Next-generation proteomics: towards an integrative view of proteome dynamics

A. F. Maarten Altelaar[1,2], Javier Munoz[1,2,3]* and Albert J. R. Heck[1,2]*

Abstract | Next-generation sequencing allows the analysis of genomes, including those representing disease states. However, the causes of most disorders are multifactorial, and systems-level approaches, including the analysis of proteomes, are required for a more comprehensive understanding. The proteome is extremely multifaceted owing to splicing and protein modifications, and this is further amplified by the interconnectivity of proteins into complexes and signalling networks that are highly divergent in time and space. Proteome analysis heavily relies on mass spectrometry (MS). MS-based proteomics is starting to mature and to deliver through a combination of developments in instrumentation, sample preparation and computational analysis. Here we describe this emerging next generation of proteomics and highlight recent applications.

[1]*Biomolecular Mass Spectrometry and Proteomics, Bijvoet Center for Biomolecular Research and Utrecht Institute for Pharmaceutical Sciences, Utrecht University, Padualaan 8, 3584 CH Utrecht, The Netherlands.*
[2]*Netherlands Proteomics Centre, Padualaan 8, 3584 CH Utrecht, The Netherlands.*
[3]*Present address: Proteomics Unit, Spanish National Cancer Research Centre (CNIO), 28029 Madrid, Spain.*
**These authors contributed equally to this work.*
Correspondence to A.J.R.H.
e-mail: a.j.r.heck@uu.nl
doi:10.1038/nrg3356
Published online
4 December 2012

Cells are composed of a huge collection of molecules of distinct nature that act coordinately through highly intricate mechanisms, allowing processes such as self-replication and adaptation to external perturbations to occur. Understanding how cells function and communicate is one of the fundamental goals in molecular biology. The discovery that DNA contains all of the genetic instructions that are necessary to create an organism led to the principal dogma of molecular biology, which described a unidirectional flow of information from DNA to RNA to proteins[1]. Through recent findings, this dogma has been challenged. Epigenetic marks, alternative splicing, non-coding RNAs (including microRNAs (miRNAs)), protein–protein interaction (PPI) networks and post-translational modifications (PTMs) represent some key examples of how genotype and phenotype are not uniquely directed by information that is present on the genome[2–5] (FIG. 1). The global analysis of proteins, which are the key functional entities in the cell, arguably forms the principal level of information required to understand how cells function; such analysis is referred to as proteomics[6]. However, gathering information at the proteome level has turned out to be challenging, when comparing it to data collection at the genome and transcriptome levels.

Global protein analysis poses a tough analytical challenge, in part owing to the highly diverse physicochemical properties of amino acids, which are the building blocks of proteins. Furthermore, compared to the genome, the proteome is complemented by alternative splicing and diverse protein modifications and degradation, and the complexity is further amplified by the interconnectivity of proteins into complexes and signalling networks that are highly divergent in time and space. In recent years, proteomics technologies — particularly mass spectrometry (MS)-based protein identification — have matured immensely through cumulative technological advances in instrumentation, sample preparation and computational analysis[7] (BOX 1; FIG. 2). Whereas the sequencing and identification of an individual protein was a major challenge a decade ago, the identification and quantification of nearly all expressed proteins is now achievable in a single experiment. Likewise, 10 years ago, barely a few hundred phosphosites could be identified by MS-based phosphoproteomics[8], whereas now more than 30,000 phosphosites can be quantitatively monitored[9,10]. Following these multiple advances, we refer to this current technology as 'next-generation proteomics' to reflect this ability to characterize almost complete proteomes[11–13]; this mirrors the comprehensive coverage of DNA and RNA species by next-generation nucleic-acid-sequencing methods. This maturation of MS-based proteomics is starting to deliver answers to important biological questions.

Here, we describe recent breakthroughs and strategies that have enabled next-generation proteomics to be

Liquid chromatography coupled to mass spectrometry (LC–MS). High-performance liquid chromatography is coupled to mass spectrometry to separate the peptide mixture in liquid phase on the basis of hydrophobic interactions with the C_{18} stationary phase of the chromatography column (C_{18} refers to the length of the alkyl chains that decorate the chromatographic beads).

developed; we focus on MS-based proteomics and highlight recent applications and the biological insights that they have permitted. We start by addressing progress in proteomics technologies, covering both advances in mass spectrometry as well as sample handling (for example, peptide separation and enrichment). Next, we review applications in quantitative expression proteomics and how such data complement information gathered at the gene and transcript levels. We then discuss the diversity of PTMs and then follow this with a look at PPI and signalling networks and how they dynamically respond to perturbations. The need for computational tools to organize and to mine all data will be addressed separately (BOX 2). We end by discussing the clinical applications of MS-based proteomics and envisaging future trends in the field.

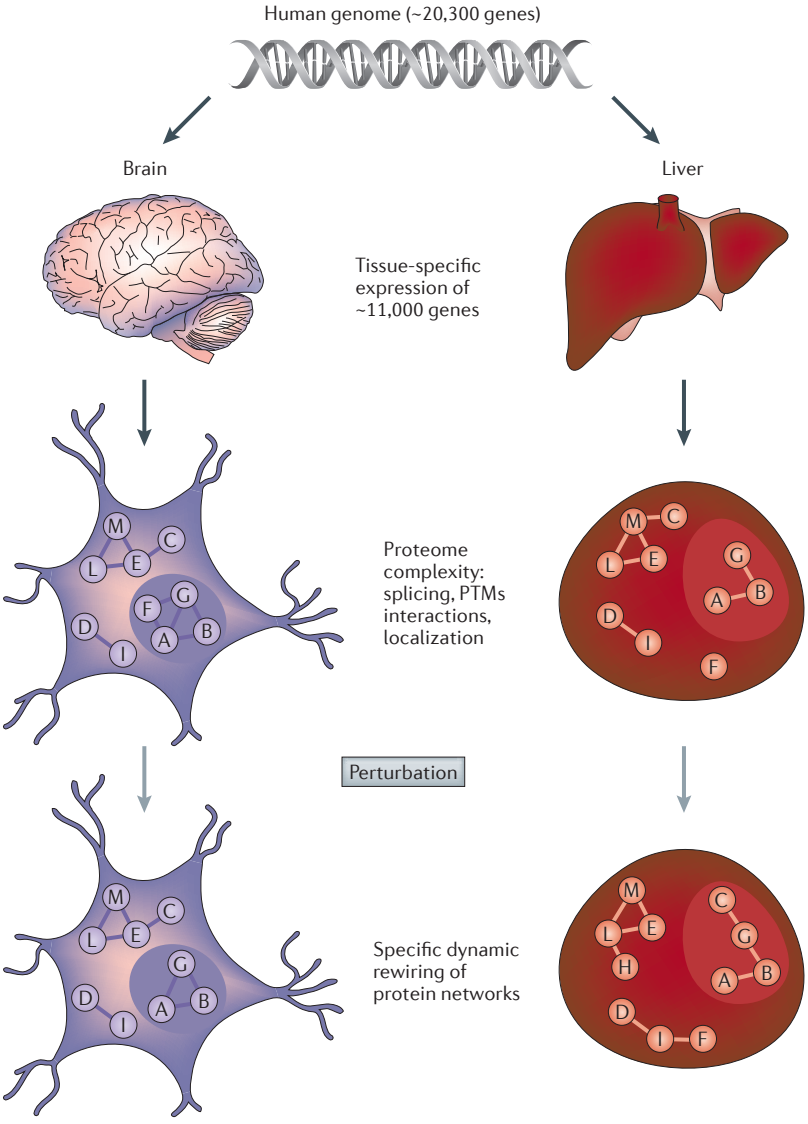

Figure 1 | **The diverse and dynamic mechanisms of proteome regulation provide a higher order of complexity to the human genome.** The human genome contains approximately 20,300 genes. The specific expression of a subset of the genome (~11,000 genes) determines the molecular backbone of the cellular phenotype (that is, the tissue cell types). A much higher order of complexity is achieved by the intricate mechanisms of protein regulation, including splicing variants, post-translational modifications (PTMs), protein–protein interactions (PPIs) and subcellular localization. This creates time-dependent tissue- and organelle-specific protein networks that respond differently to perturbations (for example, ageing or drug treatment).

Mass-spectrometry-based proteomics

Methodology. Typically, the first step in any proteomics experiment (FIG. 2; BOX 1) is sample lysis and protein extraction from the cells, tissues or bodily fluids followed by proteolysis into peptides. Depending on the aim of the study, the samples can be pre-fractionated (for example, using ion-exchange chromatography) or specific subpopulations of peptides (for example, PTMs) can be enriched by different means (for example, affinity resins and specific antibodies). These fractions are then analysed one-by-one by reversed-phase liquid chromatography coupled to mass spectrometry (LC–MS), in which selected peptides are fragmented by tandem mass spectrometry (MS/MS). MS and MS/MS spectra are then used as inputs for database search engines to identify the corresponding peptide sequences. Finally, the assigned peptide sequences are assembled into proteins, and the obtained data are statistically validated, often through decoy search strategies in which the MS/MS spectra are competitively matched against random databases to estimate the rate of false positive identifications[14–16]. Alternative approaches are being developed that alleviate the need for the availability of protein databases by inferring the peptide sequences directly from obtained MS/MS spectra[17,18].

Label-free protein quantification. Popular quantitative strategies use either label-free approaches or stable isotope labelling. Label-free quantification through spectral counting and/or signal intensity of the detected peptides seems to be the least cumbersome way to obtain quantitative information. It can provide robust and precise relative protein expression information when done using replicate measurements under stringent conditions with minimal analysis variation[19,20]. Care has to be taken with absolute copy number determinations of proteins as, for instance, the choice of the protease at the start of the experiment might greatly influence the outcome[21]. Selected reaction monitoring (SRM) is an emerging quantification strategy that targets specific proteins of interest and that is useful for validation of protein expression changes[22]. SRM assays reach a high level of specificity through the monitoring of the fairly unique combination of peptide mass-to-charge (m/z) ratio and multiple diagnostic peptide fragment ions. Carrying out SRM-type quantification is more elaborate than conventional label-free quantification because of the constraint that detailed prior information is required about the behaviour of several peptides from every protein of interest, throughout the entire LC–MS workflow (for example, LC retention or MS/MS fragmentation). When this criterion has been fulfilled and a successful assay has been constructed, accurate quantification can be achieved for the selected proteins over a high dynamic range and for a reasonable number of samples[23,24]. The precision of SRM is determined by the ability to reduce the variation

Box 1 | **Advancements in MS-based proteomics**

To achieve high proteome identification and quantification coverage, implementation of several steps in the workflow are required (FIG. 2), and recent technological and methodological advancements have greatly increased the depth of proteome coverage and sample throughput, as outlined below.

- A current trend in proteomics research is to target increasingly complex samples from ever-smaller sample amounts. For example, this can be done when looking at specific cell populations using fluorescence-activated cell sorting (FACS) or tissue micro-dissection, which requires increased sensitivity, resolution and speed in liquid chromatography coupled to mass spectrometry (LC–MS) analysis.
- Reduction of sample complexity is achieved by fractionation of the sample before LC–MS using orthogonal technologies. Popular chromatographic methods include ion-exchange chromatography and fractionation based on peptide isoelectric point. More recent developments include the use of stationary phases, such as hydrophilic interaction liquid chromatography (HILIC)[147], which can be efficiently miniaturized to obtain a high sensitivity[148].
- Further reduction of complexity is achieved by reversed-phase chromatography directly coupled to MS analysis. Chromatographic capacity and resolution have been increased through the use of longer analytical columns and smaller particle sizes (increasing surface area) in combination with longer gradients[40]. Current approaches typically use C_{18} reversed-phase liquid chromatography under nanoflow and ultra-high-pressure regimes.
- The resulting improved chromatography can then be exploited by advanced sensitive and fast MS platforms. Here, an increase in sensitivity is obtained through improved ion inlets and transfer optics and through accelerated sequencing speeds, up to 20 Hz, which are aimed at targeting all eluting peptides for tandem MS (MS/MS)[149,150].
- Finally, the amount of information retrieved can be further increased by improving identification efficiency[151]; this can be accomplished by using high mass resolution and mass accuracy[7] and using fragmentation technologies that are complementary to ion-trap collision-induced dissociation (CID), such as electron transfer dissociation (ETD)[152] and higher-energy collisional activation (HCD)[153].
- An ultimate goal in quantitative proteomics is the parallel detection of many different samples, such as multiple time points after perturbation, including both technical and biological replicates that provide statistical power with minimum overall experiment time. A step in this direction is termed 'hyperplexing', in which SILAC (triplex) and isobaric labelling (hexaplex) strategies were combined to follow protein expression changes in 18 samples in a single analysis[154].

Tandem mass spectrometry (MS/MS). A type of mass spectrometry in which ions are selectively isolated and then fragmented. The mass-to-charge ratio of each molecular fragment is measured and used for structural characterization.

Selected reaction monitoring (SRM). Protein quantification obtained by monitoring the specific combination of precursor and fragment ion mass-to-charge ratios of several selected peptides per protein.

Collision-induced dissociation (CID). Fragmentation of molecular ions in the collision cell through increasing the molecules, kinetic energy followed by collisions with neutral molecules (often helium, nitrogen or argon).

introduced before MS analysis. Therefore, SRM experiments are typically constrained to single LC–MS runs without pre-fractionation or enrichment, and this limitation currently hampers the quantification of low-abundance proteins or PTMs, especially in highly complex backgrounds. Alternative approaches have recently been reported that use a pseudo-SRM methodology in which either all ions are fragmented (SWATH)[25] or on-the-fly prediction of peptide elution is used to target a preselected array of proteins[26], thus offering new means to tackle the complexity problem.

Label-based protein quantification. Errors in quantification that occur through variation in sample handling can be minimized when differential stable isotopes are introduced in the samples to be compared in order to create peptide isotopomers that can be distinguished in the output spectra owing to their distinctive mass. In recent years, several strategies have been developed that use the incorporation of stable isotopes[27], for which minimal variability is achieved using metabolic labelling in cell culture[28] or even in whole organisms[29,30]. A drawback of metabolic labelling is that primary cells or human tissue samples are difficult to culture and to label, although this may be circumvented by constructing an isotope-labelled reference sample from cultured cells as an internal standard[31]. A more versatile alternative is the subsequent use of chemical stable isotope labelling at the protein or peptide level, which will be applicable to almost any sample and may approach similar quantitative depth and precision to cellular metabolic labelling. Quantification is usually achieved at the MS level, except for chemical labelling using isobaric chemical labels, in which quantification is based on MS/MS reporter ions. A cost-effective example of MS-level quantification is dimethyl labelling[32], which allows highly efficient and precise quantification of all sample types, from primary cells to tissues[33,34]. Quantification at the MS/MS level can be multiplexed (up to octoplex), thus permitting the analysis of multiple perturbations in parallel[35]. However, a current unresolved issue with isobaric labels is the potential co-isolation of more than one peptide species, which may lead to a distortion of the quantification results[36]. This phenomenon can be partly overcome by a double-isolation event[37,38], although it comes at a cost of efficiency.

Overall, it is clear that proteomics technologies are continuously evolving. The tremendous increase in proteome coverage obtained over the past decade cannot be attributed to a single breakthrough, and there is ample space for further developments. In BOX 1 and FIG. 2, we summarize some of the latest developments in comprehensive shotgun proteomics. There is certainly not a single ideal workflow, and the experiment undertaken and the questions being addressed will largely define the most suitable workflow.

Proteome expression profiling

The large depth of information that is now provided by fairly routine use of MS-based methodologies has positioned proteomics as one of the most powerful tools in biology. Mapping the entire set of proteins in a biological system and understanding how they readjust in space and time on perturbation is one of the primary goals in proteomics (FIG. 3).

Characterizing proteomic repertoires. Only a few years ago, in 2008, extensive sample pre-fractionation followed by several days of MS acquisition time allowed the complete proteome of yeast to be identified[39]. Four years later, comparable results were obtained in a single MS run (~240 minutes) without the need for pre-fractionation[40]. This 100-fold increase in throughput demonstrates the quantitative leap that modern MS has achieved.

Despite great advances in high-throughput RNA sequencing (RNA-seq), the question of how many genes encode protein products in human cells remains unanswered[41]. This uncertainty is partially due to the fact the genome is extensively transcribed into non-coding RNAs[42], of which the biological importance remains to be explored[43]. Recently, several studies showed that current MS-based technologies are able to identify ~11,000 expressed protein-coding genes in cultured human cells

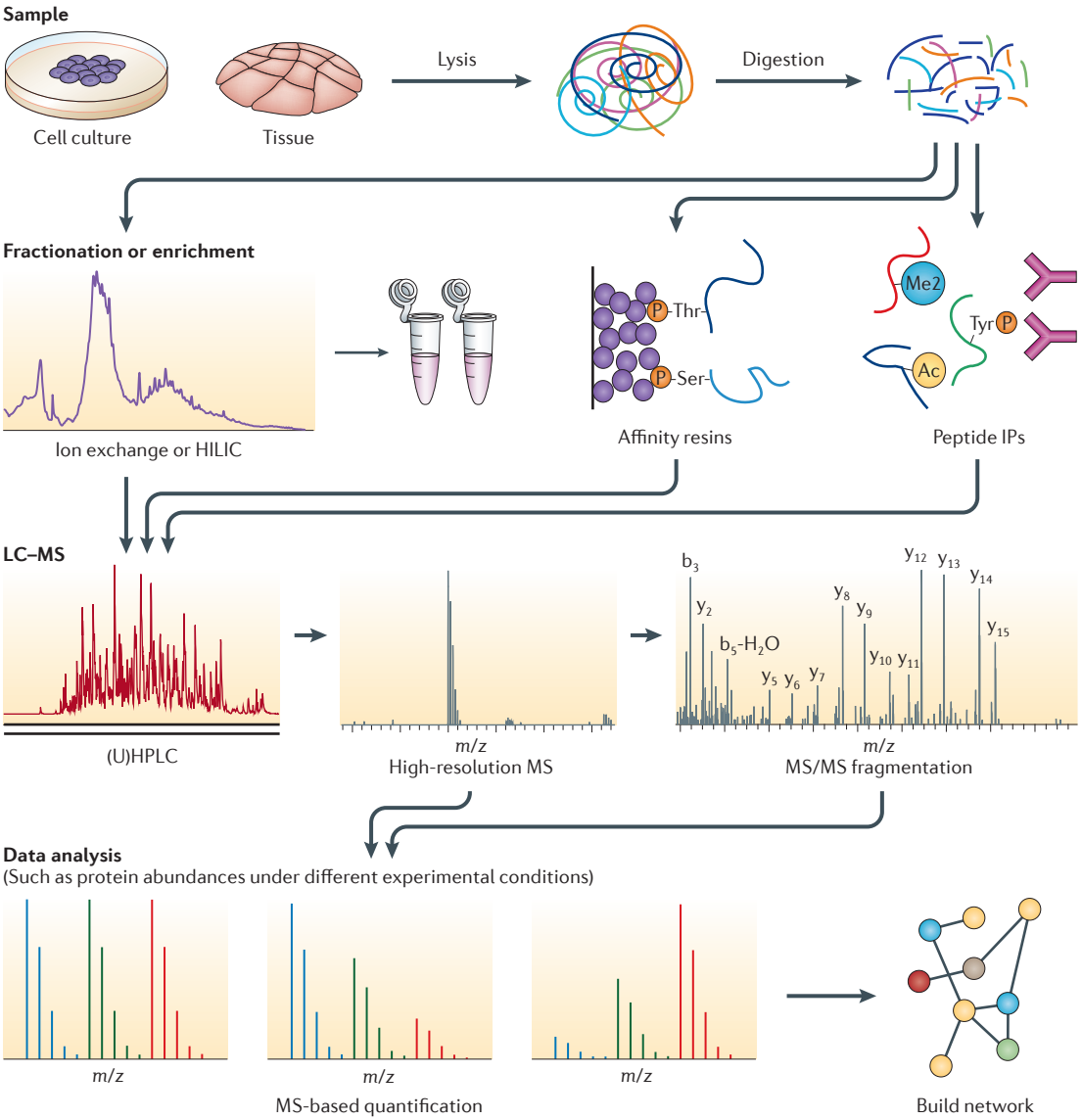

Figure 2 | **Generalized mass-spectrometry-based proteomics workflow.** In the vast majority of proteomics experiments, the extracted proteins are first digested into peptides, usually by trypsin, to create molecular species that are much more manageable in both sample handling and mass spectrometry (MS) analysis. Enzymatic digestion of a full proteome generates hundreds of thousands of peptides: this sample complexity is not directly compatible with the MS analysis. Therefore, the first step in the proteomics workflow is most often directed at reducing the sample complexity either by sample prefractionation or enrichment. In sample prefractionation, the peptide population is fractionated according to its physicochemical properties, such as charge, isoelectric point, hydrophobicity or combinations of these. It is essential here that the chosen fractionation technique is orthogonal to the liquid chromatography (LC) separation just before MS analysis. Alternatively, specific subsets of the sample can be targeted through enrichment of peptides containing modifications (for example, phosphorylation (P), dimethylation (Me2) or acetylation (Ac)) using affinity-based resins or antibody-based immunoprecipitation (IP). These prefractionated or enriched samples are then introduced to the LC system for an additional separation step to reduce complexity further; this can be achieved by using, for example, ultra-high-performance liquid chromatography (UHPLC). Ideally, all peptides eluted from the LC are queried by the mass spectrometer, but in everyday practice, many peptides elute simultaneously and compete for efficient ionization: highly abundant species can suppress the ionization of co-eluting less abundant species, thus preventing their MS analysis. After ionization, peptide precursor ions are introduced into the mass spectrometer, which records their mass-to-charge (m/z) ratio with high accuracy. For identification, single precursors are selected (on the basis of observed intensity) and subjected to a tandem MS (MS/MS) event — most commonly, collision-induced dissociation (CID) — to generate characteristic fragment ions for the selected precursor. Peaks labelled 'b' are amino-terminal fragment ions, and peaks labelled 'y' are carboxy-terminal fragment ions. The combination of precursor m/z and its fragment ions is then matched to known peptide sequences from large protein databases using search algorithms such as Mascot or SEQUEST. Finally, data are quantified (either relatively or absolutely); for example the three data analysis panels in the figure might represent three experimental conditions in which the different proteins (represented by different coloured lines) may or may not change in abundance. These protein abundances are then interpreted and visualized in the context of the biological system under study. HILIC, hydrophilic interaction liquid chromatography.

SWATH
Increments of 25 Da are isolated across the mass range of interest, and all ions within the mass window are simultaneously fragmented. The resulting fragments are analysed at a high resolution and afterwards are matched to a peptide fragment library.

Isobaric chemical labels
Chemical labels used in mass spectrometry that have identical molecular mass but that fragment during tandem mass spectrometry into reporter ions of different masses.

Ultra-high-performance liquid chromatography
(UHPLC). High-performance liquid chromatography carried out under extended pressure regimes (typically up to a 1,000 bar), allowing the use of smaller stationary phase particle sizes, increasing interaction volumes and thus separation power.

Box 2 | **Assessing and disseminating proteomics data**

Mass spectrometry (MS)-based proteomics is producing a vast amount of data regarding the expression, post-translational modifications (PTMs) and interactions of thousands of proteins. This information has to be delivered to the scientific community in an appropriate and curated form that is easily retrievable and interpretable. The public availability of proteomics data will also safeguard quality standards in the field. A first level to distribute proteomics data involves the long-term storage of the unprocessed raw data. This is important, given that improved computational methods will allow data reanalysis in the future. Tranche is one of the few public repositories that can currently handle this type of data, and it is based on an encrypted peer-to-peer system that stores data across multiple servers worldwide. However, raw data are closed formats (that is, they are proprietary of MS manufacturers), which hampers their dissemination. Therefore, efforts are being made towards the standardization of formats that keep all of the relevant information. The PRIDE database at the European Bioinformatics Institute exemplifies this determination, as it allows not only the storage of standard MS data formats (that is, XML) but also the associated peptide and protein identifications. Furthermore, the inclusion of additional information (for example, species, fragmentation techniques and proteases) allows the global meta-analysis of proteomic data sets. The GPM, PeptideAtlas and NCBIpeptidome are similar examples in which the ProteomeXchange aims now to provide a single point of submission to all of these initiatives. The next level is represented by dedicated databases that include specific information on the PTMs of proteins. PhosphoSitePlus, Phospho.ELM and Phosida are some popular resources that contain several thousands of phosphorylation, ubiquitylation, acetylation and *N*-glycosylation sites. The availability of all of this site-specific information gives rise to more specific databases, such as NetPhorest or PhosphoMotif Finder, that catalogue kinase linear sequence motifs. Likewise, databases such as String, Intact and BioGRID are fed with the thousands of protein–protein interactions reported in current affinity purification followed by MS (AP–MS) studies, and this allows reconstruction of global protein networks with a high level of detail. The last level of proteomic data annotation is supported by databases such as UniProt, which accommodate information from all of these different inputs (the Human Protein Reference Database and the recent neXtprot are specifically dedicated to the human proteome). Although a risk of annotating false positives exists, the experimental evidence by multiple studies could be used as a criterion to accept only confident results.

before saturation of the methodology, indicating that full proteomes can now be sampled[11,12], allowing the exploration of tissue-specific, cell-specific and developmental-stage-specific protein expression. In addition, the RNA sequences derived from next-generation sequencing platforms might contain exciting discoveries, such as the presence of new splice variants and even new genes. However, great precautions must be taken when dealing with such a large amount of information, as it could be prone to false-positive results. The current high-throughput of MS technology can therefore be exploited to confirm some of these genomic findings, as experimental evidence at the protein level (for example, the identification of a peptide from a new gene) will provide unequivocal proof for such new gene forms. The use of genomic and proteomic information to re-annotate gene sequence databases is known as proteogenomics. A proof of concept of such applications was given by MS sequencing to confirm translation of pseudogenes, to identify new splice variants and even to identify new protein-coding genes[44]. Moreover, MS has also been used to assess the debated pervasiveness of RNA editing[45] by showing that at least some of the discrepancies between RNA and DNA sequences are expressed at the protein level[45], rather than solely being artefacts of nucleotide sequencing[46].

Proteogenomics
The use of proteomics data, which are often derived from mass spectrometry analysis, to improve gene annotations.

Induced pluripotent stem cells
(iPSCs). Somatic cells that have been reprogrammed to a pluripotent state, which is highly similar to that of embryonic stem cells, through ectopic expression of four transcription factors (namely, OCT4, SOX2, MYC and KLF4).

Embryonic stem cells
(ESCs). Pluripotent cells that can be derived from the inner cell mass of the blastocyst-stage embryo.

Comparative profiling. System-wide analyses of differential gene expression represent a powerful approach to identify and to understand the molecular causes of many biological and pathological processes. Like transcriptomics, which allows comparison of mRNA levels between samples, implementation of quantification strategies in the proteomic workflow allows relative protein quantification with high accuracy[19,47].

The breakthrough discovery that somatic cells can be reprogrammed to an embryo-like state termed induced pluripotent stem cells (iPSCs) has opened new avenues in the field of regenerative medicine[48]. However, before iPSCs are used in the clinic, molecular comparisons between iPSCs and their embryonic stem cell (ESC) counterparts are necessary, given that such differences may have an impact on their potential therapeutic use. Using transcriptomics, it was found that iPSCs possessed a unique gene expression profile that distinguished them from ESCs[49]. However, it has also been shown that these gene expression profiles could represent signatures of laboratory-specific *in vitro* culture conditions rather than a recurrent molecular signature across different iPSC lines[50,51]. Therefore, examination of the protein levels between these cell lines would provide an important molecular angle that sheds light on this debate, especially given the fact that levels of mRNAs and proteins often poorly correlate[19,47]. Recently, two groups carried out extensive proteomic comparisons of ESCs, iPSCs and the parental fibroblasts of the iPSCs[13,52]. Both studies confirmed the high similarity of iPSCs and ESCs at the proteome level and found only a few statistically significant differences in protein levels between them. Interestingly, the set of differentially expressed proteins found by each group showed a nonsignificant overlap[53], supporting the emerging notion that differences between iPSCs and ESCs might solely reflect experimental cell-culturing conditions.

Absolute quantification. Although relative quantification reveals changes in protein levels between two or more states, it is dimensionless and is normally expressed in the form of ratios. Recent advances in MS and bioinformatics now allow the estimation of the absolute amount of proteins: that is, the copy number of proteins per cell. Using pulse labelling with nucleosides or amino acids, an elegant study was conducted to quantify turnover and expression levels of mRNA and proteins[19]. The result that mRNA levels explained only ~40% of the variability in protein levels led the authors to create a model for gene expression control that showed that protein abundance is predominantly controlled at the level of translation. These findings offer new insights into the mechanisms of gene expression control and the relationships between transcription and translation. The use of internal peptide or protein standards in combination with SRM provides another means to assess absolute copy numbers[24,54], although such methods are still prone to inaccuracy and error, such as inconsistent proteolysis of the samples and protease bias[21].

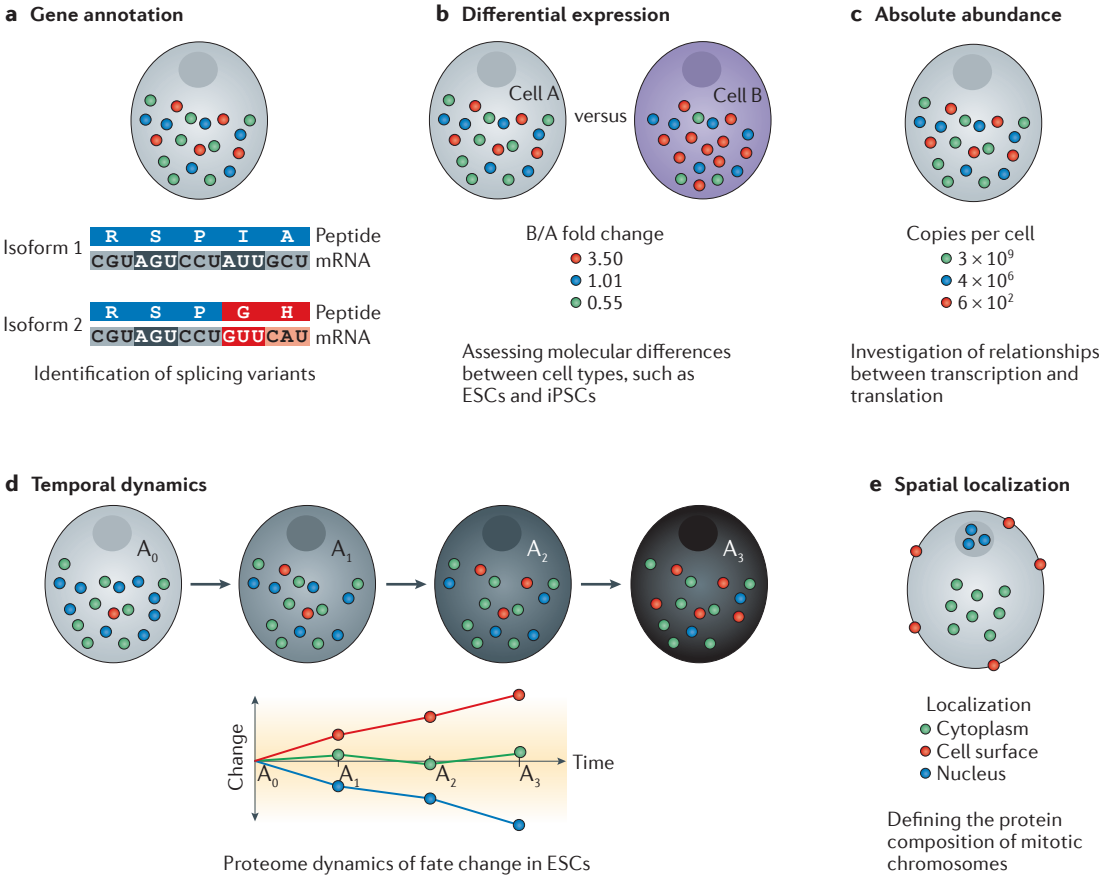

Figure 3 | **Scenarios for proteome expression profiling.** The analysis of the protein content of a sample can be achieved through various proteomics technologies. This allows obtaining biological information from several angles (as illustrative examples, the studies discussed in the text are shown for each case). **a** | The combination of high-throughput DNA or RNA sequencing (DNA-seq or RNA-seq) data with massively parallel peptide sequencing data can reveal, for instance, alternative splicing events and can be used to annotate or reannotate genomes. **b** | Proteomes can be quantitatively compared between two or more samples, enabling differential protein expression analyses to be carried out. **c** | When done under controlled conditions, proteomics data can be used to estimate the absolute abundance of proteins (that is, the number of protein copies per cell). **d** | Advances in multiplexing allow monitoring proteome dynamics of complex biological processes in time. **e** | Fractionation of cellular organelles followed by mass spectrometry analysis represents a unique approach to describe localization of proteins. ESC, embryonic stem cell; iPSC, induced pluripotent stem cell.

Dynamic and spatial proteomic characterization. Although static snapshots of proteomes provide valuable information, on perturbation of the steady-state conditions, the proteome dynamically responds in space and time. For instance, a multi-disciplinary study was conducted in differentiating ESCs[47] and differentiation-associated changes were found in the levels of hundreds of proteins, thus potentially improving our knowledge of cell-fate decisions. Furthermore, spatial distribution of proteins can be obtained using methods that purify subcellular compartments. Whereas imaging technologies allow higher-resolution and single-cell analyses, MS can be used in a 'discovery mode' to identify new protein components of organelles: for example, novel centromere-associated proteins[55]. Ideally, integration of all these approaches would allow us to obtain a clearer picture of the 'living' proteome and how it responds to external cues[56].

Deciphering post-translational regulation

PTMs are key regulators of protein activity and involve the reversible covalent modification of proteins by small chemical groups, lipids or even small proteins. In addition, proteins can be cleaved by proteases, and the chemical nature of amino acids can be modified. Taking into account the number of expressed protein-coding human genes (~11,000), the array of PTMs available (more than 200), the number of potentially modified residues, the dynamic nature and the often low stoichiometry of these modifications, one realizes the magnitude of the analytical challenge to identify and to localize the sites of these modifications. Advancements in MS methodologies, as outlined in BOX 1, have also greatly improved the analysis of PTMs. In particular, electron transfer dissociation (ETD) leaves labile PTMs intact on the peptide backbone, assisting in site assignment. MS is currently the most powerful tool for studying PTMs in system-wide approaches.

Electron transfer dissociation
(ETD). Fragmentation of molecular cations by the transferal of an electron.

Phosphoproteomics. Probably owing to its ubiquitous role in almost any biological process, phosphorylation is still the best characterized PTM. Phosphopeptide enrichment strategies, such as immobilized metal ion affinity chromatography (IMAC)[57,58], titanium dioxide (TiO_2) chromatography[59] or phosphotyrosine immunoprecipitation[60], have unlocked the analysis of phosphorylation as shown by a diverse range of studies in different types of tissues[61], disease states[62,63] and cell lines[52,64]. Most conventional enrichment strategies have a bias towards acidic peptide sequences. However, methods to counteract the under-representation of basic phosphopeptides include Ti^{4+}-IMAC-based phosphopeptide enrichment[65] or the use of antibodies selectively to immunoprecipitate peptides containing the target sequences of particular basophilic kinases of interest[66].

In 2003, MS-based studies identified ~300 phosphosites, and this was considered to be breakthrough progress at the time; now, MS technology can identify thousands of phosphosites with high precision[61]. This growing qualitative information is being annotated in databases[67], and predictions suggest that more than half of the proteome might be regulated by phosphorylation at an astonishing ~500,000 sites[9]. An obvious question arises concerning the functional relevance of all these sites. Kinase promiscuity could be seen as a natural mechanism to tune protein function. In the absence of a functional disadvantage, a phosphosite might persist as there will be no evolutionary pressure to mutate it to an amino acid that cannot be phosphorylated[67]. Determination of the stoichiometry of the modification might address this issue. Studies in yeast have revealed that fully phosphorylated sites (indicative of constitutive kinase activity or absent phosphatase activity)[68] are more evolutionarily conserved[69], implying functional relevance. Additionally, future analyses of phosphorylation turnover may provide a basis to prioritize phosphosites for functional characterization, as residues with long half-lives may indicate housekeeping functions, whereas sites with short half-lives may pinpoint the action of highly regulatory kinase–phosphatase pairs. Despite all of this progress, system-wide analyses are still severely hampered by the lack of appropriate analysis software, which is currently gene-centric and does not accommodate the dynamic analysis of multiple PTMs on a single protein.

To understand signalling pathways, quantitative phosphoproteomics is used to monitor the transient nature of phosphorylation events globally[64,70] and in specific organelles[71]. However, these analyses require careful interpretation. Changes in the abundance of a phosphopeptide could be caused by variations in the total protein abundance. Therefore, parallel quantification of proteins is required[72], especially in later time windows (for example, >12 hours following a perturbation), in which gene expression changes might distort phosphorylation ratios.

Understanding how kinases (and phosphatases) regulate all of these sites is also important. Phosphopeptide sequences occurring *in vivo* that are identified in large-scale phosphoproteomics studies[61] can be bioinformatically analysed to find over-represented amino acids that flank the phosphorylation site, providing clues on linear sequence motifs and, therefore, kinase specificity[73]. Complementary large-scale *in vitro* kinase assays are also used to refine these consensus motifs further[74]. In addition, kinase inhibitors have been used to discover new substrates for several kinases[63,75]. Finally, the generation of phospho-specific antibodies of candidate sites allows localization of a phosphorylation reaction with subcellular resolution; for example, this can reveal asymmetric phosphorylation events (those that do not occur uniformly throughout the cell) as a consequence of kinase compartmentalization[30].

Other PTMs. As we start to probe the large implications of phosphorylation, other PTMs that until recently remained elusive have become assessable by MS-based proteomics, revealing some new concepts in biology. Peptide-centric immunoprecipitation experiments are becoming successful for lysine acetylation[76] and ubiquitylation[77]. Interestingly, although lysine acetylation of histones is widely appreciated to be a gene-regulatory mechanism, it was found that acetylation targets thousands of non-histone proteins, implying a regulatory role beyond chromatin status regulation[76].

Alternatively, for ubiquitylation and sumoylation, a strategy can be used in which an affinity tag is fused to ubiquitin[78] or to SUMO conjugates[79]. This construct is transfected into cells of interest, so that endogenous protein targets become affinity-tagged during ubiquitylation or sumoylation and are then isolated by co-immunoprecipitation. Glycosylation is another widespread PTM, although analysis is challenging because of its labile character and range of glycostructures. There are two types of glycosylation: *N*-linked (that is, attached to asparagine) and *O*-linked (that is, attached to serine or threonine). *N*-linked glycosylation has primarily been studied so far: this can be enriched at the peptide and protein level by using either lectin-affinity-based columns or hydrophilic interaction liquid chromatography (HILIC)[80]. For example, a method was developed that used an elegant combination of lectin affinity chromatography with filter-aided sample preparation (FASP) to improve analysis of *N*-linked glycosylations on membrane proteins[81]. Using such methods, thousands of *N*-glycosylation[81] and ubiquitylation[77,82] sites are being identified, and their functionality is being better understood.

The coexistence of different modifications in the same protein may indicate the existence of PTM crosstalk, of which currently only a few functional relations have been reported. For example, acetylation of two lysine residues, which are mapped by MS, in the human gene expression regulator SMAD7 (also known as MADH7) prevents ubiquitylation of these residues and thereby blocks degradation of SMAD7 (REF. 83). As another example, the pluripotency-associated transcription factor SOX2 is phosphorylated on three serine residues; this induces its sumoylation[64] and thus impairs its DNA-binding properties[84]. Such crosstalk results in a sophisticated communication between PTMs in which sequential, mutually exclusive or

Immobilized metal ion affinity chromatography (IMAC). A metal ion with an affinity for analytes to be enriched (often phosphopeptides) is fixed to an insoluble matrix and serves as the adsorption centre, allowing complexation.

Basophilic kinases Kinases that have a preference for basic amino acids in the sequence motifs of their substrates.

Lectin A carbohydrate-binding protein that is involved in various biological recognition phenomena.

Hydrophilic interaction liquid chromatography (HILIC). Normal-phase chromatography with water-miscible mobile phases to separate hydrophilic compounds, such as proteins and peptides. Typically, the order of elution is the opposite of that obtained with reversed-phase chromatography.

Filter-aided sample preparation (FASP). Generation of tryptic peptides from crude lysates for liquid chromatography coupled to mass spectrometry (LC–MS) analysis within a filtration device, allowing analysis of detergent lysed cells and tissues.

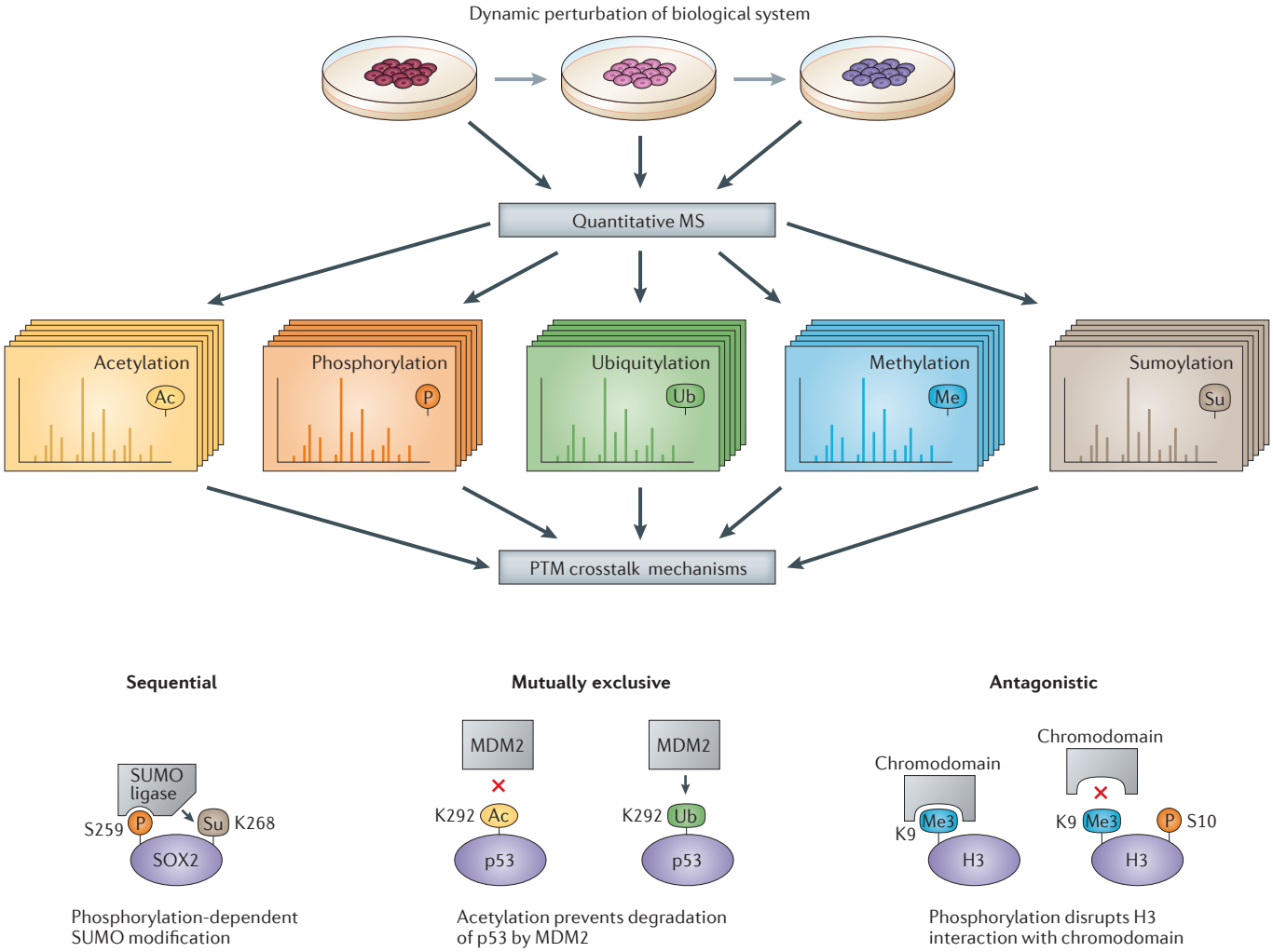

Figure 4 | **Large-scale mass spectrometry data sets of post-translational modifications (PTMs) allow analysis of PTM crosstalk.** Mass spectrometry (MS)-based proteomics represents a unique tool to identify and quantitatively to monitor global and site-specific changes of post-translational modifications (for example, acetylation, phosphorylation and ubiquitylation). Often, many of these regulatory events coexist in the same proteins. Integration of these analyses will serve as a valuable resource to reveal mechanisms of post-translational modification (PTM) crosstalk (for example, sequential, exclusive and antagonistic). Examples depicted are sequential: the phosphorylation of the transcription factor SOX2 in three serine residues induces its sumoylation[64], which ultimately impairs its DNA-binding properties[84]; mutually exclusive: acetylation of p53 in a lysine residue prevents the ubiquitylation of the same site by MDM2 and thereby proteasomal degradation of p53 (REF. 155); and antagonistic: trimethylation of histone H3 is required for binding of the heterochromatin protein 1 (HP1) chromodomain, and this interaction is antagonized by an adjacent phosphorylation site catalysed by aurora kinase B (AURKB)[156].

antagonistic mechanisms seem to regulate protein functions[85] (FIG. 4). In the light of these observations, several PTMs are now simultaneously profiled, including *O*-linked *N*-acetyl glucosamine (GlcNAc) and phosphorylation[86], as well as acetylation and phosphorylation[87,88]. The high degree of PTM complexity that is being revealed by the exponential growth in PTM data sets reveals a much higher order of complexity than was previously anticipated as most of the proteins are modified by different PTMs; this complexity forces us to change our classical view of signalling networks from a linear information flow into a highly intricate and multidirectional regulatory network.

Protein–protein interactions and network biology

Understanding the diverse and dynamic proteome requires the construction of charts of physical interactions. Proteins often interact with each other in stable or transient multi-protein complexes of distinct composition, with an estimated 130,000 binary interactions in the human interactome, most of which remain to be mapped[89]. Moreover, proteins can interact with other molecules, such as RNA[20] or metabolites[90]. These complexes have essential roles in regulatory processes, signalling cascades and cellular functions, and loss of the ability to interact can cause loss of function[91,92].

Characterization of protein–protein interactions by affinity purification–mass spectrometry. The analysis of PPIs was initially driven by yeast two-hybrid (Y2H) studies, but it has more recently been complemented by using affinity purification (AP) of a protein of interest ('bait') followed by MS to identify its interaction partners ('prey'; a method termed AP–MS)[93,94]. High-throughput Y2H assays can generate broad maps of binary PPIs, irrespective of protein abundances, including those that connect different complexes and those that are of a highly transient nature, which are difficult to target using alternative approaches. However, the quality of Y2H data sets has been controversial, and different Y2H systems have been shown to detect markedly different interactions in the same interactome, requiring tools to determine the confidence of the interactions[89,94–96]. Confident interactions obtained by Y2H reflect a different part of the interactome than do interactions obtained by AP–MS, providing orthogonal information on cellular PPI networks[95]. AP–MS can delineate the dynamics of interactions (using quantitative MS workflows) at almost physiological conditions, thus explaining its success for the determination of protein complex compositions. Recent AP–MS reports have described the global landscape of PPIs in various species, such as yeast[92,97] and fruitflies[98], and of proteins from various biological processes, such as deubiquitylating enzymes[99], kinases and phosphatases[100], and the autophagy system[101]. In a recent study, AP–MS was used to draw the interaction landscape of all 18 HIV-1 proteins with their human host proteins and protein complexes[102]. In an accompanying study, the authors showed that one of the newly discovered HIV–host PPIs is required for preserving HIV-1 infectivity[103].

It is essential in AP–MS experiments that the 'bait' protein and associated 'prey' proteins can be efficiently purified under physiological conditions. When antibodies are available, endogenous protein immunoprecipitation can be accomplished, as shown in a large study in which 3,290 immunoprecipitation experiments were carried out to chart the human co-regulator complex network[104]. As antibodies can be cumbersome to use and are not available for all bait proteins of interest, affinity tagging of a bait protein is a common alternative. Control of expression levels of the tagged bait protein is important, as overexpression can lead to altered localization and can cause the formation of non-physiological protein associations. Tandem affinity purification (TAP)[105] has been successful in mapping global PPIs in yeast[92,97]. Here, the intrinsic homologous recombination capability of yeast directly introduces the TAP tag at a chosen endogenous locus, so that the natural regulatory mechanisms control the expression of the TAP fusion protein. In many other organisms, recombination is less efficient, and alternative approaches are required. An interesting alternative is to use transgenes that are cloned using bacterial artificial chromosomes (BACs)[106] and that contain all of their endogenous regulatory sequences. PPIs were studied during human cell division by combining BACs with a modified version of TAP[107], in which one of the tags was replaced by GFP, thereby allowing both localization and affinity purification (LAP)[108]. This enabled the authors to observe proteins associating with cellular components, such as centrosomes and spindles, ultimately leading to the discovery of several new subunits in complexes that are essential for cell division[107]. In contrast to the large range of soluble protein complexes targeted with AP–MS, complexes containing membrane proteins have been less well studied owing to compatibility issues between the used AP–MS protocols and the hydrophobicity of membrane proteins. With the inclusion of non-denaturing detergents in TAP extraction and purification procedures, these restrictions have been overcome, allowing a study to be conducted of 1,726 membrane PPIs in yeast[109].

For a more global analysis, affinity enrichment of single bait proteins can be circumvented by combining chromatographic separation of intact protein complexes with high-resolution quantitative MS[110,111]. With this combination, 291 protein complexes were analysed in the human interactome, and SILAC quantification was used to follow global changes occurring on stimulation with epidermal growth factor (EGF)[111]. In another study, a combination of several biochemical fractionation technologies was used to identify 13,993 endogenous human PPIs, containing 3,006 proteins in 622 putative complexes in cytoplasmic and nuclear extracts[110].

The absence or presence of PTMs greatly affects PPIs, as demonstrated by the epigenetic regulation of transcription through specific histone marks and their 'readers'. A highly efficient method for identifying these readers and their respective complexes was applied in a human cell line: modified histone peptides were used to pull down binding proteins, which were then assigned to protein complexes by AP–MS[28]. Furthermore, besides the role of acetylation in epigenetic regulation, the double acetylation of human structural maintenance of chromosomes protein 3 (SMC3) was shown to be crucial for recruiting the protein sororin to the cohesion complex[112] to ensure proper sister chromatid cohesion and chromosome segregation.

Dynamic and quantitative AP–MS. In addition to global analysis, quantitative AP–MS can reveal highly relevant information on the dynamics of PPIs. For instance, the abundance and dynamics of subunits of the human protein phosphatase 2A (PP2A) network was studied by introducing isotope-labelled reference peptides during sample preparation to determine the amount of bait protein present in the purification[113]. The authors were then able to determine the dynamic association of PP2A with protein phosphatase methylesterase 1 (PPME1) under different conditions. Likewise, quantitative MS was used to determine the architecture of the cullin–RING ubiquitin ligase network on deneddylation, after which AQUA peptides were used to determine the occupancy of individual subunits in the network[114]. A targeted approach called AP–SRM was developed and followed 90 growth factor receptor bound protein 2 (GRB2)-interacting proteins on cellular stimulation by EGF; it revealed both constitutive complexes and transient complexes that form only after growth-factor stimulation[115].

Tandem affinity purification
(TAP). A process in which a protein is carboxy-terminally tagged with a peptide containing a calmodulin-binding peptide, a TEV protease cleavage site and protein A. The protein is first purified using immunoglobulin-G-coated beads that bind protein A. The protein fusion is then cleaved from the gene of interest by the TEV protease and purified.

AQUA peptides
A precisely known amount of a synthetic tryptic peptide, which corresponds to a peptide of interest in the sample, with one stable-isotope-labelled amino acid that is used to determine absolute protein amounts.

Quantitative AP–MS analysis of the adenomatous polyposis coli (APC)–AXIN1 destruction complex, which degrades cytosolic β-catenin, revealed a new view on its regulation by WNT signalling[116]. Unlike a previously proposed model, the complex does not dissociate on WNT activation but instead suppresses degradation of β-catenin, leading to accumulation of β-catenin within the complex and thus complex saturation. These examples highlight the strength of AP–MS in characterizing the dynamics of PPIs and thus in providing new insights into essential biological processes.

Technical challenges. Certain improvements can still be made to AP–MS studies. For example, it remains difficult to distinguish true interactions from co-purifications of background proteins. The most common solution is to carry out bait-free control purifications (such as using an empty vector or RNA inteference of the bait protein) in parallel with the bait purification. Measured interactors are considered to be nonspecific if they occur equally strongly in both bait and control conditions[117]. In addition, several computational methods have been developed to assist in the assignment of true and false interactions and have recently been reviewed[118]. More complicated are the remaining questions of how accurately to identify: highly transient interactions; proteins and protein binding sites that make direct contact within a complex; stoichiometry of the complex; and finally the directionality and function of the interactions. Such information is essential to aid in our understanding of how complex interaction networks operate. Potentially, the answer lies in the use of crosslinkers to capture the native state of interacting proteins under different conditions, pinpointing direct interaction partners and capturing transient interactions[119]. Identification of the crosslinked sites could potentially reveal binding sites and, in combination with MS analysis of intact complexes, additional structural information can be obtained[120,121], as was recently shown for PP2A complexes[122]. Finally, these advancements can be deployed to unravel the networks underlying disease, for which protein mutations or altered expression of proteins or their PTMs compromise interactions. Here, additional developments of computational tools are required that allow modelling of protein network behaviour under changing conditions, as inferred from quantitative AP–MS data, which is reviewed elsewhere[123,124].

Clinical applications

Integrative omics profiling. As outlined above, the immense complexity of molecular and cellular processes hampers the identification and interpretation of the molecular causes of disease. For example, the 3,000 human genes found to be mutated — a staggering number — in relation to known disorders[125] and the observed 1,000–10,000 somatic substitutions in the genomes of most adult cancers[126] rarely result either in an understanding of the molecular mechanisms that drive the disease or in the development of therapeutics. To realize such goals, we have to gather integrative views of molecular processes. Several examples are emerging, showing the strength of combining technologies to unravel these entwined processes. For instance, the physiological state of a single person was monitored for 14 months, combining genomic, transcriptomic, proteomic, metabolomic and autoantibody profiles[127]. This study is a proof-of-principle of integrative personal medicine termed an 'integrative personal omics profile' and revealed various medical risks, including onset of type 2 diabetes.

Another example of the successful combination of multiple technologies is the search for remedies against drug resistance in cancer therapy. Cancers are highly complex and are often driven by gene mutations that result in constitutively active signalling pathways controlling proliferation and growth. Inhibiting single nodes in such pathways can lead to rewiring to alternative pathways, thereby inducing drug resistance. Such dynamic rewiring was uncovered by generating a system-wide view of signalling networks, gene expression profiles and cellular phenotypes in combination with mathematical models to find relationships between the data[128]. From the integrative data, the authors delineated that a sequential application of two types of anticancer drugs had the highest efficacy.

Clinical biomarkers. One of the most challenging applications of proteomics is the identification of protein biomarkers with prognostic or diagnostic value (FIG. 5). Recent technological advances, as reviewed above, have materialized in the design of comprehensive pipelines that integrate discovery and validation phases, enabling plasma biomarkers to be identified for different types of disease[129,130]. Although several successful biomarkers have been introduced for clinical use, many (if not most) claimed biomarkers have a limited reliability or remain without proper validation[131], leading to scepticism among clinicians. The primary shortcomings of many biomarker studies are a lack of proper controls in the discovery phase, the use of appropriate statistical tools for biomarker definition and the need for independent validation steps in large patient cohorts to certify the legitimacy of the biomarker unambiguously [132,133]; such weaknesses lead to claimed biomarkers that are seldom directly related to the disease biology. An example of a successful biomarker that links to the disease biology and that has been validated in a large patient cohort is the protease inhibitor elafin. This protein was shown to have both prognostic and diagnostic value in acute graft-versus-host disease after allogeneic bone marrow transplant[134]. Another demonstrative example is the report of protein signatures associated with human lung cancer[135]. The success of this study was achieved through a stringent selection of markers that underpin tumour biology and through circumventing one common pitfall in the search for biomarkers: the use of healthy controls. Comparing healthy control samples with disease samples can result in processes accompanying a disease — for example, inflammation — to be the main source of protein differences between the samples[132]. Bearing this in mind, the authors compared four lung cancer mouse

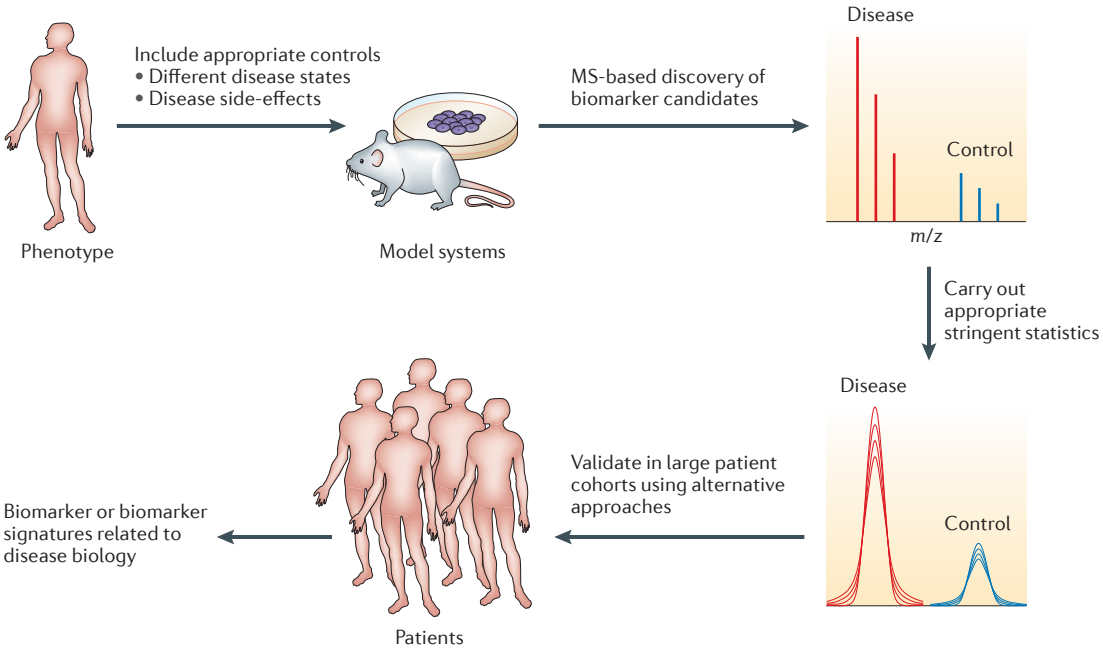

Figure 5 | **Generalized workflow for the identification, validation and stratification of protein-based biomarker signatures.** Mass spectrometry (MS)-based proteomics is used for an in-depth quantitative profiling of the proteome of a disease model and its appropriate control systems. After applying stringent statistics, a set of putative proteins is defined that may be used as a phenotype signature. Using more targeted approaches, either MS-based (for example, selected reaction monitoring (SRM)) or antibody-based, these markers are validated in large patient cohorts. Ideally, the biological associations between the signature proteins and the disease phenotype are further biochemically corroborated to confirm that the biomarker has a direct mechanistic role in the disease.

models with four alternative cancer models (namely, pancreatic, ovarian, prostate and breast cancers) and also included two models of inflammation. The specific protein signatures underlying lung cancer tumour biology could subsequently be validated in two independent human patient cohorts.

Heterogeneity, driver events and personalized cancer therapies. Expression profiles of biopsy samples can reflect the molecular signature of a disease and thus can be used to design personalized treatments. However, tissues are formed by different cell types, and this can complicate the analysis of the cell population of interest. As a solution, ultra-sensitive proteomics[136] can be coupled to laser-capture micro-dissection to isolate a particular cell population. MS methodologies have sufficiently advanced to cope with these small sample sizes, such that reasonable proteome coverage can be obtained from a few hundred cells[137] (BOX 1). This opens new avenues to study phenomena such as intra-tumour heterogeneity. Often, *in vitro* cell lines poorly recapitulate the conditions existing *in vivo*. This issue can now be partially solved by fluorescence-activated cell sorting (FACS)-based isolation of the cell type of interest from primary tissue samples. This allows the study of biological processes *ex vivo*; for example, this method was used to define a core of stem-cell-specific genes in the small intestine[34].

The deregulation of mechanisms controlling PTMs can have severe consequences. For instance, mutations in receptor tyrosine kinases (RTKs) can cause the cellular processes that they regulate to be constitutively active. In fact, many tyrosine kinases are oncogenes, and malfunctioning survival or differentiation mechanisms will ultimately lead to malignant transformation. Therefore, identification of the kinases that are active in certain tumours is indispensable, as this activation is not always directly caused by genetic mutation. Immunoprecipitation followed by MS was used to identify phosphotyrosine peptides in 41 lung cancer cell lines and 150 tumours[138]. The resulting phosphotyrosine signatures were used to classify the samples and allowed the authors to identify several novel kinases activated in lung cancer, including novel ALK and ROS fusion proteins. In an alternative approach, selective peptide immunoprecipitation experiments were used against phosphorylated substrates of the phosphoinositide 3-kinase (PI3K) and mitogen-activated protein kinase (MAPK) pathways[139]. They found that activation of the PI3K–AKT pathway — either through loss of PTEN function or through an activating mutation in the catalytic subunit of PI3K — could be detected by monitoring a phosphorylation site downstream in the pathway. Phosphoprotein biomarkers uncovered by this novel approach could then be used to identify PI3K–AKT pathway activation without sequencing multiple genes in the pathway and also to predict the efficacy of pharmacological inhibitors of AKT, 3-phosphoinositide-dependent protein kinase 1 (PDK1), PI3K or mammalian target or rapamycin (mTOR), demonstrating the potential for personalized therapy.

Fluorescence-activated cell sorting
(FACS). A method in which dissociated and individual living cells are sorted in a liquid stream according to the intensity of fluorescence that they emit as they pass through a laser beam. Sources of fluorescence include labelled antibodies that allow cell sorting on the basis of the expression of cell-surface molecules.

Organoids
Multicellular structures that resemble organs in architecture and function.

Mass cytometry
Cells bound to antibody–isotope conjugates are sprayed as single-cell droplets into an inductively coupled plasma mass spectrometer, creating a quantifiable response profile.

Ribosome profiling
Qualitative and quantitative sequencing of the RNA attached to ribosomes as a signature of genes that are expressed.

Conclusions and future perspectives

Next-generation proteomics allows a much more in-depth view of the proteome in all its facets. As the core technology, MS will remain a main player in this arena. Future advances in MS-based proteomics technologies will focus on: getting the relevant proteomics data in less analysis time[40]; reducing the quantity of material required; and allowing in-depth analysis of homogenous cell populations (for example, FACS-sorted cells)[34] or micro-dissected tissue, with the ultimate aim being single-cell proteome analysis[136]. Other challenges that remain are the transferal of existing technologies from application on laboratory cell cultures to *in vivo* tissue and primary cells, in which technical and biological variability are much trickier to control[140]. In the future, patient-derived iPSCs[141] and organoids[142] may be used for proteomic analysis as potentially valuable intermediates that are between laboratory-cultured cells and patient-derived tissue.

In response to the developments towards personalized medicine, snapshots of personalized proteomes are required as well[127]. Although unbiased MS-based methods are rapidly becoming faster and more comprehensive, targeted analysis of proteins of interest — for example, by SRM — will allow multi-level platforms that may soon compete in clinical environments with existing analyses, such as enzyme-linked immunosorbent assay (ELISA) tests[143]. In this context, mass cytometry[144], which couples an alternative FACS approach (using isotopes) with MS, already provides a competitive highly multiplexed platform for conventional FACS.

The rapid development of high-throughput technologies in the past decade, which is linked to a reduction in their costs, opens up new possibilities to interrogate a biological system at multiple regulatory levels and simultaneously offers us an unprecedented vision. A remaining challenge is to integrate further proteome data with data generated at other levels, such as genomes, transcriptomes and metabolomes[145], in which bioinformatics will have a key role. Early endeavours are the direct analysis of proteome data versus deep-sequenced DNA and RNA data and comparison of proteome data with data obtained from ribosome profiling[146]. These also rapidly maturing DNA- and RNA-sequencing methods are perhaps more comprehensive in covering whole genomes than proteome profiling; nevertheless, they still lack the ability to interrogate some crucial levels of regulation in gene expression and network biology, such as post-translational modifications. Therefore, MS-based proteomics will remain a methodology that is widely in demand for at least the next decade, generating data that are highly complementary and, in some aspects, unique to other post-genome platforms. Integrated biology approaches are essential for addressing system-wide biological questions. However, routine integration will require the maturation and alignment of diverse post-genome technologies, as well as crosstalk between different scientific communities. Efficient integration of all these technologies will ultimately result in next-generation systems biology, thus delivering meaningful biological insights.

1. Crick, F. Central dogma of molecular biology. *Nature* **227**, 561–563 (1970).
2. Frenkel-Morgenstern, M. *et al.* Chimeras taking shape: potential functions of proteins encoded by chimeric RNA transcripts. *Genome Res.* **22**, 1602–1611 (2012).
3. Salmena, L., Poliseno, L., Tay, Y., Kats, L. & Pandolfi, P. P. A ceRNA hypothesis: the Rosetta Stone of a hidden RNA language? *Cell* **146**, 353–358 (2011).
4. Baker, M. Proteomics: the interaction map. *Nature* **484**, 271–275 (2012).
5. Bensimon, A., Heck, A. J. & Aebersold, R. Mass spectrometry-based proteomics and network biology. *Annu. Rev. Biochem.* **81**, 379–405 (2012).
6. Cox, J. & Mann, M. Is proteomics the new genomics? *Cell* **130**, 395–398 (2007).
7. Cox, J. & Mann, M. Quantitative, high-resolution proteomics for data-driven systems biology. *Annu. Rev. Biochem.* **80**, 273–299 (2011).
8. Ficarro, S. B. *et al.* Phosphoproteome analysis by mass spectrometry and its application to *Saccharomyces cerevisiae*. *Nature Biotech.* **20**, 301–305 (2002).
9. Lemeer, S. & Heck, A. J. R. The phosphoproteomics data explosion. *Curr. Opin. Chem. Biol.* **13**, 414–420 (2009).
10. Lundby, A. *et al.* Quantitative maps of protein phosphorylation sites across 14 different rat organs and tissues. *Nature Commun.* **3**, 876 (2012).
11. Nagaraj, N. *et al.* Deep proteome and transcriptome mapping of a human cancer cell line. *Mol. Syst. Biol.* **7**, 548 (2011).
In this paper, 10,255 different proteins are identified in a human cell line by MS, and a paired RNA-seq data set is used to demonstrate, for the first time, nearly full proteome coverage.
12. Beck, M. *et al.* The quantitative proteome of a human cell line. *Mol. Syst. Biol.* **7**, 549 (2011).
13. Munoz, J. *et al.* The quantitative proteomes of human-induced pluripotent stem cells and embryonic stem cells. *Mol. Syst. Biol.* **7**, 550 (2011).
This paper reports a deep proteome comparison of ESCs and iPSCs, in which only a few significant protein changes were found, suggesting the absence of a recurrent molecular signature in iPSCs.
14. Nesvizhskii, A. I., Vitek, O. & Aebersold, R. Analysis and validation of proteomic data generated by tandem mass spectrometry. *Nature Methods* **4**, 787–797 (2007).
15. Käll, L., Canterbury, J. D., Weston, J., Noble, W. S. & MacCoss, M. J. Semi-supervised learning for peptide identification from shotgun proteomics datasets. *Nature Methods* **4**, 923–925 (2007).
16. Elias, J. E. & Gygi, S. P. Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nature Methods* **4**, 207–214 (2007).
17. Altelaar, A. F. M. *et al.* Database independent proteomics analysis of the ostrich and human proteome. *Proc. Natl Acad. Sci. USA* **109**, 407–412 (2012).
18. Ma, B. & Johnson, R. *De novo* sequencing and homology searching. *Mol. Cell. Proteomics* **11**, O111.014902 (2012).
19. Schwanhausser, B. *et al.* Global quantification of mammalian gene expression control. *Nature* **473**, 337–342 (2011).
This paper presents a comprehensive quantitative analysis of transcription and translation rates, indicating that protein abundance is mainly regulated at the translation level.
20. Castello, A. *et al.* Insights into RNA biology from an atlas of mammalian mRNA-binding proteins. *Cell* **149**, 1393–1406 (2012).
This study describes an approach for the comprehensive analysis of RNA-binding proteins using cross-linking and MS, increasing our knowledge on the mechanisms that determine RNA synthesis and degradation.
21. Peng, M. *et al.* Protease bias in absolute protein quantitation. *Nature Methods* **9**, 524–525 (2012).
22. Picotti, P. & Aebersold, R. Selected reaction monitoring-based proteomics: workflows, potential, pitfalls and future directions. *Nature Methods* **9**, 555–566 (2012).
23. Addona, T. A. *et al.* Multi-site assessment of the precision and reproducibility of multiple reaction monitoring-based measurements of proteins in plasma. *Nature Biotech.* **27**, 633–641 (2009).
24. Picotti, P., Bodenmiller, B., Mueller, L. N., Domon, B. & Aebersold, R. Full dynamic range proteome analysis of *S. cerevisiae* by targeted proteomics. *Cell* **138**, 795–806 (2009).
25. Gillet, L. C. *et al.* Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis. *Mol. Cell. Proteomics* **11**, O111.016717 (2012).
26. Bailey, D. J. *et al.* Instant spectral assignment for advanced decision tree-driven mass spectrometry. *Proc. Natl Acad. Sci. USA* **109**, 8411–8416 (2012).
27. Altelaar, A. F. M. *et al.* Benchmarking stable isotope labeling based quantitative proteomics. *J. Proteomics* 22 Oct 2012 (doi:10.1016/j.jprot.2012.10.009).
28. Vermeulen, M. *et al.* Quantitative interaction proteomics and genome-wide profiling of epigenetic histone marks and their readers. *Cell* **142**, 967–980 (2010).
29. Krüger, 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).
30. Galli, M. *et al.* aPKC phosphorylates NuMA-related LIN-5 to position the mitotic spindle during asymmetric division. *Nature Cell Biol.* **13**, 1132–1138 (2011).
31. Geiger, T., Cox, J., Ostasiewicz, P., Wisniewski, J. R. & Mann, M. Super-SILAC mix for quantitative proteomics of human tumor tissue. *Nature Methods* **7**, 383–385 (2010).

32. Boersema, P. J., Raijmakers, R., Lemeer, S., Mohammed, S. & Heck, A. J. R. Multiplex peptide stable isotope dimethyl labeling for quantitative proteomics. *Nature Protoc.* **4**, 484–494 (2009).
33. Klimmeck, D. *et al.* Proteomic cornerstones of hematopoietic stem cell differentiation: distinct signatures of multipotent progenitors and myeloid committed cells. *Mol. Cell. Proteomics* **11**, 286–302 (2012).
34. Muñoz, J. *et al.* The Lgr5 intestinal stem cell signature: robust expression of proposed quiescent '+4' cell markers. *EMBO J.* **31**, 3079–3091 (2012).
This paper uses a combination of FACS, proteomics and transcriptomics to define a signature of genes highly expressed in intestinal stem cells.
35. Bantscheff, M. *et al.* Quantitative chemical proteomics reveals mechanisms of action of clinical ABL kinase inhibitors. *Nature Biotech.* **25**, 1035–1044 (2007).
36. Mertins, P. *et al.* iTRAQ labeling is superior to mTRAQ for quantitative global proteomics and phosphoproteomics. *Mol. Cell. Proteomics* **11**, M111.014423 (2012).
37. Wenger, C. D. *et al.* Gas-phase purification enables accurate, multiplexed proteome quantification with isobaric tagging. *Nature Methods* **8**, 933–935 (2011).
38. Ting, L., Rad, R., Gygi, S. P. & Haas, W. MS3 eliminates ratio distortion in isobaric multiplexed quantitative proteomics. *Nature Methods* **8**, 937–940 (2011).
39. de Godoy, L. M. F. *et al.* Comprehensive mass-spectrometry-based proteome quantification of haploid versus diploid yeast. *Nature* **455**, 1251–1254 (2008).
40. Nagaraj, N. *et al.* System-wide perturbation analysis with nearly complete coverage of the yeast proteome by single-shot ultra HPLC runs on a bench top Orbitrap. *Mol. Cell. Proteomics* **11**, M111.013722 (2012).
41. Clamp, M. *et al.* Distinguishing protein-coding and noncoding genes in the human genome. *Proc. Natl Acad. Sci. USA* **104**, 19428–19433 (2007).
42. Efroni, S. *et al.* Global transcription in pluripotent embryonic stem cells. *Cell Stem Cell* **2**, 437–447 (2008).
43. Clark, M. B. *et al.* The reality of pervasive transcription. *PLoS Biol.* **9**, e1000625 (2011).
44. Brosch, M. *et al.* Shotgun proteomics aids discovery of novel protein-coding genes, alternative splicing, and 'resurrected' pseudogenes in the mouse genome. *Genome Res.* **21**, 756–767 (2011).
This paper demonstrates the huge potential of MS-based proteomics in combination with genomic information to validate, to refine and even to identify new protein-coding loci.
45. Li, M. *et al.* Widespread RNA and DNA sequence differences in the human transcriptome. *Science* **333**, 53–58 (2011).
46. Kleinman, C. L. & Majewski, J. Comment on 'Widespread RNA & DNA sequence differences in the human transcriptome'. *Science* **335**, 1302 (2012).
47. Lu, R. *et al.* Systems-level dynamic analyses of fate change in murine embryonic stem cells. *Nature* **462**, 358–362 (2009).
48. Takahashi, K. & Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126**, 663–676 (2006).
49. Chin, M. H. *et al.* Induced pluripotent stem cells and embryonic stem cells are distinguished by gene expression signatures. *Cell Stem Cell* **5**, 111–123 (2009).
50. Guenther, M. G. *et al.* Chromatin structure and gene expression programs of human embryonic and induced pluripotent stem cells. *Cell Stem Cell* **7**, 249–257 (2010).
51. Newman, A. M. & Cooper, J. B. Lab-specific gene expression signatures in pluripotent stem cells. *Cell Stem Cell* **7**, 258–262 (2010).
52. Phanstiel, D. H. *et al.* Proteomic and phosphoproteomic comparison of human ES and iPS cells. *Nature Methods* **8**, 821–827 (2011).
53. Benevento, M. & Munoz, J. Role of mass spectrometry-based proteomics in the study of cellular reprogramming and induced pluripotent stem cells. *Expert Rev. Proteomics* **9**, 379–399 (2012).
54. Beynon, R. J., Doherty, M. K., Pratt, J. M. & Gaskell, S. J. Multiplexed absolute quantification in proteomics using artificial QCAT proteins of concatenated signature peptides. *Nature Methods* **2**, 587–589 (2005).
55. Ohta, S. *et al.* The protein composition of mitotic chromosomes determined using multiclassifier combinatorial proteomics. *Cell* **142**, 810–821 (2010).
56. Lamond, A. I. *et al.* Advancing cell biology through proteomics in space and time (PROSPECTS). *Mol. Cell. Proteomics* **11**, O112.017731 (2012).
57. Posewitz, M. C. & Tempst, P. Immobilized gallium(III) affinity chromatography of phosphopeptides. *Anal. Chem.* **71**, 2883–2892 (1999).
58. Villén, J. & Gygi, S. P. The SCX/IMAC enrichment approach for global phosphorylation analysis by mass spectrometry. *Nature Protoc.* **3**, 1630–1638 (2008).
59. Pinkse, M. W. H., Uitto, P. M., Hilhorst, M. J., Ooms, B. & Heck, A. J. R. Selective isolation at the femtomole level of phosphopeptides from proteolytic digests using 2D-NanoLC-ESI-MS/MS and titanium oxide precolumns. *Anal. Chem.* **76**, 3935–3943 (2004).
60. Rush, J. *et al.* Immunoaffinity profiling of tyrosine phosphorylation in cancer cells. *Nature Biotech.* **23**, 94–101 (2005).
61. Huttlin, E. L. *et al.* A tissue-specific atlas of mouse protein phosphorylation and expression. *Cell* **143**, 1174–1189 (2010).
This resource paper presents a large collection of phosphorylation sites (36,000 unique sites) in nine mouse tissues, suggesting that ubiquitous phosphoproteins often display tissue-specific phosphorylation patterns that tune protein activity to fulfil the specific functions of each tissue.
62. Chevrier, N. *et al.* Systematic discovery of TLR signaling components delineates viral-sensing circuits. *Cell* **147**, 853–867 (2011).
63. Yu, Y. *et al.* Phosphoproteomic analysis identifies Grb10 as an mTORC1 substrate that negatively regulates insulin signaling. *Science* **332**, 1322–1326 (2011).
64. Van Hoof, D. *et al.* Phosphorylation dynamics during early differentiation of human embryonic stem cells. *Cell Stem Cell* **5**, 214–226 (2009).
65. Zhou, H. *et al.* Enhancing the identification of phosphopeptides from putative basophilic kinase substrates using Ti (IV) based IMAC enrichment. *Mol. Cell. Proteomics* **10**, M110.006452 (2011).
66. Moritz, A. *et al.* Akt-RSK-S6 kinase signaling networks activated by oncogenic receptor tyrosine kinases. *Sci. Signal.* **3**, ra64 (2010).
67. Lienhard, G. E. Non-functional phosphorylations? *Trends Biochem. Sci.* **33**, 351–352 (2008).
68. Wu, R. *et al.* A large-scale method to measure absolute protein phosphorylation stoichiometries. *Nature Methods* **8**, 677–683 (2011).
69. Tan, C. S. H. & Bader, G. D. Phosphorylation sites of higher stoichiometry are more conserved. *Nature Methods* **9**, 317; author reply 318 (2012).
70. Rigbolt, K. T. G. *et al.* System-wide temporal characterization of the proteome and phosphoproteome of human embryonic stem cell differentiation. *Sci. Signal.* **4**, rs3 (2011).
71. Keck, J. M. *et al.* A cell cycle phosphoproteome of the yeast centrosome. *Science* **332**, 1557–1561 (2011).
72. Wu, R. *et al.* Correct interpretation of comprehensive phosphorylation dynamics requires normalization by protein expression changes. *Mol. Cell. Proteomics* **10**, M111.009654 (2011).
73. Schwartz, D. & Gygi, S. P. An iterative statistical approach to the identification of protein phosphorylation motifs from large-scale data sets. *Nature Biotech.* **23**, 1391–1398 (2005).
74. Xue, L. *et al.* Sensitive kinase assay linked with phosphoproteomics for identifying direct kinase substrates. *Proc. Natl Acad. Sci. USA* **109**, 5615–5620 (2012).
75. Kettenbach, A. N. *et al.* Quantitative phosphoproteomics identifies substrates and functional modules of Aurora and Polo-like kinase activities in mitotic cells. *Sci. Signal.* **4**, rs5 (2011).
76. Choudhary, C. *et al.* Lysine acetylation targets protein complexes and co-regulates major cellular functions. *Science* **325**, 834–840 (2009).
This is the first large-scale analysis of lysine acetylation identifying 3,600 sites, showing that the regulatory scope of this modification goes beyond gene expression regulation (that is, histones), and it is comparable with that of other major PTMs.
77. Xu, G., Paige, J. S. & Jaffrey, S. R. Global analysis of lysine ubiquitination by ubiquitin remnant immunoaffinity profiling. *Nature Biotech.* **28**, 868–873 (2010).
78. Argenzio, E. *et al.* Proteomic snapshot of the EGF-induced ubiquitin network. *Mol. Syst. Biol.* **7**, 462 (2011).
79. Galisson, F. *et al.* A novel proteomics approach to identify SUMOylated proteins and their modification sites in human cells. *Mol. Cell. Proteomics* **10**, M110.004796 (2011).
80. Kolarich, D., Jensen, P. H., Altmann, F. & Packer, N. H. Determination of site-specific glycan heterogeneity on glycoproteins. *Nature Protoc.* **7**, 1285–1298 (2012).
81. Zielinska, D. F., Gnad, F., Wiśniewski, J. R. & Mann, M. Precision mapping of an *in vivo N*-glycoproteome reveals rigid topological and sequence constraints. *Cell* **141**, 897–907 (2010).
82. Wagner, S. A. *et al.* A proteome-wide, quantitative survey of *in vivo* ubiquitylation sites reveals widespread regulatory roles. *Mol. Cell. Proteomics* **10**, M111.013284 (2011).
83. Grönroos, E., Hellman, U., Heldin, C.-H. & Ericsson, J. Control of Smad7 stability by competition between acetylation and ubiquitination. *Mol. Cell* **10**, 483–493 (2002).
84. Tsuruzoe, S. *et al.* Inhibition of DNA binding of Sox2 by the SUMO conjugation. *Biochem. Biophys. Res. Commun.* **351**, 920–926 (2006).
85. Seet, B. T., Dikic, I., Zhou, M.-M. & Pawson, T. Reading protein modifications with interaction domains. *Nature Rev. Mol. Cell. Biol.* **7**, 473–483 (2006).
86. Wang, Z. *et al.* Extensive crosstalk between *O*-GlcNAcylation and phosphorylation regulates cytokinesis. *Sci. Signal.* **3**, ra2 (2010).
87. van Noort, V. *et al.* Cross-talk between phosphorylation and lysine acetylation in a genome-reduced bacterium. *Mol. Syst. Biol.* **8**, 571 (2012).
This paper investigates the interplay of lysine acetylation and phosphorylation in the genome-reduced bacterium *Mycoplasma pneumonia*, showing occurring coexistence and confirming crosstalk between them.
88. Beli, P. *et al.* Proteomic investigations reveal a role for RNA processing factor THRAP3 in the DNA damage response. *Mol. Cell* **46**, 212–225 (2012).
89. Venkatesan, K. *et al.* An empirical framework for binary interactome mapping. *Nature Methods* **6**, 83–90 (2009).
90. Li, X., Gianoulis, T. A., Yip, K. Y., Gerstein, M. & Snyder, M. Extensive *in vivo* metabolite-protein interactions revealed by large-scale systematic analyses. *Cell* **143**, 639–650 (2010).
91. Kocher, T. & Superti-Furga, G. Mass spectrometry-based functional proteomics: from molecular machines to protein networks. *Nature Methods* **4**, 807–815 (2007).
92. Gavin, A.-C. *et al.* Proteome survey reveals modularity of the yeast cell machinery. *Nature* **440**, 631–636 (2006).
93. Gingras, A.-C., Gstaiger, M., Raught, B. & Aebersold, R. Analysis of protein complexes using mass spectrometry. *Nature Rev. Mol. Cell Biol.* **8**, 645–654 (2007).
94. von Mering, C. *et al.* Comparative assessment of large-scale data sets of protein-protein interactions. *Nature* **417**, 399–403 (2002).
95. Yu, H. *et al.* High-quality binary protein interaction map of the yeast interactome network. *Science* **322**, 104–110 (2008).
96. Braun, P. *et al.* An experimentally derived confidence score for binary protein-protein interactions. *Nature Methods* **6**, 91–97 (2009).
97. Krogan, N. J. *et al.* Global landscape of protein complexes in the yeast *Saccharomyces cerevisiae*. *Nature* **440**, 637–643 (2006).
98. Guruharsha, K. G. *et al.* A protein complex network of *Drosophila melanogaster*. *Cell* **147**, 690–703 (2011).
99. Sowa, M. E., Bennett, E. J., Gygi, S. P. & Harper, J. W. Defining the human deubiquitinating enzyme interaction landscape. *Cell* **138**, 389–403 (2009).
100. Breitkreutz, A. *et al.* A global protein kinase and phosphatase interaction network in yeast. *Science* **328**, 1043–1046 (2010).
101. Behrends, C., Sowa, M. E., Gygi, S. P. & Harper, J. W. Network organization of the human autophagy system. *Nature* **466**, 68–76 (2010).
102. Jager, S. *et al.* Global landscape of HIV-human protein complexes. *Nature* **481**, 365–370 (2012).
This is an elegant demonstration of the potential of AP–MS, identifying 497 HIV–human PPIs, thereby improving our understanding of how the host machinery is manipulated during the course of HIV infection.
103. Jager, S. *et al.* Vif hijacks CBF-β to degrade APOBEC3G and promote HIV-1 infection. *Nature* **481**, 371–375 (2012).

104. Malovannaya, A. *et al.* Analysis of the human endogenous coregulator complexome. *Cell* **145**, 787–799 (2011).
105. Rigaut, G. *et al.* A generic protein purification method for protein complex characterization and proteome exploration. *Nature Biotech.* **17**, 1030–1032 (1999).
106. Poser, I. *et al.* BAC TransgeneOmics: a high-throughput method for exploration of protein function in mammals. *Nature Methods* **5**, 409–415 (2008).
107. Hutchins, J. R. A. *et al.* Systematic analysis of human protein complexes identifies chromosome segregation proteins. *Science* **328**, 593–599 (2010).
108. Cheeseman, I. M. & Desai, A. A. Combined approach for the localization and tandem affinity purification of protein complexes from metazoans. *Sci. STKE* **2005**, pl1 (2005).
109. Babu, M. *et al.* Interaction landscape of membrane-protein complexes in *Saccharomyces cerevisiae*. *Nature* **489**, 585–589 (2012).
110. Havugimana, P. C. *et al.* A census of human soluble protein complexes. *Cell* **150**, 1068–1081 (2012).
This paper presents a high number of protein complexes identified through the use of multiple biochemical fractionation technologies combined with MS.
111. Kristensen, A. R., Gsponer, J. & Foster, L. J. A high-throughput approach for measuring temporal changes in the interactome. *Nature Methods* **9**, 907–909 (2012).
112. Nishiyama, T. *et al.* Sororin mediates sister chromatid cohesion by antagonizing Wapl. *Cell* **143**, 737–749 (2010).
113. Wepf, A., Glatter, T., Schmidt, A., Aebersold, R. & Gstaiger, M. Quantitative interaction proteomics using mass spectrometry. *Nature Methods* **6**, 203–205 (2009).
114. Bennett, E. J., Rush, J., Gygi, S. P. & Harper, J. W. Dynamics of Cullin-RING ubiquitin ligase network revealed by systematic quantitative proteomics. *Cell* **143**, 951–965 (2010).
115. Bisson, N. *et al.* Selected reaction monitoring mass spectrometry reveals the dynamics of signaling through the GRB2 adaptor. *Nature Biotech.* **29**, 653–658 (2011).
This paper shows the first example of combined AP–SRM in a detailed analysis of a key adaptor protein GRB2 and its interacting proteins, showing context-specific and time-dependent networks.
116. Li, V. S. W. *et al.* Wnt signaling through inhibition of β-catenin degradation in an intact Axin1 complex. *Cell* **149**, 1245–1256 (2012).
117. Vermeulen, M. *et al.* Selective anchoring of TFIID to nucleosomes by trimethylation of histone H3 lysine 4. *Cell* **131**, 58–69 (2007).
118. Nesvizhskii, A. I. Computational and informatics strategies for identification of specific protein interaction partners in affinity purification mass spectrometry experiments. *Proteomics* **12**, 1639–1655 (2012).
119. Leitner, A. *et al.* Probing native protein structures by chemical cross-linking, mass spectrometry, and bioinformatics. *Mol. Cell. Proteomics* **9**, 1634–1649 (2010).
120. Stengel, F., Aebersold, R. & Robinson, C. V. Joining forces: integrating proteomics and cross-linking with the mass spectrometry of intact complexes. *Mol. Cell. Proteomics* **11**, R111.014027 (2012).
121. Heck, A. J. R. Native mass spectrometry: a bridge between interactomics and structural biology. *Nature Methods* **5**, 927–933 (2008).
122. Herzog, F. *et al.* Structural probing of a protein phosphatase 2A network by chemical cross-linking and mass spectrometry. *Science* **337**, 1348–1352 (2012).
123. Ghosh, S., Matsuoka, Y., Asai, Y., Hsin, K.-Y. & Kitano, H. Software for systems biology: from tools to integrated platforms. *Nature Rev. Genet.* **12**, 821–832 (2011).
124. Ideker, T. & Krogan, N. J. Differential network biology. *Mol. Syst. Biol.* **8**, 565 (2012).
125. Vidal, M., Cusick, M. E. & Barabasi, A. L. Interactome networks and human disease. *Cell* **144**, 986–998 (2011).
126. Stratton, M. R. Exploring the genomes of cancer cells: progress and promise. *Science* **331**, 1553–1558 (2011).
127. Chen, R. *et al.* Personal omics profiling reveals dynamic molecular and medical phenotypes. *Cell* **148**, 1293–1307 (2012).
128. Lee, M. J. *et al.* Sequential application of anticancer drugs enhances cell death by rewiring apoptotic signaling networks. *Cell* **149**, 780–794 (2012).
129. Addona, T. A. *et al.* A pipeline that integrates the discovery and verification of plasma protein biomarkers reveals candidate markers for cardiovascular disease. *Nature Biotech.* **29**, 635–643 (2011).
130. Whiteaker, J. R. *et al.* A targeted proteomics-based pipeline for verification of biomarkers in plasma. *Nature Biotech.* **29**, 625–634 (2011).
131. Poste, G. Bring on the biomarkers. *Nature* **469**, 156–157 (2011).
132. Liotta, L. A. & Petricoin, E. Cancer biomarkers: closer to delivering on their promise. *Cancer Cell* **20**, 279–280 (2011).
133. Mischak, H. *et al.* Recommendations for biomarker identification and qualification in clinical proteomics. *Sci. Transl. Med.* **2**, 46ps42 (2010).
134. Paczesny, S. *et al.* Elafin is a biomarker of graft-versus-host disease of the skin. *Sci. Transl. Med.* **2**, 13ra2 (2010).
135. Taguchi, A. *et al.* Lung cancer signatures in plasma based on proteome profiling of mouse tumor models. *Cancer Cell* **20**, 289–299 (2011).
136. Altelaar, A. F. M. & Heck, A. J. R. Trends in ultrasensitive proteomics. *Curr. Opin. Chem. Biol.* **16**, 206–213 (2012).
137. Wiśniewski, J. R., Ostasiewicz, P. & Mann, M. High recovery FASP applied to the proteomic analysis of microdissected formalin fixed paraffin embedded cancer tissues retrieves known colon cancer markers. *J. Proteome Res.* **10**, 3040–3049 (2011).
138. Rikova, K. *et al.* Global survey of phosphotyrosine signaling identifies oncogenic kinases in lung cancer. *Cell* **131**, 1190–1203 (2007).
This paper describes a combination of phosphotyrosine enrichment and MS in a large-scale survey, identifying known and novel oncogenic kinases in lung cancer.
139. Andersen, J. N. *et al.* Pathway-based identification of biomarkers for targeted therapeutics: personalized oncology with PI3K pathway inhibitors. *Sci. Transl. Med.* **2**, 43ra55 (2010).
140. Hawkridge, A. M. & Muddiman, D. C. Mass spectrometry-based biomarker discovery: toward a global proteome index of individuality. *Annu. Rev. Anal. Chem.* **2**, 265–277 (2009).
141. Wu, S. M. & Hochedlinger, K. Harnessing the potential of induced pluripotent stem cells for regenerative medicine. *Nature Cell Biol.* **13**, 497–505 (2011).
142. Sato, T. *et al.* Single Lgr5 stem cells build crypt-villus structures *in vitro* without a mesenchymal niche. *Nature* **459**, 262–265 (2009).
143. Schiess, R., Wollscheid, B. & Aebersold, R. Targeted proteomic strategy for clinical biomarker discovery. *Mol. Oncol.* **3**, 33–44 (2009).
144. Bendall, S. C. *et al.* Single-cell mass cytometry of differential immune and drug responses across a human hematopoietic continuum. *Science* **332**, 687–696 (2011).
145. Nurse, P. & Hayles, J. The cell in an era of systems biology. *Cell* **144**, 850–854 (2011).
146. Ingolia, N. T., Ghaemmaghami, S., Newman, J. R. S. & Weissman, J. S. Genome-wide analysis *in vivo* of translation with nucleotide resolution using ribosome profiling. *Science* **324**, 218–223 (2009).
147. Boersema, P. J., Divecha, N., Heck, A. J. R. & Mohammed, S. Evaluation and optimization of ZIC-HILIC-RP as an alternative MudPIT strategy. *J. Proteome Res.* **6**, 937–946 (2007).
148. Di Palma, S. *et al.* Highly sensitive proteome analysis of FACS-sorted adult colon stem cells. *J. Proteome Res.* **10**, 3814–3819 (2011).
149. Andrews, G. L., Simons, B. L., Young, J. B., Hawkridge, A. M. & Muddiman, D. C. Performance characteristics of a new hybrid quadrupole time-of-flight tandem mass spectrometer (TripleTOF 5600). *Anal. Chem.* **83**, 5442–5446 (2011).
150. Michalski, A. *et al.* Mass spectrometry-based proteomics using Q Exactive, a high-performance benchtop quadrupole Orbitrap mass spectrometer. *Mol. Cell. Proteomics* **10**, M111.011015 (2011).
151. Frese, C. K. *et al.* Improved peptide identification by targeted fragmentation using CID, HCD and ETD on an LTQ-Orbitrap Velos. *J. Proteome Res.* **10**, 2377–2388 (2011).
152. Syka, J. E. P., Coon, J. J., Schroeder, M. J., Shabanowitz, J. & Hunt, D. F. Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proc. Natl Acad. Sci. USA* **101**, 9528–9533 (2004).
153. Olsen, J. V. *et al.* Higher-energy C-trap dissociation for peptide modification analysis. *Nature Methods* **4**, 709–712 (2007).
154. Dephoure, N. & Gygi, S. P. Hyperplexing: a method for higher-order multiplexed quantitative proteomics provides a map of the dynamic response to rapamycin in yeast. *Sci. Signal.* **5**, rs2 (2012).
155. Ito, A. *et al.* MDM2-HDAC1-mediated deacetylation of p53 is required for its degradation. *EMBO J.* **21**, 6236–6245 (2002).
156. Hirota, T., Lipp, J. J., Toh, B.-H. & Peters, J.-M. Histone H3 serine 10 phosphorylation by aurora B causes HP1 dissociation from heterochromatin. *Nature* **438**, 1176–1180 (2005).

Acknowledgements
This work was supported by the Netherlands Proteomics Centre, by the PRIME-XS project (grant number 262067, funded by the European Union 7th Framework Program) and by the Netherlands Organization for Scientific Research (NWO)-supported large-scale proteomics facility Proteins@Work (project 184.032.201).

Competing interests statement
The authors declare no competing financial interests.

FURTHER INFORMATION
Albert J. R. Heck's homepage: http://www.hecklab.nl
BioGRID: http://thebiogrid.org
The GPM: http://www.thegpm.org
Human Protein Reference Database: http://www.hprd.org
Human Protein Reference Database PhosphoMotif Finder: http://www.hprd.org/PhosphoMotif_finder
Intact: http://www.ebi.ac.uk/intact
Matrix Science — Mascot: http://www.matrixscience.com/search_form_select.html
NetPhorest: http://netphorest.info
NeXtprot: http://www.nextprot.org
PeptideAtlas: http://www.peptideatlas.org
Phosida: http://www.phosida.com
Phospho.ELM: http://phospho.elm.eu.org
PhosphoSitePlus: http://www.phosphosite.org
PRIDE: http://www.ebi.ac.uk/pride
ProteomeXchange: http://www.proteomeexchange.org
SEQUEST: http://fields.scripps.edu/sequest/index.html
String: http://string-db.org
Tranche: https://proteomecommons.org
UniProt: http://www.uniprot.org

ALL LINKS ARE ACTIVE IN THE ONLINE PDF