TECHNOLOGIES AND TECHNIQUES

Multidimensional proteomics for cell biology

Mark Larance and Angus I. Lamond

Abstract | The proteome is a dynamic system in which each protein has interconnected properties — dimensions — that together contribute to the phenotype of a cell. Measuring these properties has proved challenging owing to their diversity and dynamic nature. Advances in mass spectrometry-based proteomics now enable the measurement of multiple properties for thousands of proteins, including their abundance, isoform expression, turnover rate, subcellular localization, post-translational modifications and interactions. Complementing these experimental developments are new data analysis, integration and visualization tools as well as data-sharing resources. Together, these advances in the multidimensional analysis of the proteome are transforming our understanding of various cellular and physiological processes.

Proteins form the structural fabric of cells and underpin all metabolic processes and regulatory mechanisms. Protein properties, including abundance levels, protein–protein interactions, post-translational modifications (PTMs), subcellular localization patterns and protein synthesis and degradation rates, are all highly dynamic and can change rapidly during the course of biological processes, such as cell proliferation, cell migration, endocytosis and development. Therefore, understanding protein structure–function relationships in cell biology not only requires the identification of proteins but also the detailed analysis of the protein properties that constitute the dimensions of the proteome (FIG. 1a).

Until recently, studying the dynamic behaviour of the proteome has most often focused on the analysis of a single main parameter, such as protein abundance or PTMs. For example, a recent analysis identified >30,000 distinct sites of phosphorylation[1], whereas other studies have identified ~19,000 sites of ubiquitylation on ~5,000 proteins[2]. Although these are landmark studies, a limitation of such single dimension analyses is that much crucial biological information is lost through the averaging of quantitative data from the different cellular pools of a protein.

With this in mind, new mass spectrometry-based proteomic approaches specifically aim to combine improvements in instrumentation and analytical procedures with experimental designs focused on annotating the proteome with high resolution and multidimensional biological information. Thus, multiple protein properties — for example, protein degradation, synthesis and turnover rates[3–6] — are examined either in parallel or sequentially with other protein properties, such as subcellular localization, total protein abundance, tissue distribution[7,8] and protein isoforms or variants[4,9–11]. Using these approaches, we can distinguish pools of proteins that behave differently, for example, in separate subcellular compartments and/or cell cycle stages (FIG. 1b).

Most mass spectrometry-based proteomics experiments have so far used the bottom-up workflow[12], in which proteins are identified by detecting peptides generated after protease cleavage. One disadvantage of this method is that the peptides identified and quantified might not have all come from a single protein species. For example, the same peptide might have originated from multiple protein isoforms and/or from distinct functional pools of the same protein. This can potentially lead to incorrect conclusions if the data are interpreted as representing the behaviour of a single polypeptide in the cell, when in reality they correspond to an average value of two or more distinct polypeptide species that have different characteristics.

Multidimensional analysis of the proteome is now possible thanks to major improvements in the sensitivity and the resolution of mass spectrometry instrumentation and the associated advances in technologies for sample preparation and data analysis[13]. For example, mass spectrometry-based proteomics provides the ability to analyse a large fraction of the population of endogenous, untagged proteins in cells and organisms, avoiding the time, cost and technical limitations that are inherent to either the construction and analysis of

Laboratory for Quantitative Proteomics, Centre for Gene Regulation and Expression, College of Life Sciences, University of Dundee, Dow Street, Dundee DD1 5EH, UK.
e-mails: a.i.lamond@dundee.ac.uk; m.larance@dundee.ac.u
doi:10.1038/nrm3970
Published online 10 April 2015

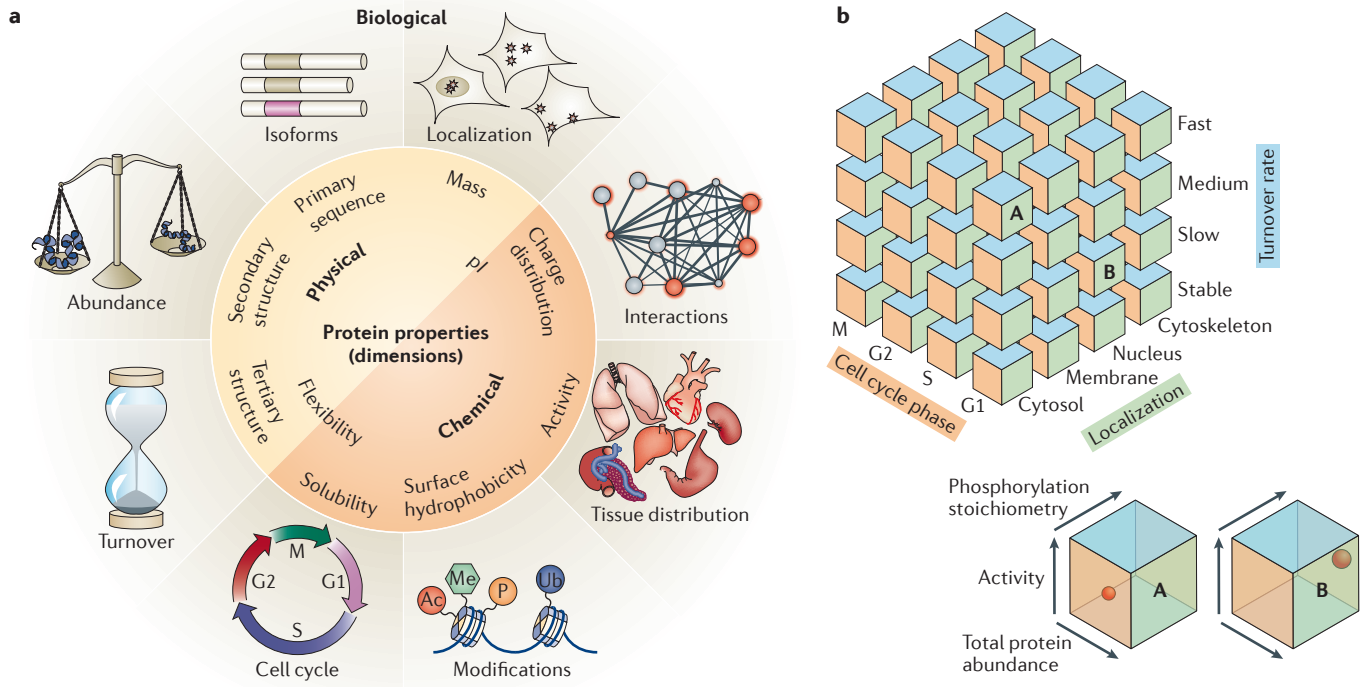


Figure 1 | **Multidimensional proteome analysis of cells and tissues.** **a** | Proteins can have many different properties (dimensions) that are either largely physically (yellow shaded area), chemically (orange shaded area) or biologically (beige shaded area) relevant. Shown in this figure are some of the properties that we think are most important for cell biology research and those that need to be taken into consideration when developing new separation methods for multidimensional analysis. **b** | A series of stacked cubes is shown, each of which contains a discrete pool of proteomics data that correspond to the value of each dimension (localization, cell cycle phase and turnover rate). For each cube (see expanded cubes) we can analyse other dimensions such as protein activity, total protein abundance and phosphorylation stoichiometry. Together these visually represent an approach for the multidimensional analysis of protein data. The spheres inside the expanded cubes A and B represent a specific protein of interest that in the G1 phase of the cell cycle may exist in either the cytosol or in the nucleus, with fast and slow turnover rates, respectively. These different pools (cubes) of the same protein have different properties, including increased protein abundance, phosphorylation and activity in the nuclear pool (cube B) compared with the cytosolic pool (cube A).

large quantities of tagged proteins or the generation of protein-specific antibodies. No other current technologies come close to mass spectrometry in terms of combined throughput, sensitivity, dynamic range and speed of data acquisition. Such speed and sensitivity has, for example, facilitated the analysis of the whole proteome of an organism in less than 24 hours[14,15]. In addition, a mass spectrometer can enhance quantitative accuracy by identifying and quantifying many peptides from each protein in a single experiment, which is equivalent to carrying out independent measurements with many different antibodies to quantify each protein.

Most studies have used one of three main methodologies for the relative quantification of samples: label-free quantification[5,16–18], *in vivo* metabolic stable-isotope labelling (stable isotope labelling by amino acids in cell culture (SILAC)[15, 19], N-labelling[20] and NeuCode SILAC[21]), or stable-isotope labelling using chemical tags that are covalently attached *in vitro* (dimethyl-labelling[22,23], tandem mass tags (TMTs)[24] and isobaric tags for relative and absolute quantification (iTRAQ)[25] (see Supplementary information S1 (box)). Continued improvements in mass spectrometry methods have also increased the flexibility of analysis. In particular, the development of data-independent acquisition (DIA) strategies[26] complement the widely used data-dependent acquisition methods that randomly sample the population of peptides injected into the mass spectrometer. An advantage of DIA strategies is the ability to systematically analyse the peptide population in an extract, which facilitates a more consistent quantitative analysis across large numbers of samples. DIA has recently been used for example, to characterize changes in the macrophage proteome in response to HIV-1 infection[27], the cyclin-dependent kinase 4 interactome[28] and the plant Golgi apparatus[29].

Another burgeoning area of development in the proteomics field explores ways of improving the analysis, the visualization and the sharing of the resulting large single and multidimensional data sets. This extends beyond the need for archiving raw mass spectrometry data sets from the published literature to finding better ways to share and to integrate proteome-level information with other genomic and transcriptomic data sets and to making these data more accessible to the wider biological community. For example, online resources are being developed to visualize, via intuitive interfaces, data derived from complex time-dependent biological responses that may affect thousands of proteins. This facilitates interactive access to the data as well as providing new mechanistic insights into protein function revealing and

Label-free quantification
Protein quantification without exogenous stable isotope labelling, using data derived either from the number of tandem mass spectrometry (MS/MS) spectra, the number of peptides identified and/or the intensity of each peptide observed.

Data-independent acquisition
(DIA). Otherwise known as 'SWATH'; a technique to acquire mass spectrometry data in predefined m/z windows across an entire liquid chromatography tandem mass spectrometry (LC-MS/MS) analysis for consistent quantification across many samples.

Table 1 | **Analysing the dimensions of the proteome**

Dimension	Examples of techniques used	Refs
Abundance (absolute and relative)	Label-free quantitation	5,16–18
	SILAC	19
	^{15}N-labelling	20
	NeuCode SILAC	21
	Dimethyl-labelling	22,23
	TMT	24
	iTRAQ	25
Cell cycle regulation	Centrifugal elutriation	124
	Chemical inhibitors of cell cycle regulators	125
	FACS (for DNA content or phase-specific markers)	126
Tissue distribution	Dissection	95,127
	FACS (for cell-type-specific markers)	126
Interactions	Affinity-enrichment (endogenous immunoprecipitation or tagged fusion protein pull-down)	63–67
	Protein correlation profiling	9,70,71
	Proximity-labelling	39,68
Post-translational modifications	Affinity enrichment: TiO_2	128,129
	Affinity enrichment: IMAC	128,130
	Modification-specific antibodies	90,131–133
	Chromatography: IEX	87
	Chromatography: HILIC	94
	Chromatography: ERLIC	134
Localization	Centrifugation	3,43,135
	Protein correlation profiling	38,44
	Proximity-labelling	39
	Detergent solubility	4
Turnover	Metabolic pulse-labelling	3,5,6,55
	Cycloheximide treatment	4
Isoform expression	• High sequence coverage to identify isoform-specific peptides • Targeted mass spectrometry analysis may be used to detect isoform-specific peptides	136,137
Solubility	Thermal denaturation followed by differential centrifugation	138
Activity	Analogue-sensitive kinases	139
	Activity-dependent binding domains	140
Tertiary Structure	Protease sensitivity	141
	Crosslinking	77,78

ERLIC, electrostatic repulsion hydrophilic interaction chromatography; FACS, fluorescence-associated cell sorting; HILIC, hydrophilic interaction chromatography; IEX, ion-exchange chromatography; IMAC, immobilized metal affinity chromatography; iTRAQ, isobaric tags for relative and absolute quantification; LC-MS/MS, liquid chromatography followed by tandem mass spectrometry; SILAC, stable isotope labelling by amino acids in cell culture; TiO_2, titanium dioxide; TMT, tandem mass tag.

how cells and organisms respond to stimuli at a system-wide level. However, data integration and sharing still presents considerable challenges because of the highly multidimensional and dynamic nature of proteomes and the complexity of the tools needed to postulate biological mechanisms from these data.

In this Review, we discuss recent advances in the multidimensional analysis of the proteome, focusing on biological dimensions of proteins, such as their subcellular localization, turnover, interaction partners and PTMs. We also discuss data analysis, sharing resources, and data integration and visualization tools, and highlight how multidimensional proteomic analysis affects our understanding of various cellular processes.

Biological dimensions

The importance of the multidimensional analysis of protein properties to understanding cell and tissue biology has been shown in the study of several cellular processes. Many of these studies have been aided by developments in mass spectrometry-based analysis, enabling higher sensitivity and a higher dynamic range of quantification[30–33]. In addition, over the past decade, biochemical and cell biological fractionation, such as chromatography or centrifugation-based separations, have increased in efficiency and resolution. Thus, multiple separations can now more easily be combined for sequential multidimensional proteome analysis (TABLE 1).

Protein subcellular localization. A major influence on the function of a protein is its distribution within the cell. The localization of certain proteins can also affect the properties of others, such as their interactions (including with substrates), their degradation rates and their PTMs. For example, in the absence of phosphorylation, the human forkhead box protein O1 (FOXO1) transcription factor is localized on chromatin, to either activate or repress transcription from specific genes[34,35]; however, phosphorylation of FOXO1 sequesters it in the cytoplasm, where it interacts with 14-3-3 proteins[36] (among others) and is subject to an altered rate of degradation[37]. This paradigm has been observed for many individual proteins, which highlights the importance of analysing the proteome as a multidimensional system.

The number of methods for the analysis of protein subcellular localization in cells and tissues is large and diverse. Each method may have advantages or disadvantages depending on the cell or the tissue type and/or the targets to be analysed. The subcellular fractionation methods most usually combined with mass spectrometry-based analysis include differential centrifugation and either equilibrium gradient centrifugation or non-equilibrium gradient centrifugation. These techniques are used for either the isolation of specific organelles or for protein correlation profiling (PCP)[38], for detergent solubility fractionation[4] and for endogenous biotin tagging[39]. The most common problem associated with any biochemical subcellular fractionation is the artefactual post-cell lysis redistribution of proteins to other fractions[40,41]. This may occur either because of the large dilution of the cellular contents, as a result of using a lysis buffer with non-physiological salt concentration, as a result of using inappropriate detergents or because of aggressive cell lysis procedures. Such cross-contamination issues are hard to avoid, especially during proteome-wide analysis, and it is likely that comparing results from multiple methods in parallel will be necessary

Differential centrifugation
Separation of particles on the basis of size and density using several steps of pelleting by centrifugation at increasing *g* force.

to generate an accurate picture of protein localization *in vivo*. Conversely, the use of endogenous biotin-tagging[39], in which cells express a biotin ligase enzyme tagged with a specific-subcellular localization signal, thus enabling the biotinylation and the subsequent purification of the proteins within the target compartment, would avoid some of the problems of mislocalization during fractionation as the biotinylation occurs before cell lysis. Combined with more traditional biochemical fractionation, it can provide complementary information on the localization of all proteins within a subcellular compartment of interest.

The resolution and the quantitative accuracy of PCP[38,42] can be enhanced through the use of *in vitro* chemical labelling of fractionated proteins, a strategy known as localization of organelle proteins by isotope tagging (LOPIT)[43]. The resolution of LOPIT has recently been expanded by using a larger number of isotope labelling reagents[44]. Such PCP methods have the power to differentiate large organelles, small intracellular vesicle populations and even large complexes, such as ribosomes, purely on the basis of their density. A key advantage of this approach is that it does not require any individual organelle or complex to be purified to homogeneity. In addition, it can be used to analyse proteins derived from any cell or tissue without requiring metabolic incorporation of stable isotope labels into living cells or organisms.

Several studies have now combined the proteome-wide analysis of protein subcellular localization with the analysis of other parameters, such as protein synthesis and degradation rates[3,4,45], stress responses[46,47], cell type-specific expression[48], PTMs[49,50] and developmental stage[51]. These studies were able to identify cellular responses that would otherwise have been obscured without the separate subcellular compartment analysis. For example, the late endosomal and lysosomal adaptor and MAPK and mTOR activator 4 (LAMTOR4; also known as C7ORF59) protein was shown to be rapidly degraded in the cytosolic, but not in membrane-associated, nuclear or cytoskeletal fractions[4]. These data highlight the importance of the combined analysis of multiple protein properties to distinguish quantitative data derived from distinct pools of a given cellular protein that behave differently.

Protein turnover. The rate of protein turnover is a combination of synthesis and degradation rates. The protein synthesis rate can be affected by many parameters, including mRNA abundance, localization and translation efficiency. Protein degradation is typically controlled by different factors, one being subcellular localization. For example, mitochondrial proteins can be degraded by pathways that are different to those that degrade cytosolic proteins[52]. Protein PTMs can also either trigger or inhibit protein degradation. This is shown by the cell cycle-dependent phosphorylation of cyclin E, which can trigger its ubiquitin-dependent degradation[53]. By contrast, acetylation of Lys residues in a protein can block ubiquitin-dependent degradation[54].

Classic methods for measuring protein turnover, such as pulse–chase radioactive labelling or cycloheximide treatment to inhibit protein synthesis, have now been adapted for the analysis of protein turnover at the proteome-wide level and have been combined with the measurement of other protein properties, including subcellular localization[3,4,6,45], tissue distribution[55] and/or protein–protein interactions[56] (FIG. 2). Many nuclear proteins were recently identified in rat tissues that were either very slowly degraded or never degraded, which is in stark contrast to the degradation rate of most proteins in many other subcellular compartments[6]. Such effects would have been difficult to observe without combining subcellular fractionation and protein turnover analysis.

Previous studies have also identified many proteins that are unstable only in specific subcellular compartments. For example, free ribosomal proteins are rapidly degraded in the nucleoplasmic compartment, but when assembled into ribosomal subunits and exported to the cytoplasm they acquire long half-lives[3,45]. Such data may also provide clues to the mechanisms by which proteins are rapidly turned over. For example, proteins that are rapidly depleted and that are also found in the endoplasmic reticulum, Golgi and endosomes are likely to either be secreted proteins or destined for degradation via the lysosome or the endoplasmic reticulum-associated protein degradation (ERAD) pathways. By contrast, cytosolic or nuclear proteins that are rapidly depleted will probably be degraded by the various proteasome complexes.

An interesting aspect to the property of protein turnover rate is that newly synthesized proteins must undergo many changes to become fully functional, including proper folding and often PTMs and binding to other proteins. However, the detection of newly synthesized proteins by mass spectrometry-based proteomics is challenging because the pool of pre-existing protein is much larger than the fraction of newly synthesized protein. Over the past few years, several groups have pioneered the use of amino acid mimetics, such as azidohomoalanine (AHA; which replaces Met residues)[57–61]. This amino acid can be fed to cells in short pulses, such that only newly synthesized proteins will contain it. The ability to easily purify proteins that have incorporated such mimetics from cell lysates, using bio-orthogonal click reactions, can greatly improve the sensitivity and the speed of analysis of newly synthesized proteins (BOX 1).

Protein interactions. Proteins do not usually function in isolation; most interact non-covalently with other molecules of the same protein or with other proteins. Given the large size of most metazoan proteomes, the systematic analysis of all protein–protein interactions taking place in a cell is a daunting task. In general, there are three mass spectrometry-based approaches for the global analysis of protein complexes and protein–protein interactions (FIG. 3).

The first and most widely used approach involves variations on the affinity pull-down (immunoprecipitation) strategy (FIG. 3a). In this method, a specific protein (and its binding partners) is isolated either by using antibodies against the endogenous protein or by ectopically

Equilibrium gradient centrifugation
Separation of particles along a gradient on the basis of their density, using a large centrifugal force until the particles reach equilibrium at the point in the gradient of the same density as their own.

Non-equilibrium gradient centrifugation
Similar to equilibrium gradient centrifugation but the application of centrifugal force is stopped before the particles reach equilibrium.

Protein correlation profiling
(PCP). The clustering of protein profiles to predict components in a particular protein complex or cellular localization.

Endoplasmic reticulum-associated protein degradation
(ERAD). A proteasome-dependent protein degradation pathway for the degradation of endoplasmic reticulum proteins.

Click reactions
Cycloaddition reactions involving chemical groups that are not found in nature, typically azide or alkyne groups. Their incorporation into cellular proteins enables labelling with biotin or fluorescent tags via the cycloaddition reaction.

expressing it tagged with, for example, GFP or with short peptides (FLAG, MYC or haemagglutinin), followed by immunoprecipitation using tag-specific antibodies[62]. Such approaches, in conjunction with high-throughput cDNA cloning strategies, have been used to examine protein complexes on a proteome-wide scale[63–67]. A considerable advantage of these methods is the very high sensitivity that can be achieved for low abundance complexes. However, so far, studies using affinity approaches to analyse the global network of protein–protein interactions have generally not been combined with measurement of other protein properties. This is in no small part owing to the time and the costs of carrying out such analyses, which typically involve >10,000 pull-downs for even a single experimental condition.

The second approach is *in vivo* proximity labelling (FIG. 3b), for example, on the basis of the ectopic expression of a protein of interest fused to either a promiscuous biotin-ligase derived from bacteria (the BioID method)[68] or to a peroxidase enzyme capable of activating biotin–phenol (the APEX method)[39]. Once activated, the biotin is rapidly and covalently conjugated to nearby Lys (in the case of BioID) or to Tyr (in the case of APEX) residues. This facilitates the subsequent enrichment of potential interacting proteins using a streptavidin pull-down. Given the high-affinity interaction between biotin and streptavidin, one advantage of this approach is that the streptavidin pull-down enables the use of stringent buffers and extensive washing to maximise the purity of the preparation. The proximity ligation approach has also been extended to study plasma membrane proteins exposed to the extracellular milieu. This is carried out by incubating intact cells with antibody–peroxidase conjugates that will specifically interact with a plasma membrane protein for the subsequent conjugation of biotin–tyramide molecules to proximal proteins[69]. Nonetheless, such methods involve the targeted expression of the recombinant ligase and are therefore more suited to application in cell lines rather than for primary cells and tissues because of the relative ease of transfecting or infecting cell lines.

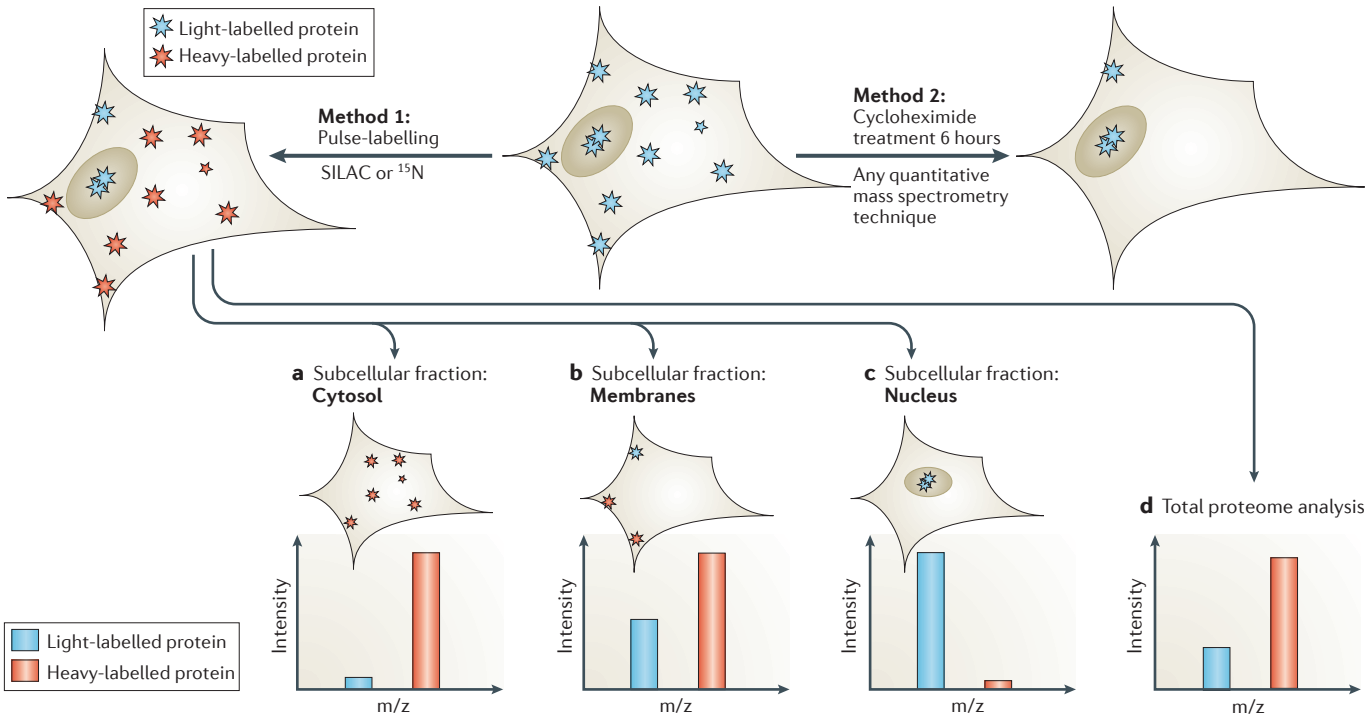
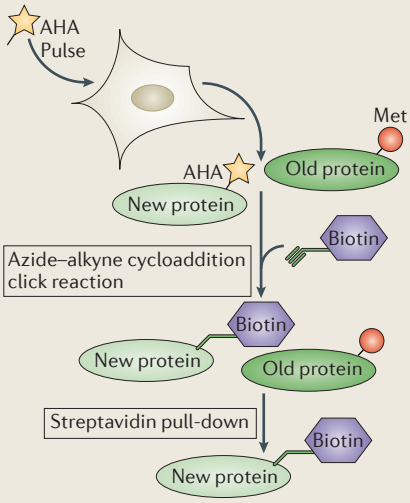


Figure 2 | **Methods for protein turnover analysis.** Proteome-wide turnover is typically measured using one of the two approaches shown in this figure. Method 1 involves pulse-labelling of amino acids using either stable isotope labelling by amino acids in cell culture (SILAC) or ^{15}N-labelling. The cells start with proteins (stars) containing 'light' stable isotopes and for various periods of time are switched into media with a 'heavy' isotope that is stably incorporated into specific amino acids and thus labels newly synthesize proteins. Proteins with rapid turnover rates (for example, cytosolic proteins (part **a**)) will rapidly incorporate high levels of the heavy isotope, whereas protein pools that have slower turnover rates (from the membrane (part **b**) and the nucleus (part **c**) in this example) will show slower rates of heavy isotope incorporation. The ratio between light- and heavy-labelled peptides, which can be extracted using mass spectrometry-based analysis, is a measure of the rate of turnover for each peptide detected and thereby for each protein. Method 2 involves the use of cycloheximide to block protein synthesis in live cells for various periods of time. The comparison of protein abundance between untreated and treated cells enables a calculation of the depletion rate of a protein in the cells, which may occur as a result of its degradation, its secretion or both. The comparison of protein abundances may use any quantitative mass spectrometry technique, such as label-free analysis and isobaric-tag labelling. For methods 1 or 2, the use of cellular fractionation (into subcellular compartments in this case) can greatly increase the information gained compared with the analysis of total cell lysates (part **d**). The protein depicted in this figure is effectively stable in the nuclear compartment (part **c**), has a slow turnover rate in the membrane-associated pool (part **b**) and a fast turnover rate in the cytosol (part **a**). When total cell lysates are examined, such turnover differences can be masked by the pools of protein that are most abundant. In this example, the fast turnover of the cytosolic pool masks the stable nuclear fraction when total cell lysates are examined. m/z, the mass to charge ratio of each peptide ion as measured by mass spectrometers.

Box 1 | **AHA and click chemistry**

Feeding azido-homoalanine (AHA) to cells in short pulses results in the incorporation of this amino acid into proteins, replacing Met, and thus enables the labelling of newly synthesized proteins (see the figure). The presence of the azide group in the side chain of this amino acid facilitates the covalent modification of AHA-containing proteins *in vitro* with affinity reagents that contain groups such as alkynes through cycloaddition reactions (click reactions). Commonly, an affinity group such as biotin is covalently attached to the AHA, resulting in tagged proteins that can subsequently be enriched by streptavidin-mediated pull-down.



The third approach for the analysis of global protein–protein interactions is based on variations of PCP[9,70,71] (FIG. 3c). These techniques use either chromatography or density gradient centrifugation to separate native protein complexes according to size, density, shape, charge and/or hydrophobicity. In this approach, cell extracts are isolated and fractionated under conditions that are designed to preserve protein–protein interactions within complexes. Fractions are then collected across the entire elution or gradient profile before being individually processed to generate peptides and analysed by liquid chromatography followed by tandem mass spectrometry (LC-MS/MS) to identify their constituent proteins. Protein elution (gradient) profiles can then be generated for each protein and compared with other profiles by computational clustering and other approaches to identify putative interacting proteins on the basis of similarities in their elution profiles.

A potential advantage of the PCP approach is that hundreds to thousands of protein complexes can be simultaneously and rapidly analysed. Combined PCP–mass spectrometry approaches have been used to characterize the interactome of several cell lines in combination with other protein properties or cellular states, such as protein isoforms, PTMs[9] or the activation of signalling cascades[71]. A recent study also showed that combining multiple chromatographic separations could increase the resolution between different protein complexes[70]. A disadvantage of the PCP approach is that currently only soluble complexes with interactions that are not markedly weakened by the buffers used can be analysed.

The experimental approaches outlined above can predict that complexes contain certain proteins but not whether they interact with each other directly or indirectly. Such information can instead be generated using protein crosslinking experiments[72,73]. Advances in crosslinking chemistry, mass spectrometry methods and data analysis tools[74] have recently enabled the successful mapping of direct protein–protein[75–77] and protein–RNA[78] interactions. These data can also provide structural insights into the assembly of protein complexes, particularly when the three-dimensional structures of individual protein components are already known[79–82]. Nonetheless, considerable challenges must still be overcome for such crosslinking analyses to be carried out on a scale even remotely approaching a proteome-wide level, not least because of the bioinformatics difficulty in identifying crosslinked peptides using mass spectrometry fragmentation methods. This includes the difficulty of accurately estimating the false-discovery rate for the identification of each co-fragmented peptide sequence, which can be derived from any protein in the original crosslinked protein mixture.

Protein PTMs. PTMs constitute a key mechanism for influencing the properties of many proteins, such as turnover rate, localization and interactions. The modifications can be introduced into, and removed from, proteins in a very rapid manner to affect protein function. Most studies of the effect and the roles of PTMs have concentrated on the analysis of a fairly small number of PTMs, such as the reversible phosphorylation of Ser, Thr and Tyr residues, or the acetylation or ubiquitylation of Lys residues. These studies have provided crucial insights into the regulation of protein properties during fundamental processes, such as cell cycle progression[83,84], and into signalling mechanisms that affect cellular pathways, such as the epidermal growth factor signalling cascade[85].

The analysis of protein phosphorylation, acetylation, glycosylation and ubiquitylation has been greatly facilitated by key developments in peptide fractionation using column-based chromatography and affinity matrices able to pull down modified peptides in batch formats[86]. Modified peptides can be considerably enriched compared to non-modified peptides in chromatographic separations, for example, on the basis of charge (by ion-exchange (IEX) chromatography, such as strong anion exchange[87]) as a result of the added negative charge of the phosphate group or groups. Separations on the basis of hydrophilic interactions, such as hydrophilic interaction liquid chromatography (HILIC), are also able to enrich phosphorylated and certain other modified peptides[88] because of the altered hydrophilicity provided by the modified residues. These separations are usually combined with batch-format affinity enrichment, for example, using acetyl-Lys-specific antibodies for acetylated peptides[89,90]. Chemical probes are also being generated that can be incorporated into proteins instead of the modification group to facilitate the easy extraction of the modified protein or peptide using bio-orthogonal click reactions[91]. For example, protein *O*-GlcNAcylation can be studied using modified precursor metabolites such as peracetylated *N*-azidoacetyl-glucosamine, which can be included in cell culture media to be used by cells during *O*-GlcNAcylation of proteins[92].

Increasingly, PTM analysis has been combined with the analysis of other protein properties such as protein–protein interactions[9,93], developmental stage[94] and tissue distribution[95]. These additional separations have enabled the observation of proteins that, although present in multiple complexes, are phosphorylated only in one

of them[9], or of proteins with tissue-dependent variations in sites and stoichiometry of phosphorylation[95]. Such distinctions would have been difficult to disentangle without biological dimension separation.

The crosstalk between different PTMs is also important as each modification could be considered to be a new protein property. This is exemplified by the complex set of modifications that may be imparted to histones — sometimes termed the histone code — in order to modulate chromatin function[96]. Other biological systems, such as signalling networks[97] and developmental pathways[98], also show crosstalk between different PTMs on the same protein. Notably, recently developed methods facilitate the sequential analysis of same-sample mixtures of peptides altered by different PTMs[99], thereby enabling more protein properties to be analysed and cross-correlated.

New evidence showing the importance of less well-characterized PTMs, including citrullination of Arg[100] and hydroxylation of Pro[101] residues, has also emerged. For example, citrullination of a single Arg residue on the H1 linker histone excludes it from chromatin interactions, ultimately leading to chromatin decondensation, which regulates pluripotency[100]. Hydroxylation of a specific Pro residue on centrosomal protein of 192 kDa (CEP192) targets it for ubiquitylation and subsequent degradation; this

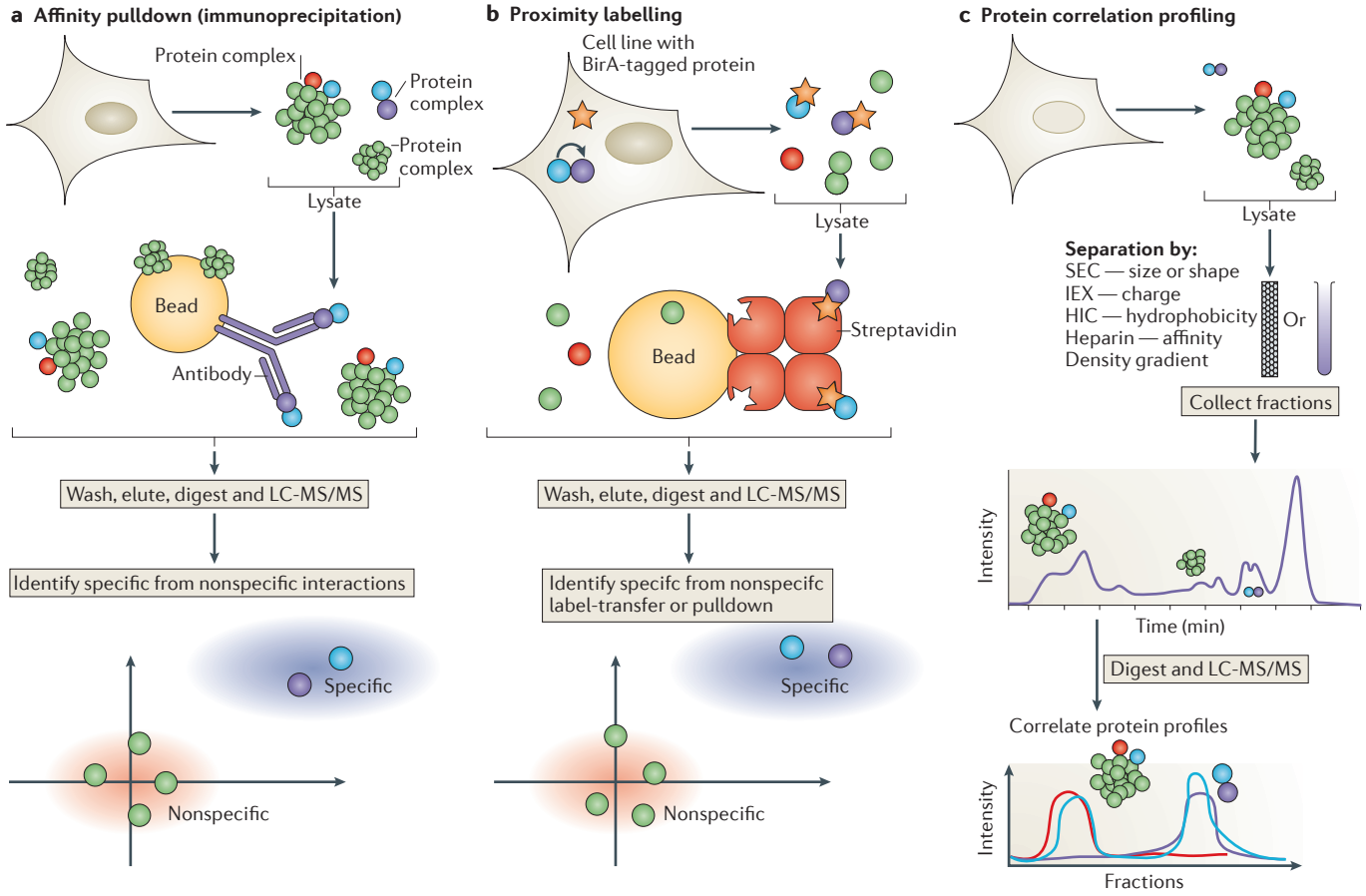


Figure 3 | **Approaches for the analysis of protein interactions.** There are three main approaches for unbiased analysis of protein–protein interactions. **a** | The first approach, based on affinity pull-down and isolation, uses either specific antibodies to an endogenous protein or to a tagged version of a protein to specifically isolate the protein of interest and its interacting partners. Protein complexes are eluted for subsequent analysis by digestion and liquid chromatography followed by tandem mass spectrometry (LC-MS/MS), before statistical approaches are applied to identify specific- (purple and blue circles) from nonspecific- (green circles) binding partners present in the eluted mixture. **b** | The second approach is based on proximity labelling, in which cell lines are constructed that ectopically express a protein of interest (blue circle) fused to either a promiscuous biotin ligase or a peroxidase enzyme. These enzymes can then covalently transfer biotin labels (orange star) to proteins in close proximity (purple circle), which are potential interacting proteins. The cells can then be lysed and the biotinylated proteins specifically isolated using streptavidin-conjugated beads. Similar to the procedure described in part **a**, the isolated proteins are digested and analysed by LC-MS/MS and statistical tests are applied to identify specific (blue and purple circles) versus nonspecific (green circles) label transfer or pull-down. **c** | The third approach uses protein correlation profiling with various biochemical methods, such as chromatography and density gradient centrifugation, to separate endogenous protein complexes according to size, density, charge or hydrophobicity, assuming that interacting proteins will co-elute. This analysis may involve a single type of separation or may involve multiple forms of separation, either sequentially or in parallel. Following the collection of fractions from each separation, the digestion of proteins and their identification by LC-MS/MS analysis, an elution profile for each detected protein is generated and compared with that of other proteins. Clustering algorithms can then identify co-eluting proteins and infer the protein complexes in the lysate. Additional information can also be obtained such as the size or the density of each identified complex. HIC, hydrophobic interaction chromatography; IEX, ion-exchange chromatography; SEC, size-exclusion chromatography.

is an important mechanism for regulating centrosomal duplication and consequently cell cycle progression[101]. This is a similar role to that of targeted Pro hydroxylation in the regulated degradation of the transcription regulator hypoxia-inducible factor 1α (HIF1α)[102].

In summary, the examples above highlight the multidimensional properties of proteins and show how they are highly interconnected and important for protein function.

Data analysis and sharing

As seen with the rapid expansion of genome and transcriptome data arising from technical advances in the instrumentation for high-throughput DNA and RNA sequencing, the current growth in the number of deep proteome studies means that the cell biology community needs consistent ways to process, analyse and share these large proteomic data sets. Below, we review some of the technologies and the resources that are currently available for the analysis and sharing of proteomics data, and we discuss the potential biological insights that can be gained by using them.

Processing software. A large range of both commercial and academic software tools have been developed for the analysis of raw mass spectrometry data files. The diversity of the software solutions partly reflects the rapidly expanding range of mass spectrometry methods in use, as well as the large range of mass spectrometry instrumentation on the market. Examples of free-to-use software for mass spectrometry analysis that have become widely used in the cell biology community include the MaxQuant[16,103,104], Skyline[105], COMPASS[106] and Census[107,108] packages.

MaxQuant provides tools for label-free, SILAC-based and reporter ion-based analyses, all within the same package. It includes both a search engine (Andromeda) for peptide identification and a range of tools for quantification and statistical analysis. The upsurge in the popularity of strategies for the data-independent acquisition-based analysis of protein samples has been aided by software tools that are able to extract quantitative information from such data sets, as exemplified by the Skyline package. Skyline enables the analysis of selected reaction monitoring (SRM), DIA and many other types of data. The Census software[107,108] (and associated tools) provides an alternative, multifunctional package that, in particular, enables the analysis of data sets derived from ^{15}N-labelling experiments. The development of mass spectrometry instruments that can generate third-stage tandem mass spectrum (MS3)-level quantitative reporter ion data in TMT and iTRAQ experiments (see Supplementary information S1 (box)) brings new analytical challenges that have not yet been addressed in most of the non-commercial proteomics software tools. Nevertheless, both the COMPASS[106] and MaxQuant software packages provide the capability for peptide identification and quantification, including the ability to analyse MS3-level reporter ion data.

Data analysis. The shareware and open-source software packages described above, along with the many commercial software solutions that provide similar functionalities, typically provide tools specifically for the identification (and, in some cases, the quantification) of peptides on the basis of the analysis of raw mass spectrometry data. However, software tools are also needed for the downstream statistical evaluation of experimental data after peptides and proteins have been identified and quantified. For example, evaluating which proteins show the most significant change either in abundance or in other properties in response to specific stimuli or predicting protein–protein interactions on the basis of the co-elution of two proteins in a chromatography-based separation. Several statistical analysis packages are available for these purposes, including SPSS and R[109], with R providing the associated Bioconductor[110] suite of tools for bioinformatics analysis. R is free to use and combines the ability to carry out database, statistical and graphing functions and can handle very large multidimensional data sets. In addition, many biology-specific packages are available for R, such as pRoloc, which is a tool for the analysis of protein localization using the protein correlation-profiling approach[111]. However, one potential disadvantage of statistical packages such as R is that they have a considerable learning curve before most cell biologists can take full advantage of the features they offer.

Creating a wider base of user-friendly tools, without such a steep learning curve, for the interactive analysis and integration of multidimensional data that is suitable for broad use in the cell biology community is an important area for future development. Illustrative examples of such software packages are the SAINT[112] package and ProHits[113], each of which uses advanced statistical analysis to differentiate between true and false-positive interactors in immunoprecipitation and affinity pull-down studies. In addition, the ComplexQuant[114] package, which provides a pipeline for the analysis of chromatography-based data for interaction analysis, also fulfils a specialized statistical analysis role to estimate the probabilities of two proteins interacting.

Data visualization and integration. Although we propose in this Review that a multidimensional view of protein-level properties is important for understanding protein function, there is also a great need to integrate these data with other information. For example, there are a wealth of pre-existing data sets that contain complementary information, such as genomic sequence variations, microRNA and mRNA abundance patterns, and in-depth functional annotation of genes in the literature from 'low-throughput' experiments, for example, as provided by gene ontology terms. Integrating these diverse data sets requires user-friendly tools with informative graphical outputs suitable for subsequent publication. One example of such a resource is DAVID[115,116] (Database for Annotation, Visualization and Integrated Discovery), which can annotate user data sets with a large range of stored prior knowledge. The prior knowledge of DAVID includes Gene Ontology terms, transcription factor-binding sites, disease association, protein interactions and cellular pathways, to name only a few. Although these data are rich and each association is strongly supported by statistical evidence, there is still a need to improve built-in visualization tools for the data output of DAVID.

Selected reaction monitoring
(SRM). A mass spectrometry method to focus the instrument on a specific fragment ion derived from a peptide ion of interest. Methods can be generated to analyse many fragment ions from the same peptide and many peptide ion precursors in a single liquid chromatography tandem mass spectrometry (LC-MS/MS) analysis.

Third-stage tandem mass spectrum
(MS3). A spectrum acquired after further fragmentation of isolated peptide fragments from a tandem mass spectrometry analysis.

Table 2 | **Data handling and sharing resources**

Resource	Key features	Strengths	Comments	Refs
Protein dimension annotation resources				
Encyclopedia of Proteome Dynamics	Graphical display of protein dimension data and diverse dimensions analysed	Diverse dimensions analysed, including multidimensional data sets	Single laboratory as source of data	4
Human Protein Atlas	Proteomic annotation of protein abundance and localization	High sensitivity, and localization is analysed within many tissues	Antibody-based, single laboratory as source of data and human data only	122
Human Proteome Map	Human proteome data annotated for tissue abundance	Whole pathway or protein family analysis	One protein dimension and human data only	7
The MaxQuant Database	Proteome data tabulated and graph for label-free abundance measurement	Quality control parameters for mass spectrometry acquisition are presented	Single laboratory as source of data	142
The Multi-Omics Profiling Expression Database	Enables search and visualizations of protein data derived from multiple species	Many different experiments can be visualized and compared	Chromosome-centric	143
The proteins across organisms database	Absolute protein abundance values determined across many organisms and tissues	Abundance histogram and STRING integration	Basic user interface	144
Phosphorylation site database	Provides data on phosphorylation, acetylation and *N*-glycosylation of proteins, including EGF-treated, cell cycle regulated, kinome-related data sets	Diverse dimensions analysed with high depth of coverage for phosphorylation data sets	Single laboratory as source of data	145
ProteomicsDB	Human proteome data annotated for tissue abundance and mass spectrometry spectra shown	Tissue protein abundance pattern and mass spectrometry spectral annotation and multi-protein analysis	Human data only	8
Mass spectrometry-based raw proteomics data repositories				
Chorus	Offers storage, search and visualization of mass spectrometry-based proteomics data files	Well-developed search and mass spectrometry data file visualization	Limited public mass spectrometry data included	–
Global Proteome Machine Database	Enables search and visualization of mass spectrometry data derived from many species	Rich graphical interface for mass spectrometry data visualization	MS2 spectral validation emphasized	146, 147
ProteomeXchange Consortium Includes the Proteomics Identifications, PeptideAtlas, PeptideAtlas SRM Experiment Library and Mass spectrometry Interactive Virtual Environment	Enables centralized submission of mass spectrometry raw data and associated files for shotgun and targeted mass spectrometry analyses	New interface for submission and download of data; managed by the EBI	Requires visiting consortium member sites for the visualization of mass spectrometry data files	148

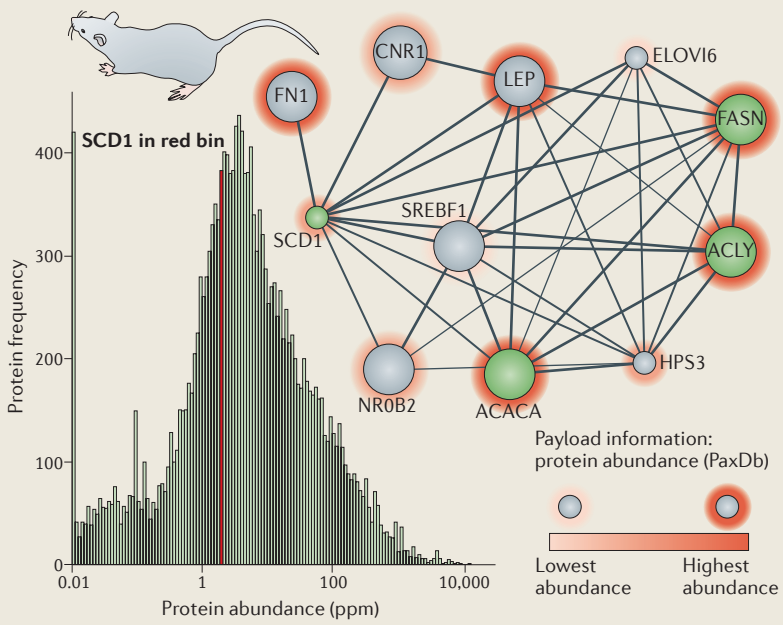
EBI, European Bioinformatics Institute; EGF, epidermal growth factor; MS2, tandem mass spectrometry; SRM, selected reaction monitoring; STRING, search tool for the retrieval of interacting genes/proteins.

The Cytoscape[117] platform is a useful tool that can provide integration of external databases with user data sets and present these analyses in rich graphical formats. Many groups have provided apps for use in Cytoscape, which facilitate multidimensional analysis of gene function. One such application is the GSEA[118] (Gene Set Enrichment Analysis) software tool. GSEA enables statistical testing of the enrichment of any list of genes or proteins within a user data set on the basis of quantified changes detected in an experiment, such as protein abundance. These data can be output from GSEA for rich visualization in the Cytoscape app called EnrichmentMap[119]. Such visualization can help the user to derive otherwise hidden biological meaning from their data.

Data-sharing resources. Effective sharing of proteomic data that can be easily searched and accessed by the cell biology community is another area in which there is great scope for further development of resources. As part of the ProteomeXchange consortium, the Proteomics Identifications (PRIDE) database, which is managed by the European Bioinformatics Institute (EBI), currently provides a central resource for sharing raw mass spectrometry files from published proteomics experiments. The raw files can be freely downloaded and re-analysed. However, although this is useful, particularly to specialist mass spectrometry and bioinformatics groups, it is not a convenient way for most cell biologists to interact with large-scale proteomics data. There are still major challenges in providing the data outputs of multidimensional proteomics experiments in convenient, searchable formats.

To help meet this need, a growing number of curated, online databases now provide access to a range of processed proteomics data that can be searched and linked with information from other large-scale, online resources,

Box 2 | **Adding extra dimensions to proteome data with the STRING database**



The search tool for the retrieval of interacting genes/proteins (STRING) database provides an extraordinary wealth of data derived from many protein databases and literature resources for the analysis of interactions (physical or genetic) between proteins. As an example, the proteins across organisms database (PaxDB), which enables the user to evaluate the abundance (in parts per million (ppm)) of proteins from diverse organisms and tissues, is integrated with the STRING[120] database. This enables the user to search for a protein of interest — mouse acyl-CoA desaturase 1 (SCD1) in the example shown in the figure — and view its abundance measurements (left) in the context of the abundances of all proteins known to interact with it (see the figure). In addition, STRING contains all the known gene ontology terms for each protein and, in this case, the proteins associated with the 'metabolism' gene ontology term have been highlighted by green nodes (right). Together, these data could provide clues to regulatory elements within a pathway and add value to the data provided by each database.

ACACA, acetyl-CoA carboxylase; ACLY, ATP-citrate pro-S lyase; CNR1, cannabinoid receptor 1; ELOVL6, elongation of very long chain fatty acids protein 6; FASN, fatty acid synthase; FN1, fibronectin; HSP3, heat shock protein 3; LEP, leptin; NR0B2, nuclear receptor subfamily 0 group B member 2; SREBF1, sterol regulatory element-binding protein 1.

such as DNA and RNA '-omics' data, metabolic pathways and three-dimensional protein structure data (TABLE 2). We have recently introduced such a resource, the 'Encyclopedia of Proteome Dynamics' (REF. 4), to share our data with the community in a web-based and user-friendly interface (see Supplementary information S2 (figure)). These resources also include protein–protein interaction networks, as exemplified by the search tool for the retrieval of interacting genes/proteins (STRING) database[120] (BOX 2), and tissue-specific protein expression patterns in humans (see Supplementary information S2 (figure)), as determined both by mass spectrometry-based proteomics[7,8] and antibody-based detection[121,122].

Laboratory information management system (LIMS). A database to store experimental data and associated metadata, typically including details of experimental design.

TABLE 2 summarizes some of the currently available online databases that provide cell biologists with access to proteome-wide information. A more widespread adoption of consistent, searchable formats for the sharing of proteomics data is required to make sure that published information is of maximum potential use for the cell biology community. The widespread use of consistent experimental metadata formats would also enhance the ability to compare data published by different groups and to contribute to the establishment of integrated, multi-dimensional data sets that encompass information from many separate experiments and data sources[123].

As well as the need for new tools for the analysis of proteomics data, there is also a need for a laboratory information management system (LIMS) software that is better accustomed to proteomics experiments. Such software could simplify the management and sharing of large-scale proteomics data sets, including the incorporation of detailed and consistent metadata that are needed to facilitate comparisons between data sets.

Conclusions

In this Review, we described the advent of new approaches that combine state-of-the-art proteomic and cell biology methods, which are now rendering a system-wide view of protein properties amenable to direct measurement and analysis. This includes improvements in the mass spectrometry-based methods used to identify and to quantify peptides and their cognate proteins with high sensitivity and accuracy. Coupled with this are innovative methods to separate proteins, either from different cellular organelles, post-translationally modified states or interaction networks, which in combination will facilitate mechanistic insight into protein function and regulation. The combination of these cell biology methods with mass spectrometry analysis will ultimately yield vast volumes of data that need to be stored, analysed and presented to the community in a clear and interactive format.

This new ability to evaluate a dynamic, multidimensional view of the proteome will not only be crucial to enhance our understanding of basic cellular physiology and regulation but also for future advances in medicine and drug development. Further improvements in mass spectrometry instrumentation also continue to make the underlying technologies easier to use and more cost-effective. Paired with this is the need for further co-development of data-handling and sharing tools that facilitate the convenient management, analysis and dissemination of these data. We predict a bright future for expanding the application of mass spectrometry-based proteomics methods to researchers in the cell biology community, and we anticipate that further advances in mass spectrometry-based multidimensional proteomics will greatly influence all aspects of future cell biology experimentation.

1. Batth, T. S., Francavilla, C. & Olsen, J. V. Off-line high-pH reversed-phase fractionation for in-depth phosphoproteomics. *J. Proteome Res.* **13**, 6176–6186 (2014).
2. Kim, W. *et al.* Systematic and quantitative assessment of the ubiquitin-modified proteome. *Mol. Cell.* **44**, 325–340 (2011).
3. Boisvert, F. O. M. *et al.* A quantitative spatial proteomics analysis of proteome turnover in human cells. *Mol. Cell. Proteomics* **11**, M111.011429 (2012).
4. Larance, M., Ahmad, Y., Kirkwood, K. J., Ly, T. & Lamond, A. I. Global subcellular characterization of protein degradation using quantitative proteomics. *Mol. Cell. Proteom.* **12**, 638–650 (2013).
5. Schwanhausser, B. *et al.* Global quantification of mammalian gene expression control. *Nature* **473**, 337–342 (2011).
6. Toyama, B. H. *et al.* Identification of long-lived proteins reveals exceptional stability of essential cellular structures. *Cell* **154**, 971–982 (2013).

7. Kim, M. S. *et al.* A draft map of the human proteome. *Nature* **509**, 575–581 (2014).
8. Wilhelm, M. *et al.* Mass-spectrometry-based draft of the human proteome. *Nature* **509**, 582–587 (2014).
9. Kirkwood, K. J., Ahmad, Y., Larance, M. & Lamond, A. I. Characterization of native protein complexes and protein isoform variation using size-fractionation-based quantitative proteomics. *Mol. Cell. Proteomics* **12**, 3851–3873 (2013).
10. Wu, L. *et al.* Variation and genetic control of protein abundance in humans. *Nature* **499**, 79–82 (2013).
11. Zhang, B. *et al.* Proteogenomic characterization of human colon and rectal cancer. *Nature* **513**, 382–387 (2014).
12. Yates, J. R., Ruse, C. I. & Nakorchevsky, A. Proteomics by mass spectrometry: approaches, advances, and applications. *Annu. Rev. Biomed. Engineer.* **11**, 49–79 (2009).
13. Breker, M. & Schuldiner, M. The emergence of proteome-wide technologies: systematic analysis of proteins comes of age. *Nature Rev. Mol. Cell Bio.* **15**, 453–464 (2014).
14. Hebert, A. S. *et al.* The one hour yeast proteome. *Mol. Cell. Proteomics* **13**, 339–347 (2014).
15. Kulak, N. A., Pichler, G., Paron, I., Nagaraj, N. & Mann, M. Minimal, encapsulated proteomic-sample processing applied to copy-number estimation in eukaryotic cells. *Nature Methods.* **11**, 319–324 (2014).
16. Cox, J. *et al.* MaxLFQ allows accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction. *Mol. Cell. Proteomics* http://dx.doi.org/10.1074/mcp.M113.031591 (2014).
17. Silva, J. C., Gorenstein, M. V., Li, G. Z., Vissers, J. P. & Geromanos, S. J. Absolute quantification of proteins by LCMSE: a virtue of parallel MS acquisition. *Mol. Cell. Proteomics* **5**, 144–156 (2006).
18. Ishihama, Y. *et al.* Exponentially modified protein abundance index (emPAI) for estimation of absolute protein amount in proteomics by the number of sequenced peptides per protein. *Mol. Cell. Proteomics* **4**, 1265–1272 (2005).
19. Ong, S. E. *et al.* 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 (2002).
20. Washburn, M. P., Ulaszek, R., Deciu, C., Schieltz, D. M. & Yates, J. R. Analysis of quantitative proteomic data generated via multidimensional protein identification technology. *Anal. Chem.* **74**, 1650–1657 (2002).
21. Hebert, A. S. *et al.* Neutron-encoded mass signatures for multiplexed proteome quantification. *Nature Methods* **10**, 332–334 (2013).
22. Boersema, P. J., Raijmakers, R., Lemeer, S., Mohammed, S. & Heck, A. J. R. Multiplex peptide stable isotope dimethyl labeling for quantitative proteomics. *Nature Protoc.* **4**, 484–494 (2009).
23. Hsu, J. L., Huang, S. Y., Chow, N. H. & Chen, S. H. Stable-isotope dimethyl labeling for quantitative proteomics. *Anal. Chem.* **75**, 6843–6852 (2003).
24. Thompson, A. *et al.* Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. *Anal. Chem.* **75**, 1895–1904 (2003).
25. Ross, P. L. *et al.* Multiplexed protein quantitation in *Saccharomyces cerevisiae* using amine-reactive isobaric tagging reagents. *Mol. Cell. Proteomics* **3**, 1154–1169 (2004).
26. Gillet, L. C. *et al.* Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis. *Mol. Cell. Proteomics* **11**, O111.016717 (2012).
27. Haverland, N. A., Fox, H. S. & Ciborowski, P. Quantitative proteomics by SWATH-MS reveals altered expression of nucleic acid binding and regulatory proteins in HIV-1-infected macrophages. *J. Proteome Res.* **13**, 2109–2119 (2014).
28. Lambert, J. P. *et al.* Mapping differential interactomes by affinity purification coupled with data-independent mass spectrometry acquisition. *Nature Methods* **10**, 1239–1245 (2013).
29. Nikolovski, N., Shliaha, P. V., Gatto, L., Dupree, P. & Lilley, K. S. Label free protein quantification for plant Golgi protein localisation and abundance. *Plant Physiol.* **166**, 1033–1043 (2014).
30. Ahrens, C. H., Brunner, E., Qeli, E., Basler, K. & Aebersold, R. Generating and navigating proteome maps using mass spectrometry. *Nature Rev. Mol. Cell. Biol.* **11**, 789–801 (2010).
31. Bensimon, A., Heck, A. J. & Aebersold, R. Mass spectrometry-based proteomics and network biology. *Annu. Rev. Biochem.* **81**, 379–405 (2012).
32. Richards, A. L., Merrill, A. E. & Coon, J. J. Proteome sequencing goes deep. *Curr. Opin. Chem. Biol.* **24**, 11–17 (2014).
33. Walther, T. C. & Mann, M. Mass spectrometry-based proteomics in cell biology. *J. Cell Biol.* **190**, 491–500 (2010).
34. Guo, S. *et al.* Phosphorylation of serine 256 by protein kinase B disrupts transactivation by FKHR and mediates effects of insulin on insulin-like growth factor-binding protein-1 promoter activity through a conserved insulin response sequence. *J. Biol. Chem.* **274**, 17184–17192 (1999).
35. Rena, G., Guo, S., Cichy, S. C., Unterman, T. G. & Cohen, P. Phosphorylation of the transcription factor forkhead family member FKHR by protein kinase B. *J. Biol. Chem.* **274**, 17179–17183 (1999).
36. Brunet, A. *et al.* Akt promotes cell survival by phosphorylating and inhibiting a Forkhead transcription factor. *Cell* **96**, 857–868 (1999).
37. Huang, H. *et al.* Skp2 inhibits FOXO1 in tumor suppression through ubiquitin-mediated degradation. *Proc. Natl Acad. Sci. USA* **102**, 1649–1654 (2005).
38. Andersen, J. S. *et al.* Proteomic characterization of the human centrosome by protein correlation profiling. *Nature* **426**, 570–574 (2003).
39. Rhee, H. W. *et al.* Proteomic mapping of mitochondria in living cells via spatially restricted enzymatic tagging. *Science* **339**, 1328–1331 (2013).
40. Press, M. F., Xu, S. H., Wang, J. D. & Greene, G. L. Subcellular-distribution of estrogen-receptor and progesterone-receptor with and without specific ligand. *Am. J. Pathol.* **135**, 857–864 (1989).
41. Tinnikov, A. A. & Samuels, H. H. A novel cell lysis approach reveals that caspase-2 rapidly translocates from the nucleus to the cytoplasm in response to apoptotic stimuli. *PloS ONE* **8**, e61085 (2013).
42. Foster, L. J. *et al.* A mammalian organelle map by protein correlation profiling. *Cell* **125**, 187–199 (2006).
43. Dunkley, T. P. J., Watson, R., Griffin, J. L., Dupree, P. & Lilley, K. S. Localization of organelle proteins by isotope tagging (LOPIT). *Mol. Cell. Proteomics* **3**, 1128–1134 (2004).
44. Christoforou, A., Arias, A. M. & Lilley, K. S. Determining protein subcellular localization in mammalian cell culture with biochemical fractionation and iTRAQ 8-plex quantification. *Methods Mol. Biol.* **1156**, 157–174 (2014).
45. Lam, Y. W., Lamond, A. I., Mann, M. & Andersen, J. S. Analysis of nucleolar protein dynamics reveals the nuclear degradation of ribosomal proteins. *Curr. Biol.* **17**, 749–760 (2007).
46. Baqader, N. O., Radulovic, M., Crawford, M., Stoeber, K. & Godovac-Zimmermann, J. Nuclear cytoplasmic trafficking of proteins is a major response of human fibroblasts to oxidative stress. *J. Proteome Res.* **13**, 4398–4423 (2014).
47. Boisvert, F. M., Lam, Y. W., Lamont, D. & Lamond, A. I. A quantitative proteomics analysis of subcellular proteome localization and changes induced by DNA damage. *Mol. Cell. Proteomics* **9**, 457–470 (2010).
48. Ziegler, Y. S. *et al.* Plasma membrane proteomics of human breast cancer cell lines identifies potential targets for breast cancer diagnosis and treatment. *PloS ONE* **9**, e102341 (2014).
49. Hebert, A. S. *et al.* Calorie restriction and SIRT3 trigger global reprogramming of the mitochondrial protein acetylome. *Mol. Cell.* **49**, 186–199 (2013).
50. Still, A. J. *et al.* Quantification of mitochondrial acetylation dynamics highlights prominent sites of metabolic regulation. *J. Biol. Chem.* **288**, 26209–26219 (2013).
51. McClatchy, D. B., Liao, L. J., Lee, J. H., Park, S. K. & Yates, J. R. Dynamics of subcellular proteomes during brain development. *J. Proteome Res.* **11**, 2467–2479 (2012).
52. Fischer, F., Hamann, A. & Osiewacz, H. D. Mitochondrial quality control: an integrated network of pathways. *Trends Biochem. Sci.* **37**, 284–292 (2012).
53. Koepp, D. M. *et al.* Phosphorylation-dependent ubiquitination of cyclin E by the SCFFbw7 ubiquitin ligase. *Science* **294**, 173–177 (2001).
54. Choudhary, C., Weinert, B. T., Nishida, Y., Verdin, E. & Mann, M. The growing landscape of lysine acetylation links metabolism and cell signalling. *Nature Rev. Mol. Cell Biol.* **15**, 536–550 (2014).
55. Claydon, A. J., Thom, M. D., Hurst, J. L. & Beynon, R. J. Protein turnover: measurement of proteome dynamics by whole animal metabolic labelling with stable isotope labelled amino acids. *Proteomics* **12**, 1194–1206 (2012).
56. Larance, M. *et al.* Characterization of MRFAP1 turnover and interactions downstream of the NEDD8 pathway. *Mol. Cell. Proteomics* **11**, M111.014407 (2012).
57. Baboo, S. *et al.* Most human proteins made in both nucleus and cytoplasm turn over within minutes. *PloS ONE* **9**, e99346 (2014).
58. Eichelbaum, K. & Krijgsveld, J. Rapid temporal dynamics of transcription, protein synthesis and secretion during macrophage activation. *Mol. Cell. Proteomics* **13**, 792–810 (2014).
59. Liang, V. *et al.* Altered proteostasis in aging and heat shock response in *C. elegans* revealed by analysis of the global and *de novo* synthesized proteome. *Cell. Mol. Life Sci.* **71**, 3339–3361 (2014).
60. Shen, W. *et al.* Acute synthesis of CPEB is required for plasticity of visual avoidance behavior in *Xenopus*. *Cell Rep.* **6**, 737–747 (2014).
61. Zhang, J., Wang, J., Ng, S., Lin, Q. & Shen, H. M. Development of a novel method for quantification of autophagic protein degradation by AHA labeling. *Autophagy* **10**, 901–912 (2014).
62. Trinkle-Mulcahy, L. *et al.* Identifying specific protein interaction partners using quantitative mass spectrometry and bead proteomes. *J. Cell Biol.* **183**, 223–239 (2008).
63. Guruharsha, K. G. *et al.* A protein complex network of *Drosophila melanogaster*. *Cell.* **147**, 690–703 (2011).
64. Gavin, A. C. *et al.* Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature* **415**, 141–147 (2002).
65. Ho, Y. *et al.* Systematic identification of protein complexes in *Saccharomyces cerevisiae* by mass spectrometry. *Nature* **415**, 180–183 (2002).
66. Hubner, N. C. *et al.* Quantitative proteomics combined with BAC TransgeneOmics reveals *in vivo* protein interactions. *J. Cell Biol.* **189**, 739–754 (2010).
67. Jager, S. *et al.* Global landscape of HIV-human protein complexes. *Nature* **481**, 365–370 (2012).
68. Roux, K. J., Kim, D. I., Raida, M. & Burke, B. A promiscuous biotin ligase fusion protein identifies proximal and interacting proteins in mammalian cells. *J. Cell Biol.* **196**, 801–810 (2012).
69. Li, X. W. *et al.* New insights into the DT40 B cell receptor cluster using a proteomic proximity labeling assay. *J. Biol. Chem.* **289**, 14434–14447 (2014).
70. Havugimana, P. C. *et al.* A census of human soluble protein complexes. *Cell* **150**, 1068–1081 (2012).
71. Kristensen, A. R., Gsponer, J. & Foster, L. J. A high-throughput approach for measuring temporal changes in the interactome. *Nature Methods* **9**, 907–909 (2012).
72. Schmidt, C. *et al.* Comparative cross-linking and mass spectrometry of an intact F-type ATPase suggest a role for phosphorylation. *Nature Commun.* **4**, http://dx.doi.org/10.1038/ncomms2985 (2013).
73. Leitner, A., Walzthoeni, T. & Aebersold, R. Lysine-specific chemical cross-linking of protein complexes and identification of cross-linking sites using LC-MS/MS and the xQuest/xProphet software pipeline. *Nature Protoc.* **9**, 120–137 (2014).
74. Fischer, L., Chen, Z. A. & Rappsilber, J. Quantitative cross-linking/mass spectrometry using isotope-labelled cross-linkers. *J. Proteom.* **88**, 120–128 (2013).
75. Brodie, N. I., Makepeace, K. A., Petrotchenko, E. V. & Borchers, C. H. Isotopically-coded short-range hetero-bifunctional photo-reactive crosslinkers for studying protein structure. *J. Proteom.* http://dx.doi.org/10.1016/j.jprot.2014.08.012 (2014).
76. Liu, F., Wu, C., Sweedler, J. V. & Goshe, M. B. An enhanced protein crosslink identification strategy using CID-cleavable chemical crosslinkers and LC/MS(n) analysis. *Proteomics* **12**, 401–405 (2012).
77. Weisbrod, C. R. *et al. In vivo* protein interaction network identified with a novel real-time cross-linked peptide identification strategy. *J. Proteome Res.* **12**, 1569–1579 (2013).
78. Kramer, K. *et al.* Photo-cross-linking and high-resolution mass spectrometry for assignment of RNA-binding sites in RNA-binding proteins. *Nature Methods* **11**, 1064–1070 (2014).
79. Bui, K. H. *et al.* Integrated structural analysis of the human nuclear pore complex scaffold. *Cell* **155**, 1233–1243 (2013).
80. Greber, B. J. *et al.* Architecture of the large subunit of the mammalian mitochondrial ribosome. *Nature* **505**, 515–519 (2014).

81. Shi, Y. *et al.* Structural characterization by cross-linking reveals the detailed architecture of a coatomer-related heptameric module from the nuclear pore complex. *Mol. Cell. Proteomics* **13**, 2927–2943 (2014).
82. Shukla, A. K. *et al.* Visualization of arrestin recruitment by a G-protein-coupled receptor. *Nature.* **512**, 218–222 (2014).
83. Dephoure, N. *et al.* A quantitative atlas of mitotic phosphorylation. *Proc. Natl Acad. Sci. USA* **105**, 10762–10767 (2008).
84. Olsen, J. V. *et al.* Quantitative phosphoproteomics reveals widespread full phosphorylation site occupancy during mitosis. *Sci. Signal.* **3**, ra3 (2010).
85. Sharma, K. *et al.* Ultradeep human phosphoproteome reveals a distinct regulatory nature of Tyr and Ser/Thr-based signaling. *Cell Rep.* **8**, 1583–1594 (2014).
86. Larsen, M. R., Trelle, M. B., Thingholm, T. E. & Jensen, O. N. Analysis of posttranslational modifications of proteins by tandem mass spectrometry. *Biotechniques* **40**, 790–798 (2006).
87. Wang, F. J. *et al.* Fractionation of phosphopeptides on strong anion-exchange capillary trap column for large-scale phosphoproteome analysis of microgram samples. *J. Sep. Sci.* **33**, 1879–1887 (2010).
88. Alpert, A. J. Hydrophilic-interaction chromatography for the separation of peptides, nucleic-acids and other polar compounds. *J. Chromatogr.* **499**, 177–196 (1990).
89. Choudhary, C. *et al.* Lysine acetylation targets protein complexes and co-regulates major cellular functions. *Science* **325**, 834–840 (2009).
90. Kim, S. C. *et al.* Substrate and functional diversity of lysine acetylation revealed by a proteomics survey. *Mol. Cell.* **23**, 607–618 (2006).
91. Chuh, K. N. & Pratt, M. R. Chemical methods for the proteome-wide identification of posttranslationally modified proteins. *Curr. Opin. Chem. Biol.* **24C**, 27–37 (2015).
92. Hahne, H. *et al.* Proteome wide purification and identification of *O*-GlcNAc-modified proteins using click chemistry and mass spectrometry. *J. Proteome Res.* **12**, 927–936 (2013).
93. Naegle, K. M., White, F. M., Lauffenburger, D. A. & Yaffe, M. B. Robust co-regulation of tyrosine phosphorylation sites on proteins reveals novel protein interactions. *Mol. Biosyst.* **8**, 2771–2782 (2012).
94. Edwards, A. V. G., Edwards, G. J., Schwammle, V., Saxtorph, H. & Larsen, M. R. Spatial and temporal effects in protein post-translational modification distributions in the developing mouse brain. *J. Proteome Res.* **13**, 260–267 (2014).
95. Huttlin, E. L. *et al.* A tissue-specific atlas of mouse protein phosphorylation and expression. *Cell* **143**, 1174–1189 (2010).
96. Lee, J. S., Smith, E. & Shilatifard, A. The language of histone crosstalk. *Cell.* **142**, 682–685 (2010).
97. Parker, B. L. *et al.* Structural basis for phosphorylation and lysine acetylation cross-talk in a kinase motif associated with myocardial ischemia and cardioprotection. *J. Biol. Chem.* **289**, 25890–25906 (2014).
98. Palmisano, G. *et al.* A novel method for the simultaneous enrichment, identification, and quantification of phosphopeptides and sialylated glycopeptides applied to a temporal profile of mouse brain development. *Mol. Cell. Proteomics* **11**, 1191–1202 (2012).
99. Mertins, P. *et al.* Integrated proteomic analysis of post-translational modifications by serial enrichment. *Nature Methods* **10**, 634–637 (2013).
100. Christophorou, M. A. *et al.* Citrullination regulates pluripotency and histone H1 binding to chromatin. *Nature* **507**, 104–108 (2014).
101. Moser, S. C. *et al.* PHD1 links cell-cycle progression to oxygen sensing through hydroxylation of the centrosomal protein Cep192. *Dev. Cell.* **26**, 381–392 (2013).
102. Semenza, G. L. Hypoxia-inducible factor 1 (HIF-1) pathway. *Sci. STKE.* http://dx.doi.org/10.1126/stke.4072007cm8 (2007).
103. Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nature Biotech.* **26**, 1367–1372 (2008).
104. Cox, J. *et al.* Andromeda: a peptide search engine integrated into the MaxQuant environment. *J. Proteome Res.* **10**, 1794–1805 (2011).
105. MacLean, B. *et al.* Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics* **26**, 966–968 (2010).
106. Wenger, C. D., Phanstiel, D. H., Lee, M. V., Bailey, D. J. & Coon, J. J. COMPASS: A suite of pre- and post-search proteomics software tools for OMSSA. *Proteomics* **11**, 1064–1074 (2011).
107. Park, S. K. *et al.* Census 2: isobaric labeling data analysis. *Bioinformatics* **30**, 2208–2209 (2014).
108. Park, S. K., Venable, J. D., Xu, T. & Yates, J. R. 3rd. A quantitative analysis software tool for mass spectrometry-based proteomics. *Nature Methods* **5**, 319–322 (2008).
109. R Development Core Team. *R: A Language and Environment for Statistical Computing.* (R Foundation for Statistical Computing, 2014).
110. Gentleman, R. C. *et al.* Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol.* **5**, 80 (2004).
111. Gatto, L., Breckels, L. M., Wieczorek, S., Burger, T. & Lilley, K. S. Mass-spectrometry-based spatial proteomics data analysis using pRoloc and pRolocdata. *Bioinformatics* **30**, 1322–1324 (2014).
112. Choi, H. *et al.* SAINT: probabilistic scoring of affinity purification-mass spectrometry data. *Nature Methods* **8**, 70–73 (2011).
113. Liu, G. *et al.* ProHits: integrated software for mass spectrometry-based interaction proteomics. *Nature Biotech.* **28**, 1015–1017 (2010).
114. Wan, C. H. *et al.* ComplexQuant: high-throughput computational pipeline for the global quantitative analysis of endogenous soluble protein complexes using high resolution protein HPLC and precision label-free LC/MS/MS. *J. Proteom.* **81**, 102–111 (2013).
115. Huang, D. W., Sherman, B. T. & Lempicki, R. A. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* **37**, 1–13 (2009).
116. Huang, D. W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protoc.* **4**, 44–57 (2009).
117. Shannon, P. *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504 (2003).
118. Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. USA* **102**, 15545–15550 (2005).
119. Merico, D., Isserlin, R., Stueker, O., Emili, A. & Bader, G. D. Enrichment map: a network-based method for gene-set enrichment visualization and interpretation. *PloS ONE* **5** e13984 (2010).
120. Franceschini, A. *et al.* STRING v9.1: protein-protein interaction networks, with increased coverage and integration. *Nucleic Acids Res.* **41**, D808–D815 (2013).
121. Fagerberg, L. *et al.* Analysis of the human tissue-specific expression by genome-wide integration of transcriptomics and antibody-based proteomics. *Mol. Cell. Proteomics* **13**, 397–406 (2014).
122. Uhlen, M. *et al.* Towards a knowledge-based Human Protein Atlas. *Nature Biotech.* **28**, 1248–1250 (2010).
123. Ahmad, Y. & Lamond, A. I. A perspective on proteomics in cell biology. *Trends Cell Biol.* **24**, 257–264 (2014).
124. Ly, T. *et al.* A proteomic chronology of gene expression through the cell cycle in human myeloid leukemia cells. *eLife* **3**, e01630 (2014).
125. Ly, T., Endo, A. & Lamond, A. I. Proteomic analysis of the response to cell cycle arrests in human myeloid leukemia cells. *eLife* **4**, e04534 (2015).
126. Bernas, T., Gregori, G., Asem, E. K. & Robinson, J. P. Integrating cytomics and proteomics. *Mol. Cell. Proteom.* **5**, 2–13 (2006).
127. Geiger, T. *et al.* Initial quantitative proteomic map of 28 mouse tissues using the SILAC mouse. *Mol. Cell. Proteom.* **12**, 1709–1722 (2013).
128. Engholm-Keller, K. *et al.* TiSH — a robust and sensitive global phosphoproteomics strategy employing a combination of TiO2, SIMAC, and HILIC. *J. Proteom.* **75**, 5749–5761 (2012).
129. Larsen, M. R., Thingholm, T. E., Jensen, O. N., Roepstorff, P. & Jorgensen, T. J. D. Highly selective enrichment of phosphorylated peptides from peptide mixtures using titanium dioxide microcolumns. *Mol. Cell. Proteomics* **4**, 873–886 (2005).
130. Neville, D. C. A. *et al.* Evidence for phosphorylation of serine 753 in CFTR using a novel metal-ion affinity resin and matrix-assisted laser desorption mass spectrometry. *Protein Sci.* **6**, 2436–2445 (1997).
131. Boisvert, F. M., Cote, J., Boulanger, M. C. & Richard, S. A proteomic analysis of arginine-methylated protein complexes. *Mol. Cell. Proteomics* **2**, 1319–1330 (2003).
132. Xu, G. Q., Paige, J. S. & Jaffrey, S. R. Global analysis of lysine ubiquitination by ubiquitin remnant immunoaffinity profiling. *Nature Biotech.* **28**, 868–873 (2010).
133. Lamoliatte, F. *et al.* Large-scale analysis of lysine SUMOylation by SUMO remnant immunoaffinity profiling. *Nature Commun.* **5**, 5409 (2014).
134. Alpert, A. J. Electrostatic repulsion hydrophilic interaction chromatography for isocratic separation of charged solutes and selective isolation of phosphopeptides. *Anal. Chem.* **80**, 62–76 (2008).
135. Prior, M. J. *et al.* Quantitative proteomic analysis of the adipocyte plasma membrane. *J. Proteome Res.* **10**, 4970–4982 (2011).
136. Ritorto, M. S., Cook, K., Tyagi, K., Pedrioli, P. G. A. & Trost, M. Hydrophilic strong anion exchange (hSAX) chromatography for highly orthogonal peptide separation of complex proteomes. *J. Proteome Res.* **12**, 2449–2457 (2013).
137. Gilar, M., Olivova, P., Daly, A. E. & Gebler, J. C. Two-dimensional separation of peptides using RP-RP-HPLC system with different pH in first and second separation dimensions. *J. Sep. Sci.* **28**, 1694–1703 (2005).
138. Savitski, M. M. *et al.* Tracking cancer drugs in living cells by thermal profiling of the proteome. *Science* **346**, 1255784 (2014).
139. Banko, M. R. *et al.* Chemical genetic screen for AMPKα2 substrates uncovers a network of proteins involved in mitosis. *Mol. Cell.* **44**, 878–892 (2011).
140. Zhang, C. C. *et al.* Development and application of a quantitative multiplexed small GTPase activity assay using targeted proteomics. *J. Proteome Res.* **14**, 967–976 (2015).
141. Feng, Y. H. *et al.* Global analysis of protein structural changes in complex proteomes. *Nature Biotech.* **32**, 1036–1044 (2014).
142. Schaab, C., Geiger, T., Stoehr, G., Cox, J. & Mann, M. Analysis of high accuracy, quantitative proteomics data in the MaxQB database. *Mol. Cell. Proteomics* **11**, M111.014068 (2012).
143. Montague, E. *et al.* MOPED 2.5 — an integrated multi-omics resource: multi-omics profiling expression database now includes transcriptomics data. *Omics* **18**, 335–343 (2014).
144. Wang, M. *et al.* PaxDb, a database of protein abundance averages across all three domains of life. *Mol. Cell. Proteom.* **11**, 492–500 (2012).
145. Gnad, F., Gunawardena, J. & Mann, M. PHOSIDA 2011: the posttranslational modification database. *Nucleic Acids Res.* **39**, D253–D260 (2011).
146. Craig, R., Cortens, J. P. & Beavis, R. C. Open source system for analyzing, validating, and storing protein identification data. *J. Proteome Res.* **3**, 1234–1242 (2004).
147. Shanmugam, A. K., Yocum, A. K. & Nesvizhskii, A. I. Utility of RNA-seq & GPMDB protein observation frequency for improving the sensitivity of protein identification by tandem MS. *J. Proteome Res.* **13**, 4113–4119 (2014).
148. Vizcaíno, J. A. *et al.* ProteomeXchange provides globally coordinated proteomics data submission and dissemination. *Nature Biotechnol.* **32**, 223–226 (2014).

Acknowledgements
A.I.L. is a Wellcome Trust Principal Research Fellow and M.L. is a Royal Society of Edinburgh Scottish Government Personal Research Fellow.

Competing interests statement
The authors declare no competing interests.

FURTHER INFORMATION
EPD: http://peptracker.com/epd
Chorus: https://chorusproject.org/pages/index.html
Cytoscape: http://apps.cytoscape.org
GPMDB: http://gpmdb.thegpm.org
Human Protein Atlas: http://www.proteinatlas.org
Human Proteome Map: http://humanproteomemap.org
MaxQB: http://maxqb.biochem.mpg.de/mxdb
MOPED: http://moped.proteinspire.org
PaxDB: http://pax-db.org
Phosida: http://www.phosida.com
ProteomeXchange Consortium: http://www.proteomexchange.org
ProteomicsDB: https://www.proteomicsdb.org

SUPPLEMENTARY INFORMATION
See online article: S1 (box) | S2 (figure)

ALL LINKS ARE ACTIVE IN THE ONLINE PDF