PORTLAND PRESS

Review Article

Principles of phosphoproteomics and applications in cancer research

Luke Higgins[1], **Henry Gerdes**[1] and **Pedro R. Cutillas**[1,2,3]

[1]Cell Signaling and Proteomics Group, Centre for Genomics and Computational Biology, Barts Cancer Institute, Queen Mary University of London, London, U.K.; [2]Alan Turing Institute, The British Library, London, U.K.; [3]Digital Environment Research Institute, Queen Mary University of London, London, U.K.

Correspondence: Pedro R. Cutillas (p.cutillas@qmul.ac.uk)

OPEN ACCESS

Phosphorylation constitutes the most common and best-studied regulatory post-translational modification in biological systems and archetypal signalling pathways driven by protein and lipid kinases are disrupted in essentially all cancer types. Thus, the study of the phosphoproteome stands to provide unique biological information on signalling pathway activity and on kinase network circuitry that is not captured by genetic or transcriptomic technologies. Here, we discuss the methods and tools used in phosphoproteomics and highlight how this technique has been used, and can be used in the future, for cancer research. Challenges still exist in mass spectrometry phosphoproteomics and in the software required to provide biological information from these datasets. Nevertheless, improvements in mass spectrometers with enhanced scan rates, separation capabilities and sensitivity, in biochemical methods for sample preparation and in computational pipelines are enabling an increasingly deep analysis of the phosphoproteome, where previous bottlenecks in data acquisition, processing and interpretation are being relieved. These powerful hardware and algorithmic innovations are not only providing exciting new mechanistic insights into tumour biology, from where new drug targets may be derived, but are also leading to the discovery of phosphoproteins as mediators of drug sensitivity and resistance and as classifiers of disease subtypes. These studies are, therefore, uncovering phosphoproteins as a new generation of disruptive biomarkers to improve personalised anti-cancer therapies.

Introduction

Cancer is caused by mutations in genes that, in normal cells, regulate fundamental cell biological processes such as bioenergetic metabolism, apoptosis and lineage identity [1]. However, genetic mutations, copy number alterations and transcriptional over/under-expression, while important, are not the only mechanisms by which the biological activities of oncogenes and tumour suppressor genes are dysregulated in cancer cells [2]. Virtually all druggable oncogenic processes are controlled by signalling networks regulated by post-translational protein modifications (PTMs), the most common and best studied of which is the addition of a phosphate group to serine, threonine or tyrosine residues, although other amino acids can also be phosphorylated [3]. Protein phosphorylation, which is catalysed by protein kinases and opposed by protein phosphatases, cooperates with other regulatory mechanisms (e.g. second messengers and protein–protein interactions (PPIs)) and enzymes (e.g. GTPases, phosphatases and lipid kinases), which together regulate intracellular signalling fluxes and virtually all cell biological functions. In normal cells, these kinase-driven signalling pathways are activated by extracellular signals provided by environmental cues (e.g. growth factors and hormones), direct cell–cell interactions and by mechanical and other stresses. In cancer, these pathways may be aberrantly activated as a result of genetic and epigenetic alterations. In addition, the microenvironment is often corrupted with the infiltration of immune and aberrant mesenchymal cells [4,5]. This tumour-associated stroma activates kinase signalling pathways in neighbouring cancer cells [6,7] and

further influences their malignant phenotype which becomes imprinted through epigenetic mechanisms [8,9]. Thus heterotypic communication between cancer and stromal cells, which together make up a tumour, can lead to the constitutive activation of oncogenic signalling pathways in the absence of genetic alterations [6,7]. In immunologically 'cold' tumours such as prostate cancer, dysregulated kinase signalling can occur due to genetic and/or epigenetic alterations that activate oncogenic drivers, thus representing tangible therapeutic targets with potential for precision medicine approaches [10].

The range and heterogeneity of molecular mechanisms that can lead to oncogenic pathway activation make the study of the phosphoproteome critical for a comprehensive understanding of cancer biology. Indeed, since each phosphorylation is, by definition, the product of kinase activity, the phosphoproteome provides readouts of all kinase-driven pathway activities in a given tumour. Thus, phosphoproteomic data capture oncogenic pathway activation irrespective of the mechanism by which such activation occurred — genetic, epigenetic or micro-environmental [11].

By providing a more direct measurement of cell processes, phosphoproteomic characterisation is enabling the identification of new pharmacological anti-cancer targets. In addition, quantification of phosphosites is being used to identify disease subgroups not captured by genetic assays [12–14], and to investigate the mechanisms of intrinsic [15–17] and acquired [18–20] resistance to targeted drugs, thus potentially leading to more effective personalised anti-cancer therapies [14,20–28].

The importance that the phosphoproteome has to illuminate oncogenic signalling has spurred intense research into technological development for its characterisation. To highlight the magnitude of the challenge, it has been estimated that the human phosphoproteome contains 13 000 phosphoproteins and 230 000 phosphorylation sites [29], most of which are present at substoichiometric amounts and are thus undetectable by standard mass spectrometry proteomics methods (i.e. without phosphoenrichment). Another issue is that most of these sites are not annotated with functional information or with the kinase(s) that catalyse their incorporation on proteins. The study of the phosphoproteome is also obscured by prolific kinase interactivity and rapid signalling dynamics. To untangle this, improvements in scan rates in modern mass spectrometers and refined sample preparation techniques have increased the depth, throughput and profiling capabilities of mass spectrometry proteomics/phosphoproteomics [30]. This, together with the development of computational concepts for the analysis and biological interpretation of phosphoproteomics data, is currently leading to a boom in the use of this technique for cancer kinome profiling.

Here, we review the current state of phosphoproteomics in cancer research with a focus on mass spectrometry-based approaches, and on introducing core concepts to new users, including methodologies and challenges that the field still faces. Finally, we review phosphoproteomic studies that illustrate how this technique is contributing, and can contribute in the future, to our understanding of cancer cell signalling and biology.

Methods for phosphoproteomic interrogation of cancer
Mass spectrometry interrogation of the proteome/phosphoproteome

Whilst immunoassays, such as reverse phase protein arrays (RPPA), allow the detection of 100s of analytes, it is easy to recognise the exciting potential of untargeted mass spectrometry-based methods which allow the quantification of 10 000s of phosphorylation sites in an untargeted manner. Indeed, it is using these technologies we now appreciate the complexity and heterogeneity of proteomic/phosphoproteomic profiles and signalling circuitries, with specific variations noted between patients and even cells [31–33].

At the simplest level, mass spectrometry is the measurement of the mass-to-charge ratio (m/z) of ions in the gas phase. The acquired m/z spectra are used to infer the analyte constitution. For the study of the phosphoproteome/proteome 'shotgun' or 'bottom-up' proteomics is most conventionally used and currently provides the most comprehensive profiling approach. In brief, proteins/phosphoproteins are digested into peptides (some of which will be post-translationally modified), separated using chromatography and sampled into the mass spectrometer (Figure 1). Whilst mass spectrometry analysis of whole proteins is possible, a method named top-down proteomics, approaches for the analysis of undigested proteins are immature compared with bottom-up methods as these suffer from complications that arise from chromatography of native proteins, decreased sensitivity and less developed bioinformatics approaches [34].

Bottom-up phosphoproteomics requires the use of high resolution and high mass accuracy mass spectrometry provided by time of flight (TOF) or Orbitrap instruments. TOF mass analysers use the ion's deflection

Sample preparation for MS
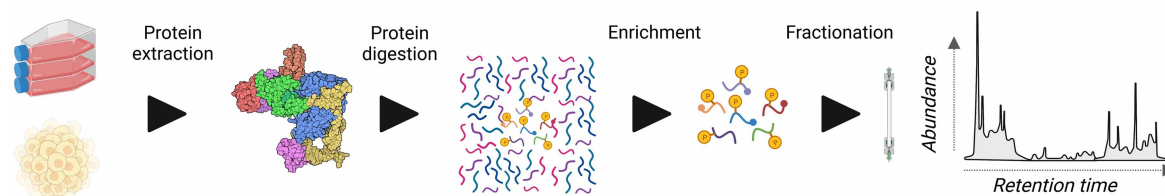
MS/MS data acquisition and analysis
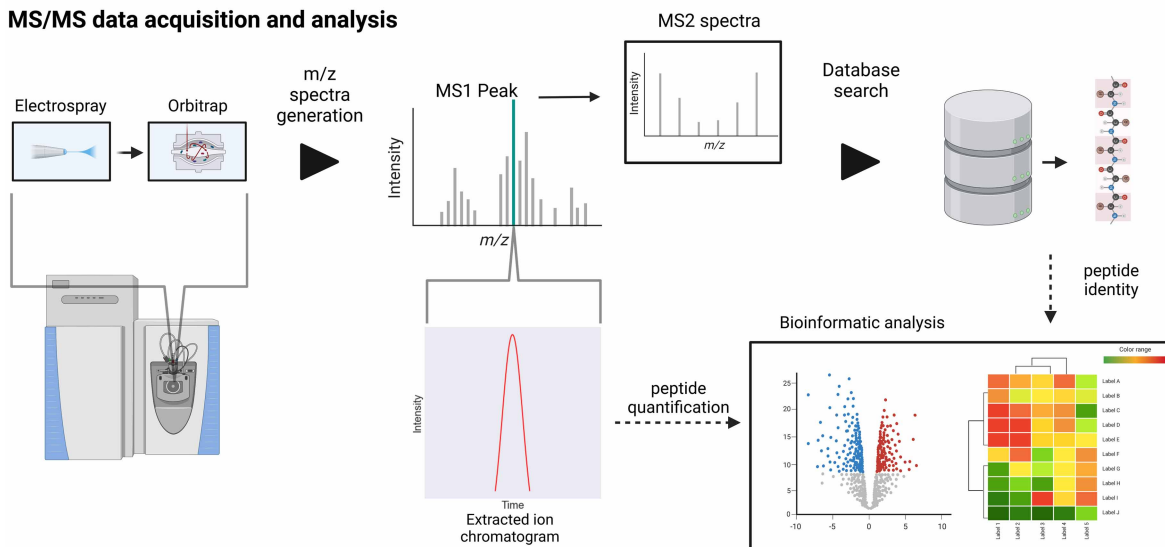
Figure 1. Liquid chromatography-tandem mass spectrometry (LC–MS/MS) workflow for data-dependent analysis (DDA) phosphoproteomics.

Upper panel, phosphoproteomic sample preparation. Proteins are extracted from samples through cell lysis and digested using proteases, commonly trypsin. Peptide fraction is subsequently enriched for phosphopeptides, conventionally using TiO_2, and separated using reversed-phase liquid chromatography before MS/MS analysis. Lower panel, MS/MS data acquisition and analysis. Following liquid chromatography, eluting peptides are ionised by electrospray and sampled by Orbitrap or TOF to generate MS1 spectra. Precursor ions are then selected and then fragmented. MS2 spectra are then generated from the second round of mass analysis. MS1 peaks can be isolated (extracted ion chromatogram) and used for quantification, whilst MS2 spectra enables peptide identification.

time as the informant of it mass. On the other hand, the Orbitrap mass analyser obtains ion m/z by tracking ion orbit frequency around the central electrode [35]. In proteomics/phosphoproteomics, these instruments are conventionally used with other ion optics that select and fragment peptide ions for tandem mass spectrometry (MS/MS or MS2) analysis. In this, a MS1 scan is used to obtain the mass of unfragmented peptides and for their quantification (in the case of conventional label-free analysis). Ions are then selected for subsequent fragmentation generating MS2 data used to infer the primary structure of peptides (Figure 1, lower panel). Peptide precursor ion *m/z* ratio in MS1 and MS2 fragment spectra allows for accurate identification of phosphopeptides.

Many fragmentation methods exist for MS2 spectra generation and the development of such tools still is an active area of research [36]. Collision-induced dissociation (CID), which applies gas-phase collision with neutral gas molecules to fragment gas-phase peptide ions, is the standard fragmentation method present in essentially all commercial tandem mass spectrometers. Therefore, CID has been widely adopted for phosphoproteomics research. Whilst in general efficient, 'neutral loss' (i.e. loss of phosphate in the phosphorylated peptide fragment ions during CID) can obscure precise identification of the phosphorylated site in cases where more than one phosphorylatable amino acid residue is present in the peptide's sequence [37]. To address this, inspirited by electron-capture dissociation (ECD) which preserve the site of modification [38], electron transfer dissociation (ETD) was developed and incorporated into tandem mass spectrometers based on ion traps. These

softer fragmentation methods are based on the transfer of electrons from radical anions to the polypeptide chain which as a consequence dissociate at the C–N bond [39,40]. ETD is also particularly useful for top- and middle-down proteomics used to characterise large proteins without or with just partial digestion [41,42] and the technique is particularly useful for the analysis of the histone PTMs [43]. ETD has been combined with higher-energy collision dissociation (HCD), a technique known as EThcD, by applying HCD following an electron transfer reaction [44], and this technique is particularly useful for the analysis of large proteins including membrane proteins [45].

Quantification from mass spectrometry data may be achieved using label-based or label-free approaches. Tandem mass tag (TMT) labelling uses isobaric labels consisting of heavy stable isotopes, and enables sample multiplexing which can expedite mass spectrometry analysis by quantifying reported ions in MS2 data [46]. Another labelling strategy consists of stable isotope labelling by/with amino acids in cell culture (SILAC) involving the culture of cells in media containing labelled amino acids [47]. A limitation of label-based approaches is that there is a limit to the number of samples and replicates that can be analysed in an experiment and that quantification is always relative to analytes present in other samples, thus limiting their utility for translational research and ultimately clinical implementation. Thus, label-free quantification has proven popular in phosphoproteomics, after showing that these approaches are effective for quantitative analysis of phosphopeptides [48–50], although a limitation has been the requirement of specialised data acquisition strategies and software.

Sample preparation strategies for label-free phosphoproteomics

Phosphoproteomic experiments require phosphopeptide enrichment methods to reduce complexity and maximise deconvolution whilst retaining the biological information held within the sample [51]. Thus, much effort in the field has strived towards optimised sample processing procedures which do not influence the phosphoproteome before analysis or compromise profiling depth [52].

As mentioned, shotgun proteomics requires proteins extracted from samples to undergo digestion into peptides using peptidases such as trypsin, but others are available and can yield different peptide pools. Logically, the choice of digestive enzyme will influence the resulting peptide pool due to its cleavage site specificity [30]. However, phosphorylated residues can impact possible cleavage sites and thus impact digestion and detectability due to peptide length and ionisation efficiency [53]. Online tools which predict protein cleavage are able to assist in the making this decision, such as PeptideCutter [54].

Following digestion, phosphopeptide enrichment steps are used to increase the proportion of phosphopeptides identified in an experiment. The most popular phosphopeptide isolation methods take advantage of the negative ionic charge of the phosphate group. Immobilised metal affinity chromatography (IMAC) uses an affinity-based phenomenon, where the negative charge of phosphate groups is exposed to metal cations in the form of chelators. With potentially decreased specificity in comparison to other affinity methods, its applicability can be broad [55]. Drawbacks of this method include the risk of phosphopeptide leaching and binding of acidic amino acids such as glutamic and aspartic acid [55].

Metal oxide affinity chromatography (MOAC) applies the same affinity-based property as IMAC. Here, phosphopeptides are extracted by ionic binding between metal oxides and the negatively charged phosphate groups. Most commonly, titanium dioxide (TiO_2) is the agent used for this approach, although, a range of metal oxides are available, Aryal and Ross demonstrated that between ZrO_2 and TiO_2, TiO_2 displayed better selectivity for the isolation of peptides with multiple phosphorylation sites [56] whereas ZrO_2 has been reported to perform better for single site phosphorylated peptides [57].

Whilst working through similar mechanisms, MOAC is now more commonly used. Furthermore, the use of MOAC columns has been demonstrated to be more stable [58] and more selective [59] than IMAC. Both techniques enrich for Serine, Threonine and Tyrosine phosphorylation sites at the ratios that they are present in cells (70 : 30 : 2, respectively). When the aim is to specifically analyse the phospho-tyrosine phosphoproteome, immunoprecipitation of peptides using anti-pTyr antibodies has been demonstrated to provide greater coverage of this particular sub-phosphoproteome [60].

For many phosphoproteomic workflows, the sample preparation procedure is more laborious than the assay setting the biological context of the experiment. Phosphoenrichment is a particularly laborious task if done manually. However, the development of robotic liquid handlers for this purpose essentially solves this issue and also help with the reproducibility and throughput of the technique [61,62].

Label-free Tandem mass spectrometry data acquisition

Peptide identification and quantification from MS data require the use of computationally demanding analytical pipelines. Traditionally, proteomic approaches involve acquiring MS2 spectra using data-dependent acquisition (DDA) methodologies which select the most abundant m/z peaks in MS1 spectra for fragmentation at a given time in the LC separation. The selection for MS2 is based on precursor ions' abundance ranking in MS1 spectra. Investigators will set the number of peptides which will be selected by DDA based on experimental considerations and the scan rates of the instrument that is being used — modern mass spectrometers have scan rates that range from 10 to 100 Hertz. These selected ions are then fragmented and the resulting MS2 spectra are used to identify the analyte with a database search that aims to match fragment spectra to the theoretical fragmentation of all proteins in databases or to spectral libraries [63–67]. This approach allows sensitive real-time detection of peptides/phosphopeptides. The MS1 scan dependency enables the acquisition of spectra of sufficient quality and quantity for deep phosphoproteome coverage [30,32,68,69].

Whilst a proven useful approach, a limitation of DDA includes a bias towards more abundant peptides due to preferential selection of intense ions for MS2 analysis, and variability between the peptides that are selected for fragmentation in runs which can arise from minute scan time variations and other stochastic processes (Figure 2). These issues with undersampling lead to missing peptide identifications, a problem that increases with sample complexity [70]. To solve this problem, peptides and phosphopeptides identified in at least one LC–MS/MS run can be searched in other runs at the MS1 level. This approach consists of assembling a database (or library) of peptides identified in a given experiment; the elution profiles (extracted ion chromatograms) are then obtained for each of these peptides across all LC–MS/MS runs that are compared in the experiment, thus significantly alleviating the problem of undersampling in DDA [48,71–73].

Another way of reducing the impact that undersampling in DDA has on proteomic and phosphoproteomics data reproducibility involves the use of data independent acquisition (DIA) analysis [74–76]. Advances in computational processing tools and processing power have overcome the challenges which hampered the widespread application of this technique. DIA has the potential to enhance MS data pipelines and broaden reliable analyte quantification, particularly when used in combination with cutting-edge mass spectrometers with increased duty cycles aided by ion mobility ion optics [77,78]. During DIA acquisition, ions are not fragmented in an intensity-selective manner, rather all precursor ions within a selected m/z range are fragmented thus removing bias towards more abundant ions and reducing the stochasticity of MS2 ion generation (Figure 2). This approach has been demonstrated to decrease missing data points in proteomics datasets, although this comes at the cost of reduced MS2 spectral interpretability [79,80] requiring more complex computational analysis to deconvolute MS2 spectra into peptide identities.

Initially, DIA was not widely adopted in phosphoproteomic studies due to the requirements of mass spectrometers with fast scan rates, which are now becoming available, and the computational strain that accompanied
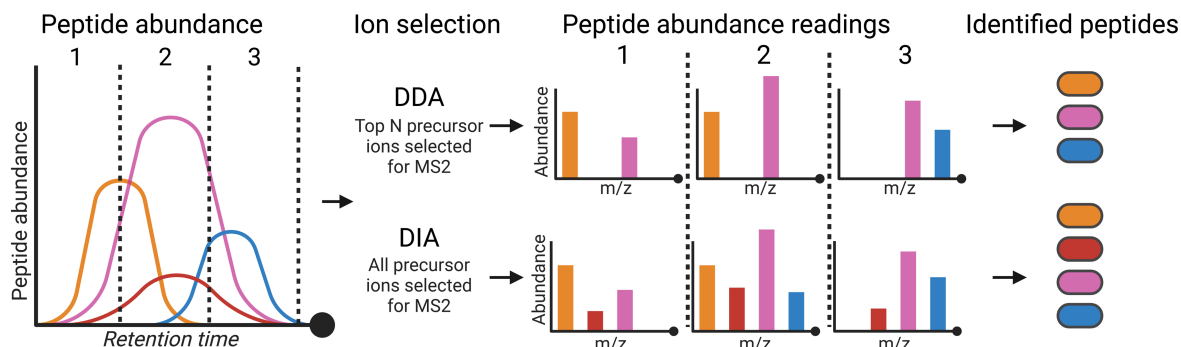
Figure 2. Overview of DIA and DDA approaches in proteomics.
Peptide abundance varies throughout the LC gradient and, as peptides elute from the column at specific time points, these are sampled by MS (represented as windows 1, 2 and 3 in the cartoon) and subsequently by MS/MS analysis. In data dependent acquisition (DDA) only the top N peptides (depending on instrument scan rates) are selected by abundance. This means that some peptides (e.g. the red ion in the cartoon) may be omitted from all windows and are not identified in DDA. This is contrary to data-independent analysis (DIA), where all peptide ions are fragmented and thus present in MS2 scans.

the disentanglement of more complex fractionation spectra due to the fact that there is no direct link between single MS1 peaks and MS2 spectra. Recent developments in software for DIA data analysis [81,82] are enabling widespread use of the technique in proteome research, but the use of DIA in phosphoproteomics is less well documented, although we expect the technique to become widespread for this purpose now that robust software has become available. However, since most phosphoproteomics experiments have been performed using DDA, we focus on studies that used this approach henceforth. We refer the reader to the work of Zhang et al. [83] for further discussion on DIA.

Assigning peptides to *m/z* spectra

Following data acquisition, the complex ion spectra must be resolved into peptide/phosphopeptide identities in order to identify proteins/phosphoproteins and phosphorylation sites. In database search methods, protein sequences in an appropriate database (e.g. Uniprot restricted to the organism analysed in the experiment) undergo *in silico* digestion and fragmentation to produce a reference of artificial peptide spectra, to which the experimental spectra is matched by a searching engine that also scores and ranks the identifications.

A widespread method for this is scoring against theoretical spectra using 'shared peak count'. Pioneering algorithms using such a strategy include SEQUEST [67] which is based on an XCorr score that serves as a measure of similarity between acquired and theoretical peptide spectra. A limitation of the method was the speed at which XCorr is calculated which became a bottleneck for MS pipelines. Subsequent efforts to increase processing speed have included dot product computation of XCorr [84]. Further efforts have been built on to the SEQUEST algorithm, such as TIDE, which takes advantage software engineering approaches to further boost speed [85].

Other algorithms for database search in proteomics include the open source X!Tanden and Protein Prospector. The latter was developed in 1995 at the University of California, San Francisco to exploit the additional specificity of identification that can be obtained by reducing the mass error windows used to match experimental with theoretical peptides [65]. It has since then undergone many iterations with additional features added to the software suite [86] and continues to be commonly used today [87]. A further more recent development is a database search method known as MSFragger which was designed with the aim of increasing processing speed. Similar to SEQUEST (resulting in the tool Crux) [88] the developers of MSFragger added a peptide database index to increase pipeline speed, however, in contrast to Crux which solely sorted peptides by mass, MSFragger additionally computes a theoretical fragment ion index, leading to improved efficiency compared with other tools [89].

In contrast to the algorithms described above, which use heuristic approaches, probabilistic algorithms, such as Mascot — widely considered the industry gold standard [66,90] — and Andromeda (search engine in MaxQuant) assign scores to peptides as the probability that the match between the experiment peptide spectra and theoretical spectra has occurred by chance. Mascot uses the MOWSE algorithm [91], whilst MaxQuant [92] offers flexibility allowing the researcher to tailor their analytical pipeline.

For phosphopeptides, there is the additional challenge of identifying the phosphosite location. These issues have been reviewed recently and therefore they will not be discussed here in depth [36]. Briefly, in cases where a peptide contains more than one serine, threonine or tyrosine residue in its sequence, confident phosphosite allocation requires 'site-determining ions' to be generated from fragmentation spectra [93], which may not be present as a result of neutral loss during CID MS/MS. Different phosphorylated residues can also have different fragmentation behaviours depending on peptide residues which different scoring algorithms can be biased to [94]. To address this, scoring methods were developed, and these include the mascot delta score [94] and Ascore [93], though a wide range of tools exist [36,95]. More recently, MS/MS spectra prediction and DIA have been proposed as means to increase the accuracy of site localisation in phosphoproteomics experiments [96,97].

Peptide verification

Proteomic identification results are adjusted for the probability of false discovery using a target-decoy method where results from searching artificial peptide sequences not found in a biological context (e.g. in a reversed protein database) are compared with those of the experimental (forward) search results. Characteristics of the ideal decoy sequence have been described by Gygi's group [98]. In brief, this relates to sharing characteristics with target sequences (residue distribution, similar length and numbers at both peptide and protein level, and importantly no shared peptide sequences). Many approaches exist for generating the decoy database, such as

peptide reversal and random sequence generation. Caveats to these include the non-randomness of sequence reversal, and in contrast, random sequence generation will unlikely maintain relevant peptide characteristics such as the presence of repeated residues [99].

It has been demonstrated that simple target-decoy strategies lead to reduction in sensitivity due to asymmetry in protein matches relating to false artificial proteins and false target protein hits, resulting from the fact that incorrect target peptide matches could occur from both target proteins and decoy proteins, in contrast to decoy peptides [100,101]. To address this, Savitski et al. [102] developed the picked-target decoy strategy (picked TDS). This approach simply compares the scores of both the target peptide match and the decoy peptide match and picks the highest-scoring match for that target peptide. Using this simple strategy, the over-estimation of false positives was reduced in the aforementioned study. Other efforts to address this issue include that of The et al. [101], who developed an approach known as the Picked Protein GroupFDR approach for protein group level FDR correction, resulting in 4% more protein identifications than the default approach in the MaxQuant pipeline [101].

A specific issue for phosphoproteomics is the estimation of false phosphorylation site localisation and the increase in search space that comes with including all the possible modifications on a given peptide in the search [95]. Efforts to address this include modified decoy search strategy, such as LuciPHOr, which uses each residue in the peptide as a phosphorylation site decoy, thus pitting candidate and non-candidate residues against one another [103]. Further work by Ramsbottom et al. [104] demonstrated that Alanine or Leucine residues make sensible decoy candidates.

Bioinformatic tools and resources for phosphoproteomic data interpretation

Phosphoproteomic analysis tools

A challenge in omics analysis is biological interpretation but the issue in phosphoproteomics is particularly intricate because of the impact that phosphorylation has on biology is often ambiguous — some sites increase the activity of the protein that is modified, but many are inhibitory while for most of them there is no functional information. To address this, two main computational approaches to derive biological insights from phosphoproteomics data are employed, which we term substrate-centric and kinase-centric. The substrate-centric approach involves considering the effects that changes in individual phosphorylation sites may have on the protein that receives the phosphorylation, whereas the kinase-centric method consists of deriving readouts of kinase activity by matching kinases to their substrates or downstream targets. Both methods may be combined to analyse a given dataset.

The substrate-centric approach needs information on the effect that a given phosphorylation site may have on the activity of the protein receiving these modifications. This is a challenge because, as discussed above, while some phosphorylation sites are well characterised, in most cases the biological function of phosphorylation is ambiguous. The low conservation of phosphorylation across species suggests that most sites are non-functional [105,106], but this view has been challenged [107]. Recently, machine learning was used to infer the phosphorylation sites likely to be functional [108], and progress is also being made in predicting the sign of regulation (i.e. whether the phosphorylation is activatory or inhibitory) [109].

As for the kinase-centric approach, kinase-substrate enrichment analysis (KSEA) and similar methods infer differences in kinase activities across samples by matching kinases to their substrates and quantifying their level of enrichment relative to background phosphorylation (by for example, z-score calculation and/or Kolmogorov–Smirnov testing) [17]. In the original study, KSEA-derived kinase activities showed high concordance to western blot data [17] and changed in cells treated with signalling perturbagens as expected. KSEA is based on the analysis of validated kinase-substrate relationships and thus provides a more reliable method of characterising kinase activity than investigating individual kinase phosphorylation sites [16–18,110] or approaches such as NetworKIN [111] which computationally predicts kinase substrates based on linear phosphorylation motifs.

Since the publication of KSEA other methods that use phosphoproteomics data to infer kinase activity have been published. These have been recently reviewed and will not be discussed here in detail [112]. All these methods require a knowledge of phosphorylation sites that are signatures of kinase activities. Therefore, the challenge that all these tools face is that the knowledge of kinase-phosphosite relationships is limited to the best-studied kinases, and for many of them there is a lack of knowledge of their downstream targets, thus

precluding their analysis by this method. To solve this problem, Hijazi et al. used the known promiscuity of kinase inhibitors to map kinase downstream targets of less well-studied kinases. Kinase inhibitors developed to inhibit a given kinase also inhibit unintended enzymes, some of which are poorly studied. This was exploited to identify the combination of compounds that represent specific inhibitor fingerprints for given kinases. Phosphoproteomics of cells treated with the set of inhibitors in the fingerprint then revealed phosphorylation sites downstream of kinases [24]. The study added 6 000 new kinase-phosphosite relationships that may be used as the source of KSEA or similar methods to infer kinase activities from phosphoproteomics data. A version of KSEA that incorporates this and other databases as the source of kinase-phosphosite relationships can be accessed from github.com/CutillasLab/KSEA_plus [21].

Contrary to KSEA, which compares kinase activities across samples, other methods focus on ranking kinase activities within a sample. This information may be used for selecting the best kinase inhibitor to treat a given patient. Two of these methods have been developed: Kinase activity ranking using phosphoproteomics data (KARP) [113] and the inferred kinase activity (INKA) algorithm developed by the Jimenez laboratory [114]. INKA estimates kinase activity as a function of total kinase phosphorylation, phosphorylation of the kinase's activation loop and substrate phosphorylation, thus combining kinase and substrate-centric approaches. Using INKA, Cordeo et al. [25] selected kinase inhibitors for the treatment of acute myeloid leukaemia (AML) cell lines, highlighting its potential to inform treatment combinations.

Empirical kinase network inference

Phosphoproteomic data have been used to reconstruct kinase networks. Methods for this purpose include the use of statistical frameworks combined with known substrate-kinase relationship data. An example is provided by work by the Saez-Rodriguez laboratory [115], who developed a method called Phonemes based on logic modelling of kinase inhibitor perturbation data through Gaussian mixture models for each phosphorylation site. This involved training models with a background network (built from known/predicted kinase-substrate relationships). Application of Phonemes identified CDK regulation by MTOR, which was verified experimentally. Using a complementary approach, Clarke et al. [116] developed a network-based tool to link expression signature to upstream kinase networks. Another strategy developed by Yilmaz et al. [117] uses a heterogeneous network model, which uses circuit-based propagation in an attempt to bridge missing network connections that can arise due to under sampling of the phosphoproteome, an inevitable issue due the large number of yet unknown kinase-phosphosite interactions [118] and under sampling of phosphorylation sites that can occur in MS experiments. Thus, this approach takes advantage of current network knowledge to fill in potential gaps in the kinome network under the assumption that biological events stemming from network action will originate from associated sites. However, this may introduce some bias into new explorations due to reliance on well-characterised signalling nodes. Other groups have focussed on using KSEA to provide a global analysis of kinase signalling. Ochoa et al. [119] compiled 41 studies including drug perturbation data for 399 kinases to track kinase–kinase relationships through monitoring the activity of a broad range of kinases in response to inhibitor treatment.

Current challenges of kinase-phosphosite mapping
Overcoming bias towards the known phosphoproteome

A major hurdle in phosphoproteomics is the low of coverage of known kinase-phosphosite interactions. Indeed, it is estimated that 87% of listed kinase substrates are phosphorylated by 20% of known kinases, a distribution that seems unlikely due to large number of shared substrates between kinases [118]. This bias is propagated by the fact that targeted functional assays are lengthy and time consuming, along with the problem of confirmation bias that can exist within scientific research which is present in phosphosite databases. To address this, Hijazi et al. [24] used a chemical phosphoproteomics approach to identify direct and indirect kinase-phosphosite relationships. More recently. Johnson et al. used positional scanning peptide array analysis (PSPA), consisting of *in vitro* assays with peptide pools and individual kinases and ATP, to determine motifs determinants of kinase-substrate specificity, primarily via the negative selection on the substrate's side. Furthermore, the authors found a third of the sites tested lacked such residues, and were therefore shared with a large number of kinases. This highlights substrate availability and other kinase regulatory mechanisms' importance in coordinating specific signalling responses [120].

To document kinase-substrate relationships, there are many dedicated databases for recording such information. One of the largest databases is PhosphositePlus [121], which consists of curated information on protein PTMs for human, mouse and rat phosphoproteomes. Another database known as Signor, is also generated from curated literature of phosphorylation site information, but is stored in a directed graph format, the goal being to associate PTM to effect/function [122]. Another recent and useful metadatabase for the analysis of phosphoproteomics data is Omnipath [123]. Kinase-phosphosite relationships identified by chemical phosphoproteomics are provided in the ChemPhoPro database [24]. These are invaluable resources for the community, and the incorporation of curated experimental data from the literature, as these become available in the future, will result in more accurate signalling network construction approaches based on phosphoproteomics data.

Studies that used perturbation data can be a good source of new kinase-phosphosite relationships and also reveal kinase-kinase connexions. Perturbagens may be siRNA [124] or kinase inhibitors [24]. Bodenmiller et al. performed a kinome/phosphatase-wide genetic screen in yeast, where they individually perturbed kinase/phosphatases in yeast prior to MS analysis. They demonstrated that the number of indirect phosphosites far outnumber the predicted direct phosphosites and the interconnectivity of signalling architectures. On the other hand, as discussed above, the kinase inhibitor screen by Hijazi et al. relied on the promiscuity of kinase inhibitors to identify downstream targets of poorly studied kinases. The advantage of siRNA is its specificity, but it has the drawback that it requires long treatments times, and therefore changes in gene expression also contribute to the observed changes in phosphorylation. This does not occur in cells treated with kinase inhibitors for short times, but these approaches need complex algorithms to link kinase inhibitor specificity with their effects in the phosphoproteome. Despite these caveats, these large-scale screens, combined with mass spectrometry-based phosphoproteomics could contribute to a more balanced annotation of less well-explored kinases in commonly used reference databases.

Data variability

One of the perceived drawbacks of phosphoproteomics data is that the high instability of phosphorylated peptides, due to their rapid degradation by phosphatases and high pH, contributes to noise and variability in such datasets. For phosphoproteomics data to be clinically meaningful, it has to be resilient to the highly dynamic nature of kinase signalling, which makes this technique susceptible to environmental changes. Responses to kinase inhibitors have been characterised minutes after treatment, in *in vitro* models [125] and in cancer cells left in ambient conditions, which resulted in the activation of stress and autophagy signalling pathways [126]. However, the activation of these pathways was reduced using simple measures (by placing cells on ice prior to the lysis procedure). In addition, phosphoproteomics experiments carried out in separate laboratories show differences in peptide coverage, in addition to intra-laboratory variation from different staff members. This highlights the reactivity of phosphorylation activity and the requirement for robust sample preparation protocols to create reliable datasets. However, these issues also greatly impact other omics methodologies such as single-cell RNA-Seq, RNA-Seq and CHIP-Seq [127–129].

The impact of batch effects has been greatly diminished in omics datasets due to improvements in their throughput, which allows characterising a wider range of biological conditions, improving signal-to-noise ratio and ultimately separating variation of biologically relevant changes from data noise [128,130]. Recent advances in automation have also allowed for the development of robust high throughput desalting and phosphoenrichment procedures, requiring minute protein amounts and allowing more samples to be processed in shorter times and improving reproducibility. These advancements in turn have made phosphoproteomics approaches more suited for use in translational and clinical investigations.

Applications of phosphoproteomics in cancer research

Phosphoproteomics has numerous uses in cancer research (examples are shown in Figure 3). Among other applications, it has been used to rationalise cancer phenotypes, as a readout of gene knock-out experiments, to investigate the mode of action of anti-cancer drugs, and as a source of biomarkers for precision medicine. Below we give examples of studies that illustrate the use of phosphoproteomics in these research areas.

Rationalisation of cancer phenotypes using phosphoproteomic data

The first application of phosphoproteomics followed by KSEA focused on the investigation of determinants of drug response phenotypes in AML cell lines and primary samples treated with PI3K inhibitors [17]. Contrary to expectations, cells with higher PI3K/MTOR pathway activity were not more sensitive to PI3K/MTOR dual
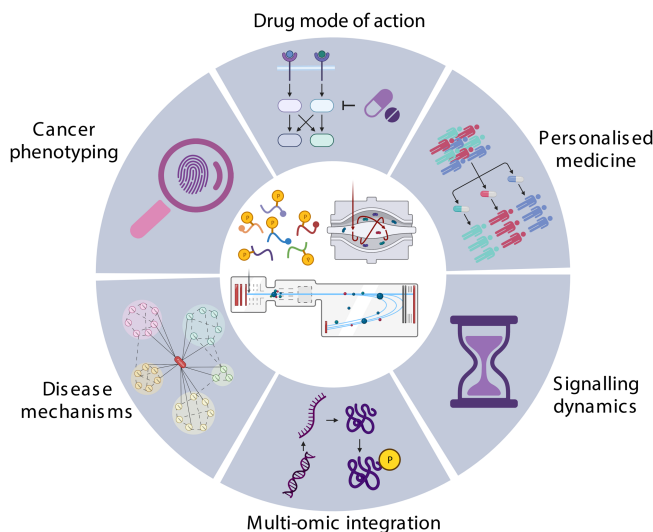
Figure 3. Application of phosphoproteomics in cancer research.

inhibitors than cells with lower activity of this pathway. Instead, activities of kinase in parallel pathways — namely PKC and ERK — were increased in resistant cells, showing that measuring the activity of compensatory pathways is needed to fully rationalise why some cancer cells respond to kinase inhibitors whereas others do not. This phenomenon was also observed in cells treated with trametinib, a MEK inhibitor. Although cells with high MEK/MAPK pathway activity responded better to trametinib than cells with low pathway activity, other compensatory pathways (including FLT3 mutations and STAT5 activation) were increased in cells that, despite activating MEK/MAPK, were resistant to inhibitors this pathway [14]. More recently, in a pan-cancer study, KSEA of phosphoproteomics data was used to estimate activities of 218 kinases across 1 110 tumours and 77 cell lines [131]. These data were matched to the activities of 292 transcription factors as well as genetic mutations and responses to therapy. An interesting observation was that overall kinase phosphorylation is, for some kinases, significantly associated with their extent of activation, as noted in a previous study [24], while TF activation is associated with their expression at the protein level. Of note, it was found that loss of function mutations was not on the whole associated with the activation of downstream signalling cascades, suggesting that, in most cases, feedback loops that suppress excessive pathway activation [132] are maintained in cancer cells. However, the activities of frequently dysregulated proteins were found to be associated with drug responses [131], thus suggesting that these could be used to determine prognosis. Taken together, these studies illustrate the complex nature of kinase signalling regulation, which underpin the idea that proxies of pathway activities (such as genetic mutation or protein expression) often fail to accurately reveal the extent of oncogenic pathways activation in cancer cells.

Several other studies have applied phosphoproteomics methods to investigate the mode of action of specific anti-cancer drugs. A study by Kazi et al. [133] applied phosphoproteomics analysis to KRAS mutant dependent and non-dependent pancreatic cancer cell lines to reveal a key role of CDK kinases in KRAS dependency. Following these results the authors validated their finding through mining CRISPR–cas9 dependency data, drug screens and western blot endpoints showed that targeting CDK hyperactivation may be an effective way of treating KRAS mutant pancreatic cancer. Similarly, Lee et al. [134] applied this approach to gauge the signalling shifts involved in the development of resistance against an EGFR/HER2 inhibitor (lapatinib) in human gastric cancer cells (SNU216 cells). Comparison between lapatinib resistant and parental SNU216 cells identified a MET-derived signalling cascade that increases the activity of the PI3K/AKT and MAPK/ERK signalling axes. Another example of phosphoproteomic interrogation is provided by Vye et al. [135], who used SILAC mass spectrometry to identify two distinct resistance mechanisms to both pazopanib and dasatinib in A204 cells, revealing that dasatinib-resistant cells upregulated insulin receptor/IGF-1R signalling as a compensatory pathway. An elegant study by Emdal et al. [19] extended this to primary AML samples where phosphoproteomic analysis of responders and non-responders to selinexor highlighted AKT as a combination target for resistant cells.

Phosphoproteomics to elucidate disease mechanisms

Phosphoproteomics is also a powerful tool to provide mechanistic insights into cancer cell biology and signalling in different tumour types. Examples of discoveries made by KSEA include the identification of protein kinase A as a mediator of an increasing cell metabolism in tumours with low FAK expression in cancer-associated fibroblasts [7], the identification of the MAPK-ROCK2 pathway being activated in cancer cells by loss of β-integrin in mural endothelial cells [136] and the discovery of MAPKAPK2 as a kinase downstream of MTOR that regulates the senescence secretome [137]. In a recent study, phosphoproteomics and KSEA showed that targeting the lysine demethylase LSD1 in AML cells activates the MEK/MAPK pathway, while inhibiting PI3K/AKT signalling. Consequently, LSD1 inhibition produced a signalling switch that primed cells for MEK treatment [138].

In a different study, Tamir et al. applied functional kinome screens for target stratification followed by phosphoproteomic interrogation. The authors performed a gain-of-function arrayed screen whereby 385 kinases and kinase-associated proteins where overexpressed in HEK-293T cells. Using this approach, an unexplored kinase BRSK1/2 showed a pivotal role in NFR2 transcription factor regulation, relevant in cancer and neurodegenerative disease. Subsequent analysis using TMT labelled phosphoproteomics of transfected BRSK1 and BRSK2 cells revealed an increased number of key phosphosites in BRSK2 transfected cells, suggesting a more central signalling role than its BRSK1 homologue [139].

Precision medicine

The phosphoproteome stands to provide a rich source of biomarkers of drug responses that can complement the information provided by genetics to advance the field of precision medicine. As an example of study, Liu et al. [140] were able to define molecular-based subtype profiles for cancer patients and signatures which were used to propose candidate drugs for more informed treatment. The volume and complex nature of biomarkers identified will require machine learning approaches to make sense of this wealth of data for advancing precision medicine and this is a highly active area of research. For example, Gerdes et al. [22] developed a machine-learning workflow which used phosphoproteomic data to rank the most effective treatments based on phosphoproteomic profiles. To this end, the authors used a novel feature selection strategy and an internal distance metric to narrow down selected markers, thereby accurately predicting drug responses in cell lines obtained from different lineages. Importantly, the models created by Gerdes et al. [141], also predicted drug responses in an independent study focusing on cholangiocarcinoma cell lines and primary samples. These proof-of-principle studies support the integration of machine learning-based precision medicine [142], although more work is needed for the confident roll out of these methods in the clinic. Despite the remaining challenges, it is clear that the phosphoproteome will be an invaluable source of information to guide the next generation of targeted and personalised therapies.

Unpacking kinome dynamics

Signalling is a dynamic process and perturbation with agonists or antagonists induces waves of kinase activation that are specific for given pathways and cell types. Thus, sampling time should be carefully considered as targeting distinct kinase signalling nodes can have temporally distinct influences on cell phenotypes [143]. Analysis techniques tailored for temporal analysis are being developed in attempt to increase time-series understanding of kinases [144]. As depicted in Figure 4, the importance of phosphorylation can be felt across the gene expression process illustrating complex and non-linear timelines in terms of relaying signals from/to the transcriptome/genome. For example, an in-depth view of the Drosophila phosphoproteome in combination with proteomics/transcriptomics, identified key kinases involved in the organism's circadian rhythm [145]. Understanding key checkpoint kinases could be used for targeting disease phenotypes. Alternatively, these temporal phosphoproteomics could pin point key signalling effectors in the acquisition of disease phenotypes and enable increasing understanding of how successful treatment leads to transition from a disease to disease-free state. As an example, an insightful study [146] used phosphoproteomics to compare kinase activation in HER2-postive gastric cancers before and after treatment with the HER2 antibody inhibitor trastuzumab. By comparing phosphoproteomic analysis of gastric cancer biopsies before and after treatment, the authors were able to gain insight into the impact of the HER2 antibody on the phosphoproteome enabling the discovery of new response biomarkers.
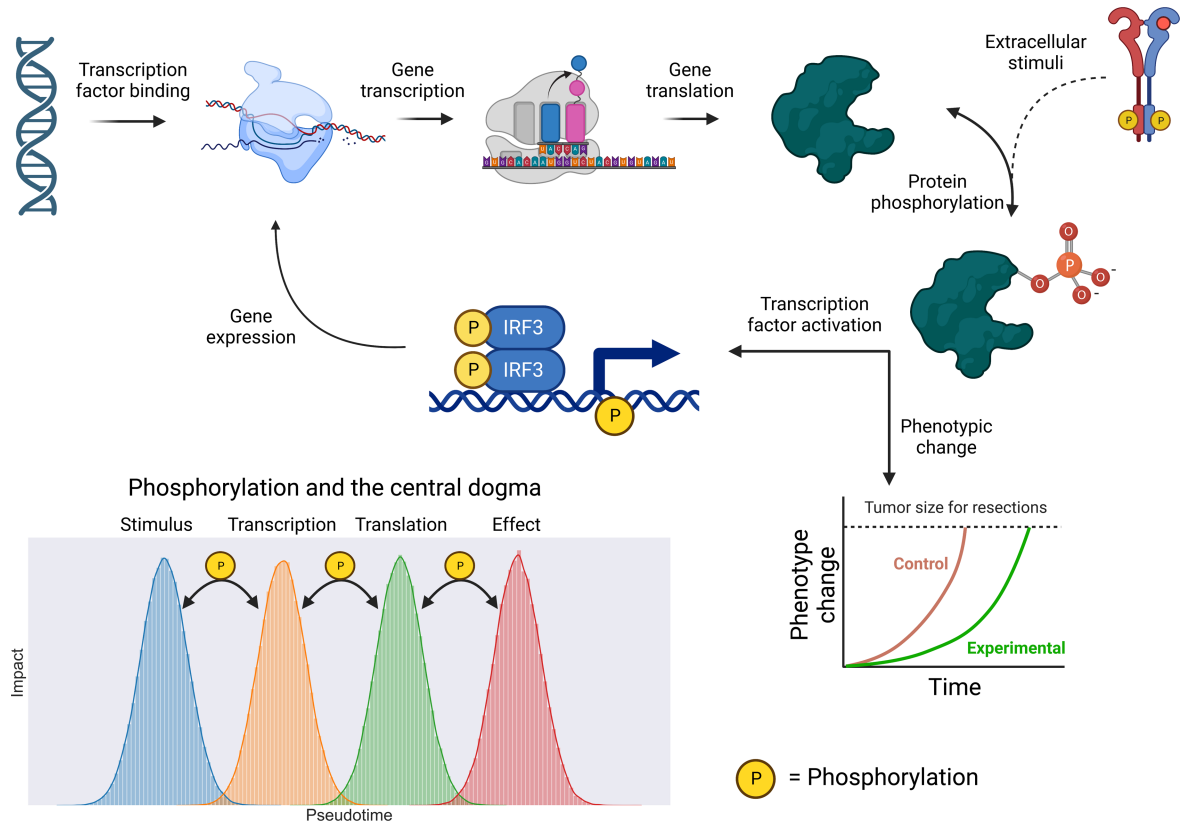
Figure 4. The Phosphoproteome acts as the cell's sensor, regulator and effector.

Protein phosphorylation has diverse impacts and roles in cell biology. Effects relayed by increase expression of exomic genes often yield protein products whose activity is regulated by phosphorylation. A prime example is transcription factors (here uses IFN regulatory factor 3 (IRF3) as an illustrative example), whereby phosphorylation can lead to increased expression of genes and drive as well as regulate changes in cell phenotype, which can occur due to external or internal stimuli. Changes in the phosphoproteome can originate from extracellular/intracellular cues and impact all stages of the central dogma in biology (lower left panel).

Phosphoproteomics' role in the omics landscape

Phosphoproteomics data have had an important role to play in the multi-omic profiling of primary clinical samples. A successful initiative known as the clinical proteomic tumour analysis consortium (CPTAC) was established in 2011 and since become a trove of proteomic and genomic data from primary tumour samples. These and other studies have provided invaluable datasets from different cancer types including breast [147], hepatocellular [148], ovarian [149], haematological [14,150], lung [151] and colorectal [152] among others. These studies are providing a wealth of information of disease subtypes and opportunities to identify new kinase drug targets [149]. As an example, Casado et al. [153] found a phosphoproteomic signature that defines two biologically distinct forms of KMT2A rearranged AML (also known as MLL-AML). This is significant because MLL-AML was considered to be a homogenous disease from a biological standpoint. Contrary to this view, phosphoproteomics data revealed two MLL-AML subgroups that differ in the phosphorylation of several proteins involved in KMT2A-mediated regulation of gene expression, including DOT1L. These differences in the biochemistry of MLL-AML cells translated to differences in responses to genotoxic and other drugs, suggesting that the identified signature may be used to select patients for therapy.

Integration of multi-omic data is key for making the most of such complex datasets but, until recently, software for this purpose was not available and researchers have used simple approaches such as the correlation between protein and RNA expression [154]. In response to this, computational tools have been developed to integrate different omics datasets. One such tool includes MiNETi [155], a computational pipeline that

integrates proteomics, phosphoproteomics and transcriptomics to construct a multi-network spanning graph. This is achieved by sequential network updates starting with a proteomics-based PPI network, updated with phosphorylation network data and finally transcriptional networks [155]. Other multi-omic analysis tools include COSMOS, which also connects transcription factors and kinase activity. COSMOS performs this by constructing prior knowledge network spanning the multi-omic space [156].

Conclusions

There is a clear role for phosphoproteomics to play in the interrogation of oncogenic mechanisms and for drug development. Improvements in technology offer new opportunities to derive insights into cellular signalling and regulatory biology with unprecedented depth, accentuating the interwoven nature of signalling mechanisms. However, challenges remain in the field of phosphoproteomics. For example, research resources for phosphoproteomics lag behind those of nucleotide-based 'omics' due to the highly specialised nature of the technology and the need for expensive equipment, and the potential to translate phosphoproteomics biomarkers into clinical assays has not yet been demonstrated. In the future, further advances will come from improvements in mass spectrometers with enhanced sensitivity and shorter duty cycles that allow the application of new forms of data analysis (e.g. DIA) and from the development of computational methods to better interrogate phosphoproteomics data in the context of kinase-phosphosite relationships and the biological function of the phosphorylation sites that can be profiled with this technique. Together, these developments are enabling the analysis of the phosphoproteome as a means to tailor drug treatments to patients, for drug development and for the elucidation of mechanisms that govern normal and disease biology.

Competing Interests

The authors declare that there are no competing interests associated with the manuscript.

CRediT Author Contribution

Pedro R. Cutillas: Conceptualisation, Supervision, Funding acquisition, Investigation, Writing — original draft, Writing — review and editing. **Luke Higgins**: Investigation, Visualisation, Writing — original draft, Writing — review and editing. **Henry Gerdes**: Investigation, Visualisation, Writing — review and editing.

Funding

Acknowledgements

Abbreviations

AML, acute myeloid leukaemia; CID, collision-induced dissociation; DDA, data-dependent acquisition; DIA, data-independent acquisition; ETD, electron transfer dissociation; HCD, higher-energy collision dissociation; IMAC, immobilised metal affinity chromatography; INKA, inferred kinase activity; KSEA, kinase-substrate enrichment analysis; MOAC, metal oxide affinity chromatography; PPI, protein–protein interaction; PTMs, post-translational protein modifications; SILAC, stable isotope labelling by/with amino acids in cell culture; TiO_2, titanium dioxide; TMT, tandem mass tag; TOF, time of flight.

References

1   Hanahan, D. and Weinberg, R.A. (2011) Hallmarks of cancer: the next generation. *Cell* **144**, 646–674 https://doi.org/10.1016/j.cell.2011.02.013
2   Marine, J.C., Dawson, S.J. and Dawson, M.A. (2020) Non-genetic mechanisms of therapeutic resistance in cancer. *Nat. Rev. Cancer* **20**, 743–756 https://doi.org/10.1038/s41568-020-00302-4
3   Hardman, G., Perkins, S., Brownridge, P.J., Clarke, C.J., Byrne, D.P., Campbell, A.E. et al. (2019) Strong anion exchange-mediated phosphoproteomics reveals extensive human non-canonical phosphorylation. *EMBO J.* **38**, e100847 https://doi.org/10.15252/embj.2018100847
4   Joshi, S.K., Nechiporuk, T., Bottomly, D., Piehowski, P.D., Reisz, J.A., Pittsenbarger, J. et al. (2021) The AML microenvironment catalyzes a stepwise evolution to gilteritinib resistance. *Cancer Cell* **39**, 999–1014.e8 https://doi.org/10.1016/j.ccell.2021.06.003
5   Vermeulen, L., De Sousa, E.M.F., van der Heijden, M., Cameron, K., de Jong, J.H., Borovski, T. et al. (2010) Wnt activity defines colon cancer stem cells and is regulated by the microenvironment. *Nat. Cell Biol.* **12**, 468–476 https://doi.org/10.1038/ncb2048

6   Straussman, R., Morikawa, T., Shee, K., Barzily-Rokni, M., Qian, Z.R., Du, J. et al. (2012) Tumour micro-environment elicits innate resistance to RAF inhibitors through HGF secretion. *Nature* **487**, 500–504 https://doi.org/10.1038/nature11183

7   Demircioglu, F., Wang, J., Candido, J., Costa, A.S.H., Casado, P., de Luxan Delgado, B. et al. (2020) Cancer associated fibroblast FAK regulates malignant cell metabolism. *Nat. Commun.* **11**, 1290 https://doi.org/10.1038/s41467-020-15104-3

8   Boulding, T., McCuaig, R.D., Tan, A., Hardy, K., Wu, F., Dunn, J. et al. (2018) LSD1 activation promotes inducible EMT programs and modulates the tumour microenvironment in breast cancer. *Sci. Rep.* **8**, 73 https://doi.org/10.1038/s41598-017-17913-x

9   Spangle, J.M., Roberts, T.M. and Zhao, J.J. (2017) The emerging role of PI3K/AKT-mediated epigenetic regulation in cancer. *Biochim. Biophys. Acta Rev. Cancer* **1868**, 123–131 https://doi.org/10.1016/j.bbcan.2017.03.002

10   Drake, J.M., Paull, E.O., Graham, N.A., Lee, J.K., Smith, B.A., Titz, B. et al. (2016) Phosphoproteome integration reveals patient-specific networks in prostate cancer. *Cell* **166**, 1041–1054 https://doi.org/10.1016/j.cell.2016.07.007

11   Cutillas, P.R. (2015) Role of phosphoproteomics in the development of personalized cancer therapies. *Proteom. Clin. Appl.* **9**, 383–395 https://doi.org/10.1002/prca.201400104

12   Casado, P., Rio-Machin, A., Miettinen, J.J., Bewicke-Copley, F., Rouault-Pierre, K., Krizsan, S. et al. (2023) Integrative phosphoproteomics defines two biologically distinct groups of KMT2A rearranged acute myeloid leukaemia with different drug response phenotypes. *Sig. Transduct. Target. Ther.* **8**, 80 https://doi.org/10.1038/s41392-022-01288-1

13   Liu, W., Xie, L., He, Y.H., Wu, Z.Y., Liu, L.X., Bai, X.F. et al. (2021) Large-scale and high-resolution mass spectrometry-based proteomics profiling defines molecular subtypes of esophageal cancer for therapeutic targeting. *Nat. Commun.* **12**, 4961 https://doi.org/10.1038/s41467-021-25202-5

14   Casado, P., Wilkes, E.H., Miraki-Moud, F., Hadi, M.M., Rio-Machin, A., Rajeeve, V. et al. (2018) Proteomic and genomic integration identifies kinase and differentiation determinants of kinase inhibitor sensitivity in leukemia cells. *Leukemia* **32**, 1818–1822 https://doi.org/10.1038/s41375-018-0032-1

15   Alcolea, M.P., Casado, P., Rodriguez-Prados, J.C., Vanhaesebroeck, B. and Cutillas, P.R. (2012) Phosphoproteomic analysis of leukemia cells under basal and drug-treated conditions identifies markers of kinase pathway activation and mechanisms of resistance. *Mol. Cell. Proteom.* **11**, 453–466 https://doi.org/10.1074/mcp.M112.017483

16   Casado, P., Alcolea, M.P., Iorio, F., Rodriguez-Prados, J.C., Vanhaesebroeck, B., Saez-Rodriguez, J. et al. (2013) Phosphoproteomics data classify hematological cancer cell lines according to tumor type and sensitivity to kinase inhibitors. *Genome Biol.* **14**, R37 https://doi.org/10.1186/gb-2013-14-4-r37

17   Casado, P., Rodriguez-Prados, J.C., Cosulich, S.C., Guichard, S., Vanhaesebroeck, B., Joel, S. et al. (2013) Kinase-substrate enrichment analysis provides insights into the heterogeneity of signaling pathway activation in leukemia cells. *Sci. Signal.* **6**, rs6 https://doi.org/10.1126/scisignal.2003573

18   Wilkes, E.H., Terfve, C., Gribben, J.G., Saez-Rodriguez, J. and Cutillas, P.R. (2015) Empirical inference of circuitry and plasticity in a kinase signaling network. *Proc. Natl Acad. Sci. U.S.A.* **112**, 7719–7724 https://doi.org/10.1073/pnas.1423344112

19   Emdal, K.B., Palacio-Escat, N., Wigerup, C., Eguchi, A., Nilsson, H., Bekker-Jensen, D.B. et al. (2022) Phosphoproteomics of primary AML patient samples reveals rationale for AKT combination therapy and p53 context to overcome selinexor resistance. *Cell Rep.* **40**, 111177 https://doi.org/10.1016/j.celrep.2022.111177

20   Dermit, M., Casado, P., Rajeeve, V., Wilkes, E.H., Foxler, D.E., Campbell, H. et al. (2017) Oxidative stress downstream of mTORC1 but not AKT causes a proliferative defect in cancer cells resistant to PI3K inhibition. *Oncogene* **36**, 2762–2774 https://doi.org/10.1038/onc.2016.435

21   Casado, P., Hijazi, M., Gerdes, H. and Cutillas, P.R. (2022) Implementation of clinical phosphoproteomics and proteomics for personalized medicine. *Methods Mol. Biol.* **2420**, 87–106 https://doi.org/10.1007/978-1-0716-1936-0_8

22   Gerdes, H., Casado, P., Dokal, A., Hijazi, M., Akhtar, N., Osuntola, R. et al. (2021) Drug ranking using machine learning systematically predicts the efficacy of anti-cancer drugs. *Nat. Commun.* **12**, 1850 https://doi.org/10.1038/s41467-021-22170-8

23   Heward, J., Konali, L., D'Avola, A., Close, K., Yeomans, A., Philpott, M. et al. (2021) KDM5 inhibition offers a novel therapeutic strategy for the treatment of KMT2D mutant lymphomas. *Blood* **138**, 370–381 https://doi.org/10.1182/blood.2020008743

24   Hijazi, M., Smith, R., Rajeeve, V., Bessant, C. and Cutillas, P.R. (2020) Reconstructing kinase network topologies from phosphoproteomics data reveals cancer-associated rewiring. *Nat. Biotechnol.* **38**, 493–502 https://doi.org/10.1038/s41587-019-0391-9

25   Cordo, V., Meijer, M.T., Hagelaar, R., de Goeij-de Haas, R.R., Poort, V.M., Henneman, A.A. et al. (2022) Phosphoproteomic profiling of T cell acute lymphoblastic leukemia reveals targetable kinases and combination treatment strategies. *Nat. Commun.* **13**, 1048 https://doi.org/10.1038/s41467-022-28682-1

26   Cucchi, D.G.J., Van Alphen, C., Zweegman, S., Van Kuijk, B., Kwidama, Z.J., Al Hinai, A. et al. (2021) Phosphoproteomic characterization of primary AML samples and relevance for response toward FLT3-inhibitors. *Hemasphere* **5**, e606 https://doi.org/10.1097/hs9.0000000000000606

27   van Alphen, C., Cloos, J., Beekhof, R., Cucchi, D.G.J., Piersma, S.R., Knol, J.C. et al. (2020) Phosphotyrosine-based phosphoproteomics for target identification and drug response prediction in AML cell lines. *Mol. Cell. Proteom.* **19**, 884–899 https://doi.org/10.1074/mcp.RA119.001504

28   Zagorac, I., Fernandez-Gaitero, S., Penning, R., Post, H., Bueno, M.J., Mouron, S. et al. (2018) In vivo phosphoproteomics reveals kinase activity profiles that predict treatment outcome in triple-negative breast cancer. *Nat. Commun.* **9**, 3501 https://doi.org/10.1038/s41467-018-05742-z

29   Vlastaridis, P., Kyriakidou, P., Chaliotis, A., Van de Peer, Y., Oliver, S.G. and Amoutzias, G.D. (2017) Estimating the total number of phosphoproteins and phosphorylation sites in eukaryotic proteomes. *Gigascience* **6**, 1–11 https://doi.org/10.1093/gigascience/giw015

30   Riley, N.M. and Coon, J.J. (2016) Phosphoproteomics in the age of rapid and deep proteome profiling. *Anal. Chem.* **88**, 74–94 https://doi.org/10.1021/acs.analchem.5b04123

31   Nita-Lazar, A., Saito-Benz, H. and White, F.M. (2008) Quantitative phosphoproteomics by mass spectrometry: past, present, and future. *Proteomics* **8**, 4433–4443 https://doi.org/10.1002/pmic.200800231

32   Harsha, H.C. and Pandey, A. (2010) Phosphoproteomics in cancer. *Mol. Oncol.* **4**, 482–495 https://doi.org/10.1016/j.molonc.2010.09.004

33   Cutillas, P.R. and Jorgensen, C. (2011) Biological signalling activity measurements using mass spectrometry. *Biochem. J.* **434**, 189–199 https://doi.org/10.1042/BJ20101974

34   Armirotti, A. and Damonte, G. (2010) Achievements and perspectives of top-down proteomics. *Proteomics* **10**, 3566–3576 https://doi.org/10.1002/pmic.201000245

35   Hu, Q., Noll, R.J., Li, H., Makarov, A., Hardman, M. and Graham Cooks, R. (2005) The Orbitrap: a new mass spectrometer. *J. Mass Spectrom.* **40**, 430–443 https://doi.org/10.1002/jms.856

36 Potel, C.M., Lemeer, S. and Heck, A.J.R. (2019) Phosphopeptide fragmentation and site localization by mass spectrometry: an update. *Anal. Chem.* **91**, 126–141 https://doi.org/10.1021/acs.analchem.8b04746

37 Boersema, P.J., Mohammed, S. and Heck, A.J. (2009) Phosphopeptide fragmentation and analysis by mass spectrometry. *J. Mass Spectrom.* **44**, 861–878 https://doi.org/10.1002/jms.1599

38 Shi, S.D., Hemling, M.E., Carr, S.A., Horn, D.M., Lindh, I. and McLafferty, F.W. (2001) Phosphopeptide/phosphoprotein mapping by electron capture dissociation mass spectrometry. *Anal. Chem.* **73**, 19–22 https://doi.org/10.1021/ac000703z

39 Z, F.W. and M, R.K.N. (1998) Electron capture dissociation of multiply charged protein cations. *J. Am. Chem. Soc.* **120**, 3265–3266 https://doi.org/10.1021/ja973478k

40 Syka, J.E.P., Coon, J.J., Schroeder, M.J., Shabanowitz, J. and Hunt, D.F. (2004) Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proc. Natl Acad. Sci. U.S.A.* **101**, 9528–9533 https://doi.org/10.1073/pnas.0402700101

41 Catherman, A.D., Skinner, O.S. and Kelleher, N.L. (2014) Top down proteomics: facts and perspectives. *Biochem. Biophys. Res. Commun.* **445**, 683–693 https://doi.org/10.1016/j.bbrc.2014.02.041

42 Riley, N.M., Sikora, J.W., Seckler, H.S., Greer, J.B., Fellers, R.T., LeDuc, R.D. et al. (2018) The value of activated ion electron transfer dissociation for high-throughput top-down characterization of intact proteins. *Anal. Chem.* **90**, 8553–8560 https://doi.org/10.1021/acs.analchem.8b01638

43 Lu, C., Coradin, M., Porter, E.G. and Garcia, B.A. (2021) Accelerating the field of epigenetic histone modification through mass spectrometry-based approaches. *Mol. Cell. Proteom.* **20**, 100006 https://doi.org/10.1074/mcp.R120.002257

44 Frese, C.K., Altelaar, A.F., van den Toorn, H., Nolting, D., Griep-Raming, J., Heck, A.J. et al. (2012) Toward full peptide sequence coverage by dual fragmentation combining electron-transfer and higher-energy collision dissociation tandem mass spectrometry. *Anal. Chem.* **84**, 9668–9673 https://doi.org/10.1021/ac3025366

45 Cohn, W., Huguet, R., Zabrouskov, V. and Whitelegge, J. (2021) Dissociation strategies to maximize coverage of alpha-helical domains in top-down mass spectrometry of integral membrane proteins. *J. Am. Soc. Mass Spectrom.* **32**, 1380–1387 https://doi.org/10.1021/jasms.1c00031

46 Thompson, A., Schäfer, J., Kuhn, K., Kienle, S., Schwarz, J., Schmidt, G. et al. (2003) Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. *Anal. Chem.* **75**, 1895–1904 https://doi.org/10.1021/ac0262560

47 Ong, S.E., Blagoev, B., Kratchmarova, I., Kristensen, D.B., Steen, H., Pandey, A. et al. (2002) Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol. Cell. Proteom.* **1**, 376–386 https://doi.org/10.1074/mcp.m200025-mcp200

48 Cutillas, P.R., Geering, B., Waterfield, M.D. and Vanhaesebroeck, B. (2005) Quantification of gel-separated proteins and their phosphorylation sites by LC-MS using unlabeled internal standards: analysis of phosphoprotein dynamics in a B cell lymphoma cell line. *Mol. Cell. Proteom.* **4**, 1038–1051 https://doi.org/10.1074/mcp.M500078-MCP200

49 Montoya, A., Beltran, L., Casado, P., Rodriguez-Prados, J.C. and Cutillas, P.R. (2011) Characterization of a TiO(2) enrichment method for label-free quantitative phosphoproteomics. *Methods* **54**, 370–378 https://doi.org/10.1016/j.ymeth.2011.02.004

50 Casado, P. and Cutillas, P.R. (2011) A self-validating quantitative mass spectrometry method for assessing the accuracy of high-content phosphoproteomic experiments. *Mol. Cell. Proteom.* **10**, M110 003079 https://doi.org/10.1074/mcp.M110.003079

51 Urban, J. (2022) A review on recent trends in the phosphoproteomics workflow. From sample preparation to data analysis. *Anal. Chim. Acta* **1199**, 338857 https://doi.org/10.1016/j.aca.2021.338857

52 Low, T.Y., Mohtar, M.A., Lee, P.Y., Omar, N., Zhou, H. and Ye, M. (2021) Widening the bottleneck of phosphoproteomics: evolving strategies for phosphopeptide enrichment. *Mass Spectrom. Rev.* **40**, 309–333 https://doi.org/10.1002/mas.21636

53 Beltran, L., Casado, P., Rodriguez-Prados, J.C. and Cutillas, P.R. (2012) Global profiling of protein kinase activities in cancer cells by mass spectrometry. *J. Proteom.* **77**, 492–503 https://doi.org/10.1016/j.jprot.2012.09.029

54 Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S.E., Wilkins, M. R., Appel, R. D. et al. (2005) Protein identification and analysis tools on the ExPASy server. In *The Proteomics Protocols Handbook* (Walker, J. M., ed.), pp. 571–607, Humana Press, Totowa, NJ

55 Dunn, J.D., Reid, G.E. and Bruening, M.L. (2010) Techniques for phosphopeptide enrichment prior to analysis by mass spectrometry. *Mass Spectrom. Rev.* **29**, 29–54 https://doi.org/10.1002/mas.20219

56 Aryal, U.K. and Ross, A.R. (2010) Enrichment and analysis of phosphopeptides under different experimental conditions using titanium dioxide affinity chromatography and mass spectrometry. *Rapid Commun. Mass Spectrom.* **24**, 219–231 https://doi.org/10.1002/rcm.4377

57 Kweon, H.K. and Håkansson, K. (2006) Selective zirconium dioxide-based enrichment of phosphorylated peptides for mass spectrometric analysis. *Anal. Chem.* **78**, 1743–1749 https://doi.org/10.1021/ac0522355

58 Negroni, L., Claverol, S., Rosenbaum, J., Chevet, E., Bonneu, M. and Schmitter, J.M. (2012) Comparison of IMAC and MOAC for phosphopeptide enrichment by column chromatography. *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.* **891–892**, 109–112 https://doi.org/10.1016/j.jchromb.2012.02.028

59 Larsen, M.R., Thingholm, T.E., Jensen, O.N., Roepstorff, P. and Jorgensen, T.J. (2005) Highly selective enrichment of phosphorylated peptides from peptide mixtures using titanium dioxide microcolumns. *Mol. Cell. Proteom.* **4**, 873–886 https://doi.org/10.1074/mcp.T500007-MCP200

60 Rikova, K., Guo, A., Zeng, Q., Possemato, A., Yu, J., Haack, H. et al. (2007) Global survey of phosphotyrosine signaling identifies oncogenic kinases in lung cancer. *Cell* **131**, 1190–1203 https://doi.org/10.1016/j.cell.2007.11.025

61 Murillo, J.R., Kuras, M., Rezeli, M., Miliotis, T., Betancourt, L. and Marko-Varga, G. (2018) Automated phosphopeptide enrichment from minute quantities of frozen malignant melanoma tissue. *PLoS One* **13**, e0208562 https://doi.org/10.1371/journal.pone.0208562

62 Birk, M.S., Charpentier, E. and Frese, C.K. (2021) Automated phosphopeptide enrichment for gram-positive bacteria. *J. Proteome Res.* **20**, 4886–4892 https://doi.org/10.1021/acs.jproteome.1c00364

63 Chalkley, R.J., Baker, P.R., Medzihradszky, K.F., Lynn, A.J. and Burlingame, A.L. (2008) In-depth analysis of tandem mass spectrometry data from disparate instrument types. *Mol. Cell. Proteom.* **7**, 2386–2398 https://doi.org/10.1074/mcp.M800021-MCP200

64 Yates, J.R., Eng, J.K., Clauser, K.R. and Burlingame, A.L. (1996) Search of sequence databases with uninterpreted high-energy collision-induced dissociation spectra of peptides. *J. Am. Soc. Mass Spectrom.* **7**, 1089–1098 https://doi.org/10.1016/S1044-0305(96)00079-7

65 Clauser, K.R., Baker, P. and Burlingame, A.L. (1999) Role of accurate mass measurement (+/- 10 ppm) in protein identification strategies employing MS or MS/MS and database searching. *Anal. Chem.* **71**, 2871–2882 https://doi.org/10.1021/ac9810516

66 Perkins, D.N., Pappin, D.J., Creasy, D.M. and Cottrell, J.S. (1999) Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* **20**, 3551–3567 https://doi.org/10.1002/(SICI)1522-2683(19991201)20:18<3551::AID-ELPS3551>3.0.CO;2-2

67 Eng, J.K., McCormack, A.L. and Yates, J.R. (1994) An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J. Am. Soc. Mass Spectrom.* **5**, 976–989 https://doi.org/10.1016/1044-0305(94)80016-2

68 Ficarro, S.B., McCleland, M.L., Stukenberg, P.T., Burke, D.J., Ross, M.M., Shabanowitz, J. et al. (2002) Phosphoproteome analysis by mass spectrometry and its application to *Saccharomyces cerevisiae*. *Nat. Biotechnol.* **20**, 301–305 https://doi.org/10.1038/nbt0302-301

69 Amanchy, R., Periaswamy, B., Mathivanan, S., Reddy, R., Tattikota, S.G. and Pandey, A. (2007) A curated compendium of phosphorylation motifs. *Nat. Biotechnol.* **25**, 285–286 https://doi.org/10.1038/nbt0307-285

70 Michalski, A., Cox, J. and Mann, M. (2011) More than 100,000 detectable peptide species elute in single shotgun proteomics runs but the majority is inaccessible to data-dependent LC-MS/MS. *J. Proteome Res.* **10**, 1785–1793 https://doi.org/10.1021/pr101060v

71 Cutillas, P.R. and Vanhaesebroeck, B. (2007) Quantitative profile of five murine core proteomes using label-free functional proteomics. *Mol. Cell. Proteom.* **6**, 1560–1573 https://doi.org/10.1074/mcp.M700037-MCP200

72 Cutillas, P.R. (2017) Targeted in-depth quantification of signaling using label-free mass spectrometry. *Methods Enzymol.* **585**, 245–268 https://doi.org/10.1016/bs.mie.2016.09.021

73 Bateman, N.W., Goulding, S.P., Shulman, N.J., Gadok, A.K., Szumlinski, K.K., MacCoss, M.J. et al. (2014) Maximizing peptide identification events in proteomic workflows using data-dependent acquisition (DDA). *Mol. Cell. Proteom.* **13**, 329–338 https://doi.org/10.1074/mcp.M112.026500

74 Huang, Q., Yang, L., Luo, J., Guo, L., Wang, Z., Yang, X. et al. (2015) SWATH enables precise label-free quantification on proteome scale. *Proteomics* **15**, 1215–1223 https://doi.org/10.1002/pmic.201400270

75 Hu, A., Noble, W.S. and Wolf-Yadlin, A. (2016) Technical advances in proteomics: new developments in data-independent acquisition. *F1000Res* **5**, F1000 https://doi.org/10.12688/f1000research.7042.1

76 White, F.M. and Wolf-Yadlin, A. (2016) Methods for the analysis of protein phosphorylation-mediated cellular signaling networks. *Annu. Rev. Anal. Chem (Palo Alto Calif).* **9**, 295–315 https://doi.org/10.1146/annurev-anchem-071015-041542

77 Bekker-Jensen, D.B., Martinez-Val, A., Steigerwald, S., Ruther, P., Fort, K.L., Arrey, T.N. et al. (2020) A compact quadrupole-orbitrap mass spectrometer with FAIMS interface improves proteome coverage in short LC gradients. *Mol. Cell. Proteom.* **19**, 716–729 https://doi.org/10.1074/mcp.TIR119.001906

78 Skowronek, P., Thielert, M., Voytik, E., Tanzer, M.C., Hansen, F.M., Willems, S. et al. (2022) Rapid and in-depth coverage of the (phospho-)proteome with deep libraries and optimal window design for dia-PASEF. *Mol. Cell. Proteom.* **21**, 100279 https://doi.org/10.1016/j.mcpro.2022.100279

79 Guo, J. and Huan, T. (2020) Comparison of full-scan, data-dependent, and data-independent acquisition modes in liquid chromatography-mass spectrometry based untargeted metabolomics. *Anal. Chem.* **92**, 8072–8080 https://doi.org/10.1021/acs.analchem.9b05135

80 Liu, Y., Hüttenhain, R., Collins, B. and Aebersold, R. (2013) Mass spectrometric protein maps for biomarker discovery and clinical research. *Expert. Rev. Mol. Diagn.* **13**, 811–825 https://doi.org/10.1586/14737159.2013.845089

81 Demichev, V., Messner, C.B., Vernardis, S.I., Lilley, K.S. and Ralser, M. (2020) DIA-NN: neural networks and interference correction enable deep proteome coverage in high throughput. *Nat. Methods* **17**, 41–44 https://doi.org/10.1038/s41592-019-0638-x

82 Bruderer, R., Bernhardt, O.M., Gandhi, T., Miladinovic, S.M., Cheng, L.Y., Messner, S. et al. (2015) Extending the limits of quantitative proteome profiling with data-independent acquisition and application to acetaminophen-treated three-dimensional liver microtissues. *Mol. Cell. Proteom.* **14**, 1400–1410 https://doi.org/10.1074/mcp.M114.044305

83 Zhang, F., Ge, W., Ruan, G., Cai, X. and Guo, T. (2020) Data-independent acquisition mass spectrometry-based proteomics and software tools: a glimpse in 2020. *Proteomics* **20**, e1900276 https://doi.org/10.1002/pmic.201900276

84 Eng, J.K., Fischer, B., Grossmann, J. and Maccoss, M.J. (2008) A fast SEQUEST cross correlation algorithm. *J. Proteome Res.* **7**, 4598–4602 https://doi.org/10.1021/pr800420s

85 Diament, B.J. and Noble, W.S. (2011) Faster SEQUEST searching for peptide identification from tandem mass spectra. *J. Proteome Res.* **10**, 3871–3879 https://doi.org/10.1021/pr101196n

86 Chalkley, R.J., Baker, P.R., Huang, L., Hansen, K.C., Allen, N.P., Rexach, M. et al. (2005) Comprehensive analysis of a multidimensional liquid chromatography mass spectrometry dataset acquired on a quadrupole selecting, quadrupole collision cell, time-of-flight mass spectrometer: II. New developments in protein prospector allow for reliable and comprehensive automatic analysis of large datasets. *Mol. Cell. Proteom.* **4**, 1194–1204 https://doi.org/10.1074/mcp.D500002-MCP200

87 Rangel-Angarita, V., Mahoney, K.E., Ince, D. and Malaker, S.A. (2022) A systematic comparison of current bioinformatic tools for glycoproteomics data. *bioRxiv* 2022.2003.2015.484528 https://doi.org/10.1101/2022.03.15.484528

88 Park, C.Y., Klammer, A.A., Käll, L., MacCoss, M.J. and Noble, W.S. (2008) Rapid and accurate peptide identification from tandem mass spectra. *J. Proteome Res.* **7**, 3022–3027 https://doi.org/10.1021/pr800127y

89 Kong, A.T., Leprevost, F.V., Avtonomov, D.M., Mellacheruvu, D. and Nesvizhskii, A.I. (2017) MSFragger: ultrafast and comprehensive peptide identification in mass spectrometry-based proteomics. *Nat. Methods* **14**, 513–520 https://doi.org/10.1038/nmeth.4256

90 Mann, M. and Wilm, M. (1994) Error-tolerant identification of peptides in sequence databases by peptide sequence tags. *Anal. Chem.* **66**, 4390–4399 https://doi.org/10.1021/ac00096a002

91 Pappin, D.J., Hojrup, P. and Bleasby, A.J. (1993) Rapid identification of proteins by peptide-mass fingerprinting. *Curr. Biol.* **3**, 327–332 https://doi.org/10.1016/0960-9822(93)90195-t

92 Cox, J. and Mann, M. (2008) Maxquant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* **26**, 1367–1372 https://doi.org/10.1038/nbt.1511

93 Beausoleil, S.A., Villén, J., Gerber, S.A., Rush, J. and Gygi, S.P. (2006) A probability-based approach for high-throughput protein phosphorylation analysis and site localization. *Nat. Biotechnol.* **24**, 1285–1292 https://doi.org/10.1038/nbt1240

94 Savitski, M.M., Lemeer, S., Boesche, M., Lang, M., Mathieson, T., Bantscheff, M. et al. (2011) Confident phosphorylation site localization using the mascot delta score. *Mol. Cell. Proteom.* **10**, M110003830 https://doi.org/10.1074/mcp.M110.003830

95 Baker, P.R., Trinidad, J.C. and Chalkley, R.J. (2011) Modification site localization scoring integrated into a search engine. *Mol. Cell. Proteom.* **10**, M111.008078 https://doi.org/10.1074/mcp.M111.008078

96 Yang, Y., Horvatovich, P. and Qiao, L. (2021) Fragment mass spectrum prediction facilitates site localization of phosphorylation. *J. Proteome Res.* **20**, 634–644 https://doi.org/10.1021/acs.jproteome.0c00580

97   Srinivasan, A., Sing, J.C., Gingras, A.C. and Rost, H.L. (2022) Improving phosphoproteomics profiling using data-independent mass spectrometry. *J. Proteome Res.* **21**, 1789–1799 https://doi.org/10.1021/acs.jproteome.2c00172

98   Elias, J.E. and Gygi, S.P. (2010) Target-decoy search strategy for mass spectrometry-based proteomics. *Methods Mol. Biol.* **604**, 55–71 https://doi.org/10.1007/978-1-60761-444-9_5

99   Moosa, J.M., Guan, S., Moran, M.F. and Ma, B. (2020) Repeat-preserving decoy database for false discovery rate estimation in peptide identification. *J. Proteome Res.* **19**, 1029–1036 https://doi.org/10.1021/acs.jproteome.9b00555

100   Reiter, L., Claassen, M., Schrimpf, S.P., Jovanovic, M., Schmidt, A., Buhmann, J.M. et al. (2009) Protein identification false discovery rates for very large proteomics data sets generated by tandem mass spectrometry. *Mol. Cell. Proteom.* **8**, 2405–2417 https://doi.org/10.1074/mcp.M900317-MCP200

101   The, M., Samaras, P., Kuster, B. and Wilhelm, M. (2022) Re-analysis of ProteomicsDB using an accurate, sensitive and scalable false discovery rate estimation approach for protein groups. *Mol. Cell. Proteom.* **100437**. https://doi.org/10.1016/j.mcpro.2022.100437

102   Savitski, M.M., Wilhelm, M., Hahne, H., Kuster, B. and Bantscheff, M. (2015) A scalable approach for protein false discovery rate estimation in large proteomic data sets. *Mol. Cell. Proteom.* **14**, 2394–2404 https://doi.org/10.1074/mcp.M114.046995

103   Fermin, D., Walmsley, S.J., Gingras, A.C., Choi, H. and Nesvizhskii, A.I. (2013) LuciPHOr: algorithm for phosphorylation site localization with false localization rate estimation using modified target-decoy approach. *Mol. Cell. Proteom.* **12**, 3409–3419 https://doi.org/10.1074/mcp.M113.028928

104   Ramsbottom, K.A., Prakash, A., Riverol, Y.P., Camacho, O.M., Martin, M.J., Vizcaíno, J.A. et al. (2022) Method for independent estimation of the false localization rate for phosphoproteomics. *J. Proteome Res.* **21**, 1603–1615 https://doi.org/10.1021/acs.jproteome.1c00827

105   Beltrao, P., Albanese, V., Kenner, L.R., Swaney, D.L., Burlingame, A., Villen, J. et al. (2012) Systematic functional prioritization of protein posttranslational modifications. *Cell* **150**, 413–425 https://doi.org/10.1016/j.cell.2012.05.036

106   Lienhard, G.E. (2008) Non-functional phosphorylations? *Trends Biochem. Sci.* **33**, 351–352 https://doi.org/10.1016/j.tibs.2008.05.004

107   Tan, C.S., Jorgensen, C. and Linding, R. (2010) Roles of 'junk phosphorylation' in modulating biomolecular association of phosphorylated proteins? *Cell Cycle* **9**, 1276–1280 https://doi.org/10.4161/cc.9.7.11066

108   Ochoa, D., Jarnuczak, A.F., Vieitez, C., Gehre, M., Soucheray, M., Mateus, A. et al. (2020) The functional landscape of the human phosphoproteome. *Nat. Biotechnol.* **38**, 365–373 https://doi.org/10.1038/s41587-019-0344-3

109   Invergo, B.M., Petursson, B., Akhtar, N., Bradley, D., Giudice, G., Hijazi, M. et al. (2020) Prediction of signed protein kinase regulatory circuits. *Cell Syst.* **10**, 384–396.e9 https://doi.org/10.1016/j.cels.2020.04.005

110   Hernandez-Armenta, C., Ochoa, D., Gonçalves, E., Saez-Rodriguez, J. and Beltrao, P. (2017) Benchmarking substrate-based kinase activity inference using phosphoproteomic data. *Bioinformatics* **33**, 1845–1851 https://doi.org/10.1093/bioinformatics/btx082

111   Linding, R., Jensen, L.J., Pasculescu, A., Olhovsky, M., Colwill, K., Bork, P. et al. (2008) NetworKIN: a resource for exploring cellular phosphorylation networks. *Nucleic Acids Res.* **36**, D695–D699 https://doi.org/10.1093/nar/gkm902

112   Piersma, S.R., Valles-Marti, A., Rolfs, F., Pham, T.V., Henneman, A.A. and Jimenez, C.R. (2022) Inferring kinase activity from phosphoproteomic data: tool comparison and recent applications. *Mass Spectrom. Rev.* e21808 https://doi.org/10.1002/mas.21808

113   Wilkes, E.H., Casado, P., Rajeeve, V. and Cutillas, P.R. (2017) Kinase activity ranking using phosphoproteomics data (KARP) quantifies the contribution of protein kinases to the regulation of cell viability. *Mol. Cell. Proteom.* **16**, 1694–1704 https://doi.org/10.1074/mcp.O116.064360

114   Beekhof, R., van Alphen, C., Henneman, A.A., Knol, J.C., Pham, T.V., Rolfs, F. et al. (2019) INKA, an integrative data analysis pipeline for phosphoproteomic inference of active kinases. *Mol. Syst. Biol.* **15**, e8250 https://doi.org/10.15252/msb.20188250

115   Terfve, C.D., Wilkes, E.H., Casado, P., Cutillas, P.R. and Saez-Rodriguez, J. (2015) Large-scale models of signal propagation in human cells derived from discovery phosphoproteomic data. *Nat. Commun.* **6**, 8033 https://doi.org/10.1038/ncomms9033

116   Clarke, D.J.B., Kuleshov, M.V., Schilder, B.M., Torre, D., Duffy, M.E., Keenan, A.B. et al. (2018) Expression2kinases (X2K) Web: linking expression signatures to upstream cell signaling networks. *Nucleic Acids Res.* **46**, W171–W179 https://doi.org/10.1093/nar/gky458

117   Yılmaz, S., Ayati, M., Schlatzer, D., Çiçek, A.E., Chance, M.R. and Koyutürk, M. (2021) Robust inference of kinase activity using functional networks. *Nat. Commun.* **12**, 1177 https://doi.org/10.1038/s41467-021-21211-6

118   Needham, E.J., Parker, B.L., Burykin, T., James, D.E. and Humphrey, S.J. (2019) Illuminating the dark phosphoproteome. *Sci. Signal.* **12**, eaau8645 https://doi.org/10.1126/scisignal.aau8645

119   Ochoa, D., Jonikas, M., Lawrence, R.T., El Debs, B., Selkrig, J., Typas, A. et al. (2016) An atlas of human kinase regulation. *Mol. Syst. Biol.* **12**, 888 https://doi.org/10.15252/msb.20167295

120   Johnson, J.L., Yaron, T.M., Huntsman, E.M., Kerelsky, A., Song, J., Regev, A. et al. (2023) An atlas of substrate specificities for the human serine/threonine kinome. *Nature* **613**, 759–766 https://doi.org/10.1038/s41586-022-05575-3

121   Hornbeck, P.V., Chabra, I., Kornhauser, J.M., Skrzypek, E. and Zhang, B. (2004) Phosphosite: a bioinformatics resource dedicated to physiological protein phosphorylation. *Proteomics* **4**, 1551–1561 https://doi.org/10.1002/pmic.200300772

122   Licata, L., Lo Surdo, P., Iannuccelli, M., Palma, A., Micarelli, E., Perfetto, L. et al. (2020) SIGNOR 2.0, the SIGnaling network open resource 2.0: 2019 update. *Nucleic Acids Res.* **48**, D504–D510 https://doi.org/10.1093/nar/gkz949

123   Turei, D., Korcsmaros, T. and Saez-Rodriguez, J. (2016) Omnipath: guidelines and gateway for literature-curated signaling pathway resources. *Nat. Methods* **13**, 966–967 https://doi.org/10.1038/nmeth.4077

124   Bodenmiller, B., Wanka, S., Kraft, C., Urban, J., Campbell, D., Pedrioli, P.G. et al. (2010) Phosphoproteomic analysis reveals interconnected system-wide responses to perturbations of kinases and phosphatases in yeast. *Sci. Signal.* **3**, rs4 https://doi.org/10.1126/scisignal.2001182

125   Lafont, E., Draber, P., Rieser, E., Reichert, M., Kupka, S., de Miguel, D. et al. (2018) TBK1 and IKKε prevent TNF-induced cell death by RIPK1 phosphorylation. *Nat. Cell Biol.* **20**, 1389–1399 https://doi.org/10.1038/s41556-018-0229-6

126   Casado, P., Bilanges, B., Rajeeve, V., Vanhaesebroeck, B. and Cutillas, P.R. (2014) Environmental stress affects the activity of metabolic and growth factor signaling networks and induces autophagy markers in MCF7 breast cancer cells. *Mol. Cell. Proteom.* **13**, 836–848 https://doi.org/10.1074/mcp.M113.034751

127   Leek, J.T., Scharpf, R.B., Bravo, H.C., Simcha, D., Langmead, B., Johnson, W.E. et al. (2010) Tackling the widespread and critical impact of batch effects in high-throughput data. *Nat. Rev. Genet.* **11**, 733–739 https://doi.org/10.1038/nrg2825

128   Li, T., Zhang, Y., Patil, P. and Johnson, W.E. (2021) Overcoming the impacts of two-step batch effect correction on gene expression estimation and inference. *Biostatistics* kxab039 https://doi.org/10.1093/biostatistics/kxab039

129   Sprang, M., Andrade-Navarro, M.A. and Fontaine, J.-F. (2022) Batch effect detection and correction in RNA-seq data using machine-learning-based automated assessment of quality. *BMC Bioinform.* **23**, 279 https://doi.org/10.1186/s12859-022-04775-y

130   Kim, H.J., Kim, T., Hoffman, N.J., Xiao, D., James, D.E., Humphrey, S.J. et al. (2021) Phosr enables processing and functional analysis of phosphoproteomic data. *Cell Rep.* **34**, 108771 https://doi.org/10.1016/j.celrep.2021.108771

131   Sousa, A., Dugourd, A., Memon, D., Petursson, B., Petsalaki, E., Saez-Rodriguez, J. et al. (2023) Pan-Cancer landscape of protein activities identifies drivers of signalling dysregulation and patient survival. *Mol. Syst. Biol.* e10631 https://doi.org/10.15252/msb.202110631

132   Labrie, M., Brugge, J.S., Mills, G.B. and Zervantonakis, I.K. (2022) Therapy resistance: opportunities created by adaptive responses to targeted therapies in cancer. *Nat. Rev. Cancer* **22**, 323–339 https://doi.org/10.1038/s41568-022-00454-5

133   Kazi, A., Chen, L., Xiang, S., Vangipurapu, R., Yang, H., Beato, F. et al. (2021) Global phosphoproteomics reveal CDK suppression as a vulnerability to KRas addiction in pancreatic cancer. *Clin. Cancer Res.* **27**, 4012–4024 https://doi.org/10.1158/1078-0432.CCR-20-4781

134   Lee, Y.Y., Kim, H.P., Kang, M.J., Cho, B.K., Han, S.W., Kim, T.Y. et al. (2013) Phosphoproteomic analysis identifies activated MET-axis PI3K/AKT and MAPK/ERK in lapatinib-resistant cancer cell line. *Exp. Mol. Med.* **45**, e64 https://doi.org/10.1038/emm.2013.115

135   Vyse, S., McCarthy, F., Broncel, M., Paul, A., Wong, J.P., Bhamra, A. et al. (2018) Quantitative phosphoproteomic analysis of acquired cancer drug resistance to pazopanib and dasatinib. *J. Proteom.* **170**, 130–140 https://doi.org/10.1016/j.jprot.2017.08.015

136   Wong, P.P., Munoz-Felix, J.M., Hijazi, M., Kim, H., Robinson, S.D., De Luxan-Delgado, B. et al. (2020) Cancer burden is controlled by mural cell-beta3-integrin regulated crosstalk with tumor cells. *Cell* **181**, 1346–1363.e21 https://doi.org/10.1016/j.cell.2020.02.003

137   Herranz, N., Gallage, S., Mellone, M., Wuestefeld, T., Klotz, S., Hanley, C.J. et al. (2015) mTOR regulates MAPKAPK2 translation to control the senescence-associated secretory phenotype. *Nat. Cell Biol.* **17**, 1205–1217 https://doi.org/10.1038/ncb3225

138   Pedicona, F., Casado, P., Hijazi, M., Gribben, J.G., Rouault-Pierre, K. and Cutillas, P.R. (2022) Targeting the lysine-specific demethylase 1 rewires kinase networks and primes leukemia cells for kinase inhibitor treatment. *Sci. Signal.* **15**, eabl7989 https://doi.org/10.1126/scisignal.abl7989

139   Tamir, T.Y., Bowman, B.M., Agajanian, M.J., Goldfarb, D., Schrank, T.P., Stohrer, T. et al. (2020) Gain-of-function genetic screen of the kinome reveals BRSK2 as an inhibitor of the NRF2 transcription factor. *J. Cell Sci.* **133**, jcs241356 https://doi.org/10.1242/jcs.241356

140   Liu, Z., Liu, Y., Qian, L., Jiang, S., Gai, X., Ye, S. et al. (2021) A proteomic and phosphoproteomic landscape of KRAS mutant cancers identifies combination therapies. *Mol. Cell* **81**, 4076–4090.e4078 https://doi.org/10.1016/j.molcel.2021.07.021

141   Khorsandi, S.E., Dokal, A.D., Rajeeve, V., Britton, D.J., Illingworth, M.S., Heaton, N. et al. (2021) Computational analysis of cholangiocarcinoma phosphoproteomes identifies patient-specific drug targets. *Cancer Res.* **81**, 5765–5776 https://doi.org/10.1158/0008-5472.CAN-21-0955

142   Wang, Y., Carter, B.Z., Li, Z. and Huang, X. (2022) Application of machine learning methods in clinical trials for precision medicine. *JAMIA Open* **5**, ooab107 https://doi.org/10.1093/jamiaopen/ooab107

143   Zhang, J.D., Koerner, C., Bechtel, S., Bender, C., Keklikoglou, I., Schmidt, C. et al. (2011) Time-resolved human kinome RNAi screen identifies a network regulating mitotic-events as early regulators of cell proliferation. *PLoS One* **6**, e22176 https://doi.org/10.1371/journal.pone.0022176

144   Yang, P., Zheng, X., Jayaswal, V., Hu, G., Yang, J.Y. and Jothi, R. (2015) Knowledge-based analysis for detecting key signaling events from time-series phosphoproteomics data. *PLoS Comput. Biol.* **11**, e1004403 https://doi.org/10.1371/journal.pcbi.1004403

145   Wang, C., Shui, K., Ma, S., Lin, S., Zhang, Y., Wen, B. et al. (2020) Integrated omics in Drosophila uncover a circadian kinome. *Nat. Commun.* **11**, 2710 https://doi.org/10.1038/s41467-020-16514-z

146   Hirano, H., Abe, Y., Nojima, Y., Aoki, M., Shoji, H., Isoyama, J. et al. (2022) Temporal dynamics from phosphoproteomics using endoscopic biopsy specimens provides new therapeutic targets in stage IV gastric cancer. *Sci. Rep.* **12**, 4419 https://doi.org/10.1038/s41598-022-08430-7

147   Mertins, P., Mani, D.R., Ruggles, K.V., Gillette, M.A., Clauser, K.R., Wang, P. et al. (2016) Proteogenomics connects somatic mutations to signalling in breast cancer. *Nature* **534**, 55–62 https://doi.org/10.1038/nature18003

148   Gao, Q., Zhu, H., Dong, L., Shi, W., Chen, R., Song, Z. et al. (2019) Integrated proteogenomic characterization of HBV-related hepatocellular carcinoma. *Cell* **179**, 561–577.e22 https://doi.org/10.1016/j.cell.2019.08.052

149   Tong, M., Yu, C., Zhan, D., Zhang, M., Zhen, B., Zhu, W. et al. (2019) Molecular subtyping of cancer and nomination of kinase candidates for inhibition with phosphoproteomics: reanalysis of CPTAC ovarian cancer. *EBioMedicine* **40**, 305–317 https://doi.org/10.1016/j.ebiom.2018.12.039

150   Chen, Y.J., Roumeliotis, T.I., Chang, Y.H., Chen, C.T., Han, C.L., Lin, M.H. et al. (2020) Proteogenomics of non-smoking lung cancer in east Asia delineates molecular signatures of pathogenesis and progression. *Cell* **182**, 226–244.e217 https://doi.org/10.1016/j.cell.2020.06.012

151   Stewart, P.A., Welsh, E.A., Slebos, R.J.C., Fang, B., Izumi, V., Chambers, M. et al. (2019) Proteogenomic landscape of squamous cell lung cancer. *Nat. Commun.* **10**, 3578 https://doi.org/10.1038/s41467-019-11452-x

152   Vasaikar, S., Huang, C., Wang, X., Petyuk, V.A., Savage, S.R., Wen, B. et al. (2019) Proteogenomic analysis of human colon cancer reveals new therapeutic opportunities. *Cell* **177**, 1035–1049.e1019 https://doi.org/10.1016/j.cell.2019.03.030

153   Casado, P., Rio-Machin, A., Miettinen, J.J., Bewicke-Copley, F., Rouault-Pierre, K., Krizsan, S. et al. (2023) Phosphoproteomics defines two biologically distinct groups of KMT2A rearranged acute myeloid leukaemia with different drug response phenotypes. *Signal. Transduct. Target Ther.* **8**, 80 https://doi.org/10.1038/s41392-022-01288-1

154   Tang, W., Zhou, M., Dorsey, T.H., Prieto, D.A., Wang, X.W., Ruppin, E. et al. (2018) Integrated proteotranscriptomics of breast cancer reveals globally increased protein-mRNA concordance associated with subtypes and survival. *Genome Med.* **10**, 94 https://doi.org/10.1186/s13073-018-0602-x

155   Santra, T., Herrero, A., Rodriguez, J., von Kriegsheim, A., Iglesias-Martinez, L.F., Schwarzl, T. et al. (2019) An integrated global analysis of compartmentalized HRAS signaling. *Cell Rep.* **26**, 3100–3115.e3107 https://doi.org/10.1016/j.celrep.2019.02.038

156   Dugourd, A., Kuppe, C., Sciacovelli, M., Gjerga, E., Gabor, A., Emdal, K.B. et al. (2021) Causal integration of multi-omics data with prior knowledge to generate mechanistic hypotheses. *Mol. Syst. Biol.* **17**, e9730 https://doi.org/10.15252/msb.20209730