Applying mass spectrometry-based proteomics to genetics, genomics and network biology

Matthias Gstaiger[*‡] *and Ruedi Aebersold*[*‡§||]

Abstract | The systematic and quantitative molecular analysis of mutant organisms that has been pioneered by studies on mutant metabolomes and transcriptomes holds great promise for improving our understanding of how phenotypes emerge. Unfortunately, owing to the limitations of classical biochemical analysis, proteins have previously been excluded from such studies. Here we review how technical advances in mass spectrometry-based proteomics can be applied to measure changes in protein abundance, posttranslational modifications and protein–protein interactions in mutants at the scale of the proteome. We finally discuss examples that integrate proteomics data with genomic and phenomic information to build network-centred models, which provide a promising route for understanding how phenotypes emerge.

Mass spectrometry
An analytical technique for the identification of the chemical composition of compounds on the basis of the mass to charge ratios of charged particles.

*Institute of Molecular Systems Biology, Swiss Federal Institute of Technology (ETH) Zurich, 8093 Zurich, Switzerland. ‡Competence Center for Systems Physiology and Metabolic Diseases, Swiss Federal Institute of Technology (ETH) Zurich, 8093 Zurich, Switzerland. §Faculty of Science, University of Zurich, 8006 Zurich, Switzerland. ||Institute for Systems Biology, 1221 N. 34th Street, Seattle, Washington 98103, USA.
e-mails: gstaiger@imsb.biol.ethz.ch; aebersold@imsb.biol.ethz.ch
doi:10.1038/nrg2633

Since the rediscovery of Gregor Mendel's findings in the early twentieth century we have known that the observable traits of an organism are largely controlled by inherited functional units. In the years that have followed, much effort and progress have been made in genetics and functional genomics to understand the molecular basis of inheritance and how it translates into a particular phenotype. However, the simple hope that genomes can be viewed as blueprints from which phenotypes can be inferred has not been fulfilled.

The term phenotype is generally used to describe the observable traits of an organism (for example, morphology, size, physiology and behaviour) that emerge from its genotype and the specific environmental conditions. As the sensitivity of cell biological and biochemical methods has increased, the term has been extended to various aspects of cellular and molecular organization. Although phenotypic manifestations at the different scales of biological complexity are interconnected, it is conceivable that genetic mutations introduce perturbations in complex molecular systems, and phenotypes emerge at a higher level of biological organization from this perturbed molecular state.

Given the complexity of cellular systems, techniques have been developed over the years that allow the comprehensive analysis of molecular components. The most progress has been made in obtaining information on metabolites and mRNA at the systems level. Unfortunately, the global analysis of proteins was exceedingly difficult in the past. However, systematic information on proteins is crucial as first, proteins can interact with all of the other classes of molecular components, including other proteins, and second, most of the genes for which phenotypes have been described encode proteins. Mutations can affect proteins at various levels, including their abundance, their pattern of posttranslational modifications (PTMs) and their propensity to transiently or stably interact with other components in the cell. Changes in protein properties are not confined to the mutated protein itself as, in many cases, mutations elicit indirect responses that also affect the properties of other proteins.

The classical biochemical analysis of proteins has been a daunting and time-consuming task in the past, and has yielded incomplete data sets that provided little quantitative information. In addition, changes in protein abundance cannot be simply inferred from DNA microarray data as mRNA abundance poorly correlates with protein abundance[1,2]. Recently, and partly as a consequence of the advances in genome sequencing, enormous progress has been achieved in mass spectrometry (MS)-based proteomics. Despite limited proteome coverage in most existing studies, this is the only currently available method to systematically characterize molecular alterations at the protein level. For the first time, MS-based proteomics allows the identification and

© 2009 Macmillan Publishers Limited. All rights reserved

quantitative profiling of organismal proteomes and the systematic analysis of protein modifications and interactions, offering a new range of opportunities for geneticists and network biologists to improve existing models of how phenotypes emerge.

In this Review we will discuss the gap between genotypes and phenotypes in light of the recent advances in MS-based proteomics. We will introduce the basic analytical concepts of MS-based protein identification and quantification that are used for mutant proteome profiling. Examples from existing studies will be presented that show how mutant proteome profiling in many species, including *Mus musculus*, *Caenorhabditis elegans* and *Saccharomyces cerevisiae,* has provided important insights into the inheritance of protein abundance, as well as the use of this technique for the identification of new disease biomarkers. We will present recent techniques for the global profiling of PTMs of proteins and discuss how they can be used to infer candidate pathways that are implicated in the emergence of phenotypes. Mutations might affect functionally important interactions between proteins, and we will explain how they can be measured using affinity purification–mass spectrometry (AP–MS). Finally, we will conclude with a discussion of the integration of annotated MS-based proteomics data with genomic and phenomic information as a promising framework for inferring the molecular networks that underlie the emergence of phenotypes.

Protein identification by MS shotgun proteomics

MS-based protein identification provides useful information on the relationship between the genome and the proteome and is the basis for profiling mutant proteomes using quantitative MS. To date, the most widely used method for protein identification is referred to as MS shotgun proteomics. Analagous to the shotgun sequencing approach in genomics, this term describes a method for systematic protein identification from complex samples using a combination of liquid chromatography (LC) separation of peptides generated by trypsin digestion and their subsequent analysis by tandem mass spectrometry (MS/MS)[3]. The basic concept of protein identification by MS shotgun proteomics is explained in FIG. 1.

Ideally, proteome profiling should allow us to identify and quantify all of the protein components in a cell. How many proteins can we detect in a typical shotgun MS experiment? Despite much progress, the comprehensive characterization of complex protein samples is still a major challenge and requires extensive resources in terms of time, sample amounts and instrumentation. Two major factors challenge comprehensive protein identification by shotgun MS: the complexity of cellular proteomes and the high dynamic range at which proteins are expressed in biological samples. It is estimated that in human cells, approximately 100,000 different protein isoforms are likely to be expressed from the 20,325 annotated protein-coding genes (see Further Information for a link to the UniProt website). After digestion using trypsin, these proteins produce highly complex peptide mixtures. The concentrations of proteins can vary over more than four orders of magnitude in yeast[4] to an estimated ten orders of magnitude for human body fluids, including serum[5]. However, the dynamic range of detection in an MS instrument typically ranges between 1,000–10,000. This leads primarily to the redundant identification of highly abundant proteins and often precludes the identification of protein species that have a low abundance from complex samples.

Despite these limitations, recent instrumental advances combined with advances in fractionation techniques have been used for the comprehensive proteomic analysis of several genetic model organisms, including *S. cerevisiae*, *Drosophila melanogaster, C. elegans* and *Arabidopsis thaliana*[1,6–8]. The coverage achieved in these studies typically ranges between 50% and 80% of the proteins encoded by the genome. By combining diverse biological samples that represent different developmental stages, activity states and organs, and by using multidimensional biochemical fractionation, one such study identified 9,124 proteins from *D. melanogaster,* which corresponds to 63% of the predicted ORFs[8]. Data from this study and others confirmed existing gene models or suggested alternative gene models. These data also revealed the existence of novel protein-coding genes that had not been

Affinity purification–mass spectrometry
A method for the analysis of protein complexes that combines purification of protein complexes using affinity reagents and mass spectrometry.

Tandem mass spectrometry
This combines two mass spectrometers: one (MS1) for the detection and selection of precursor ions, which is followed by a second (MS2) for the analysis of fragment ion spectra generated from selected precursor ions after collision-induced fragmentation. The information from the fragment ion spectra is used for peptide identification.

Dynamic range
The ratios between the highest and lowest possible ion intensities in a mass spectrum for which accurate masses can be determined by a mass spectrometer.

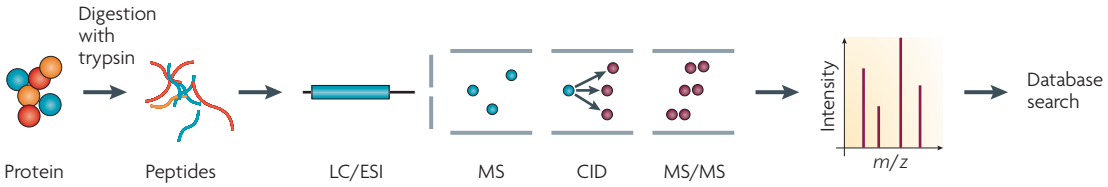


Figure 1 | **Mass spectrometry-based protein identification.** Protein samples are digested with trypsin and the resulting peptide mixtures are fractionated by reversed-phase liquid chromatography (LC). The fractionated peptide solution is subjected to an electric potential, which causes a spray to be formed, leading to the desolvation and ionization of the peptides (electrospray ionization; ESI). Mass to charge (*m*/*z*) ratios are measured from peptide ions that pass the collision cell without fragmentation in the mass spectrometer (MS). Specific ions are randomly selected for collision-induced dissociation (CID) with neutral gas molecules (for example, helium) and the resulting fragment ions are measured in the second mass analyser in tandem mass spectrometry (MS/MS). The MS precursor ion intensities obtained in the first stage can be used for peptide quantification, whereas MS/MS fragment ion information from the second stage contains sequence information that can be compared with sequences from *in silico* digested protein sequence databases for peptide and subsequent protein identification.

© 2009 Macmillan Publishers Limited. All rights reserved

identified by gene prediction algorithms and provided evidence for the expression of specific splice isoforms at the protein level[6–8].

Quantitative MS

The comparative analysis of molecular phenotypes depends on quantitative MS technologies to detect the relative and absolute protein abundance changes that result from genetic perturbations. Two main approaches are currently being used for MS-based quantitative proteomics: differential isotopic labelling and label-free quantification (BOX 1). The two approaches have different advantages, isotope-labelling methods are thought to measure protein abundance with higher accuracy, whereas label-free approaches have a greater dynamic range and achieve higher levels of proteome coverage[9].

Differential isotope labelling. MS quantification on the basis of differential isotopic labelling of peptides and proteins builds on the theory of stable isotope dilution. This theory states that the relative signal intensity obtained in a mass spectrometer of two analytes that are chemically identical but have different stable isotope composition represents the relative abundance of the two analytes in the sample. Protein abundance in wild-type and mutant samples (or for any other condition) can be analysed in a single liquid chromatography–tandem mass spectrometry (LC–MS/MS) experiment on the basis of observable mass shifts caused by differential isotope labelling. Two principal workflows are typically applied: *in vitro* labelling of isolated proteins and peptides or *in vivo* incorporation of isotope-labelled amino acids through metabolic labelling, which is also referred to as stable isotope labelling with amino acids in cell culture (SILAC; for details see BOX 1).

Label-free quantification. Label-free quantitative MS methods are based either on spectral counting or on peptide precursor ion intensities that are obtained using the first mass spectrometer (MS1) of a tandem mass spectrometer. Spectral counting was introduced as a semi-quantitative method for the analysis of shotgun MS data at a moderate mass resolution for which alternative non-isotopic quantification methods are not applicable[10–12]. This method is based on the assumption that the rate at which a peptide precursor ion is selected for fragmentation in a mass spectrometer is correlated to its abundance. For relative protein quantification, the spectral counts are then averaged into a protein abundance index. The method depends on the quality of the MS/MS peptide identification and therefore problems for protein quantification might arise if specific, correctly identified peptides map to more than one protein. Although the method works reliably for large and abundant proteins, the number of peptides observed from small proteins and low abundance proteins is often insufficient for accurate quantification by spectral counting.

MS-based label-free quantification is based on the accurate mass and time tag approach[13] and builds on the alignment of high-mass accuracy spectra that are obtained from the analysis of wild-type and mutant samples by separate LC–MS/MS experiments (BOX 1). Peptides are identified across different LC runs based on their specific retention time coordinates and precise mass to charge (m/z) values, which in principle allows the quantification of all of the peptides detected from a biological sample that are within the sensitivity range of a MS analyser, independent of MS/MS acquisition.

Analysis of global proteome changes

Genetic mutations can affect the structure and the abundance of proteins. Importantly, mutations can also have indirect systemic effects on the abundance of other cellular proteins. Abundance profiling of mutant proteomes by MS might therefore provide important information for the identification of processes and pathways that are involved in establishing mutant phenotypes. It is still early days for these types of studies, as substantial experimental efforts are required and the proteome coverage has been far from complete in most of the studies that have been carried out. Nevertheless, the results so far are encouraging and show that MS-based proteomics can provide information on the molecular phenotype that is not accessible using other methods. Applications of published studies include assessing the genetic basis of protein abundance, elucidating the effects of specific mutants and biomarker discovery. TABLE 1 summarizes recent studies that have carried out global MS-based profiling of mutant proteomes.

The genetic basis of protein abundance. To study the inheritance of protein abundance the proteomes of two distinct parental strains of *S. cerevisiae* and 98 segregants were compared using label-free quantification[14]. The results showed that protein abundances, as for mRNA abundances, represent quantitative traits, and that variation in protein abundance among segregants is primarily due to their genetic differences. The Mann group recently compared proteomes from haploid and diploid strains of *S. cerevisiae* using a SILAC approach[1]. These authors quantified >4,000 yeast proteins, showing for the first time that, in principle, almost complete proteome coverage can be reached by MS-based quantitative proteome profiling if substantial efforts are made during sample fractionation and MS analysis. Notably, only 196 proteins showed significant changes in abundance between haploid and diploid yeast proteomes. Among the group of haploid-specific proteins, the authors found key components of the pheromone signalling pathway (Ste2, Ste4, Ste18, Gpa1, Ste5, Fus3 and Ste12) that are required for the mating of haploid cells.

Proteomic profiling of mutants. Proteomic profiling of mutants has been used to identify new components of molecular pathways and to shed light on the processes that contribute to the generation of a phenotype. In an attempt to identify new targets of the insulin signalling pathway, John Yates' group analysed the global protein profiles from *C. elegans* mutants of the insulin-like growth factor 1 receptor gene *daf-2* (REF. 15). Using samples derived from ^{15}N-labelled wild-type worms as a reference, 86 proteins were shown to be upregulated

Liquid chromatography–tandem mass spectrometry
Liquid chromatography is used in MS-based proteomics to separate peptides in complex mixtures primarily on the basis of their charge or hydrophobicity using strong cation exchange or reversed-phase chromatography columns.

© 2009 Macmillan Publishers Limited. All rights reserved

Box 1 | **Mutant proteome profiling by quantitative mass spectrometry**

Isotope-labelling approaches

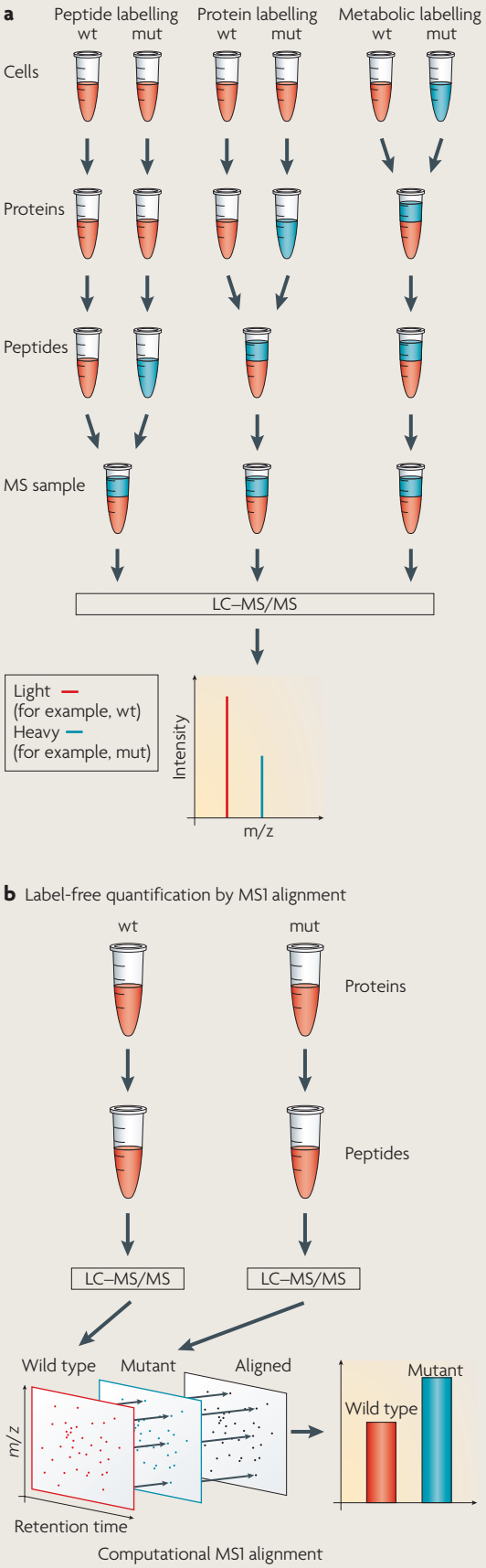
As shown in part **a** in the figure, differential labelling of proteins or peptides with heavy or light isotopes (indicated in red or blue) can be done *in vitro* or by the incorporation of isotope-labelled amino acids by metabolic labelling *in vivo*. For *in vitro* labelling, wild-type (wt) and mutant (mut) samples are prepared separately and isolated proteins or peptides are differentially labelled with heavy or light versions of isotope-tagging reagents, mainly through their sulphhydryl (for example, isotope-coded affinity tags)[71] or amine groups (for example, isotope-coded protein labels)[73]. Differential labelling introduces a characteristic mass shift, which can be used to determine the MS1 peptide ratios between pairs of heavy and light peptides. Peptide labelling with recently introduced isobaric tags for relative and absolute quantitation, which as the name indicates, keep the mass of the differentially labelled precursor ions of a given peptide constant but allow quantification after tandem mass spectrometry (MS/MS) analysis on the basis of sample-specific reporter ion intensities from up to eight different samples in a single liquid chromatography–tandem mass spectrometry (LC–MS/MS) experiment[74,75].

The use of synthetic isotope-labelled reference peptides for absolute quantification that was pioneered by Desiderio *et al.*[76] has been extended to proteomic studies by Steve Gygi and colleagues[77]. In this approach, known amounts of synthetic isotope-labelled reference peptides, which correspond to proteotypic peptides of the proteins to be analysed, are added to the samples before LC–MS/MS analysis for absolute quantification of proteins.

Stable isotope labelling with amino acids in cell culture is an *in vivo* isotope labelling method that is becoming increasingly popular[78]. Wild-type and mutant cells are grown in media that contains either light or heavy isotope versions of lysine or arginine, which yield differentially labelled proteomes. The entire labelling process occurs at the beginning of the experiment, which has the advantage that samples can be combined at early steps to avoid errors that can be introduced when samples are separately processed. As the method is limited to cells or organisms that can be metabolically labelled, it is not generally applicable to human tissues and body fluids.

Label-free quantification from aligned MS1 spectra

For label-free quantification (part **b** in the figure) wild-type and mutant proteomes are analysed by separate LC–MS/MS experiments and the MS1 spectra are computationally aligned to calculate the relative protein abundance changes on the basis of the signal intensities of extracted ion chromatograms from aligned peptide features. This reduces the undersampling problem that is known to occur with MS/MS-based approaches and results in a dynamic range of three to four orders of magnitude[79]. Newer hybrid MS instruments (LTQ FT and LTQ Orbitrap) offer the option to simultaneously record MS signal intensities and identify peptides using MS/MS. These two types of information can be combined by recently developed computational approaches[80–82]. The number of peptides that can be mapped across different LC–MS/MS experiments therefore depends on the accuracy of the peptide masses that are determined by the mass analyser and reproducibility of the LC system. Strategies for signal normalization and for correcting variations in LC performances have been developed and are now integrated in automated computational platforms for label-free MS analysis[83].



© 2009 Macmillan Publishers Limited. All rights reserved

Table 1 | **Examples of global profiling of mutant proteomes using quantitative mass spectrometry**

Organism	Genetic model	Method	Refs
Yeast	Parental versus segregants	Label free	14
Yeast	Haploid versus diploid	SILAC	1
Fly	RNAi-mediated knockdown of ISWI	SILAC	69
Worm	*daf-2*$^{-/-}$	^{15}N metabolic labelling	15
Mouse	*Pdx1-CreKras*^{G12D} *Ink4a/Arf*$^{lox/lox}$	^{13}C acrylamide labelling	24
Mouse	*Camk2a*$^{-/-}$	iTRAQ	70
Mouse	*Fermt3*$^{-/-}$	SILAC	20
Mouse	*Tp53*^{K317R}	ICAT	16
Mouse	Huntington's disease	ICAT	18
Mouse	*Fmr1*$^{-/-}$	SILAC	19
Mouse	Triple transgenic (*PS1*^{M146V}; *APP*^{Swe}; *MAPT*^{P301L}) Alzheimer's model	iTRAQ	17
Rat	Rat1/Myc	ICAT	71

APP, amyloid precursor protein; *Camk2a*, calcium/calmodulin-dependent protein kinase IIα; *daf-2*, abnormal dauer formation 2; *Fermt3*, fermitin family homolog 3 (also known as kindlin 3); *Fmr1*, fragile X mental retardation 1; ICAT, isotope-coded affinity tag; iTRAQ, isobaric tag for relative and absolute quantitation; *MAPT*, microtubule-associated protein tau; *Pdx1*, pancreatic and duodenal homeobox 1; *PS1*, presenilin 1; SILAC, stable isotope labelling with amino acids in cell culture.

or downregulated in *daf-2* mutants. Changes in protein abundance were confirmed by targeted proteomics using selected reaction monitoring (SRM). The subsequent genetic analysis of a subset of these proteins showed that they had specific roles in DAF-2-dependent processes, such as dauer formation and aging. In another study, global proteome analysis following the ionizing irradiation of thymocytes isolated from wild-type and *Tp53*^{K317R} knock-in mice indicated that lysine 317 acetylation has a role in the modulation of the p53-mediated DNA damage response[16]. Other studies have analysed tissues and cells from mutant mouse models of human diseases, such as Alzheimer's disease[17], Huntington's disease[18] and fragile X syndrome[19], and have identified mutant-specific proteomic changes that can be subjected to further functional analysis.

The metabolic labelling of entire mammalian organisms offers an alternative approach to the quantitative analysis of mutant proteomes[20–22]. In a recent study by the Mann group, mice were fed with a $^{13}C_6$ lysine diet, which led to almost 100% incorporation of the isotope-labelled amino acid into the mouse proteome[20]. Similar to the SILAC experiments described above, isotope-labelled proteomes isolated from wild-type and mutant mouse tissue can be compared using whole-proteome profiling to detect differences in the abundance of specific proteins. The feasibility of the approach has been confirmed by proteomic profiling of mice that are deficient for β1-integrin 3, β-parvin 3 and kindlin 3, by analysing platelets, heart tissue and red blood cells, respectively. In the case of the kindlin3-deficient mice (*Fermt3*$^{-/-}$), this approach showed that several proteins (ankyrin 1, band 4.1, adducin 2 and dematin) were almost completely absent in *Fermt3*$^{-/-}$erythrocyte membranes, which together with additional functional and morphological data provide an explanation for the anaemia phenotype that has been observed in newborn *Fermt3*$^{-/-}$ mice.

Selected reaction monitoring
This is a sensitive mass spectrometry-based method for targeted proteomics that is based on the measurement of precursor–fragment ion pairs (transitions) of proteotypic peptides.

Mutant proteome profiling for biomarker discovery. In search of early disease biomarkers, some studies have focused on changes in protein abundance in serum proteomes from mice that carry disease-causing mutations that are frequently found in human disorders[23–25]. In one such study, the Hanash group compared serum proteomes from healthy control mice and *Pdx1-CreKras*^{G12D}*Ink4a/Arf*$^{lox/lox}$ mutant mice that develop pancreatic intraepithelial neoplasia and ductal adenocarcinomas . Protein profiling was carried out by multistep fractionation and labelling of cysteine-containing peptides with ^{13}C isotope-labelled acrylamide[24]. Among the 1,442 identified proteins, 621 could be quantified and 165 proteins were upregulated in cancer serum samples compared with samples from control individuals. Subsequent ELISA (enzyme-linked immunosorbent assay) analysis identified similar changes in protein abundance among the corresponding human orthologous proteins, which provided additional support that they were of interest as biomarker candidates for early-stage disease profiling in patients with pancreatic cancer. Similar approaches for biomarker discovery are likely to identify biomarkers for other disease phenotypes.

Phenotypic profiling using targeted MS

So far, changes in mutant proteomes have been analysed using shotgun LC–MS/MS technologies, which are biased towards the discovery of the most abundant and easily observable proteomic changes. However, biologically relevant molecular responses are often below the detection limits of shotgun MS-based proteomics. More recently, targeted proteomics workflows have been introduced to overcome these shortcomings of shotgun MS proteomics[26–28]. Targeted proteomics allow the selective detection and quantification of predetermined peptide ions, which are analogous to mRNA profiling using DNA microarrays (FIG. 2).

© 2009 Macmillan Publishers Limited. All rights reserved

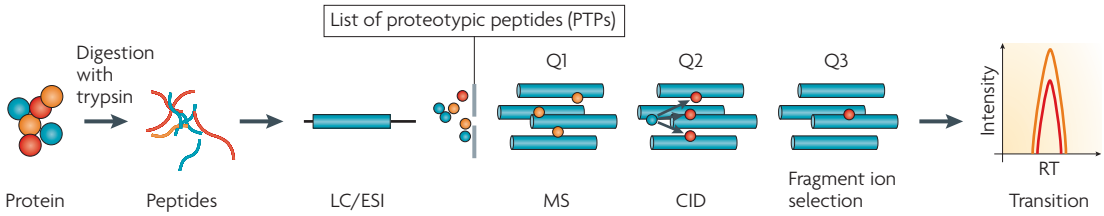


Figure 2 | **Targeted proteomics.** Targeted protein identification can be achieved by selected reaction monitoring using triple quadrupole instruments. Peptide mixtures are fractionated and ionized by electrospray ionization (ESI) and liquid chromatography (LC). The mass spectrometer can be programmed with a list of target proteotypic peptides. A specific peptide ion will be selected in the first quadrupole (Q1). A second mass filter in the third quadrupole (Q3) allows filtering of the corresponding fragment ions following collision-induced dissociation (CID) in the second quadrupole (Q2). The resulting precursor fragment ion pairs (transitions) are highly specific for a given proteotypic peptide and are therefore unique for a given protein in the analysed proteome. By including isotope-labelled versions of the proteotypic peptides that are being analysed, it is possible to simultaneously determine the absolute amounts of the protein. The approximate retention time (RT) information can be used to restrict the time for the detection of a specific transition and therefore allows the detection of multiple peptide ions in one measurement, a process that is referred to as scheduled selected reaction monitoring.

To tailor targeted MS experiments towards a specific biological question (for example, how is a known signalling pathway affected in the mutant phenotype?), the increasing amounts of information on pathways, protein interactions, gene expression changes and gene ontologies can be accessed using public databases to establish a list of target proteins. Next, proteotypic peptides for the chosen target proteins can be selected using either software tools to predict their sequences[29] or accessed from existing MS data depositories, which can also provide the corresponding *m*/*z* values and chromatographic retention times of these peptides[30,31]. The subsequent identification and quantification of proteotypic peptides are based on precursor and fragment ion *m*/*z* ratios (which are also referred to as transitions) that are highly specific for a particular peptide. The whole process is referred to as SRM or multiple reaction monitoring (MRM) (FIG. 2). A variant of this approach that is referred to as scheduled SRM allows the quantitative analysis of several hundred proteins in a single MS experiment[28,32].

In summary, the process of targeted MS analysis can be viewed as a western blot using a mass spectrometer, which in principle allows the detection and quantification of all of the proteins for which information on observable proteotypic peptides is available. Targeted MS is currently the most sensitive MS detection method and allows the detection of yeast proteins that are present at <50 copies per cell from total yeast lysate or of proteins with <10 copies per cell after 2-dimensional peptide fractionation[31,33]. In combination with isotope-labelled reference peptides, absolute quantitative information on protein abundance can be obtained with high sensitivity. However, the increase in sensitivity and reproducibility of targeted MS analysis comes at a cost: SRM analysis is currently limited to a few hundred peptides that can be analysed in a single SRM experiment and requires careful assay optimization and validation. However, the substantial increase in sensitivity, accuracy and reproducibility that has been shown by recent SRM studies[31,34–36] sets the stage for hypothesis-driven protein profiling of mutant proteomes with an unprecedented sensitivity in the near future.

Proteotypic peptide
Proteotypic peptides can be observed by mass spectrometry and uniquely identify a specific protein or a specific isoform of a protein.

Analysis of posttranslational changes in mutants

Covalent modifications of proteins (for example, phosphorylation, acetylation, methylation, glycosylation, ubiquitin and ubiquitin-like modification) play important parts in the control of cellular functions. Depending on the cellular state, cell type and environmental condition these modifications control the activity, localization and stability of proteins and their interactions with other macromolecules. Functional genomics cannot provide experimental evidence for protein modifications other than protein sequence information for the *in silico* prediction of candidate sites of modification. Over 200 different PTMs have been described for proteins[37] and the list is still increasing.

Several sensitive techniques for the global and targeted analysis of PTMs of proteins by MS have been introduced (for detailed reviews, see REFS 38,39). Protein modifications cause distinct changes in peptide mass, which can be measured in a mass spectrometer. For example, phosphorylation of a tyrosine-containing peptide or its acetylation on lysine causes a discrete mass increase of 80 Da or 42 Da, respectively, which can be detected in precursor and fragment ion spectra. Tryptic cleavage of ubiquitinylated proteins results in a glycine–glycine modification with a corresponding mass increase of 114 Da. The fragmentation of a modified peptide can also release modification-specific neutral-loss ions, which are detectable in MS/MS spectra. For example, phosphoserine-containing peptides lose phosphoric acid after collision-induced dissociation fragmentation, which causes a characteristic neutral-loss peak of 98 Da in the corresponding MS/MS spectrum. Computer programmes are continuously improved for the systematic scanning and annotation of PTMs from MS and MS/MS data.

The analysis of PTMs by MS is particularly challenging as often only a small fraction of proteins harbours a specific modification. This problem has been addressed by the specific enrichment of modified proteins and

© 2009 Macmillan Publishers Limited. All rights reserved

peptides to increase the dynamic range and sensitivity in PTM analysis by MS. Some PTMs can be enriched by derivatization of protein modifications to make them accessible to chemical solid-phase capture techniques, and other PTMs can be purified using, for example, metal affinity chromatography (such as titanium dioxide or immobilized metal affinity chromatography; IMAC) or antibodies that are specific for a given modification. Despite their potential, only a few studies have used these techniques to systematically study PTMs in mutant proteomes. Here, we describe global phosphoproteome analysis in more detail, as the most substantial progress has been achieved in this area.

Protein phosphorylation plays a central part in the regulation of almost all eukaryotic cellular processes. Data on the phosphorylation status of a cellular proteome are therefore informative as they suggest which proteins are regulated and what signalling networks might be activated in a given mutant phenotype. The size of the human phosphoproteome is unknown but it can be estimated from the >50,000 known sites, which map to >8,000 proteins present in public databases (see Further Information for a link to the PhosphoSitePlus website), that >100,000 phosphorylated sites may exist in the human proteome. The complexity of organismal phosphoproteomes and the problem of the dynamic range that is associated with the temporal dynamics of protein phosphorylation make global phosphoproteome analysis impractical for classical biochemical workflows.

This situation changed when phosphopeptide-enrichment techniques were developed and combined with quantitative MS. These combined methods now allow the global identification and quantification of thousands of phosphorylation sites[40,41]. All of the approaches for global analysis of phosphoproteins combine steps for the enrichment of phosphorylated proteins or peptides and strategies to reduce the sample complexity before MS (FIG. 3). Additional strategies that build on the improved accuracy of modern MS instruments and alternative methods for the fragmentation of phosphopeptides (for example, electron transfer dissociation) are available to further improve the accuracy and sensitivity of phophopeptide identification by LC–MS. Several studies recently showed that thousands of new protein phosphorylation sites could be identified from low sample amounts[42,43], and these studies provide fascinating new opportunities for also studying phosphoproteome changes in mutant cells.

In one of the few studies that have applied MS-based phosphoproteomics to the analysis of mutant organisms, the roles of the PI3K-related Mec1 and Tel1 kinases in DNA damage checkpoint control have been analysed in yeast[44]. DNA damage was induced in wild-type and *mec1* and *tel1* mutant yeast strains and differential phosphorylation was analysed after IMAC phosphopeptide enrichment using differential isotope peptide labelling and quantitative LC–MS analysis. The authors found that 62 out of the 2,689 phosphosites identified were specifically affected in the kinase-mutant strains. Follow-up experiments using *in vitro* kinase assays indicated a panel of new *in vivo* kinase substrates, including

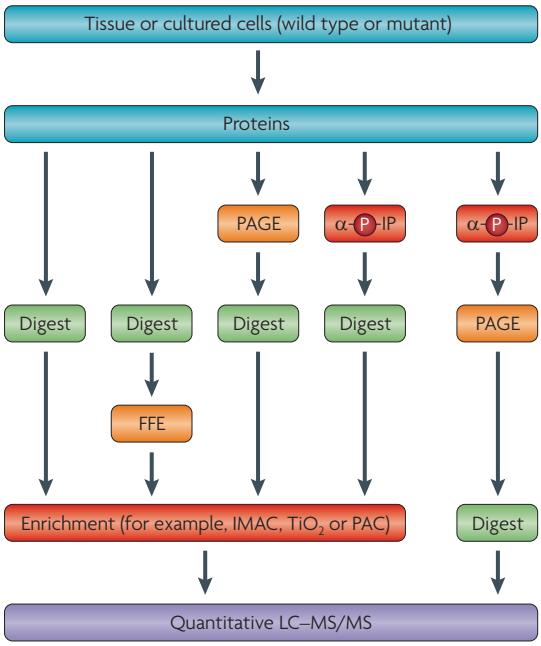


Figure 3 | **Strategies for global phosphoproteome profiling of mutant proteomes.** Proteins purified from mutant and wild-type control cells are processed using the methods indicated. Several methods to reduce sample complexity (orange boxes), including one-dimensional polyacrylamide gel electrophoresis (PAGE) and free-flow electrophoresis (FFE), are used at different steps, in combination with phosphopeptide-enrichment techniques (red boxes; titanium dioxide (TiO2) spheres, immobilized metal affinity chromatography (IMAC) and phosphoramidate chemistry (PAC)) to improve the overall proteome coverage. These approaches are typically combined with quantitative mass spectrometry (MS) workflows for measuring the relative and absolute changes in mutant phosphoproteomes. IP, immunoprecipitation; LC–MS/MS, liquid chromatography–tandem mass spectrometry; P, phosphorylation site.

proteins such as Rad9 and Rad17, which are linked to the DNA damage checkpoint pathway. Yeast offers the unique possibility to profile all kinase- and phosphatase-deletion mutants. It is expected that similar studies in yeast and other genetic systems will provide important new insights into functional kinase and phosphatase substrate networks. The approach is not limited to the analysis of mutated kinase and phosphatase genes as protein phosphorylation can be regarded as a sensitive indicator for the pathways that are engaged in the cellular responses that follow a genetic perturbation.

Global profiling of mutant PTMs by MS is still in its infancy as techniques for the analysis of PTMs have just been developed. So far, phosphoproteome profiling is the only method that has been successfully applied to genetic mutants. Extending such analyses to other PTMs promises to provide new opportunities for the molecular analysis of mutants. Because many PTMs are highly regulated by cell signalling, future studies also need to provide kinetic information about global PTM changes in mutants using quantitative MS approaches.

© 2009 Macmillan Publishers Limited. All rights reserved

Analysis of altered protein interactions

Almost all proteins function in the context of specific interactions with other proteins. Published interaction data from systematic protein interaction studies and classical biochemical analysis can be accessed using public databases. Unfortunately, the existing databases often lack important information about which cellular conditions the respective interactions were observed under and are incomplete and almost non-existent for mutant proteins in higher eukaryotes.

During the past few years, many AP–MS workflows have been introduced that allow the efficient and sensitive detection of protein-binding partners[45,46]. Protein complexes are purified either using antibodies that recognize the endogenously expressed protein (or its mutant form) or using an affinity tag that is fused to the protein of interest, and are subsequently analysed by MS-based proteomics as described above. In cases in which a mutation maps to a protein of unknown function, AP–MS/MS can provide valuable biochemical insights based on the known functions of the binding partners identified. Mutations have been found that disrupt or increase protein–protein interactions, which can have profound effects on the control of a range of cellular processes. Not surprisingly, such mutations are not uncommon in the 2,000 genes that have known mutations in human diseases[47]. Recently, several experimental strategies have been introduced that combine affinity purification with quantitative MS to measure relative[48] and absolute[49] changes in protein interactions, which will facilitate the analysis of mutant proteins (FIG. 4a).

Using AP-MS, many human tumour suppressor gene or oncogene products have been linked to known signalling pathways and their roles in oncogenesis have been elucidated on the basis of altered protein interaction data. For example, a recent AP–MS study showed that the Wilms' tumour suppressor protein, WTX, is part of a complex that contains β-catenin, axin 1, β-transducin repeat-containing protein 2 (β-TRCP2) and adenomatous polyposis coli (APC), in which it seems to control tumour growth through the degradation of β-catenin[50], suggesting a model for tumour suppression by WTX through the β-catenin pathway.

Mutations often have pleiotropic effects, which are hard to understand if proteins are considered as having only one function. An increasing number of proteins have been shown to play multiple parts in cellular regulation[51]. There is an increasing amount of AP–MS/MS data to suggest that proteins, such as the well-studied protein serine/threonine phosphatase PP2A, can be partitioned in multiple protein complexes, each of which has different cellular roles[45,52,53]. Such studies could therefore be applied to understand the pleiotropic phenotypes that are caused by genetic mutations.

Another question that can at least partly be addressed by AP–MS is, how do mutations in different genes produce identical or related phenotypes? In *S. cerevisiae*, in which protein interactions have been mapped by AP–MS/MS on a proteome-wide scale[54–57], it is now possible to integrate protein interaction data with genetic data that have been obtained from large-scale phenotypic profiling experiments (FIG. 4b). In many cases, the occurrence of common phenotypes can be explained by the fact that genes with similar phenotypes encode proteins that interact with one another to form a functional

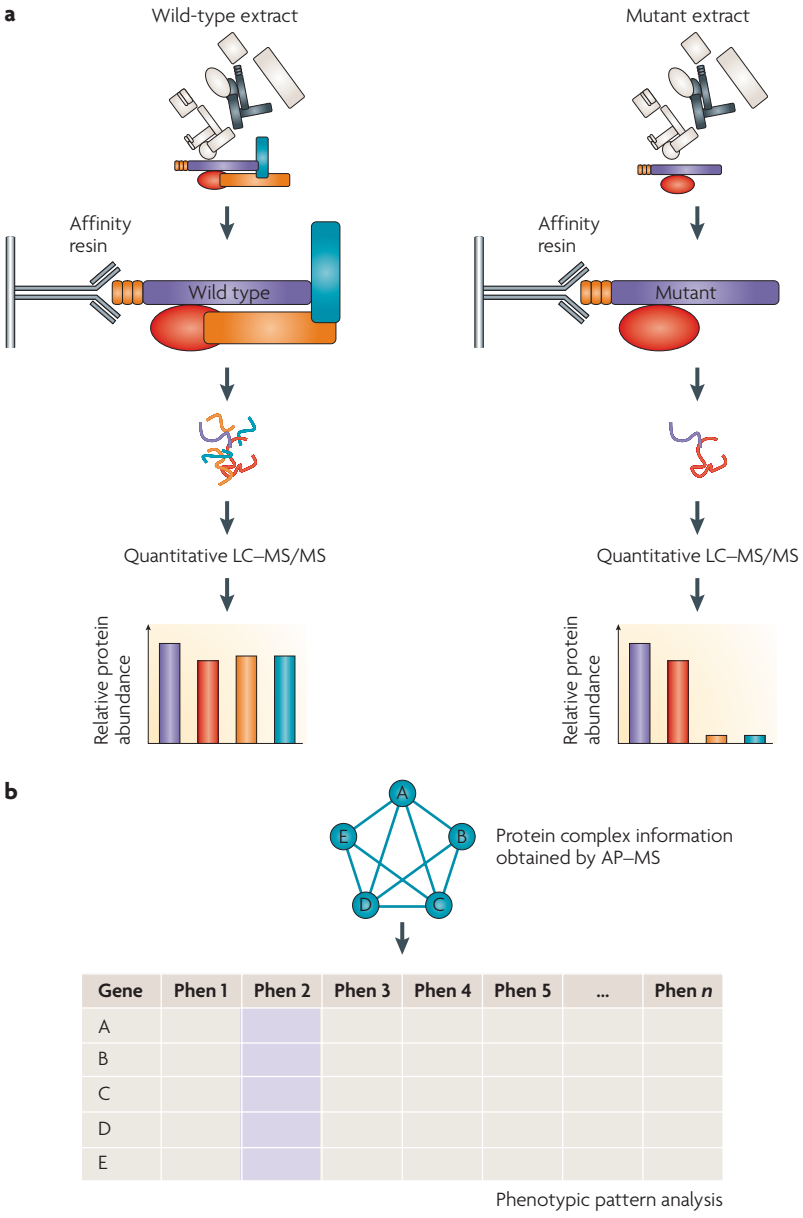


Figure 4 | **Mass spectrometry-based interaction proteomics for the analysis of mutant phenotypes. a** | Analysis of mutant interaction proteomes by quantitative affinity purification–mass spectrometry (AP–MS). Complexes containing the wild-type or the corresponding mutant bait protein are purified by affinity purification. Changes in the stochiometries of the interacting proteins that are caused by the mutation can be measured by including known amounts of isotope-labelled reference peptides, which correspond to the proteotypic peptides of the bait and prey proteins. This is followed by quantitative liquid chromatography–tandem mass spectrometry (LC-MS/MS) analysis[84]. **b** | Protein interaction information can be used to explain the occurrence of similar phenotypes. In the hypothetical example shown, the genes that encode components of protein complexes (blue nodes) show a similar phenotypic pattern when they are mutated. Phen, phenotype. Images modified, with permission, from *Nature* REF. 54 © (2006) Macmillan Publishers Ltd. All rights reserved.

© 2009 Macmillan Publishers Limited. All rights reserved

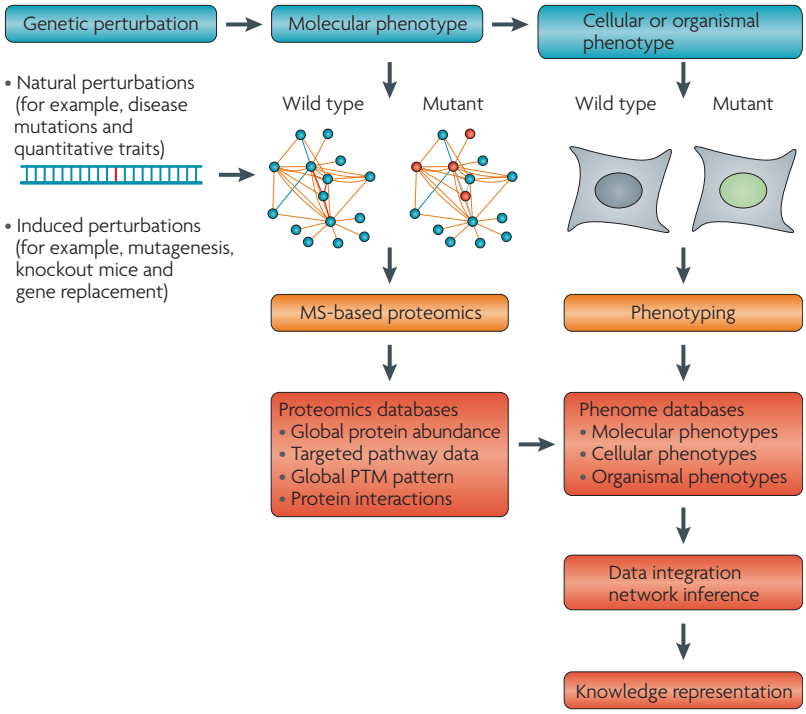


Figure 5 | **Using proteomic data in network biology.** A data integration strategy for inference of genetically perturbed molecular networks based on proteomics, phenomics and functional genomics data. MS, mass spectrometry; PTM, posttranslational modification.

protein complex[54]. It has also been found that synthetic lethal interactions are not randomly distributed in the yeast interactome but are associated with multiprotein complexes[58]. Although the density of the data on physical and genetic interactions is currently insufficient for network inference analysis in higher eukaryotes, it is expected that the integration of systematic AP–MS data with phenotypic information will identify similar relationships in the near future, and may therefore also help to explain phenotypic data in higher organisms.

Integrating proteomic and phenomic data

As discussed in the previous sections, proteomics information can be used to improve existing gene models, to profile molecular phenotypes at the levels of protein abundance, PTMs and protein–protein interactions, and to obtain specific pathway information using targeted MS strategies. As for other 'omics' strategies, the challenge lies in obtaining knowledge from the data that are collected. A promising approach includes the reconstruction of functional molecular networks through the integration of high-quality information from functional genomics and proteomics data (FIG. 5).

Techniques for the successful integration of large data sets have been proposed and reviewed recently[59,60]. Few examples exist, however, that show the value of integrating genetic information with MS-based proteomics data. An interesting example has been presented recently in which the integration of genetics and proteomics data has been successfully used to dissect the function of the small ubiquitin-related modifier (SUMO) system[61]. Yeast SUMO pathway mutants were subjected to synthetic genetic array (SGA) screening against a genome-wide collection of viable yeast mutants. The resulting genetic interaction data were integrated with a proteomics data set, which included known SUMO conjugation targets and data from AP–MS analyses. The analysis of these integrated data led to the development of a molecular network model for the SUMO pathway that is linked to >15 cellular processes. Using a similar approach, another study functionally dissected a network of protein complexes that are involved in chromosome biology on the basis of epistatic miniarray profiles and systematic AP–MS/MS information[61].

A few pioneering studies have integrated proteomics and functional genomics data for subsequent network inference analysis with the goal of improving the prediction of cancer phenotypes[63,64]. This approach is based on the assumption that the modularity of oncogenic networks is altered in transformed cells. The Wrana laboratory recently studied how the modularity of oncogenic pathways is changed in two cohorts of patients with breast cancer that have either a good or poor prognosis[64]. The authors used curated protein interaction information and large mRNA expression data sets to identify protein hubs and their interacting partners that have different coexpression patterns in groups of patients with breast cancer with either a good or poor prognosis. They identified 256 hub proteins that showed an altered gene expression correlation of their binding partners in the patient group with a poor breast cancer prognosis. These results suggest that altered network modularity can be used as a prognostic signature for cancers. The increasing coverage of human protein–protein interaction data sets by future large-scale AP–MS studies will increase the predictive performance of such integrative approaches.

Previous studies showed that the successful use of MS-based proteomics crucially depends on the proper annotation and accessibility of these data through public databases. Several initiatives have led to standardized machine-readable formats for the description of MS experiments using controlled vocabularies and the exchange of MS data[65,66]. Both processed and raw data from proteomics experiments can be accessed through many proteomics databases (see Further information for a list of useful websites). These standardizing efforts will continue to allow the efficient integration of MS experimental data into existing networks of genotype and phenotype databases[67]. Such efforts to create a cyberinfrastructure for complex biological data[68] will be important in the near future as more phenotypic and proteomics data will become available for bridging the genotype–phenotype gap.

Conclusion

Phenotypes originate from genetically perturbed molecular networks in mutant cells. Functional genomics and

Synthetic genetic array
This has been primarily applied to yeast and is a technology for the high-throughput analysis of genetic interactions. Yeast deletion strains are crossed with each other to systematically generate double mutant strains. The resulting growth phenotypes are determined based on the size of the resulting double mutant colonies.

© 2009 Macmillan Publishers Limited. All rights reserved

metabolomics were, and still are, instrumental in the ongoing process to understand and model molecular networks. However, these approaches fail to take into account an essential class of network components — proteins. In this Review, we have shown that the progress in MS-based proteomics now allows systems-wide and hypothesis-driven analysis of mutant proteomes. However, this is still a new area of research and there are still limitations to addressing the full complexity and dynamic nature of cellular proteomes.

New analytical concepts, such as targeted proteomics, are being explored to overcome many of the existing limitations, and MS instrumentation will continue to improve the sensitivity and accuracy of current MS measurements. It will be equally important to develop an effective computational framework for the integration of proteomics data with phenomic and functional genomic information to reconstruct molecular networks and how they are perturbed in mutant cells — an essential step in bridging the genotype–phenotype gap.

1. de Godoy, L. M. *et al.* Comprehensive mass-spectrometry-based proteome quantification of haploid versus diploid yeast. *Nature* **455**, 1251–1254 (2008).
2. Ideker, T. *et al.* Integrated genomic and proteomic analyses of a systematically perturbed metabolic network. *Science* **292**, 929–934 (2001).
3. Wolters, D. A., Washburn, M. P. & Yates, J. R. 3rd. An automated multidimensional protein identification technology for shotgun proteomics. *Anal. Chem.* **73**, 5683–5690 (2001).
4. Ghaemmaghami, S. *et al.* Global analysis of protein expression in yeast. *Nature* **425**, 737–741 (2003).
5. Anderson, N. L. *et al.* The human plasma proteome: a nonredundant list developed by combination of four separate sources. *Mol. Cell. Proteomics* **3**, 311–326 (2004).
6. Schrimpf, S. P. *et al.* Comparative functional analysis of the *Caenorhabditis elegans* and *Drosophila melanogaster* proteomes. *PLoS Biol.* **7**, e48 (2009).
7. Baerenfaller, K. *et al.* Genome-scale proteomics reveals *Arabidopsis thaliana* gene models and proteome dynamics. *Science* **320**, 938–941 (2008).
8. Brunner, E. *et al.* A high-quality catalog of the *Drosophila melanogaster* proteome. *Nature Biotechnol.* **25**, 576–583 (2007).
9. Bantscheff, M., Schirle, M., Sweetman, G., Rick, J. & Kuster, B. Quantitative mass spectrometry in proteomics: a critical review. *Anal. Bioanal. Chem.* **389**, 1017–1031 (2007).
This review gives a good overview of the various quantitative MS approaches that are currently applied for quantitative proteomics.
10. MacCoss, M. J., Wu, C. C., Liu, H., Sadygov, R. & Yates, J. R. 3rd. A correlation algorithm for the automated quantitative analysis of shotgun proteomics data. *Anal. Chem.* **75**, 6912–6921 (2003).
11. Liu, H., Sadygov, R. G. & Yates, J. R. 3rd. A model for random sampling and estimation of relative protein abundance in shotgun proteomics. *Anal. Chem.* **76**, 4193–4201 (2004).
12. Colinge, J., Chiappe, D., Lagache, S., Moniatte, M. & Bougueleret, L. Differential proteomics via probabilistic peptide identification scores. *Anal. Chem.* **77**, 596–606 (2005).
13. Strittmatter, E. F., Ferguson, P. L., Tang, K. & Smith, R. D. Proteome analyses using accurate mass and elution time peptide tags with capillary LC time-of-flight mass spectrometry. *J. Am. Soc. Mass Spectrom.* **14**, 980–991 (2003).
14. Foss, E. J. *et al.* Genetic basis of proteome variation in yeast. *Nature Genet.* **39**, 1369–1375 (2007).
15. Dong, M. Q. *et al.* Quantitative mass spectrometry identifies insulin signaling targets in *C. elegans*. *Science* **317**, 660–663 (2007).
16. Jenkins, L. M. *et al.* Quantitative proteomics analysis of the effects of ionizing radiation in wild type and p53^{K317R} knock-in mouse thymocytes. *Mol. Cell. Proteomics* **7**, 716–727 (2008).
17. Martin, B. *et al.* iTRAQ analysis of complex proteome alterations in 3xTgAD Alzheimer's mice: understanding the interface between physiology and disease. *PLoS ONE* **3**, e2750 (2008).
18. Chiang, M. C. *et al.* Systematic uncovering of multiple pathways underlying the pathology of Huntington disease by an acid-cleavable isotope-coded affinity tag approach. *Mol. Cell. Proteomics* **6**, 781–797 (2007).
19. Liao, L., Park, S. K., Xu, T., Vanderklish, P. & Yates, J. R. 3rd. Quantitative proteomic analysis of primary neurons reveals diverse changes in synaptic protein content in *fmr1* knockout mice. *Proc. Natl Acad. Sci. USA* **105**, 15281–15286 (2008).
20. Kruger, M. *et al.* SILAC mouse for quantitative proteomics uncovers kindlin-3 as an essential factor for red blood cell function. *Cell* **134**, 353–364 (2008).
In this paper an approach based on the metabolic labelling of entire mice with isotope-labelled amino acids is presented for MS-based profiling of mouse mutant proteomes.
21. Wu, C. C., MacCoss, M. J., Howell, K. E., Matthews, D. E. & Yates, J. R. 3rd. Metabolic labeling of mammalian organisms with stable isotopes for quantitative proteomic analysis. *Anal. Chem.* **76**, 4951–4959 (2004).
22. McClatchy, D. B., Liao, L., Park, S. K., Venable, J. D. & Yates, J. R. Quantification of the synaptosomal proteome of the rat cerebellum during post-natal development. *Genome Res.* **17**, 1378–1388 (2007).
23. Alagaratnam, S. *et al.* Serum protein profiling in mice: identification of factor XIIIa as a potential biomarker for muscular dystrophy. *Proteomics* **8**, 1552–1563 (2008).
24. Faca, V. M. *et al.* A mouse to human search for plasma proteome changes associated with pancreatic tumor development. *PLoS Med.* **5**, e123 (2008).
A quantitative MS approach in combination with mouse genetics identified candidate biomarkers for early pancreatic tumour development.
25. Hung, K. E. *et al.* Comprehensive proteome analysis of an Apc mouse model uncovers proteins associated with intestinal tumorigenesis. *Cancer Prev. Res.* **2**, 224–233 (2009).
26. Kuster, B., Schirle, M., Mallick, P. & Aebersold, R. Scoring proteomes with proteotypic peptide probes. *Nature Rev. Mol. Cell Biol.* **6**, 577–583 (2005).
27. Malmstrom, J., Lee, H. & Aebersold, R. Advances in proteomic workflows for systems biology. *Curr. Opin. Biotechnol.* **18**, 378–384 (2007).
28. Lange, V., Picotti, P., Domon, B. & Aebersold, R. Selected reaction monitoring for quantitative proteomics: a tutorial. *Mol. Syst. Biol.* **4**, 222 (2008).
This tutorial gives a good introduction into the emerging field of targeted proteomics using SRM.
29. Mallick, P. *et al.* Computational prediction of proteotypic peptides for quantitative proteomics. *Nature Biotechnol.* **25**, 125–131 (2007).
30. Deutsch, E. W., Lam, H. & Aebersold, R. PeptideAtlas: a resource for target selection for emerging targeted proteomics workflows. *EMBO Rep.* **9**, 429–434 (2008).
31. Picotti, P. *et al.* A database of mass spectrometric assays for the yeast proteome. *Nature Methods* **5**, 913–914 (2008).
32. Stahl-Zeng, J. *et al.* High sensitivity detection of plasma proteins by multiple reaction monitoring of N-glycosites. *Mol. Cell. Proteomics* **6**, 1809–1817 (2007).
33. Picotti, P., Bodenmiller, B., Mueller, L. N., Domon, B. & Aebersold, R. Full dynamic range proteome analysis of *S. cerevisiae* by targeted proteomics. *Cell* 6 Aug 2009 [epub ahead of print].
This paper illustrates how SRM targeted proteomics can be used to resolve the changes in the abundance of yeast metabolic enzymes during metabolic transitions.
34. Wolf-Yadlin, A., Hautaniemi, S., Lauffenburger, D. A. & White, F. M. Multiple reaction monitoring for robust quantitative proteomic analysis of cellular signaling networks. *Proc. Natl Acad. Sci. USA* **104**, 5860–5865 (2007).
35. Lange, V. *et al.* Targeted quantitative analysis of *Streptococcus pyogenes* virulence factors by multiple reaction monitoring. *Mol. Cell. Proteomics* **7**, 1489–1500 (2008).
36. Addona, T. A. *et al.* Multi-site assessment of the precision and reproducibility of multiple reaction monitoring-based measurements of proteins in plasma. *Nature Biotechnol.* **27**, 633–641 (2009).
A multilaboratory study that shows the high reproducibility and sensitivity of SRM-based assays for proteome profiling.
37. Krishna, R. G. & Wold, F. Post-translational modification of proteins. *Adv. Enzymol. Relat. Areas Mol. Biol.* **67**, 265–298 (1993).
38. Jensen, O. N. Interpreting the protein language using proteomics. *Nature Rev. Mol. Cell Biol.* **7**, 391–403 (2006).
This review provides a good overview on the MS-based analysis of PTMs.
39. Witze, E. S., Old, W. M., Resing, K. A. & Ahn, N. G. Mapping protein post-translational modifications with mass spectrometry. *Nature Methods* **4**, 798–806 (2007).
40. Bodenmiller, B., Mueller, L. N., Mueller, M., Domon, B. & Aebersold, R. Reproducible isolation of distinct, overlapping segments of the phosphoproteome. *Nature Methods* **4**, 231–237 (2007).
41. Olsen, J. V. *et al.* Global, *in vivo*, and site-specific phosphorylation dynamics in signaling networks. *Cell* **127**, 635–648 (2006).
42. Bodenmiller, B. *et al.* PhosphoPep — a phosphoproteome resource for systems biology research in *Drosophila* Kc167 cells. *Mol. Syst. Biol.* **3**, 139 (2007).
43. Dephoure, N. *et al.* A quantitative atlas of mitotic phosphorylation. *Proc. Natl Acad. Sci. USA* **105**, 10762–10767 (2008).
44. Smolka, M. B., Albuquerque, C. P., Chen, S. H. & Zhou, H. Proteome-wide identification of *in vivo* targets of DNA damage checkpoint kinases. *Proc. Natl Acad. Sci. USA* **104**, 10364–10369 (2007).
In this paper, a quantitative phosphoproteomics approach is described for the analysis of yeast cells that lack DNA checkpoint kinases, which identified novel *in vivo* substrates for these kinases.
45. Glatter, T., Wepf, A., Aebersold, R. & Gstaiger, M. An integrated workflow for charting the human interaction proteome: insights into the PP2A system. *Mol. Syst. Biol.* **5**, 237 (2009).
46. Rigaut, G. *et al.* A generic protein purification method for protein complex characterization and proteome exploration. *Nature Biotechnol.* **17**, 1030–1032 (1999).
47. Charbonnier, S., Gallego, O. & Gavin, A. C. The social network of a cell: recent advances in interactome mapping. *Biotechnol. Annu. Rev.* **14**, 1–28 (2008).
48. Pflieger, D. *et al.* Quantitative proteomic analysis of protein complexes: concurrent identification of interactors and their state of phosphorylation. *Mol. Cell. Proteomics* **7**, 326–346 (2008).
49. Wepf, A., Glatter, T., Schmidt, A., Aebersold, R. & Gstaiger, M. Quantitative interaction proteomics using mass spectrometry. *Nature Methods* **6**, 203–205 (2009).
50. Major, M. B. *et al.* Wilms tumor suppressor WTX negatively regulates WNT/β-catenin signaling. *Science* **316**, 1043–1046 (2007).
In this paper, an AP–MS approach was used to identify a link between the Wilms' tumour suppressor protein WTX and degradation of β-catenin.
51. Jeffery, C. J. Moonlighting proteins. *Trends Biochem. Sci.* **24**, 8–11 (1999).
52. Gingras, A. C. *et al.* A novel, evolutionarily conserved protein phosphatase complex involved in cisplatin sensitivity. *Mol. Cell. Proteomics* **4**, 1725–1740 (2005).

© 2009 Macmillan Publishers Limited. All rights reserved

53. Goudreault, M. *et al.* A PP2A phosphatase high density interaction network identifies a novel striatin-interacting phosphatase and kinase complex linked to the cerebral cavernous malformation 3 (CCM3) protein. *Mol. Cell. Proteomics* **8**, 157–171 (2009).
54. Gavin, A. C. *et al.* Proteome survey reveals modularity of the yeast cell machinery. *Nature* **440**, 631–636 (2006).
55. Gavin, A. C. *et al.* Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature* **415**, 141–147 (2002).
56. Ho, Y. *et al.* Systematic identification of protein complexes in *Saccharomyces cerevisiae* by mass spectrometry. *Nature* **415**, 180–183 (2002).
57. Krogan, N. J. *et al.* Global landscape of protein complexes in the yeast *Saccharomyces cerevisiae*. *Nature* **440**, 637–643 (2006).
58. Le Meur, N. & Gentleman, R. Modeling synthetic lethality. *Genome Biol.* **9**, R135 (2008).
59. Zhu, X., Gerstein, M. & Snyder, M. Getting connected: analysis and principles of biological networks. *Genes Dev.* **21**, 1010–1024 (2007).
60. Joyce, A. R. & Palsson, B. O. The model organism as a system: integrating 'omics' data sets. *Nature Rev. Mol. Cell Biol.* **7**, 198–210 (2006).
61. Makhnevych, T. *et al.* Global map of SUMO function revealed by protein–protein interaction and genetic networks. *Mol. Cell* **33**, 124–135 (2009).
62. Collins, S. R. *et al.* Functional dissection of protein complexes involved in yeast chromosome biology using a genetic interaction map. *Nature* **446**, 806–810 (2007).
63. Chuang, H. Y., Lee, E., Liu, Y. T., Lee, D. & Ideker, T. Network-based classification of breast cancer metastasis. *Mol. Syst. Biol.* **3**, 140 (2007).
64. Taylor, I. W. *et al.* Dynamic modularity in protein interaction networks predicts breast cancer outcome. *Nature Biotechnol.* **27**, 199–204 (2009).
Network inference using protein–protein interaction data and gene expression data identified a change in the modularity of oncogenic pathways between patients with breast cancer who had either a good or poor prognosis.
65. Pedrioli, P. G. *et al.* A common open representation of mass spectrometry data and its application to proteomics research. *Nature Biotechnol.* **22**, 1459–1466 (2004).
66. Taylor, C. F. Minimum reporting requirements for proteomics: a MIAPE primer. *Proteomics* **6** (Suppl. 2), 39–44 (2006).
67. Thorisson, G. A., Muilu, J. & Brookes, A. J. Genotype–phenotype databases: challenges and solutions for the post-genomic era. *Nature Rev. Genet.* **10**, 9–18 (2009).
68. Stein, L. D. Towards a cyberinfrastructure for the biological sciences: progress, visions and challenges. *Nature Rev. Genet.* **9**, 678–688 (2008).
69. Bonaldi, T. *et al.* Combined use of RNAi and quantitative proteomics to study gene function in *Drosophila*. *Mol. Cell* **31**, 762–772 (2008).
70. Li, K. W. *et al.* Quantitative proteomics and protein network analysis of hippocampal synapses of CaMKIIα mutant mice. *J. Proteome Res.* **6**, 3127–3133 (2007).
71. Shiio, Y. *et al.* Quantitative proteomic analysis of Myc-induced apoptosis: a direct role for Myc induction of the mitochondrial chloride ion channel, mtCLIC/CLIC4. *J. Biol. Chem.* **281**, 2750–2756 (2006).
72. Gygi, S. P. *et al.* Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nature Biotechnol.* **17**, 994–999 (1999).
73. Schmidt, A., Kellermann, J. & Lottspeich, F. A novel strategy for quantitative proteomics using isotope-coded protein labels. *Proteomics* **5**, 4–15 (2005).
74. Choe, L. *et al.* 8-plex quantitation of changes in cerebrospinal fluid protein expression in subjects undergoing intravenous immunoglobulin treatment for Alzheimer's disease. *Proteomics* **7**, 3651–3660 (2007).
75. Chong, P. K., Gan, C. S., Pham, T. K. & Wright, P. C. Isobaric tags for relative and absolute quantitation (iTRAQ) reproducibility: implication of multiple injections. *J. Proteome Res.* **5**, 1232–1240 (2006).
76. Desiderio, D. M. & Zhu, X. Quantitative analysis of methionine enkephalin and β-endorphin in the pituitary by liquid secondary ion mass spectrometry and tandem mass spectrometry. *J. Chromatogr. A* **794**, 85–96 (1998).
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. 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).
79. Callister, S. J. *et al.* Normalization approaches for removing systematic biases associated with mass spectrometry and label-free proteomics. *J. Proteome Res.* **5**, 277–286 (2006).
80. Andreev, V. P. *et al.* A new algorithm using cross-assignment for label-free quantitation with LC–LTQ-FT MS. *J. Proteome Res.* **6**, 2186–2194 (2007).
81. Mueller, L. N. *et al.* SuperHirn — a novel tool for high resolution LC–MS-based peptide/protein profiling. *Proteomics* **7**, 3470–3480 (2007).
82. Rinner, O. *et al.* An integrated mass spectrometric and computational framework for the analysis of protein interaction networks. *Nature Biotechnol.* **25**, 345–352 (2007).
83. Mueller, L. N., Brusniak, M. Y., Mani, D. R. & Aebersold, R. An assessment of software solutions for the analysis of mass spectrometry based quantitative proteomics data. *J. Proteome Res.* **7**, 51–61 (2008).
84. Wepf, A., Glatter, T., Schmidt, A., Aebersold, R. & Gstaiger, M. Quantitative interaction proteomics using mass spectrometry. *Nature Methods* **6**, 203–205 (2009).

Acknowledgements
This work was supported by the by the Swiss National Science Foundation grant 31000-10767, by federal funds from the National Heart, Lung and Blood Institute, the National Institutes of Health grant N01-HV-28179 and by SystemsX.ch, the Swiss initiative for systems biology.

DATABASES
Entrez Gene: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=gene
daf-2 | *Fermt3*
UniProtKB: http://www.uniprot.org
β-catenin | β-TRCP2 | APC | axin 1 | Fus3 | Gpa1 | Mec1 | Rad9 | Rad17 | Ste2 | Ste4 | Ste5 | Ste12 | Ste18 | Tel1 | WTX

FURTHER INFORMATION
Matthias Gstaiger's homepage: http://www.imsb.ethz.ch/researchgroup/gstaigem
Ruedi Aebersold's homepage: http://www.imsb.ethz.ch/researchgroup/rudolfa
BioGrid protein interaction database: http://www.thebiogrid.org
IntAct protein interaction database: http://www.ebi.ac.uk/intact
MRMAtlas compendium of targeted proteomics assays: http://www.mrmatlas.org
PeptideAtlas MS-based peptide data: http://www.peptideatlas.org
PeptideSieve tool for prediction of proteotypic peptides: http://tools.proteomecenter.org/wiki/index.php?title=Software:PeptideSieve
PhosphoPep MS-based data on phosphopeptides: http://www.phosphopep.org
PhosphoSitePlus: http://www.phosphosite.org
Pride repository for proteomics data: http://www.ebi.ac.uk/pride
Protein Interaction Network Analysis database (PINA): http://csbi.ltdk.helsinki.fi/pina
Trans Proteomic Pipeline (TPP) tools for MS/MS proteomics: http://tools.proteomecenter.org/wiki/index.php?title=Software:TPP
X!Tandem protein identification software: http://www.thegpm.org/TANDEM

ALL LINKS ARE ACTIVE IN THE ONLINE PDF

© 2009 Macmillan Publishers Limited. All rights reserved