

The biological impact of mass-spectrometry-based proteomics

Benjamin F. Cravatt^{1,2}, Gabriel M. Simon^{1,2} & John R. Yates III¹

In the past decade, there have been remarkable advances in proteomic technologies. Mass spectrometry has emerged as the preferred method for in-depth characterization of the protein components of biological systems. Using mass spectrometry, key insights into the composition, regulation and function of molecular complexes and pathways have been gained. From these studies, it is clear that mass-spectrometry-based proteomics is now a powerful 'hypothesis-generating engine' that, when combined with complementary molecular, cellular and pharmacological techniques, provides a framework for translating large data sets into an understanding of complex biological processes.

Of the many fascinating discoveries made by genome-sequencing projects, perhaps none is more provocative than the prediction that all prokaryotes and eukaryotes produce numerous proteins with uncharacterized or pleiotropic functions^{1,2}. Confronted with the challenge of annotating this enormous segment of the proteome, scientists have sought to expedite the characterization of proteins by developing new methods for rapid and parallel analysis. These large-scale approaches to protein science are collectively termed proteomics^{3,4}.

Proteomics encompasses diverse techniques that allow different aspects of protein structure and function to be analysed. Many proteomic methods — including protein microarrays^{5,6}, large-scale two-hybrid analyses⁷, and high-throughput protein production and crystallization⁸ — have had a marked impact on the current understanding of protein structures, activities and interactions. Among proteomic techniques, however, mass spectrometry has emerged as the main method for analysing the production and function of proteins in native biological systems^{9–11}.

Mass spectrometry has become the dominant technique for several reasons, mainly because of its unparalleled ability to acquire high-content quantitative information about biological samples of enormous complexity. The core technologies of mass-spectrometry-based proteomics, including the instrumentation and the methods for data acquisition and analysis, have been discussed in several recent reviews^{9–11} and are outlined in Box 1. Although these technologies will continue to be developed in the quest for improved sensitivity, throughput and proteome coverage, mass-spectrometry-based proteomics has now developed to the point at which it is routinely applied worldwide to address a large range of biological problems. It therefore seems an opportune time to reflect on the functional impact of mass-spectrometry-based proteomics. What has been learned about the molecular mechanisms of complex biological processes? How were successful experiments carried out? What additional methods were required to make important biological discoveries? Finally, are there lessons that might guide future applications?

In this review, we address these questions by focusing on several cases in which mass-spectrometry-based proteomics has had a crucial role in advancing our understanding of basic cellular and physiological processes. We highlight common themes that seem to have primed investigations for success, including the configuration of biologically relevant model systems (and controls), the implementation

of mass-spectrometry-based proteomics as a hypothesis-generating platform, and a commitment to focused follow-up studies that test emerging hypotheses directly. These examples underscore both the opportunities and the challenges that face the systematic integration of mass-spectrometry-based proteomic techniques into the arsenal of experimental approaches used by molecular and cellular biologists.

Mass-spectrometry-based proteomics has several biological applications. In many pioneering studies, it was used to make an inventory of the content of subcellular structures and organelles, creating valuable repositories of information about the localization of proteins in cells and tissues (see ref. 12 for a recent review). It is also emerging as a powerful way to discern higher-order structural features of protein complexes, including subunit orientation and stoichiometry (see page 973). Here, we focus on two main applications — the functional characterization of protein complexes, and the functional characterization of protein pathways — highlighting studies that have led to major advances in understanding the molecular basis of cellular and physiological processes.

Functional characterization of protein complexes

Many proteins function as components of complexes in cells and tissues¹³. Protein complexes can vary in size and composition, from megadalton assemblies of dozens of proteins (such as the ribosome and the spliceosome) to smaller clusters of a few proteins. The composition and stability of protein complexes is highly regulated in both a context-dependent manner (for example, there are cell-type-specific differences) and a time-dependent manner¹⁴. These biological variables present a challenge to researchers interested in determining the structure and the function of protein complexes. Mass-spectrometry-based proteomics, however, can address this issue in a systematic and relatively unbiased manner, often revealing surprising protein partnerships and assemblies that regulate cellular and physiological processes. In addition to the examples discussed in this section, other notable studies that have used mass-spectrometry-based proteomics to characterize protein complexes are described in refs 15–19.

A mitochondrial protein complex that links apoptosis and glycolysis

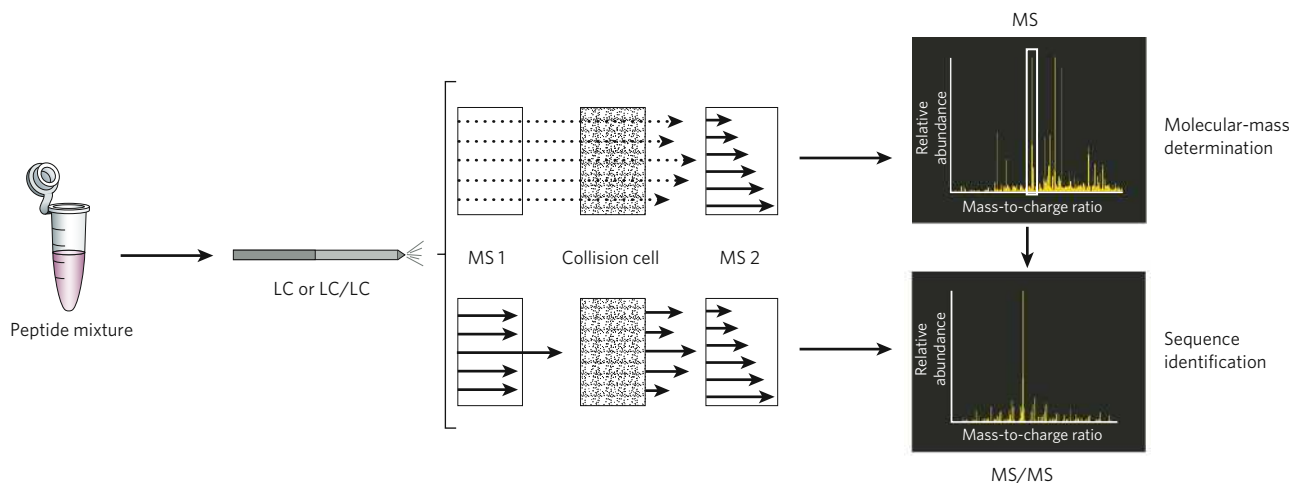
Stanley Korsmeyer and colleagues²⁰ provided an early example of the value of mass-spectrometry-based proteomics for uncovering unexpected

¹Department of Chemical Physiology, ²The Skaggs Institute for Chemical Biology, The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, California 92037, USA.

physical connections between proteins that had been thought to function in independent pathways. With the aim of mapping mitochondrial protein complexes that contain the pro-apoptotic protein Bcl-2 antagonist of cell death (BAD), they developed a gentle and efficient enrichment method to isolate these complexes from mouse liver mitochondria (Fig. 1a). A 232-kDa assembly consisting of five major proteins was identified by silver staining of proteins separated by polyacrylamide gel electrophoresis

(PAGE). The protein spots constituting this complex were then excised and analysed by liquid-chromatography–tandem mass spectrometry (LC–MS/MS). This revealed, in addition to BAD, the presence of protein phosphatase 1C (PP1C), cyclic-AMP-dependent protein kinase (PKA) and the PKA-anchoring protein WAVE1. Previous studies had implicated PKA as an important regulator of BAD, inhibiting the activity of BAD by phosphorylating multiple serine residues²¹. The authors provided

Box 1 Fundamentals of mass-spectrometry-based proteomic experiments



Mass-spectrometry-based proteomic experiments involve several steps (see Figure). A peptide mixture can be obtained from the sample of interest by proteolytic digestion of a protein mixture or of a gel band or spot, following separation by electrophoresis. These peptides are then introduced into a one-dimensional (LC) or multi-dimensional (LC/LC) liquid-chromatography system. After separation, they are eluted into an electrospray ionization tandem mass spectrometer. The mass-to-charge ratios of the peptide ions are measured first by mass spectrometry (upper panels; ions pass unperturbed through the first mass analyser and collision cell) to determine the molecular mass of each peptide. Then, each peptide ion is isolated in the first mass analyser (MS 1) and directed into a collision cell, where it collides with neutral gas molecules (for example, helium) and becomes fragmented (lower panels). The mass-to-charge ratios of the resultant fragments are measured in the second mass analyser (MS 2), producing a tandem mass spectrum (shown for the peptide ion indicated with a white rectangle in the upper panel) and, after computer analysis, an amino-acid sequence for each peptide. These steps are described in further detail below.

Sample preparation

Two general strategies are often used to prepare proteins. Proteins that have been enriched or obtained as part of an experiment can be fractionated by SDS–polyacrylamide gel electrophoresis (SDS–PAGE). Individual bands can be removed and analysed, or an entire lane can be excised and divided into 10–15 slices. Proteins in the gel slices are then digested *in situ* with trypsin, and the peptides are extracted. Extracted peptides are then analysed by mass spectrometry. Protein mixtures can also be digested directly in solution. A protein mixture is denatured by using chaotropes and then digested — sometimes by a two-step procedure that involves proteases, such as the endoprotease LysC followed by trypsin — to generate a peptide mixture that is suitable for mass-spectrometry analysis. In general, trypsin digestion is preferred to generate peptides with an arginine or lysine residue at the C terminus, but other types of enzyme, including nonspecific proteases, have also been used⁷⁸.

Liquid chromatography

Before peptide mixtures are introduced into the mass spectrometer, they are fractionated in-line with the instrument. The most common method of fractionation is reversed-phase liquid chromatography,

which separates peptides according to their hydrophobicity. To achieve the best sensitivity and efficiency of separation, microcolumns (<100 μm in length) with a small diameter tip (for example, 5 μm) are typically used. The electrospray ionization that takes place in the mass spectrometer acts like a concentration-dependent detector; therefore, the introduction of peptides in narrow peaks improves detection limits, and low flow rates (in the order of nL min^{-1}) are used to achieve this. As peptide mixtures become more complex, the introduction of a second dimension of separation can improve the resolution of separation. A good choice for a second dimension is strong cation-exchange (SCX) liquid chromatography, which separates peptides mainly on the basis of positive charges. SCX can be used off-line, and then each fraction is analysed by reversed-phase high-pressure liquid chromatography, followed by mass spectrometry. Alternatively, both the reversed-phase and SCX resins can be packed into a single column, and, by introducing buffers in series, a two-dimensional separation can be achieved before mass-spectrometry analysis.

Electrospray ionization

A potential is placed on the liquid flowing from the liquid-chromatography column through a fused silica column or needle, causing the solution to spray. The spray contains fine droplets that encompass the sample. The droplets are desolvated as they enter the mass spectrometer, by applying heat to generate ions. The efficiency of ionization depends on the chemical properties of each molecule.

Mass-spectrometry analysis

Mass spectrometers measure the mass-to-charge ratio of an ion. This is carried out by manipulating ions in electric and/or magnetic fields or by measuring their time of flight (TOF). In addition to determining the mass-to-charge ratio, the intensity of the signal obtained reflects the abundance of the ion. The abundance of ions can vary with the ionization, so samples can be labelled with stable isotopes to determine quantitatively the ratio of peptides from different 'states' (by measuring the mass-to-charge ratio and abundance). Various mass analysers are used in proteomic experiments, including ion-trap mass spectrometers, quadrupole/TOF hybrids, ion-trap/orbitrap hybrids and ion-trap/ion-cyclotron-resonance (FTMS) hybrids. Some types of mass analyser can measure the mass-to-charge ratio with high resolution (up to 150,000 $m/\Delta m$, where m denotes mass) and high mass accuracy (to <1 part per million).

strong evidence that PP1C functions to counter the effect of PKA, by dephosphorylating BAD. These data therefore led to a model in which a BAD–PKA–PP1C complex, possibly scaffolded by WAVE1, creates a local microenvironment in which the phosphorylation status of BAD can be finely controlled.

Korsmeyer and colleagues next considered the possible function(s) of this BAD-containing complex in mitochondrial physiology. First, they pursued the characterization of the fifth (and still unidentified) component of the complex. By LC–MS/MS analysis, this 50-kDa protein was identified as the glycolytic enzyme glucokinase (also known as hexokinase 4). This result was initially surprising; even though apoptosis and glycolysis are both crucial physiological processes that are linked to cell survival^{22,23}, the molecular pathways involved had been thought to function independently. The organization of glucokinase and BAD into a stable multiprotein complex in mitochondria indicated otherwise. Indeed, the authors showed that *Bad*^{-/-} mice, which lack the gene encoding BAD, had severely blunted mitochondrial glucokinase activity, glucose-driven respiration and glucose-dependent ATP production. Moreover, these mice had significantly higher blood glucose concentrations after fasting than did wild-type (*Bad*^{+/+}) mice. The effect of BAD on glucose homeostasis depended on its phosphorylation status, suggesting that other members of the BAD-containing complex (for example, PKA and PP1C) had a regulatory role.

These findings therefore indicate that the mitochondrial fraction of glucokinase — defined as that portion of the enzyme stably associated with the BAD–PKA–PP1C–WAVE1 complex (Fig. 1b) — has a disproportionate role in maintaining proper glucose metabolism (given that most of the glucokinase in a cell is cytosolic). The glucokinase- and BAD-containing mitochondrial complex was also proposed to function as an integration centre that links metabolic state and cell death. This hypothesis was supported by the finding that liver cells from *Bad*^{-/-} mice underwent less glucose-deprivation-induced apoptosis than wild-type liver cells. In summary, the mass-spectrometry-based proteomic analysis of mitochondrial BAD-containing complexes discovered an unexpected physical association between a key apoptotic protein (BAD) and a key glycolytic protein (glucokinase), thereby leading to new models to explain how cells coordinate metabolic signals and survival signals.

A transcription-factor complex relevant to trichothiodystrophy

The proper maintenance, repair and transcription of DNA requires several multiprotein complexes²⁴. Ruedi Aebersold and colleagues²⁵ were interested in fully characterizing the components of the yeast (*Saccharomyces cerevisiae*) RNA polymerase II (PolII) pre-initiation complex and established an elegant biochemical system to enrich this protein assembly (Fig. 2). They first isolated nuclear extracts from a mutant yeast strain carrying a temperature-sensitive allele of the TATA-binding protein (TBP) and incubated these proteomes with an immobilized *HIS4* promoter from yeast, which includes the TATA box, in the presence or absence of recombinant TBP. They then used isotope-coded affinity tagging (ICAT) in conjunction with LC–MS/MS²⁶ to identify proteins that were quantitatively enriched (by at least 1.9-fold) in promoter-associated fractions from TBP-containing proteomes. Nearly all of the proteins that met this criterion were known components of the PolII transcriptional machinery, with the notable exception of an open reading frame — YDR079C-A (Fig. 2) — which corresponded to a small (8 kDa) protein of unknown function.

BLAST searches (<http://www.ncbi.nlm.nih.gov/blast/Blast.cgi>) revealed homologues of YDR079C-A in several eukaryotic organisms, including humans and *Chlamydomonas reinhardtii*. Interestingly, the protein encoded by *C. reinhardtii* was known to suppress sensitivity to ultraviolet light (which can damage DNA)²⁷. This finding suggested that the protein encoded by YDR079C-A might be a component of transcription factor IIH (TFIIH), which has a role in both general transcription and repair of DNA damage. Aebersold and colleagues²⁵ confirmed this hypothesis by carrying out a further round of quantitative proteomic experiments, this time comparing proteins that bound to an epitope (FLAG)-tagged YDR079C-A protein. The proteins that were most

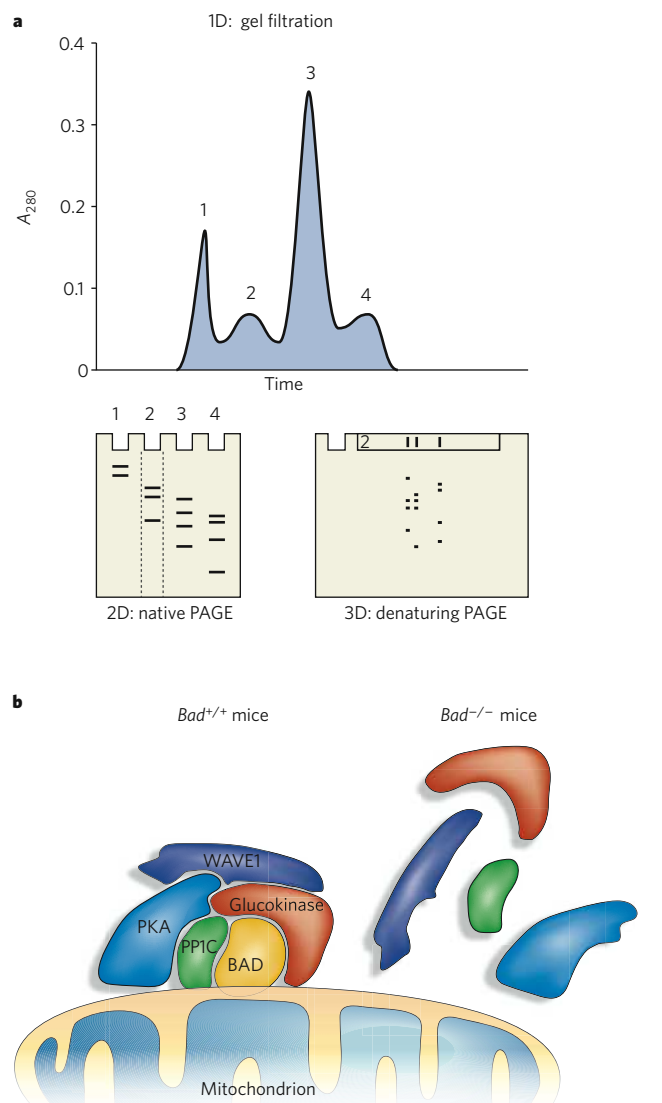


Figure 1 | Discovery of a BAD-containing protein complex that localizes to mitochondria and integrates apoptotic and glycolytic processes.

a, A complex containing BAD, PKA, PP1C, WAVE1 and glucokinase was discovered by multidimensional fractionation of mitochondrial protein complexes from *Bad*^{+/+} (wild-type) mice and *Bad*^{-/-} mice, followed by LC–MS/MS analysis²⁰. Multidimensional fractionation involved gel filtration as the first dimension (1D), native (non-denaturing) PAGE as the second dimension (2D), and denaturing PAGE (SDS–PAGE) as the third dimension (3D). **b**, In the absence of BAD, none of the members of the protein complex associates with mitochondrial membranes.

enriched after immunoprecipitation of FLAG–YDR079C-A protein with FLAG-specific antibodies were components of TFIIH. Conversely, immunoprecipitation with antibodies specific for other TFIIH components ‘pulled down’ the YDR079C-A protein. Finally, YDR079C-A protein was shown to be required for stable recruitment of TFIIH to promoters. These findings confirmed that the YDR079C-A protein is a core subunit of TFIIH, prompting the authors to rename the protein as the transcription factor subunit Tfb5.

In an amazing example of a basic scientific discovery being rapidly translated, the human homologue of Tfb5 was almost immediately shown to be the long-sought gene product that is mutated in a set of unexplained cases of trichothiodystrophy²⁸, a rare human disease characterized by brittle hair and skin photosensitivity. Most cases of trichothiodystrophy had been traced to mutations in the genes encoding the nine known components of TFIIH. However, an individual with symptoms of trichothiodystrophy but no mutations in these genes had been found several years earlier²⁹. Interestingly, cells from patients

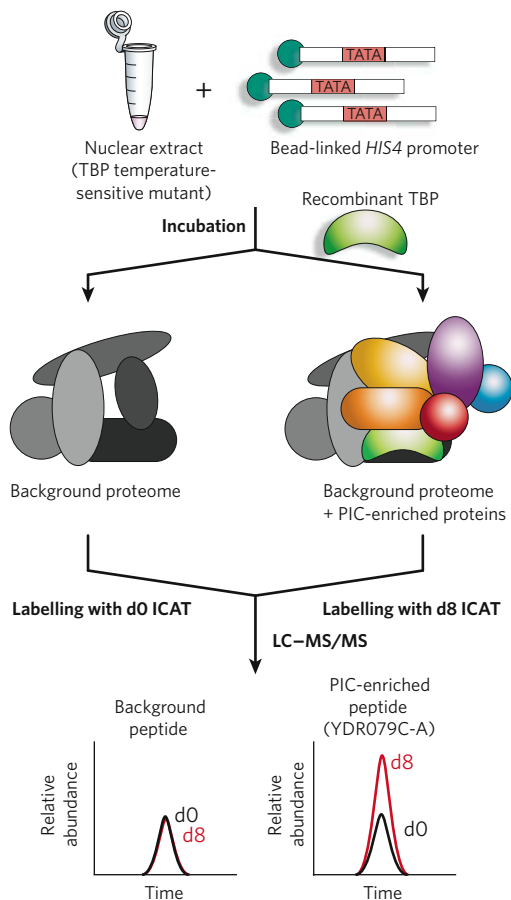


Figure 2 | Discovery of Tfb5 as the tenth subunit of TFIIF, which is involved in transcriptional and DNA-repair processes. Nuclear extracts from yeast expressing a temperature-sensitive mutant of TBP were incubated with *HIS4*-promoter-linked beads, in the presence or absence of recombinant TBP. The inclusion of recombinant TBP facilitated enrichment for proteins that are members of PolII pre-initiation complexes (PICs). Samples enriched in the absence or presence of recombinant TBP were then labelled with non-deuterated (d0) and deuterated (d8) ICAT probes, respectively, enabling LC-MS-based quantitative proteomic analysis of differentially enriched proteins. This led to the identification of YDR079C-A (subsequently named Tfb5) as a component of the transcription-factor complex TFIIF²⁵. Complementary genetic studies (not shown) confirmed that the gene encoding the human homologue of Tfb5 is mutated in rare forms of trichothiodystrophy²⁸.

with this unusual variant of trichothiodystrophy (called trichothiodystrophy A) were found to have low cellular concentrations of TFIIF²⁹. On discovery of Tfb5 as the tenth component of yeast TFIIF, Wim Vermeulen, Aebersold and colleagues sequenced the corresponding human gene in patients with trichothiodystrophy A and found inactivating mutations in four individuals from three separate families²⁸. Moreover, they showed that recombinant human TFB5 could stabilize TFIIF complexes and correct the DNA-repair defects in cells from patients with trichothiodystrophy A.

In summary, the use of mass-spectrometry-based proteomics to characterize a previously unknown component of TFIIF catalysed a remarkable bench-to-bedside-to-bench investigation that succeeded in explaining the molecular basis for a human photosensitivity syndrome. On a technical note, it is worth emphasizing that this discovery hinged on the use of quantitative profiling, which allowed the researchers to confidently identify Tfb5, despite its showing only a moderate (twofold) increase in abundance in promoter-associated samples compared with control samples. Direct LC-MS/MS analysis also proved crucial, because the small size of Tfb5 (8 kDa) precluded straightforward detection by SDS-PAGE (and might explain why this protein eluded detection by more-classical methods).

Finally, the interaction of Tfb5 with another component of TFIIF, Tfb2, was recently confirmed in a genome-wide tandem-affinity-purification study³⁰, thus underscoring the capacity of large-scale mass-spectrometry-based proteomic experiments to characterize physiologically relevant protein complexes.

A chaperone complex that regulates CFTR folding and transport

Transmembrane proteins depend on a complex range of chaperones and co-chaperones for optimal folding, localization and, ultimately, function^{31,32}. Genetic mutations that impair the folding of membrane proteins form the basis of many human diseases, including cystic fibrosis. Cystic fibrosis is mainly caused by point mutations in the gene encoding an apical membrane ATP-regulated chloride channel, which is known as the cystic fibrosis transmembrane conductance regulator (CFTR)³³. The main disease-associated mutation, $\Delta F508$ (deletion of the phenylalanine residue at position 508 of the wild-type protein), disrupts the folding of CFTR in the endoplasmic reticulum, leading to almost complete degradation of this channel³⁴ (Fig. 3a). Interestingly, however, properly folded CFTR with this mutation can traffic to the plasma membrane, where it forms a functional chloride channel. These findings suggest that rescuing the folding of $\Delta F508$ -CFTR could eventually be used to treat patients with cystic fibrosis.

Initial studies have implicated both chaperone assemblies that contain heat-shock protein 90 (HSP90) and those that contain HSP40 and HSP70 in the folding pathways for CFTR^{35,36}. William Balch, John Yates and colleagues³⁷ proposed that a more complete understanding of the chaperone assemblies that regulate CFTR folding and transport could be achieved by carrying out a proteomic analysis of the proteins associated with the channel. The authors used the shotgun LC-MS method MudPIT (multidimensional protein-identification technology)³⁸ to analyse cells expressing the gene encoding the wild-type CFTR or $\Delta F508$ -CFTR. MudPIT analysis of wild-type CFTR and $\Delta F508$ -CFTR immunoprecipitates identified nearly 200 CFTR-associated proteins (compared with controls in which nonspecific antibodies or cells lacking CFTR were used). Collectively, these proteins have been named the CFTR interactome (Fig. 3b). These proteins included known CFTR-binding chaperones, such as calnexin, HSP40-HSP70 and HSP90, as well as many previously unknown interactors.

These researchers next set out to test whether any of the newly discovered CFTR-associated proteins regulated channel folding and export from the endoplasmic reticulum³⁷. RNA-interference (RNAi)-mediated knockdown of either p23 (also known as PTGES3) or FKBP8 — two HSP90 co-chaperones that were selectively identified in $\Delta F508$ -CFTR immunoprecipitates (as determined by protein-sequence coverage and spectral counting in MudPIT experiments) — resulted in greatly reduced amounts of endoplasmic-reticulum-associated and cell-surface-associated $\Delta F508$ -CFTR (Fig. 3c). Overexpression of these co-chaperones, however, had opposite effects on $\Delta F508$ -CFTR, with an increase in p23 leading to more endoplasmic-reticulum-associated CFTR and an increase in FKBP8 leading to less. These data were thought to reflect the distinct roles of p23 and FKBP8 in modulating specific aspects of the HSP90-guided folding cycle. Notably, overexpression of either co-chaperone failed to increase the amount of cell-surface-associated $\Delta F508$ -CFTR (or wild-type CFTR), suggesting that they modulate the initial folding and stability of CFTR but do not participate in the subsequent steps that are required to deliver the folded channel to the endoplasmic-reticulum export machinery.

RNAi-mediated knockdown of a third HSP90 co-chaperone present in wild-type and $\Delta F508$ -CFTR immunoprecipitates, AHA1, substantially corrected the amount of both endoplasmic-reticulum-associated and cell-surface-associated $\Delta F508$ -CFTR (Fig. 3c). These data suggest that disruption of AHA1, unlike p23 or FKBP8, facilitates a folding pathway that favours not only stability of the channel but also coupling to the endoplasmic-reticulum export machinery. Potentially consistent with this premise, a decrease in CFTR-bound HSP90 was observed in cells in which AHA1 had been knocked down, similar to the finding from MudPIT analyses that wild-type CFTR immunoprecipitates contained less HSP90 than $\Delta F508$ -CFTR immunoprecipitates. Collectively, these data

indicate that a reduction in the amount of AHA1 might alter the kinetics of HSP90–CFTR interactions, thereby increasing the efficiency of transition from folding to export pathways. Finally, the authors showed that a considerable proportion of the cell-surface-associated $\Delta F508$ -CFTR channels were functional, as determined by chloride conductance measurements.

Mass-spectrometry-based proteomic studies of the CFTR interactome thus identified specific co-chaperone and chaperone folding pathways that seem to control mutant channel stability, cell-surface expression and function. Why might the basal chaperone machinery of a cell, or ‘chaperome’, prevent proper assembly and transport of $\Delta F508$ -CFTR, thereby exacerbating the disease phenotype? The authors³⁷ speculate that the folding energetics of $\Delta F508$ -CFTR lie outside the capacity of the normal chaperome environment, which has been evolutionarily optimized to fold wild-type proteins (and to eliminate misfolded proteins). A provocative extension of this idea is that genetic or pharmacological interventions that shift the chaperome so that it can support the folding and transport of mutant proteins could be used to treat patients with cystic fibrosis, as well as those with other protein-conformational disorders.

Functional characterization of protein pathways

One of the original and most enduring applications of mass-spectrometry-based proteomics is the comparative analysis of biological samples that differ in specific physiological or pathophysiological phenotypes (that is, comparing ‘disease’ and ‘normal’³⁹). These studies are intended to identify the minimal protein ‘signatures’ that depict and, ideally, determine the higher-order biological processes under investigation. As highlighted in this section, mass-spectrometry-based proteomics carried out in this comparative analysis mode has successfully identified previously unknown protein pathways with key roles in a wide range of biological systems.

Kinase pathways that regulate sex-specific functions in *Plasmodium*

Malaria is caused by unicellular parasites of the genus *Plasmodium*. These parasites undergo a complex series of highly regulated life-cycle transitions that allow transmission between vertebrates and mosquitoes⁴⁰. Chief among these life-cycle stages is the generation of haploid sexually differentiated (male and female) cells, termed gametocytes. In vertebrate blood, gametocytes are in a state of arrest, but on transfer to the mosquito mid-gut, they become activated and differentiate into gametes, which fertilize and, eventually, produce infectious oocysts. Despite the importance of sexual development to the transmission of *Plasmodium* spp., the proteins that distinguish male and female cells have not been systematically inventoried. Andrew Waters and colleagues set out to address this important problem through an innovative combination of advanced cell-biological models and proteomic technologies⁴¹.

Previous efforts to characterize sex-specific proteins had been confounded by a technical inability to separate and purify male and female gametocytes. Waters and colleagues overcame this difficulty by creating transgenic lines of *Plasmodium berghei* that produce green fluorescent protein (GFP) under the control of a male-specific or a female-specific promoter (from the genes encoding α -tubulin II and elongation factor 1 α , respectively) (Fig. 4a). These lines enabled male gametocytes and female gametocytes to be selectively enriched by flow cytometry (Fig. 4b). These sex-specific cell populations were then compared with one another (and with gametocyte-free (asexual) blood stages) by mass-spectrometry-based proteomics. Specifically, proteomes were separated into ten fractions by one-dimensional SDS–PAGE, and each fraction was digested with trypsin and analysed by LC–MS/MS. Sex-specific proteins were identified by comparing the number of unique ‘tryptic peptides’ in each sample.

A remarkable number of sex-specific proteins were identified: there were 236 unique proteins in male gametocytes, and 101 in female gametocytes (Fig. 4c). Analysis of the sex-specific proteomes showed clear enrichment for protein families that are functionally linked to male-gametocyte and female-gametocyte biology. For example, nearly 70% of the *Plasmodium* proteins that are annotated as DNA-replication proteins (17 of 25) were highly represented in male gametocytes, which is

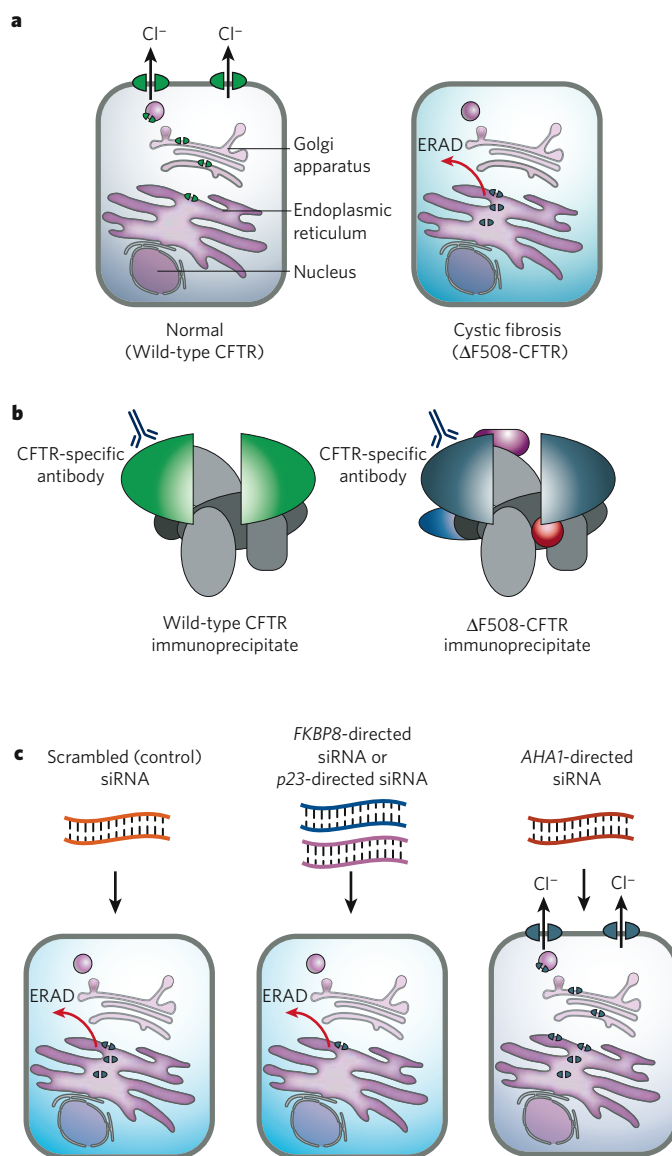


Figure 3 | Discovery of chaperone complexes that regulate CFTR folding and endoplasmic-reticulum-mediated export. **a**, The cellular fate of CFTR chloride channels is depicted. Wild-type CFTR is transported to the plasma membrane. By contrast, $\Delta F508$ -CFTR, the mutant protein present in individuals with cystic fibrosis, is degraded by endoplasmic-reticulum-mediated pathways (ERAD) before it reaches the plasma membrane. **b**, Immuno-enrichment of proteins bound to wild-type CFTR and $\Delta F508$ -CFTR identified several chaperones and co-chaperones³⁷. **c**, RNAi-mediated knockdown of these chaperones and co-chaperones was carried out, using small interfering RNA (siRNA) directed against the corresponding mRNAs. In the case of p23 and FKBP8 (centre), less $\Delta F508$ -CFTR was found in the endoplasmic reticulum, indicating that these proteins regulate the folding and stability of CFTR proteins in the endoplasmic reticulum. By contrast, in the case of AHA1 (right), more $\Delta F508$ -CFTR was found both in the endoplasmic reticulum and at the cell surface, indicating that this chaperone controls both the folding of CFTR proteins and their export from the endoplasmic reticulum³⁷.

consistent with the more extensive genome replication that these cells undergo during the gamete-activation cycle. The male-gametocyte proteome was also strongly enriched in axoneme proteins, which form the flagella required for motility of male gametes. The female-gametocyte proteome, by contrast, contained larger amounts of ribosomal and mitochondrial proteins than the male-gametocyte proteome.

Waters and colleagues next selected individual sex-specific proteins for functional analysis, to determine whether they have important

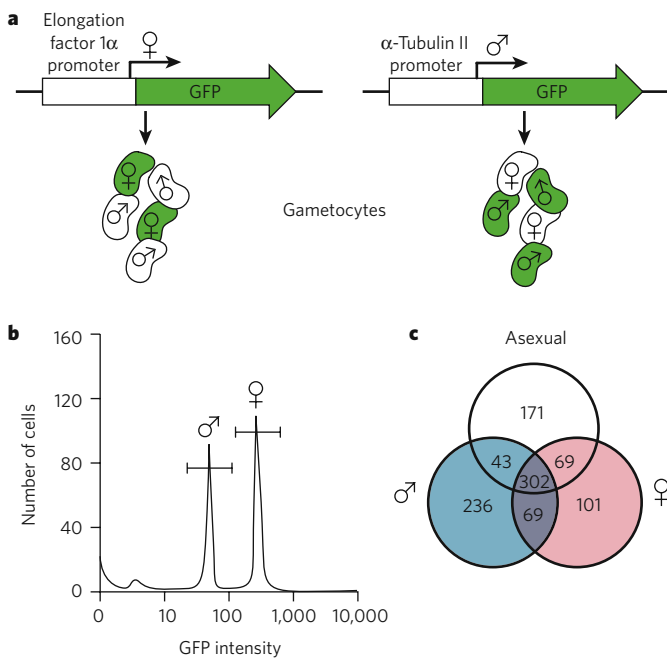


Figure 4 | Identification of male-gametocyte-specific and female-gametocyte-specific proteins in *Plasmodium berghei*. **a**, Transgenic *Plasmodium berghei* lines that produce GFP under the control of a female-gametocyte-specific promoter (from the gene encoding elongation factor 1 α) or a male-gametocyte-specific gene promoter (from the gene encoding α -tubulin II) were generated⁴¹. **b**, Enriched populations of female and male gametocytes were then obtained by flow cytometry, gating on GFP signals (shown for the line in which female gametocytes produce GFP). **c**, Comparative proteomic analysis of male-gametocyte-enriched, female-gametocyte-enriched and asexual populations identified 236, 101 and 171 proteins, respectively, that are expressed solely in each cell type. Among these proteins were identified two protein kinases that contribute to male-gametocyte-specific and female-gametocyte-specific cellular functions.

roles in male-gametocyte or female-gametocyte biology. The authors focused on two protein kinases: mitogen-activated protein kinase 2 (MAP2; accession number PB000659.00.0), which was found only in male gametocytes; and NIMA-related kinase (NEK4; accession number PB001094.00.0), which was found only in female gametocytes. Targeted disruption of the gene encoding MAP2 resulted in male gametocytes that can re-enter the cell cycle after activation and complete genome replication but fail to enter nuclear division. Disruption of the gene encoding NEK4, by contrast, did not seem to impair gamete formation but, instead, arrested zygote development. Cross-fertilization studies confirmed that the latter phenotype was due to defective female (but not male) gametes.

In summary, developing an innovative cell-biological strategy to enrich distinct sexual stages of the *P. berghei* life cycle allowed the generation of high-quality cellular models for in-depth analysis by mass-spectrometry-based proteomics. The output of the proteomic investigations was the most comprehensive inventory of sex-specific parasite proteins generated so far, including the discovery of novel protein kinases that regulate male-specific and female-specific signalling pathways. Interestingly, both MAP2 and NEK4 belong to protein-kinase subfamilies (the MAP and NEK subfamilies) that have multiple members in *Plasmodium* spp.⁴². These studies are therefore a compelling example of the value of comparative proteomics for assigning unique (that is, non-redundant) cellular functions to uncharacterized members of protein classes.

An ether-lipid signalling pathway that supports tumour pathogenesis

Cancer cells have long been suspected to have alterations in metabolism that support their malignant behaviour. Most cancer cells, for example, have a greater dependence on glycolysis than on oxidative phosphorylation for energy production, a phenomenon referred to as

the Warburg effect²². In an effort to map dysregulated biochemical pathways in cancer more globally, Benjamin Cravatt and colleagues used a chemical proteomic technology known as activity-based protein profiling (ABPP)⁴³, in conjunction with mass-spectrometry-based analytical platforms (such as MudPIT), to identify enzyme activities that are increased in aggressive cancer cell lines and primary tumours in humans⁴⁴.

In ABPP, active-site-directed probes are used to profile the functional state of enzymes directly in native proteomes⁴³. ABPP probes contain two main elements: a reactive group that binds to, and covalently labels, many enzymes from a given mechanistic class; and a reporter group, such as biotin or a fluorophore, that allows detection, enrichment and identification of probe-modified enzymes (Fig. 5a). In their initial studies, Cravatt and colleagues used fluorophosphonate-containing ABPP probes^{45,46} to profile the activities of serine hydrolases in a panel of human cancer cell lines⁴⁴. These experiments identified sets of enzyme activities that distinguished cancer cells on the basis of tissue of origin and state of aggressiveness. Chief among these enzymes was a previously uncharacterized transmembrane enzyme KIAA1363 (also known as AADACL1), increased amounts of which were found in aggressive lines from several tumour types, including breast cancer, ovarian cancer and melanoma (Fig. 5b). Cravatt and colleagues later showed by ABPP–MudPIT analysis that the activity of KIAA1363 is much higher in oestrogen-receptor-negative primary breast tumours from humans than in oestrogen-receptor-positive primary breast tumours, which are usually less aggressive, or in normal breast tissue⁴⁷.

Cravatt and colleagues next used a competitive version of ABPP⁴⁸ to develop a potent and selective inhibitor of KIAA1363, which they named AS115 (Fig. 5c). Treatment of cancer cells with this inhibitor, followed by metabolomic analysis using untargeted LC–MS methods⁴⁹, revealed that KIAA1363 regulates an unusual class of neutral lipids: the monoalkylglycerol ethers (MAGEs)⁵⁰ (Fig. 5d). Additional studies confirmed that KIAA1363 is a 2-acetyl-MAGE hydrolase, producing large amounts of MAGEs in aggressive cancer cells. These MAGEs are, in turn, converted into biologically active lysophospholipids, such as lysophosphatidic acid (LPA). By contrast, inhibiting KIAA1363 stabilizes 2-acetyl-MAGE, resulting in its conversion into another class of signalling molecule, the lipid platelet-activating factor. Finally, RNAi-mediated knockdown of the protein and activity of KIAA1363 led to a marked decrease in the amount of MAGE and LPA lipids in cancer cells, correlating with significant reductions in the migratory and tumour-forming potential of these cells (Fig. 5d).

In summary, Cravatt and colleagues used a combination of mass-spectrometry-based functional proteomic and metabolomic methods to determine that the enzyme KIAA1363 is more abundant and has a higher activity in aggressive cancer cells, where it is a key node that bridges platelet-activating factor and LPA in an ether-lipid signalling network. Considering that disruption of this network impaired cancer-cell migration and tumour growth, the KIAA1363–ether-lipid pathway probably has a key role in regulating important aspects of cancer pathogenesis.

Phosphoprotein networks involved in the DNA-damage response

Post-translational modifications constitute one of the most pervasive mechanisms for regulating protein function in cells and tissues. Protein phosphorylation, in particular, dynamically modulates numerous signalling pathways and is controlled by the complementary action of protein kinases and protein phosphatases. One of the big challenges in the post-genomic era is determining the endogenous substrates of the more than 500 protein kinases in the human proteome⁵¹. Recently, Stephen Elledge and colleagues⁵² introduced a creative mass-spectrometry-based proteomic strategy that allowed them to make a comprehensive inventory of substrates for the protein kinases ATM (ataxia telangiectasia mutated) and ATR (ATM and Rad3 related), which are involved in the DNA-damage-response pathway⁵².

Previous studies had identified about 25 ATM and/or ATR substrates, which contained an unusual consensus sequence for phosphorylation: Ser/Thr–Gln⁵³. On the basis of this information, Elledge and colleagues used a panel of 68 antibodies specific for phospho-Ser–Gln or

phospho-Thr-Gln to immunoprecipitate candidate substrates for ATM and ATR from human cells that were either exposed to ionizing radiation to induce the DNA-damage response or not irradiated (Fig. 6). These cell populations had previously been subjected to stable isotope labelling^{54,55}, so radiation-induced phosphorylation events could be quantified by LC-MS/MS analysis. Relative quantification of heavy-isotope-labelled and light-isotope-labelled phosphopeptide pairs identified 905 phosphorylation sites, across 700 proteins, that had fourfold higher signals in irradiated cells than in non-irradiated cells. Thus, in a single set of experiments, the researchers increased the number of candidate ATM and ATR substrates by more than 20-fold (from about 25 proteins to 700 proteins). The increase in phosphorylation found in irradiated cells was confirmed for several candidate substrates by immunoblotting with antibodies specific for phospho-Ser-Gln or phospho-Thr-Gln.

The researchers next examined whether the newly identified substrates have a role in the DNA-damage response, by systematically disrupting expression of the corresponding genes by RNAi. Of the 37 substrates examined, 35 were found to contribute to at least one aspect of the DNA-damage response. Although these studies do not directly test whether the phosphorylation state of the proteins is crucial for their function, the results indicate that many more proteins contribute to the DNA-damage response than was originally thought, and these proteins are subject to dynamic phosphorylation in response to DNA-damage signals. Interestingly, there were several cases in which multiple components of a given pathway were phosphorylated, leading the authors to conclude that protein kinases can increase their effect on specific signalling pathways by simultaneously phosphorylating several nodes.

The authors then rapidly mined their phosphoproteomic data sets, facilitating the functional annotation of two previously uncharacterized proteins. One of these proteins, which the authors named abraxas (also known as CCD98 and FLJ13614), was identified as a potential ATM and/or ATR substrate. Abraxas was more heavily phosphorylated in irradiated cells than in non-irradiated cells, and it formed a complex with RAP80 (also known as UIMC1) and BRCA1, which was required for resistance to DNA damage, control of the cycle checkpoint at the G2-M boundary and repair of DNA⁵⁶. The other protein, which the authors named FANCI (also known as KIAA1794), was found to form a complex with FANCD2, which then localized to chromatin in response to DNA damage⁵⁷. Interestingly, a mutation in the gene encoding FANCI had been causally linked to Fanconi's anaemia, a syndrome that impairs development and increases the risk of developing cancer.

These studies, together with others⁵⁸⁻⁶¹, underscore the rapid development of quantitative mass-spectrometry methods for mapping protein phosphorylation sites in proteomes. Similar approaches are emerging for the global analysis of other key protein modifications, including acetylation^{62,63}, methylation^{63,64}, glycosylation⁶⁵ and ubiquitylation⁶⁶. We expect that these methods will also help to improve our understanding of the role of post-translational modifications in regulating protein function in biological systems.

Insulin pathways in *Caenorhabditis elegans* dauer formation and ageing

Genetic studies in *C. elegans* have determined that the signalling pathway involving insulin and insulin-like growth factor has an important role in regulating lifespan⁶⁷. For example, disruption of the *C. elegans* receptor DAF-2, which is homologous to the mammalian receptors for both insulin and insulin-like growth factor 1, extends lifespan and increases entry to the dauer phase (a phase characterized by delayed development, which *C. elegans* can enter if environmental conditions are unfavourable early in development)⁶⁸. To understand the molecular basis of these marked changes in physiology, John Yates and colleagues carried out a quantitative mass-spectrometry-based proteomic analysis of wild-type and *daf-2* mutant strains of *C. elegans*⁶⁹.

Two forms of quantification were used: ratiometric analysis of proteomes from both wild-type *C. elegans* and *daf-2* mutants with a reference proteome corresponding to wild-type *C. elegans* fed on ¹⁵N-enriched bacteria⁷⁰; and direct spectral counting⁷¹ of unlabelled proteins in wild-type and *daf-2* mutant proteomes (Fig. 7a). Together, these methods identified 86 proteins that were differentially expressed in *daf-2* mutants, 47 that were more abundant and 39 that were less abundant than in wild-type *C. elegans*. There were good correlations between the proteomic data obtained with the two methods, indicating that either approach can provide an accurate estimate of the relative levels of proteins in two or more biological samples. The authors verified their proteomic data by selecting several proteins from wild-type strains and *daf-2* mutant strains for analysis by immunoblotting.

Interestingly, proteins that had similar changes in abundance in the *daf-2* mutant strain tended to show a functional relationship. For example, as a group, the more abundant proteins tended to have translation-elongation and lipid-transport functions, whereas the less-abundant proteins were over-represented in the categories of amino-acid biosynthesis, reactive-oxygen-species metabolism and carbohydrate metabolism. Yates and colleagues next tested whether these proteins

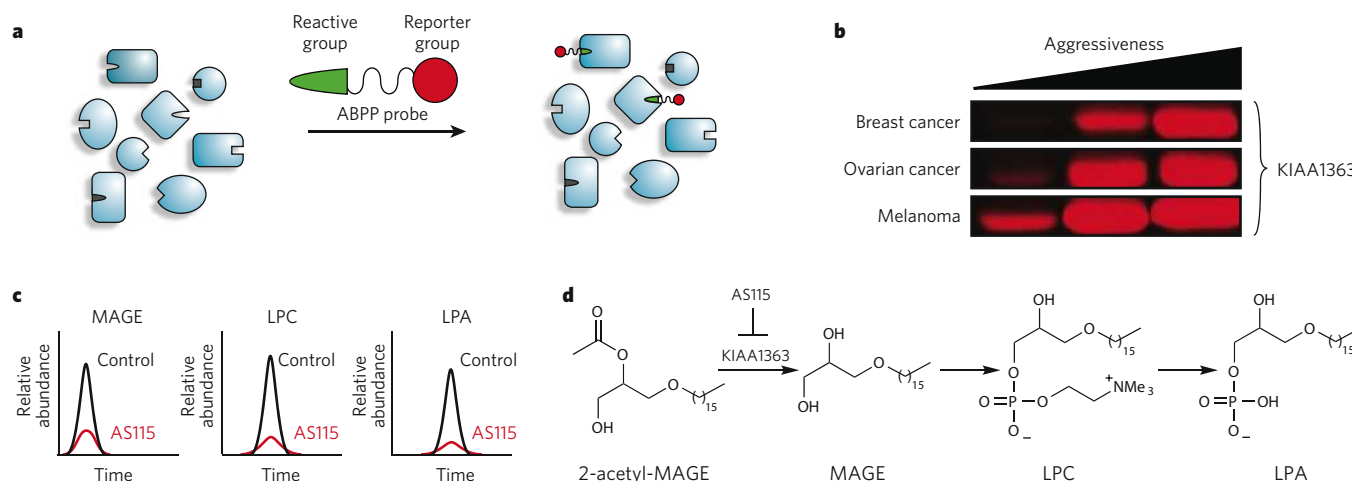


Figure 5 | Discovery of an ether-lipid signalling pathway that supports cancer pathogenesis. **a**, The general structure and mechanism of action of ABPP probes are shown, with proteins of various activities in blue and a probe with a specific reactive group. **b**, ABPP of a panel of human tumour cell lines identified an uncharacterized hydrolase, KIAA1363, in aggressive cell lines from several tumour types. The activity of KIAA1363 increased with the aggressiveness of the cell lines (as determined by in-gel fluorescence scanning of probe-labelled KIAA1363)⁴⁴. **c**, Inactivation of

KIAA1363 by the selective inhibitor AS115 (or short hairpin RNA probes) decreased the abundance of a family of ether lipids, including MAGEs and alkyl-lysophospholipids (lysophosphatidylcholine (LPC) and LPA), as determined by LC-MS analysis⁵⁰. **d**, These results suggest a model in which KIAA1363 regulates an ether-lipid pathway that proceeds from MAGEs to LPC and LPA. Disruption of this lipid network by blockade of KIAA1363 inhibited cancer-cell migration and tumour growth (not shown)⁵⁰. Me, methyl.

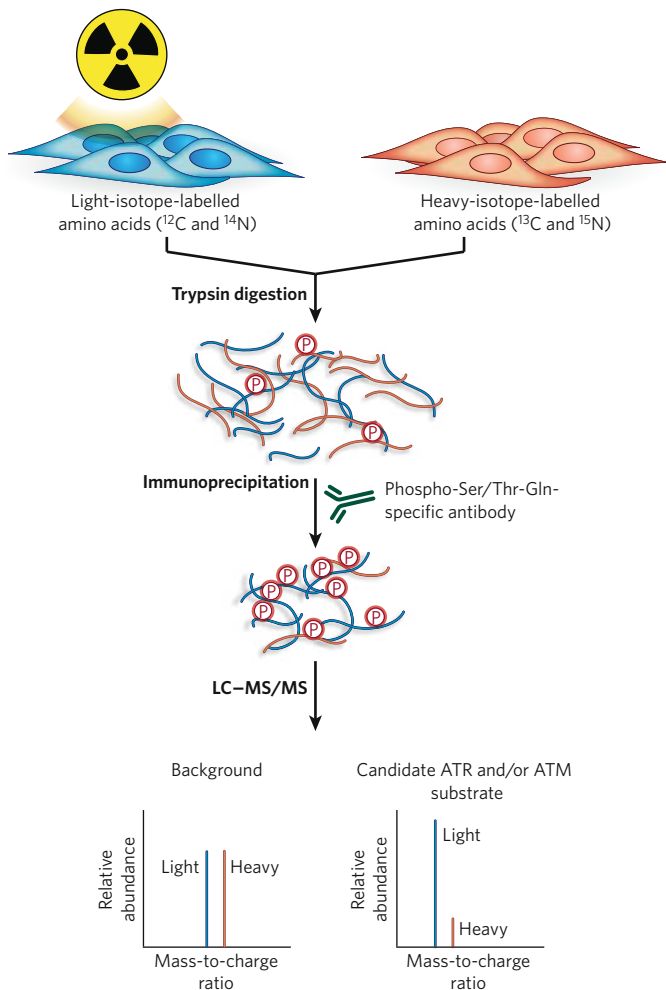


Figure 6 | Identification of candidate ATM and/or ATR substrates involved in the DNA-damage response. Cells treated with light-isotope-labelled amino acids were exposed to ionizing radiation, and cells treated with heavy-isotope-labelled amino acids were maintained under control conditions. Candidate ATM and ATR substrates were then identified by trypsin digestion of whole-cell proteomes, followed by immunoprecipitation with antibodies specific for the consensus ATM and ATR phosphorylation motif phospho-Ser/Thr-Gln, and then LC-MS/MS analysis. Phosphoproteins produced in response to irradiation were identified by ratiometric analysis of mass signals from light-isotope-labelled cells and heavy-isotope-labelled cells. Many of these proteins were found to have important roles in the DNA-damage response⁵⁷.

affected DAF-2-dependent processes such as lifespan. Curiously, RNAi-mediated knockdown in wild-type *C. elegans* of the mRNAs encoding proteins that had increased abundance in *daf-2* mutants tended to extend the lifespan further, whereas knockdown of the mRNAs encoding less-abundant proteins shortened the lifespan of wild-type *C. elegans* (Fig. 7b). These results suggest that many of the proteomic changes observed in *daf-2* mutants reflect compensatory changes in metabolic and/or signalling pathways that limit the impact of loss of DAF-2 function.

Principal among the observed compensatory pathways was TAX-6 (also known as CNA-1), the *C. elegans* orthologue of the protein phosphatase known as calcineurin A. Significantly more TAX-6 was present in *daf-2* mutants than in wild-type *C. elegans*. In addition, disruption of the gene encoding TAX-6 (*tax-6*) produced a similar phenotype to loss of DAF-2 (that is, extended lifespan and increased entry to the dauer phase). Disruption of both *tax-6* and *daf-2* resulted in even more marked phenotypes. Collectively, these data indicate that TAX-6 is part of a feedback loop that buffers the effects of DAF-2 on longevity, through compensatory mechanisms (Fig. 7c). A provocative extension

of this idea is that pharmacological strategies to block such compensatory pathways might be useful for extending the lifespan of animals. From a more technical perspective, this study, together with another study by Yates and colleagues⁷², shows that stable isotope labelling can be applied to intact organisms, as well as to cell-culture preparations, thus greatly expanding the potential applications of this quantitative mass-spectrometry-based proteomic method.

Emergent themes for mass-spectrometry-based proteomics

The studies described in this review have several common conceptual and experimental themes that are instructive for researchers interested in using mass-spectrometry-based proteomics. First, it is clear that, to ask specific biological questions, well-configured model systems need to be established. Proteomic experiments produce large amounts of data. For these data sets to deliver answers or inspire compelling hypotheses that explain the molecular basis of complex biological processes, well-designed experimental systems and controls must be incorporated into the research plan. Not surprisingly, experimental systems often involve pathophysiological states for which clinical phenotypes are well described. Using the appropriate controls allows investigators rapidly to winnow down proteomic observations to a manageable number of proteins that show changes in abundance, activity or post-translational modification in the experimental model under study.

If carried out properly, mass-spectrometry-based proteomics experiments should uncover a set of proteins associated with a specific cellular or physiological process. Testing the function of these proteins, however, requires 'targeted' follow-up studies that use complementary experimental approaches. A second theme is the emergence of RNAi as a near-universal method to perturb the production of any protein in cells and organisms, offering researchers a powerful strategy to test the function of proteins identified in proteomic experiments. RNAi also has the advantage of operating on a scale that is compatible with screening the biological function of hundreds to thousands of candidate proteins^{73,74}, making it an attractive method to rapidly validate targets discovered in large-scale proteomic endeavours. Perhaps the best way to picture the growing synergistic relationship between mass-spectrometry-based proteomic techniques and RNAi techniques is to view the former approach as a hypothesis-generating engine and the latter as a tool for testing these hypotheses. In this manner, proteomic observations can be connected to function or phenotype.

A third common theme among the studies highlighted here is that the follow-up biological experiments were carried out by the same research group as the original mass-spectrometry-based proteomics investigation. Although repositories of proteomic data are undoubtedly useful, this finding suggests that the primary biological users of proteomic information are typically the generators of these data. There are several reasons why this might be the case. First, biologists are inundated with large-scale data sets, including those that inventory transcript, protein and metabolite expression, as well as protein-protein interactions and post-translational modification state. This glut of molecular information almost certainly has a saturating effect on potential users, who may face too many candidate targets or pathways to explore. Second, potential users might be concerned about the quality of mass-spectrometry-based proteomic data (for example, the number of false-positive and false-negative results). Follow-up biological studies are not trivial in terms of cost or time, and having confidence in the quality of the data would probably lower the 'activation energy barrier' for secondary users of proteomic results. Last, it might simply take more time for secondary users to incorporate mass-spectrometry-based proteomic data sets into their biological studies, or secondary users might incorporate data from proteomic experiments mainly to validate observations from their own experiments. Thus, there might be particular issues to overcome before repositories of large-scale proteomic data influence hypothesis-driven research, which often involves highly specific objectives for which proteomic data might be too general to address. This situation should improve as new methods for mining stored proteomic data are developed. It should also be noted that it is much easier to track scientific

progress if a common authorship is preserved. We might therefore be underestimating the number of researchers who have capitalized on repositories of mass-spectrometry-based proteomic data to gain new insights into biological systems.

We have highlighted experimental commonalities among mass-spectrometry-based proteomic studies that made important biological discoveries; however, there are also some noteworthy differences. For example, several methods for protein quantification have been used: these include ICAT²⁵, stable isotope labelling of cells⁵² and organisms⁶⁹, and label-free techniques such as unique peptide number⁴¹, protein-sequence coverage³⁷ and spectral counting^{37,47,69}. Given that each of these strategies is generally successful, does there need to be a single form of data collection and analysis in quantitative mass-spectrometry-based proteomic experiments? This question can be distilled to the issue of balancing accuracy and ease of implementation. Label-free methods are the simplest and most cost-effective to carry out, but they lack the precision of isotope-labelling techniques. However, as long as researchers are committed to validating a portion of their proteomic results by using complementary techniques (for example, immunoblotting or selective-reaction monitoring), confidence in the overall data sets acquired with either method should be achievable. These validation experiments should, for example, readily identify false-positive data, which can be eliminated from further analysis. False-negative results (that is, changes that occur but are not detected) are more problematic but are almost certainly minimized as the accuracy of the quantification method increases. On this note, the discovery of Tfb5 as a component of TFIIF is worth revisiting. This protein showed only a twofold increase in ICAT signal in enriched TFIIF complexes²⁵, a signal difference that probably would not have registered as meaningful if less-accurate (label-free) quantification methods had been used. Regardless, in all of the studies highlighted here, the overall importance of the proteomic data sets was established by follow-up biological experiments.

Conclusions and future directions

Future technical challenges for mass-spectrometry-based proteomics mainly relate to the nature of proteins in biological systems. Proteins have a wide range of abundances, and this is further confounded by the myriad post-translational modifications that are dynamically regulated by cellular context and time. To capture the various states of proteins in a cell fully, proteomes must therefore be sampled in different conditions and at several time points following perturbation. Several technical aspects of mass spectrometers need to be improved to meet the demands for higher throughput and proteome coverage without sacrificing information content. First, advances in instrument scan speed would allow more frequent sampling of ions. Higher rates of sampling would translate into more tandem mass spectra acquired per unit time, which would, in turn, enable higher-resolution chromatography methods to be used. Increased sampling rates should also improve dynamic range, because lower-abundance ions are more likely to be detected. Second, coupling these changes to continued improvements in sensitivity and mass accuracy measurements, the gain in dynamic range could be multiplied. Increased resolution and mass accuracy should also strengthen confidence in peptide identifications and facilitate the discovery of protein modifications. Third, advances in 'top-down' mass spectrometry for sequence-based characterization of intact proteins can allow patterns of modifications on a protein to be correlated with specific activities or functions. At present, top-down mass spectrometry is most effective for small proteins (<25 kDa) and presents difficulties for analysing larger proteins⁷⁵. Key areas for the improvement of top-down mass spectrometry are the development of more general fragmentation methods for large proteins, and of higher-throughput and more-robust methods to introduce intact proteins into the mass spectrometer. Final issues to consider relate to the throughput and sample demands of standard mass-spectrometry-based proteomics experiments. Unbiased, global methods such as a two-dimensional liquid-chromatography-based shotgun proteomics require considerable time (several hours per sample) and material (>0.1 mg protein per sample). Using other strategies such

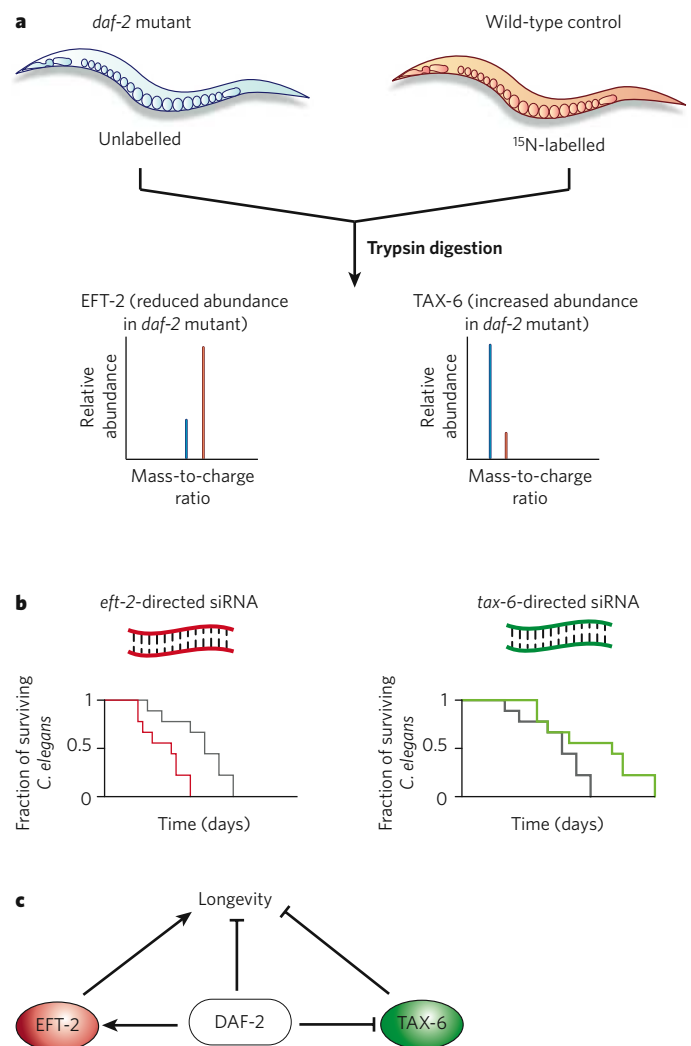


Figure 7 | Discovery of DAF-2-regulated protein pathways that modulate longevity in *Caenorhabditis elegans*. **a**, A quantitative proteomic analysis of changes in protein abundance was carried out in *daf-2* mutant *C. elegans*, by using metabolic labelling and MudPIT analysis⁶⁹. Shown are representative examples of proteins that either decreased (EFT-2) or increased (TAX-6) in abundance in *daf-2* mutants. **b**, Follow-up studies on differentially expressed proteins identified cases in which RNAi-mediated knockdown of the corresponding mRNAs decreased (EFT-2) or increased (TAX-6) the lifespan of worms. These results suggest that the proteins participate in compensatory pathways that limit the effects of *daf-2* mutation on longevity and dauer formation. **c**, A model of how DAF-2-regulated proteins participate in compensatory pathways that affect longevity was assembled from the results of these experiments.

as accurate mass tagging and single-ion reaction monitoring of peptides can increase throughput and reduce sample demands, but they limit the analysis to peptides or proteins that are known to be present in a mixture and, therefore, preclude serendipitous discoveries^{76,77}. Because mass-spectrometry-based proteomic methodology continues to develop at a rapid pace, there is much hope that these and other problems will be solved.

It is clear that biologists are becoming increasingly savvy users of mass-spectrometry instrumentation and, conversely, that mass spectrometrists are gaining familiarity with other biological techniques. We therefore expect that distinctions between these types of scientist will soon begin to lose meaning. How long might it be, for example, before mass spectrometers stand alongside centrifuges and PCR machines as core pieces of equipment in every biology lab? Wouldn't that be the ultimate sign of biological impact for this powerful analytical technology? ■

1. Whisstock, J. C. & Lesk, A. M. Prediction of protein function from protein sequence and structure. *Q. Rev. Biophys.* **36**, 307–340 (2003).
2. Galperin, M. Y. & Koonin, E. V. 'Conserved hypothetical' proteins: prioritization of targets for experimental study. *Nucleic Acids Res.* **32**, 5452–5463 (2004).
3. Zhu, H., Bilgin, M. & Snyder, M. Proteomics. *Annu. Rev. Biochem.* **72**, 783–812 (2003).
4. de Hoog, C. L. & Mann, M. Proteomics. *Annu. Rev. Genomics Hum. Genet.* **5**, 267–293 (2004).
5. MacBeath, G. Protein microarrays and proteomics. *Nature Genet.* **32** (suppl.), 526–532 (2002).
6. Hall, D. A., Ptacek, J. & Snyder, M. Protein microarray technology. *Mech. Ageing Dev.* **128**, 161–167 (2007).
7. Causier, B. Studying the interactome with the yeast two-hybrid system and mass spectrometry. *Mass Spectrom. Rev.* **23**, 350–367 (2004).
8. Stevens, R. C., Yokoyama, S. & Wilson, I. A. Global efforts in structural genomics. *Science* **294**, 89–92 (2001).
9. Aebersold, R. & Mann, M. Mass spectrometry-based proteomics. *Nature* **422**, 198–207 (2003).
10. Yates, J. R. Mass spectral analysis in proteomics. *Annu. Rev. Biophys. Biomol. Struct.* **33**, 297–316 (2004).
11. Domon, B. & Aebersold, R. Mass spectrometry and protein analysis. *Science* **312**, 212–217 (2006).
12. Andersen, J. S. & Mann, M. Organellar proteomics: turning inventories into insights. *EMBO Rep.* **7**, 874–879 (2006).
13. Cusick, M. E., Klitgord, N., Vidal, M. & Hill, D. E. Interactome: gateway into systems biology. *Hum. Mol. Genet.* **14**, R171–R181 (2005).
14. Michnick, S. W. Proteomics in living cells. *Drug Discov. Today* **9**, 262–267 (2004).
15. Neubauer, G. *et al.* Mass spectrometry and EST-database searching allows characterization of the multi-protein spliceosome complex. *Nature Genet.* **20**, 46–50 (1998).
16. Wang, Y. *et al.* BASC, a super complex of BRCA1-associated proteins involved in the recognition and repair of aberrant DNA structures. *Genes Dev.* **14**, 927–939 (2000).
17. Rout, M. P. *et al.* The yeast nuclear pore complex: composition, architecture, and transport mechanism. *J. Cell Biol.* **148**, 635–651 (2000).
18. Bouwmeester, T. *et al.* A physical and functional map of the human TNF- α /NF- κ B signal transduction pathway. *Nature Cell Biol.* **6**, 97–105 (2004).
19. Das, R. *et al.* SR proteins function in coupling RNAP II transcription to pre-mRNA splicing. *Mol. Cell* **26**, 867–881 (2007).
20. Danial, N. N. *et al.* BAD and glucokinase reside in a mitochondrial complex that integrates glycolysis and apoptosis. *Nature* **424**, 952–956 (2003).
21. Harada, H. *et al.* Phosphorylation and inactivation of BAD by mitochondria-anchored protein kinase A. *Mol. Cell* **3**, 413–422 (1999).
22. Gatenby, R. A. & Gillies, R. J. Why do cancers have high aerobic glycolysis? *Nature Rev. Cancer* **4**, 891–899 (2004).
23. Wang, X. The expanding role of mitochondria in apoptosis. *Genes Dev.* **15**, 2922–2933 (2001).
24. Coulombe, B., Jeronimo, C., Langelier, M. F., Cojocaru, M. & Bergeron, D. Interaction networks of the molecular machines that decode, replicate, and maintain the integrity of the human genome. *Mol. Cell. Proteomics* **3**, 851–856 (2004).
25. Ranish, J. A. *et al.* Identification of TFB5, a new component of general transcription and DNA repair factor IIIH. *Nature Genet.* **36**, 707–713 (2004).
26. Gygi, S. P. *et al.* Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nature Biotechnol.* **17**, 994–999 (1999).
27. Cenki, B., Petersen, J. L. & Small, G. D. REX1, a novel gene required for DNA repair. *J. Biol. Chem.* **278**, 22574–22577 (2003).
28. Giglia-Mari, G. *et al.* A new, tenth subunit of TFIIH is responsible for the DNA repair syndrome trichothiodystrophy group A. *Nature Genet.* **36**, 714–719 (2004).
29. Vermeulen, W. *et al.* Subliming concentration of TFIIH transcription/DNA repair factor causes TTD-A trichothiodystrophy disorder. *Nature Genet.* **26**, 307–313 (2000).
30. Krogan, N. J. *et al.* Global landscape of protein complexes in the yeast *Saccharomyces cerevisiae*. *Nature* **440**, 637–643 (2006).
31. Krebs, M. P., Noorwez, S. M., Malhotra, R. & Kaushal, S. Quality control of integral membrane proteins. *Trends Biochem. Sci.* **29**, 648–655 (2004).
32. Kelly, J. W. & Balch, W. E. The integration of cell and chemical biology in protein folding. *Nature Chem. Biol.* **2**, 224–227 (2006).
33. Riordan, J. R. Assembly of functional CFTR chloride channels. *Annu. Rev. Physiol.* **67**, 701–718 (2005).
34. Qu, B. H., Strickland, E. H. & Thomas, P. J. Localization and suppression of a kinetic defect in cystic fibrosis transmembrane conductance regulator folding. *J. Biol. Chem.* **272**, 15739–15744 (1997).
35. Loo, M. A. *et al.* Perturbation of Hsp90 interaction with nascent CFTR prevents its maturation and accelerates its degradation by the proteasome. *EMBO J.* **17**, 6879–6887 (1998).
36. Meacham, G. C., Patterson, C., Zhang, W., Younger, J. M. & Cyr, D. M. The Hsc70 co-chaperone CHIP targets immature CFTR for proteasomal degradation. *Nature Cell Biol.* **3**, 100–105 (2001).
37. Wang, X. *et al.* Hsp90 co-chaperone Aha1 downregulation rescues misfolding of CFTR in cystic fibrosis. *Cell* **127**, 803–815 (2006).
38. Washburn, M. P., Wolters, D. & Yates, J. R. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nature Biotechnol.* **19**, 242–247 (2001).
39. Hanash, S. Disease proteomics. *Nature* **422**, 226–232 (2003).
40. Kumar, N. *et al.* Molecular complexity of sexual development and gene regulation in *Plasmodium falciparum*. *Int. J. Parasitol.* **34**, 1451–1458 (2004).
41. Khan, S. M. *et al.* Proteome analysis of separated male and female gametocytes reveals novel sex-specific *Plasmodium* biology. *Cell* **121**, 675–687 (2005).
42. Ward, P., Equinet, L., Packer, J. & Doerig, C. Protein kinases of the human malaria parasite *Plasmodium falciparum*: the kinome of a divergent eukaryote. *BMC Genomics* **5**, 79 (2004).
43. Jessani, N. & Cravatt, B. F. The development and application of methods for activity-based protein profiling. *Curr. Opin. Chem. Biol.* **8**, 54–59 (2004).
44. Jessani, N., Liu, Y., Humphrey, M. & Cravatt, B. F. Enzyme activity profiles of the secreted and membrane proteome that depict cancer invasiveness. *Proc. Natl Acad. Sci. USA* **99**, 10335–10340 (2002).
45. Liu, Y., Patricelli, M. P. & Cravatt, B. F. Activity-based protein profiling: the serine hydrolases. *Proc. Natl Acad. Sci. USA* **96**, 14694–14699 (1999).
46. Patricelli, M. P., Giang, D. K., Stamp, L. M. & Burbaum, J. J. Direct visualization of serine hydrolase activities in complex proteome using fluorescent active site-directed probes. *Proteomics* **1**, 1067–1071 (2001).
47. Jessani, N. *et al.* A streamlined platform for high-content functional proteomics of primary human specimens. *Nature Methods* **2**, 691–697 (2005).
48. Leung, D., Hardouin, C., Boger, D. L. & Cravatt, B. F. Discovering potent and selective reversible inhibitors of enzymes in complex proteomes. *Nature Biotechnol.* **21**, 687–691 (2003).
49. Saghatelian, A. *et al.* Assignment of endogenous substrates to enzymes by global metabolite profiling. *Biochemistry* **43**, 14332–14339 (2004).
50. Chiang, K. P., Niessen, S., Saghatelian, A. & Cravatt, B. F. An enzyme that regulates ether lipid signaling pathways in cancer annotated by multidimensional profiling. *Chem. Biol.* **13**, 1041–1050 (2006).
51. Manning, G., Whyte, D. B., Martinez, R., Hunter, T. & Sudarsanam, S. The protein kinase complement of the human genome. *Science* **298**, 1912–1934 (2002).
52. Matsuoka, S. *et al.* ATM and ATR substrate analysis reveals extensive protein networks responsive to DNA damage. *Science* **316**, 1160–1166 (2007).
53. Shiloh, Y. The ATM-mediated DNA-damage response: taking shape. *Trends Biochem. Sci.* **31**, 402–410 (2006).
54. Oda, Y., Huang, K., Cross, F. R., Cowburn, D. & Chait, B. T. Accurate quantitation of protein expression and site-specific phosphorylation. *Proc. Natl Acad. Sci. USA* **96**, 6591–6596 (1999).
55. Mann, M. Functional and quantitative proteomics using SILAC. *Nature Rev. Mol. Cell Biol.* **7**, 952–958 (2006).
56. Wang, B. *et al.* Abraxas and RAP80 form a BRCA1 protein complex required for the DNA damage response. *Science* **316**, 1194–1198 (2007).
57. Smogorzewska, A. *et al.* Identification of the FANCD1 protein, a monoubiquitinated FANCD2 paralog required for DNA repair. *Cell* **129**, 289–301 (2007).
58. Olsen, J. V. *et al.* Global, *in vivo*, and site-specific phosphorylation dynamics in signaling networks. *Cell* **127**, 635–648 (2006).
59. Rush, J. *et al.* Immunoaffinity profiling of tyrosine phosphorylation in cancer cells. *Nature Biotechnol.* **23**, 94–101 (2005).
60. Jin, M. *et al.* Quantitative analysis of protein phosphorylation in mouse brain by hypothesis-driven multistage mass spectrometry. *Anal. Chem.* **77**, 7845–7851 (2005).
61. Huang, P. H. *et al.* Quantitative analysis of EGFRvIII cellular signaling networks reveals a combinatorial therapeutic strategy for glioblastoma. *Proc. Natl Acad. Sci. USA* **104**, 12867–12872 (2007).
62. Kim, S. C. *et al.* Substrate and functional diversity of lysine acetylation revealed by a proteomics survey. *Mol. Cell* **23**, 607–618 (2006).
63. Garcia, B. A., Pesavento, J. J., Mizzen, C. A. & Kelleher, N. L. Pervasive combinatorial modification of histone H3 in human cells. *Nature Methods* **4**, 487–489 (2007).
64. Ong, S. E., Mittler, G. & Mann, M. Identifying and quantifying *in vivo* methylation sites by heavy methyl SILAC. *Nature Methods* **1**, 119–126 (2004).
65. Khidekel, N. *et al.* Probing the dynamics of O-GlcNAc glycosylation in the brain using quantitative proteomics. *Nature Chem. Biol.* **3**, 339–348 (2007).
66. Peng, J. *et al.* A proteomics approach to understanding protein ubiquitination. *Nature Biotechnol.* **21**, 921–926 (2003).
67. Mukhopadhyay, A. & Tissenbaum, H. A. Reproduction and longevity: secrets revealed by *C. elegans*. *Trends Cell Biol.* **17**, 65–71 (2007).
68. Kenyon, C., Chang, J., Gensch, E., Rudner, A. & Tabtiang, R. A *C. elegans* mutant that lives twice as long as wild type. *Nature* **366**, 461–464 (1993).
69. Dong, M. Q. *et al.* Quantitative mass spectrometry identifies new insulin targets in *C. elegans*. *Science* **317**, 660–663 (2007).
70. Venable, J. D., Dong, M. Q., Wohlschlegel, J., Dillin, A. & Yates, J. R. Automated approach for quantitative analysis of complex peptide mixtures from tandem mass spectra. *Nature Methods* **1**, 39–45 (2004).
71. Liu, H., Sadygov, R. G. & Yates, J. R. A model for random sampling and estimation of relative protein abundance in shotgun proteomics. *Anal. Chem.* **76**, 4193–4201 (2004).
72. Wu, C. C., MacCoss, M. J., Howell, K. E., Matthews, D. E. & Yates, J. R. Metabolic labeling of mammalian organisms with stable isotopes for quantitative proteomic analysis. *Anal. Chem.* **76**, 4951–4959 (2004).
73. Berns, K. *et al.* A large-scale RNAi screen in human cells identifies new components of the p53 pathway. *Nature* **428**, 431–437 (2004).
74. Perrimon, N. & Mathey-Pervot, B. Applications of high-throughput RNA interference screens to problems in cell and developmental biology. *Genetics* **175**, 7–16 (2007).
75. Zabrovskov, V., Senko, M. W., Du, Y., Leduc, R. D. & Kelleher, N. L. New and automated MSⁿ approaches for top-down identification of modified proteins. *J. Am. Soc. Mass Spectrom.* **16**, 2027–2038 (2005).
76. Conrads, T. P., Anderson, G. A., Veenstra, T. D., Pasa-Tolic, L. & Smith, R. D. Utility of accurate mass tags for proteome-wide protein identification. *Anal. Chem.* **72**, 3349–3354 (2000).
77. Gerber, S. A., Rush, J., Stemman, O., Kirschner, M. W. & Gygi, S. P. Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS. *Proc. Natl Acad. Sci. USA* **100**, 6940–6945 (2003).
78. MacCoss, M. J. *et al.* Shotgun identification of protein modifications from protein complexes and lens tissue. *Proc. Natl Acad. Sci. USA* **99**, 7900–7905 (2002).

Acknowledgements We gratefully acknowledge the support of the National Institutes of Health.

Author information Reprints and permissions information is available at npg.nature.com/reprints. Correspondence should be addressed to B.F.C. (cravatt@scripps.edu) or J.R.Y. (lyates@scripps.edu).