 735–741, https://doi.org/10.1038/nbt.1558.
[109] C. Li, M. Donizelli, N. Rodriguez, H. Dharuri, L. Endler, V. Chelliah, et al., BioModels database: an enhanced, curated and annotated resource for published quantitative kinetic models, BMC Syst. Biol. 4 (2010) 92, https://doi.org/10.1186/1752-0509-4-92.
[110] N. Bonzanni, K.A. Feenstra, W. Fokkink, E. Krepska, What can formal methods bring to systems biology? International Symposium on Formal Methods, 16, 2009.
[111] B. Di Camillo, A. Carlon, F. Eduati, G.M. Toffolo, A rule-based model of insulin signalling pathway, BMC Syst. Biol. 10 (2016) 38, https://doi.org/10.1186/s12918-016-0281-4.
[112] C.D.A. Terfve, E.H. Wilkes, P. Casado, P.R. Cutillas, J. Saez-Rodriguez, Large-scale models of signal propagation in human cells derived from discovery phosphoproteomic data, Nat. Commun. 6 (2015) 8033, https://doi.org/10.1038/ncomms9033.
[113] V. Danos, C. Laneve, Formal molecular biology, Theor. Comput. Sci. 325 (2004) 69–110, https://doi.org/10.1016/j.tcs.2004.03.065.
[114] V. Danos, J. Feret, W. Fontana, R. Harmer, J. Krivine, Rule-based modelling of cellular signalling, in: L. Caires, V.T. Vasconcelos (Eds.), CONCUR 2007—Concurrency Theory, Springer Berlin Heidelberg, Berlin, Heidelberg 2007, pp. 17–41, https://doi.org/10.1007/978-3-540-74407-8_3.
[115] V. Danos, J. Feret, W. Fontana, R. Harmer, J. Krivine, Rule-based modelling and model perturbation, Transactions on Computational Systems Biology XI 2009, pp. 116–137.
[116] D.T. Gillespie, Stochastic simulation of chemical kinetics, Annu. Rev. Phys. Chem. 58 (2007) 35–55, https://doi.org/10.1146/annurev.physchem.58.032806.104637.
[117] A.G. Forbes, A. Burks, K. Lee, X. Li, P. Boutillier, J. Krivine, et al., Dynamic influence networks for rule-based models, IEEE Trans. Vis. Comput. Graph. 24 (2018) 184–194, https://doi.org/10.1109/TVCG.2017.2745280.
[118] D.P. Gamblin, S. van Kasteren, G.J.L. Bernardes, J.M. Chalker, N.J. Oldham, A.J. Fairbanks, et al., Chemical site-selective prenylation of proteins, Mol. Biosyst. 4 (2008) 558–561, https://doi.org/10.1039/b802199f.
[119] M.H. Wright, W.P. Heal, D.J. Mann, E.W. Tate, Protein myristoylation in health and disease, J. Chem. Biol. 3 (2010) 19–35, https://doi.org/10.1007/s12154-009-0032-8.
[120] D.K.G. Ma, C. Stolte, S. Kaur, M. Bain, S.I. O'Donoghue, Visual Analytics of Phosphorylation Time-Series Data on Insulin Response, AIP, 2013 185–196, https://doi.org/10.1063/1.4825010.
[121] B.F. Cravatt, G.M. Simon, J.R. Yates, The biological impact of mass-spectrometry-based proteomics, Nature 450 (2007) 991–1000, https://doi.org/10.1038/nature06525.
[122] M. Correll, D. Moritz, J. Heer, Value-Suppressing Uncertainty Palettes, Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems—CHI '18, ACM Press, New York, New York, USA 2018, pp. 1–11, https://doi.org/10.1145/3173574.3174216.
[123] N. Boukhelifa, A. Bezerianos, T. Isenberg, J. Fekete, Evaluating sketchiness as a visual variable for the depiction of qualitative uncertainty, IEEE Trans. Vis. Comput. Graph. 18 (2012) 2769–2778, https://doi.org/10.1109/TVCG.2012.220.
[124] K. Glass, M. Girvan, Annotation enrichment analysis: an alternative method for evaluating the functional properties of gene sets, Sci. Rep. 4 (2014) 4191, https://doi.org/10.1038/srep04191.
[125] H. Heberle, M.F. Carazzolle, G.P. Telles, G.V. Meirelles, R. Minghim, CellNetVis: a web tool for visualization of biological networks using force-directed layout constrained by cellular components, BMC Bioinf. 18 (2017) 395, https://doi.org/10.1186/s12859-017-1787-5.
[126] S. Frias, K. Bryan, F.S.L. Brinkman, D.J. Lynn, CerebralWeb: a Cytoscape.js plug-in to visualize networks stratified by subcellular localization, Database (Oxford) 2015 (2015) bav041, https://doi.org/10.1093/database/bav041.
[127] D. Holten, J.J. van Wijk, Force-directed edge bundling for graph visualization, Comput. Graphics Forum 28 (2009) 983–990, https://doi.org/10.1111/j.1467-8659.2009.01450.x.
[128] G. Michal, D. Schomburg (Eds.), Biochemical Pathways: An Atlas of Biochemistry and Molecular Biology, John Wiley & Sons, Inc., Hoboken, NJ, USA, 2012https://doi.org/10.1002/9781118657072.
[129] T. von Landesberger, A. Kuijper, T. Schreck, J. Kohlhammer, J.J. van Wijk, J.D. Fekete, et al., Visual analysis of large graphs: state-of-the-art and future research challenges, Comput. Graphics Forum 30 (2011) 1719–1749, https://doi.org/10.1111/j.1467-8659.2011.01898.x.
[130] O.-H. Kwon, T. Crnovrsanin, K.-L. Ma, What would a graph look like in this layout? A machine learning approach to large graph visualization, IEEE Trans. Vis. Comput. Graph. 24 (2018) 478–488, https://doi.org/10.1109/TVCG.2017.2743858.