# Multidimentional proteomics for cell biology

Larance, Mark; Lamond, Angus I.

*Published in:*
Nature Reviews Molecular Cell Biology

*DOI:*
10.1038/nrm3970

*Publication date:*
2015

*Document Version*
Peer reviewed version

# Multi-Dimensional Proteomics for Cell Biology

Mark Larance[1], Angus I. Lamond[1,#].

[1]Laboratory for Quantitative Proteomics, Centre for Gene Regulation and Expression, College of Life Sciences, University of Dundee, Dow St, Dundee, United Kingdom, DD15EH.

[#]Address correspondence to: Angus I. Lamond Tel. (+44) 01382385473; E-mail a.i.lamond@dundee.ac.uk

## Glossary

- DIA, data-independent acquisition. Otherwise known as "SWATH", a technique to acquire mass spectrometry data in pre-defined m/z windows across an entire LC-MS/MS analysis for consistent quantitation across many samples.
- Differential centrifugation. A separation of particles based on size and density using several steps of pelleting by centrifugation at increasing g-force.
- Equilibrium gradient centrifugation. A separation of particles based upon density, usually using a gradient of either sucrose, iodixanol, or caesium chloride. Samples are applied to the top of the gradient and a large centrifugal force applied 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. The clustering of protein profiles to predict components in a particular protein complex or cellular localisation. Usually this is based upon proteomic analysis of cellular organelles or protein complexes fractionated using techniques such as chromatography or centrifugation.
- ERAD, endoplasmic-reticulum (ER)-associated protein degradation. A proteasome-dependent protein degradation pathway for the destruction of ER proteins.
- 'Click' (reaction). A cycloaddition reaction, particularly between an azide and alkyne, involving chemical groups that are not found in nature. The incorporation into cellular proteins of a probe of interest containing either an azide, or alkyne group, allows subsequent specific labelling of the probe with a tag such as a biotin molecule or fluorescent dye via the cycloaddition reaction.
- SRM, selected reaction monitoring. A mass-spectrometry method used 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 LC-MS/MS analysis.
- MS3, third-stage tandem mass spectrum. Acquired after fragmentation of isolated peptide fragments from a MS2 analysis.

- PTM. Post-translational modification of proteins due to either enzymatic, or chemical reactions altering the structure of amino acids after protein synthesis.
- Label-free quantitation. The quantitation of proteins using data derived either from the number of MS/MS spectra, the number of peptides identified and/or the intensity of each peptide observed. These values can be normalised and combined in various ways, for example, taking into account protein size and the number of peptides that could theoretically be observed. These techniques involve no exogenous stable isotope labelling.
- LC-MS/MS, liquid chromatography coupled online to tandem mass spectrometry. Standard approach for bottom-up proteome analysis of peptides. Usually uses nano-flow reverse-phase chromatography coupled to an electrospray ionisation source and a mass spectrometer capable of multiple rounds of isolation, fragmentation and mass spectrum acquisition.
- LIMS, laboratory information management system. A database to store experimental data and associated metadata, typically including details of experimental design.

## Online Summary

- The proteome is complex due to the array of properties each protein may display and the constant dynamic changes in these properties. It has not been common to study many of these properties due to either expense, time-required, and/or lack of tools.
- These interconnected protein properties include subcellular localisation, interactions, turnover rate and post-translational modifications, among many others.
- Without analysing many of these properties, we will achieve only a limited understanding of many crucial biological regulatory mechanisms that involve variation in protein properties.
- The impact of proteomics on cell biology will be enhanced when we can predict interdependence of protein properties for each protein. For example, knowing that a protein with a certain modification, if localised in the cytosol, will be degraded.
- Future innovations will allow more comprehensive measurement of a wider range of protein properties.
- Data analysis and data sharing are critical to maximise the impact for cell biology of mass spectrometry based proteome-wide analyses. In particular, making such data available to cell biologists in free to access, web-based and graphically-rich formats will be important to leverage the full benefit of all the information being generated.

## Biographies

**Mark Larance is a Royal Society of Edinburgh – Scottish Government Personal Research Fellow at the University of Dundee. He studies metabolism using a combination of cell biology, biochemistry and mass spectrometry-based proteome analysis.

**Angus Lamond is Professor of Biochemistry and a Wellcome Trust Principal Research Fellow at the University of Dundee. Angus' group study gene expression and the functional organization of

mammalian cell nuclei. Their approach combines quantitative mass spectrometry based proteomics with fluorescence imaging, cell biology and computational methods.

## Preface

The proteome is a dynamic system in which each protein displays interconnected properties — dimensions — that together contribute to cell phenotype. Measuring these properties has proved challenging owing to their diversity and dynamic nature. Advances in mass-spectrometry-based proteomics are now enabling 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 visualisation tools and data-sharing resources. Together, these advances in the multi-dimensional analysis of the proteome are transforming our understanding of various cellular and physiological processes.

## Introduction

Proteins form the structural fabric of cells and underpin all metabolic processes and regulatory mechanisms. A wide range of 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. Understanding protein structure–function relationships in cell biology, therefore, not only requires the identification of proteins, but also the detailed analysis of protein properties that constitute the dimensions of the proteome (**Figure 1a**).

Until recently, studying the dynamic behaviour of the proteome has most often focused on the analysis of a single, main parameter, such as either 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]. While these are landmark studies, a limitation of such single dimension analyses is that much critical 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 (MS)-based proteomics approaches specifically aim to combine improvements in instrumentation and analytical procedures with experimental designs focused on annotating the proteome with high-resolution and multi-dimensional 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 in separate subcellular compartments and/or cell-cycle stages, and so on (**Figure 1b**).

Most mass spectrometry-based proteomics experiments to date have used the bottom-up workflow[12], in which proteins are identified by detecting peptides generated after protease cleavage.

One downside 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 from two or more distinct polypeptide species with different characteristics.

Multidimensional analysis of the proteome is now facilitated thanks to major improvements in the sensitivity and resolution of mass spectrometry instrumentation and 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 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 performing independent measurements with many different antibodies to quantify each protein.

Most studies have used one of three main methodologies for relative quantitation of samples: i.e. label-free[5, 16-18]; *in vivo* metabolic stable-isotope labelling (SILAC (stable isotope labelling by amino acids in cell culture)[19], $^{15}$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 (TMT)[24] and isobaric tags for relative and absolute quantitation (iTRAQ)[25]; see **Supplementary Information S1**). 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, allowing for a more consistent quantitative analysis across large numbers of samples. DIA has been used recently, for example, to characterize changes in the macrophage proteome in response to HIV 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, visualization and sharing of the resulting large single and multi-dimensional 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 integrate proteome-level information with other genomic and transcriptomic data sets and to make these data more accessible to the wider biological community. For example, online resources are being developed to visualise, 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 how cells and organisms respond to stimuli at a system-wide level. Data integration

and sharing still presents significant challenges, however owing to both the highly multi-dimensional 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 multi-dimensional analysis of the proteome, focusing on biological dimensions of proteins, such as their subcellular localization, turnover, interaction partners, and post-translational modifications. We also discuss data analysis, sharing-resources and data integration and visualization tools, and highlight how multi-dimensional proteomic analysis impacts our understanding of various cellular processes.

## Biological dimensions

The importance of the multi-dimensional analysis of protein properties to understanding cell and tissue biology has been demonstrated in the study of a number of cellular processes. Many of these studies have been aided by developments in mass spectrometry-based analysis allowing higher sensitivity and a higher dynamic range of quantitation[30-33]. In addition, biochemical and cell biological fractionation such as either chromatography, or centrifugation-based separations, have, over the last decade, increased in efficiency and resolution. Thus, multiple separations can now more easily be combined for sequential, multi-dimensional 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), degradation rates and post-translational modifications. 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]; phosphorylation of FOXO1, however, 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, highlighting the importance of analysing the proteome as a multi-dimensional system.

The number of methods for the analysis of protein subcellular localisation in cells and tissues is large and diverse. Each method may have advantages or disadvantages, depending on cell or 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, or non-equilibrium gradient centrifugation. These techniques are used for either the isolation of specific organelles, or for protein correlation profiling (PCP; see below)[38], detergent solubility fractionation[4] and 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 due to the large dilution of the cellular contents, to using a lysis buffer with non-physiological salt concentration, to using inappropriate detergents or to aggressive cell lysis procedures. Such cross-contamination issues are hard to avoid, especially during proteome-wide analysis and it is likely comparing results from multiple methods in parallel will be necessary to generate an accurate picture of protein localization *in vivo*. Conversely, the use of endogenous biotin-

tagging[39], where cells express a biotin ligase enzyme tagged with a specific-subcellular localization signal, thus allowing the biotinylation and subsequent purification of the proteins within the target compartment, would avoid some of the problems of mis-localization during fractionation as the biotinylation occurs prior to cell lysis. Combined with more traditional biochemical fractionation, it can provide complementary information on the localisation of all proteins within a subcellular compartment of interest.

The resolution and quantitative accuracy of PCP[38, 42] can be enhanced through the use of *in vitro* chemical labelling of fractionated proteins, a strategy known as LOPIT (localization of organelle proteins by isotope tagging)[43]. Recently, the resolution of LOPIT has 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 based on 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.

A number of 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], post-translational modifications[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/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 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 different pathways than cytosolic proteins[52]. Protein PTMs can also either trigger, or inhibit, protein degradation. This is demonstrated by the cell-cycle-dependent phosphorylation of cyclin E, which can trigger its ubiquitin-dependent degradation[53]. By contrast, acetylation of lysine residues in a protein can block ubiquitin-dependent degradation[54].

Classic methods for measuring protein turnover, such as either 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 combined with the measurement of other protein properties, including subcellular localization[3, 4, 6, 45], tissue distribution[55] and/or protein–protein interactions[56] (**Figure 2**). Recently many nuclear proteins were identified in rat tissues that were either very slowly degraded, or never degraded, 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 are also found in the endoplasmic reticulum, golgi and endosomes are likely to either be secreted proteins, or else destined for degradation via the lysosome or ERAD pathways. In contrast, cytosolic or nuclear proteins that are rapidly depleted will likely be degraded by the various proteasome complexes.

An interesting aspect to the property of protein turnover rate is that newly synthesised proteins must undergo many changes to become fully functional, including proper folding, and often post-translational modifications and binding to interaction partners. However, the detection of newly synthesised proteins by mass spectrometry-based proteomics is challenging, because the pool of pre-existing protein is much larger than the fraction of newly synthesised protein. Over the past few years, several groups have pioneered the use of amino acid mimetics, such as azido-homoalanine (AHA; which replaces methionine residues) for the analysis of newly synthesized proteins[57-61]. This amino acid can be fed to cells in short pulses, such that only newly synthesised 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 speed of analysis of newly synthesised proteins (**Box 1**).

***Protein interactions.*** Proteins do not usually function in isolation; most interact non-covalently with either 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, three mass spectrometry-based approaches exist for the global analysis of protein complexes and protein–protein interactions (**Figure 3**).

The first and most widely used approach involves variations on the affinity pull-down (immunoprecipitation) strategy (**Figure 3a**). This isolates a specific protein (and its binding partners) either by using antibodies against the endogenous protein, or by ectopically expressing it, tagged with, for example, GFP or with short peptides (e.g. FLAG, Myc or HA), followed by immunoprecipitation using anti-tag 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 significant advantage of these methods is the very high sensitivity that can be achieved for low abundance complexes. So far, however, 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 costs of performing such analyses, which typically involve >10,000 pull-downs for even a single experimental condition.

The second approach is *in vivo* proximity labelling (**Figure 3b**), for example based on the ectopic expression of a protein of interest fused to either a promiscuous biotin-ligase derived from

bacteria (the BioID method)[68], or a peroxidase enzyme capable of activating biotin-phenol (the APEX method)[39]. Once activated, the biotin is rapidly and covalently conjugated to nearby proteins via either lysines, or tyrosines, for BioID and APEX, respectively. This facilitates the subsequent enrichment of potential interacting proteins using streptavidin pull-down. Given the high affinity interaction between biotin and streptavidin, one advantage of this approach is that the streptavidin pull-down allows using 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 performed 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 therefore are more suited to application in cell lines rather than for primary cells and tissues, owing to the relative ease of transfecting/infecting cell lines.

The third approach for the analysis of global protein–protein interactions is based on variations of the technique known as protein correlation profiling[9, 70, 71] (PCP) (**Figure 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 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 coupled online to tandem mass spectrometry (LC-MS/MS) to identify their constituent proteins. Protein elution (gradient) profiles can then be generated for each protein and compared to 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 analysed simultaneously and rapidly. Combined PCP– mass spectrometry approaches have been used to characterize the interactome of a number of cell lines in combination with other protein properties, or cellular states, such as protein isoforms, post-translational modifications[9], or the activation of signalling cascades[71]. A recent study also demonstrated that combining multiple chromatographic separations can 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 significantly 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 be generated instead by 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 insight 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 performed on a scale approaching even remotely 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 post-translational modifications.*** 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 roles of PTMs have concentrated on the analysis of a relatively small number of PTMs, such as the reversible phosphorylation of serine, threonine and tyrosine residues, or the acetylation or ubiquitylation of lysine residues. These studies have provided critical 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 ubiquitination has been greatly facilitated by key developments in both peptide fractionation using column-based chromatography and affinity matrices able to pull-down modified peptides in batch formats[86]. Modified peptides may be significantly enriched compared to non-modified peptides in chromatographic separations based on charge (ion exchange chromatography such as strong anion exchange[87]), for example, due to the added negative charge of the phosphate group(s). Separations based on hydrophilic interactions, such as HILIC, are also able to enrich phosphorylated and certain other modified peptides[88] due to the altered hydrophilicity provided by the modified residues. These separations are usually combined with batch-format affinity enrichment, for example, using acetyl-lysine-specific antibodies for acetylated peptides[89, 90]. Chemical probes are also being generated that can be incorporated into proteins instead of their normal modification group to facilitate the easy extraction of the modified protein, or peptide, using bio-orthogonal "Click" reactions[91]. For example, protein O-GlcNAcylation may 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 in the 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 extra separations have allowed the observation of proteins present in multiple complexes, but where the phosphorylated form of a protein is only found within one of these complexes[9]. Also, the same protein present in different tissues can vary greatly in the sites and stoichiometry of phosphorylation[95]. Such distinctions would have been difficult to disentangle without the extra levels of biological dimension separation.

The crosstalk between different post-translational modifications is also important as each modification could be considered a new protein property. This is exemplified by the complex set of modifications that may be imparted to histones – sometime termed the histone code - in order to modulate chromatin function[96]. However, many other biological systems, such as signalling networks[97] and developmental pathways[98], also display crosstalk between different modifications on the same

protein. Notably, recent methods facilitate the sequential analysis of peptide mixtures generated from the same sample that are altered by many different post-translational modifications[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 arginine[100] and hydroxylation of proline[101] residues, has also emerged. For example, citrullination of a single arginine residue on the H1 linker histone excludes it from chromatin interactions, ultimately leading to chromatin decondensation, thereby regulating pluripotency[100]. Hydroxylation of a specific proline residue on the centrosomal protein CEP192 (centrosomal protein of 192 kDa) targets it for ubiquitylation and subsequent degradation; this is an important mechanism for regulating centrosomal duplication and consequently cell-cycle progression[101]. This is a similar role to that of targeted proline hydroxylation in the regulated degradation of the transcription regulator hypoxia-inducible factor 1α (HIF1α)[102].

In summary, the examples above highlight the multi-dimensional 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. We review below some of the technologies and resources that are currently available for the analysis and sharing of proteomics data and discuss the potential biological insight 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 variety 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 quantitation 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 allows for the analysis of selected reaction monitoring (SRM), DIA and many other data types. Meanwhile, the Census software[107, 108] (and associated tools) provides an alternative, multifunctional package that, in particular, allows for the analysis of data sets derived from $^{15}$N-labelling experiments. The development of mass spectrometry instruments that can generate MS3-level quantitative reporter ion data in TMT/iTRAQ experiments (**Supplementary Information S1**) 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 quantitation that include 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, quantification) of peptides based on 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. A number of statistical analysis packages are available for these purposes, including SPSS and R[109] with its associated Bioconductor[110] suite of tools for bioinformatics analysis. The free-to-use R program combines the ability to perform database, statistical and graphing functions and can handle very large multi-dimensional datasets. In addition, many biology-specific packages are available for R, such as pRoloc, which is a tool for the analysis of protein localisation using the protein correlation profiling approach[111]. One potential down-side of statistical packages such as R however is that they have a significant learning curve before most cell biologists can take full advantage of the features offered.

Creating a wider base of user-friendly tools, without such a steep learning curve, for the interactive analysis and integration of multi-dimensional 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 specialised statistical analysis role to estimate probabilities that two proteins interact.

***Data visualisation and integration.*** While we propose in this review that a multi-dimensional 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 is a wealth of pre-existing datasets 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 datasets 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 datasets with a large variety of stored prior-knowledge. DAVID's prior knowledge includes Gene Ontology terms, transcription factor binding sites, disease association, protein interactions and cellular pathways, to name only a few. While these data are rich and each

association is strongly backed up by statistical evidence, there still is a need to improve built-in visualisation tools for the data output by DAVID.

The Cytoscape[117] platform is a useful tool that can provide integration of external databases with user datasets and output these analyses in rich graphical formats. Many groups have provided apps for use in Cytoscape (see apps.cytoscape.org), which facilitate multi-dimensional 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 dataset, based on quantified changes detected in an experiment, such as protein abundance. These data can be output from GSEA for rich visualisation in the Cytoscape app called EnrichmentMap[119]. Such visualisation can help the user to derive otherwise hidden biological meaning from their data.

***Data-sharing resources.***

Effective sharing of proteomics data that can be easily searched and accessed by the cell biology community, is another area where 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. Major challenges remain in providing the data outputs of multi-dimensional 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, such as DNA and RNA 'omics' data, metabolic pathway and 3D protein structure data (**Table 2**). We have recently introduced such a resource, the "Encyclopedia of Proteome Dynamics"[4] to share our data with the community in a web-based and user-friendly interface (**Supplementary Information S2a**). 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 (**Supplementary Information S2b**), as determined both by mass spectrometry -based proteomics[7, 8] and antibody-based detection[121, 122]. **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 contribute to the establishment of integrated, multi-dimensional datasets 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 better accustomed to proteomics experiments. This 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

We have described here 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 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-transnationally-modified states, or interaction networks, which in combination will facilitate mechanistic insight into protein function and regulation. Ultimately the combination of these cell biology methods with mass spectrometry analysis will 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, multi-dimensional view of the proteome will not only be critical for enhancing our understanding of basic cellular physiology and regulation, but also important 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/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 anticipate that further advances in mass spectrometry -based multi-dimensional proteomics will greatly influence all aspects of future cell biology experimentation.

## Acknowledgements

AIL is a Wellcome Trust Principal Research Fellow & ML is a Royal Society of Edinburgh - Scottish Government Personal Research Fellow.

## Boxes

**Box 1. Azido-homoalanine (AHA) and click-chemistry.** Feeding azido-homoalanine (AHA) to cells in short pulses results in the incorporation of this amino acid into proteins, replacing methionine (MET), and thus enables the labelling of newly synthesized proteins. 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 (‘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.

**Box 2. Adding Extra Dimensions to Proteome Data with the STRING database.** The search tool for the retrieval of interacting genes/proteins **[CE: The forward-slash is part of the official name;**

please leave as is.] **[Au: Please ignore the comment in green – it's for the copy editors.]** (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 allows the user to search for a protein of interest — the 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. 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 blue nodes (right). Together, these data could provide clues to regulatory elements within a pathway and add value to the data provided by each database.

## Figure Legends

**Figure 1. Multi-dimensional proteome analysis of cells and tissues. a)** Proteins can display many different properties (dimensions) that are either largely physically (green shaded area), chemically (orange shaded area) or biologically (grey shaded area) relevant. Shown here 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 multi-dimensional analysis. **b**) A series of stacked cubes is shown, each of which contains a discrete pool of proteomic data that corresponds to the value of each dimension (localization, cell cycle phase and turnover rate). For each cube (see blow-ups) we can analyse other dimensions such as protein activity, total protein abundance and phosphorylation stoichiometry. Together these visually represent an approach for the multi-dimensional analysis of protein data. The pink 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 the nucleus, displaying 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).

**Figure 2. Methods for protein turnover analysis.** Proteome-wide turnover is typically measured using one of the two approaches illustrated here. Method 1 involves pulse-labelling of amino acids using either isotope labelling by amino acids in cell culture (SILAC), or $^{15}$N-labelling. The cells start with proteins (stars) containing 'light' stable isotopes (brown) and for various periods of time are switched into media with a 'heavy' isotope that is stably incorporated into specifc amino acids and thus labels newly synthesise proteins. Proteins with rapid turnover rates (here for example cytosolic proteins (**a)**) will rapidly incorporate high levels of the 'heavy' isotope, whereas protein pools that have slower turnover rates (from the membrane (**b**) and nucleus(**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 cycloheximide to block protein

synthesis in live cells for various periods of time. The comparison of protein abundance between untreated and treated cells allows a calculation of the depletion rate of a protein in the cells, which may occur either due to its degradation, 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 either Methods 1 or 2, the use of cellular fractionation (into subcellular compartments in this case) can greatly increase the information gained compared to the analysis of total cell lysates (**d**). The protein depicted here is effectively stable in the nuclear compartment (**c**), has a slow turnover rate in the membrane-associated pool (**b**) and a fast turnover rate in the cytosol (**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 is the mass to charge ratio of each peptide ion as measured by mass spectrometers.

**Figure 3. Approaches for the analysis of protein interactions.** Three main approaches exist for unbiased analysis of protein-protein interactions.

(**a**) The first approach, based on affinity pulldown and isolation, uses either specific antibodies to an endogenous protein, or 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 non-specific (green circles) binding partners present in the eluted mixture.

(**b**) The second approach is based on proximity labelling, where 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 (pink 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 (**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 non-specific label transfer or pull-down (green circles).

(**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 could 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 to 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 density of each identified complex.

Abbreviations: HIC, hydrophobic interaction chromatography; IEX, ion-exchange chromatography; LC-MS/MS, liquid chromatography – tandem mass spectrometry; SEC, size-exclusion chromatography.

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

| Dimension | Examples of techniques used |
|---|---|
| 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[126] (for DNA content or phase-specific markers) |
| Tissue distribution | Dissection[95, 127]; FACS[126] (for cell-type specific markers) |
| Interactions | Affinity-enrichment (endogenous IP or tagged fusion protein pull-down)[63-67]; Protein correlation profiling[9, 70, 71], Proximity-labelling[39, 68] |
| Post-translational modifications | Affinity enrichment (TiO2[128, 129], IMAC[128, 130], modification-specific antibodies[90, 131-133]); Chromatography (IEX[87], HILIC[94], ERLIC[134]) |
| Localisation | Centrifugation[3, 43, 135], Affinity-enrichment; 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. Typically achieved by peptide-level fractionation prior to LC-MS/MS using either ion-exchange[136], or high-pH reverse-phase[137] chromatography. Alternatively, targeted MS analysis may be used to detect isoform-specific peptides. |
| 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]. |

Abbreviations: FACS, Fluorescence-associated cell sorting; SILAC, stable isotope labelling by amino acids in cell culture; TMT, tandem mass tags; iTRAQ, isobaric tags for relative and absolute quantitation; IMAC, immobilised metal affinity chromatography; IEX, ion-exchange chromatography; HILIC, hydrophilic interaction chromatography; ERLIC, Electrostatic repulsion hydrophilic interaction chromatography.

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

| Resource | Key features | Strengths | Comments |
|---|---|---|---|
| **Protein dimension annotation resources** | | | |
| Encyclopedia of Proteome Dynamics (EPD)[4] peptracker.com/epd | Graphical display of protein dimension data; diverse dimensions analysed | Diverse dimensions analysed including multi-dimensional data sets. | Single lab as source of data |
| Human Protein Atlas[122] proteinatlas.org | Proteomic annotation of protein abundance and localization | High sensitivity; localisation is analysed within many tissues. | Antibody-based; single lab as source of data; human data only |
| Human Proteome Map[7] humanproteomemap.org | Human proteome data annotated for tissue abundance | Whole pathway or protein family analysis. | One protein dimension; human data only |

| | | | |
|---|---|---|---|
| The MaxQuant Database (MaxQB)[142] maxqb.biochem.mpg.de/mxdb | Proteome data tabulated; graph for label-free abundance measurement | Quality control parameters for mass spectrometry acquisition are presented. | Single lab as source of data |
| The Multi-Omics Profiling Expression Database (MOPED)[143] moped.proteinspire.org | Allows search and visualizations of protein data derived from multiple species | Many different experiments can be visualised and compared. | Chromosome-centric |
| The proteins across organisms database (PaxDB)[144] pax-db.org | Absolute protein abundance values determined across many organisms and tissues. | Abundance histogram; STRING integration. | Basic user interface |
| Phosphorylation site database (Phosida)[145] www.phosida.com | Provides data on phosphorylation, acetylation, and N-glycosylation of proteins; includes EGF-treated, cell-cycle regulated, kinome-related data sets | Diverse dimensions analysed with high depth of coverage for phosphorylation data sets. | Single lab as source of data |
| ProteomicsDB[8] https://www.proteomicsdb.org | Human proteome data annotated for tissue abundance; mass spectrometry spectra shown | Tissue protein abundance pattern and mass spectrometry spectral annotation; Multi-protein analysis. | Human data only |
| **Mass spectrometry -based raw proteomics data repositories** | | | |
| Chorus chorusproject.org | 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 (GPMDB)[146, 147] gpmdb.thegpm.org | Allows search and visualization of mass spectrometry data derived from many species | Rich graphical interface for mass spectrometry data visualization | MS2 spectral validation emphasized |
| ProteomeXchange Consortium proteomexchange.org Includes: The Proteomics Identifications [PRIDE], PeptideAtlas, PeptideAtlas SRM Experiment Library [PASSEL], Mass spectrometry Interactive Virtual Environment [MassIVE] | Allows 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 |

Abbreviations: EBI, European Bioinformatics Institute; EGF, epidermal growth factor; MS2, tandem mass spectrometry; SRM, selected reaction monitoring

# References


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. Proteomics.* **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**, 575-81 (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. *Annual Review of Biomedical Engineering.* **11**, 49-79 (2009).
13. Breker, M. & Schuldiner, M. The emergence of proteome-wide technologies: systematic analysis of proteins comes of age. *Nat. 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. *Nat. 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.* (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. Proteomics.* **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. *Nat. 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. *Nat. 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. *Nat. 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. *Nat. 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.* **24C**, 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. U.S.A.* **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** (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. *Shotgun Proteomics: Methods and Protocols.* **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., Moresco, J.J., Tu, P.G., Yates, J.R., 3rd & Nardulli, A.M. 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**, 284-92 (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. *Nat. Rev. Mol. Cell Bio.* **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 reports.* **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**, 901-12 (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**, 141-7 (2002).
65. Ho, Y. et al. Systematic identification of protein complexes in Saccharomyces cerevisiae by mass spectrometry. *Nature.* **415**, 180-3 (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**, 365-70 (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. *Nat. 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. *Nat. Commun.* **4**, 1985 (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. *Nat. Protoc.* **9**, 120-137 (2014).
74. Fischer, L., Chen, Z.A. & Rappsilber, J. Quantitative cross-linking/mass spectrometry using isotope-labelled cross-linkers. *J. Proteomics.* **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. Proteomics.* **doi: 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. *Nat. 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. U.S.A.* **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**, 25890-906 (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. *Nat. 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.* **2007**, cm8 (2007).
103. Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* **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. *Nat. Methods.* **5**, 319-322 (2008).
109. R Core Team. R: A Language and Environment for Statistical Computing. (R Foundation for Statistical Computing, Vienna, Austria, 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. *Nat. Methods.* **8**, 70-73 (2011).
113. Liu, G. et al. ProHits: integrated software for mass spectrometry-based interaction proteomics. *Nature Biotechnology.* **28**, 1015-7 (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. Proteomics.* **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. *Nat. 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. *P. Natl. Acad. Sci. U.S.A.* **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** (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. *Nat. Biotechnol.* **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** (2015).
126. Bernas, T., Gregori, G., Asem, E.K. & Robinson, J.P. Integrating cytomics and proteomics. *Mol. Cell. Proteomics.* **5**, 2-13 (2006).
127. Geiger, T. et al. Initial Quantitative Proteomic Map of 28 Mouse Tissues Using the SILAC Mouse. *Mol. Cell. Proteomics.* **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. Proteomics.* **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. *Nat. Biotechnol.* **28**, 868-U154 (2010).

133. Lamoliatte, F. et al. Large-scale analysis of lysine SUMOylation by SUMO remnant immunoaffinity profiling. *Nat. Commun.* **5** (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**, 55-+ (2014).
139. Banko, M.R. et al. Chemical genetic screen for AMPKalpha2 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. *Nat. Biotechnol.* **32**, 1036-+ (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. Proteomics.* **11**, 492-500 (2012).
145. Gnad, F., Gunawardena, J. & Mann, M. PHOSIDA 2011: the posttranslational modification database. *Nucleic Acids Res.* **39**, D253-260 (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 and GPMDB Protein Observation Frequency for Improving the Sensitivity of Protein Identification by Tandem MS. *J. Proteome Res.* **13**, 4113-4119 (2014).