ELSEVIER

Teaser The field of proteomics has developed quickly over the past decade and its application to cancer research has considerable potential in the area of precision medicine.

Proteomics and drug discovery in cancer

Matheus H. Dias[1,3], Eduardo S. Kitano[1,3], André Zelanis[2] and Leo K. Iwai[1]

[1] Laboratório Especial de Toxinologia Aplicada (LETA), Center of Toxins, Immune-Response and Cell Signaling (CeTICS), Instituto Butantan, São Paulo, Brazil
[2] Laboratório de Proteômica Funcional, Instituto de Ciência e Tecnologia, Universidade Federal de São Paulo (ICT-UNIFESP), São José dos Campos, São Paulo, Brazil

Proteomics has emerged as an invaluable tool in the quest to unravel the biochemical changes that give rise to the hallmarks of cancer. In this review, we present the advances and challenges facing proteomics technology as applied to cancer research, and address how the information gathered so far has helped to enhance understanding of the mechanisms underlying the disease and contributed to the discovery of biomarkers and new drug targets. We conclude by presenting a perspective on how proteomics could be applied in the future to determine prognostic biomarkers and direct strategies for effective cancer treatment.

Introduction

The viability of multicellular organisms requires that individual cells must proliferate, differentiate, quiesce, senesce, and even die on behalf of organism homeostasis. The control of cellular fate involves many levels of complexity and organization, from hormonal signaling to cell cycle checkpoints. The foundation of several diseases lies in the malfunction and/or loss of this control, leading to the disruption of system homeostasis and, eventually, death. Cancers are the most frequent examples of such aberrations in the mechanisms that control cell fate.

Malignant transformation is a multistep process in which genetic and epigenetic alterations result in the malfunction of the normal checkpoints that control cell fate. Many different routes lead to a variety of malignant phenotypes observed in human cancers. However, it is a consensus that there is a defined collection of 'abilities' that all malignant cells must acquire to develop into a cancer. These include: sustained proliferative signaling; evasion of growth suppressive signaling; resistance to cell death; limitless replication; induction of angiogenesis; invasion; and acquisition of metastatic capability [1]. Underlying all these traits, known as the hallmarks of cancer, is the genomic instability that fuels the gradual acquisition of these capabilities through Darwinian natural selection during the onset and progression of the disease, which it does by creating a diversity of phenotypes within the transformed cell population. A remarkable

Corresponding author: Iwai, L.K. (leo.iwai@butantan.gov.br)
[3] These authors contributed equally to this work.

Leo K. Iwai, Dr Iwai received his undergraduate degree in chemistry, from the University of São Paulo in Brazil in 1995. He received his MSc and PhD in molecular biology from the Federal University of São Paulo in 1999 and 2004, respectively. After post-doctoral studies at the Harvard Medical School and at the Massachusetts Institute of Technology from 2005 to 2010 and at the Institute of Cancer Research in London from 2010 to 2012, he joined the Instituto Butantan in São Paulo as a research scientist in 2010. Dr Iwai studies snake venoms and their components as potential targets for diagnostics and therapy of cancer.

Eduardo S. Kitano, Eduardo Kitano graduated in pharmacy and biochemistry from the University of São Paulo, Brazil, in 2011. He is currently a PhD candidate at the Department of Biochemistry at University of São Paulo, developing his work at the Instituto Butantan. His current research focuses on the analysis of complex protein samples by mass spectrometry and sample preparation optimization with emphasis on peptide separation techniques, such as MudPIT and StageTip. Besides his active collaborative work with several different groups within the institute and at the University of São Paulo, he has focused on the characterization of protein expression patterns in sugarcane leaves at different stages of plant development.

Matheus Dias, Dr Dias graduated in biomedical sciences in 2005 from the Bandeirante University of São Paulo (Brazil). He received his PhD in biochemistry from the University of São Paulo in 2012, studying the mechanisms underlying growth factor signaling and cell cycle control in Ras-driven malignant cells. Since 2013, he has been a postdoctoral fellow at the Instituto Butantan in São Paulo, where he has been using an interdisciplinary approach, which includes proteomic techniques and computational modeling and/or simulation, to probe the network subjacent of Ras signaling in malignant cells to uncover potentially targetable vulnerabilities in Ras-driven malignant phenotypes.

http://dx.doi.org/10.1016/j.drudis.2015.10.004

heterogeneity exists both between different cancer types and within individual tumors. This is a consequence of the many different genetic and epigenetic alterations that are able to initiate malignant transformation overlaid with the effect of genomic instability and continuous selection and/or adaptation during tumor evolution [2]. Given this inherent inter- and intratumoral heterogeneity, and the influence of Darwinian selection, it is of little surprise that strategies targeting specific cancer cell traits, such as inhibition of an overactivated signaling pathway or blocking DNA synthesis, are often undermined by the emergence of drug resistance, resulting in tumor relapse and failure to eliminate the underlying malignancy [3,4]. The emergence of resistance and the recalcitrance of many types of cancer to long-term clinical control illustrate the need for large-scale and high-throughput biological data acquisition from tumors and cancer cells to provide new therapeutic strategies and additional guidance for the application of existing treatments.

In this regard, since 2008, when the first cancer genome was fully sequenced [5], next-generation sequencing has provided a massive amount of information about mutations that could trigger and drive tumorigenesis, providing novel targets for diagnosis, prognosis, and drug discovery (reviewed in [6]). However, such studies are unable to provide a complete picture of the relevant biological changes driving tumorigenesis. In addition to mutational events, miRNA deregulation, changes in DNA methylation patterns, and post-translational modifications (PTMs) of proteins can also alter the expression of oncogenes and/or tumor suppressors. Furthermore, some mutations function in a context-dependent manner, such that their detection alone is of limited utility in directing treatment strategies. For example, drugs targeting the same genetic lesion on the *BRAF* gene (V600E) in different tumors, can lead to distinct outcomes according to the expression levels of wild-type epidermal growth factor receptor (EGFR) [7]. Moreover, although some mutations function as drivers of tumor progression, these typically exist in a background of many others passenger mutations, which can cloud the identification of pathologically relevant driver events. Therefore, the search for reliable cancer biomarkers and therapeutic targets using genomic approaches is restrained by a variety of intrinsic features of the disease.

The concept of a biomarker refers to a characteristic that can be measured as an indicator of the physiological and/or pathological process or response to therapeutic intervention [8]. In cancer, the diversity of tumor phenotypes, the relation with the tumor microenvironment, and the underlying pathologies themselves result in multiple points for creation of potential biomarkers. These include: cancer-associated genomic mutations; noncoding RNAs; expression of specific proteins and/or peptides; circulating tumor cells (CTCs); and genomic, proteomic, and metabolomic signatures. Some well-established biomarkers have proved to be valuable tools for cancer risk assessment. For example, the BRCA1 and BRCA2 mutations, which correlate with significantly increased risk for ovarian and breast cancer [9]. Similarly, EGFR mutation status correlates with sensitivity to tyrosine kinase inhibitors in lung cancer [10], and presence of CTCs can monitor the effectiveness of prostate cancer treatment [11]. In the context of tumor heterogeneity and the current trend toward targeted therapies, the identification of reliable and specific biomarkers has fundamental importance in drug discovery by allowing the characterization and classification of the patient populations who are most likely to respond to specific therapeutic agents. Given that the proteome is the result of genetic background, PTMs, environmental, and microenvironmental factors, proteomic-based approaches to biomarker discovery are able to address levels of biological complexity that cannot be addressed with genomic and transcriptomic approaches, and provide data that are complementary to these analyses.

The application of proteomics to cancer has progressed alongside technological developments in the wider field of proteomics itself. The design of new strategies for sample fractionation, labeling, processing, and analysis combined with the increased speed and sensitivity of the latest generation of mass spectrometers (MS) have increased the capacity of proteomics to identify and quantify proteins and PTMs across a wide dynamic range.

The evolution of proteomics for biomarker discovery

The first proteomic platform for studying complex diseases, including cancer, was based on 2D polyacrylamide gel electrophoresis (2D-PAGE) [12]. In this approach, proteins are separated by their isoelectric point (first dimension) and molecular mass (second dimension), providing the opportunity to analyze several hundred proteins simultaneously and characterize their expression patterns in different samples (Fig. 1a). The development of soft ionization techniques, such as electrospray ionization (ESI) [13] and matrix-assisted laser desorption/ionization (MALDI) [14], made it possible to analyze proteins and peptides by MS, enabling the identification of proteins separated by 2D-PAGE more readily than by using previous sequencing-based techniques. One of the pioneering studies in cancer using 2D-PAGE was by Hanash and co-workers, who compared lymphoblasts from patients with different subtypes of acute lymphoblastic leukemia (ALL). Twelve protein spots were found to be differentially regulated between the different ALL subtypes [15].

Despite the utility of the method, variability between replicate 2D-PAGE gels can hinder the identification of differentially regulated proteins. To overcome this limitation, 2D-fluorescence differential-in-gel electrophoresis (DIGE) technology was developed by Unlu and colleagues [16]. This technology utilizes different fluorescent dyes to label samples from different conditions. Labeled samples are then mixed and analyzed on the same gel, eliminating the effects of gel-to-gel variation from the analysis (Fig. 1b). Zhou and colleagues used 2D-DIGE to identify cancer-specific protein markers in esophageal carcinoma [17]. In this study, 58 protein spots were found to be upregulated in cancer cells compared with normal controls, while a further 107 were found to be downregulated. Although the 2D-DIGE approach has gone some way to addressing the issue of gel-to-gel variability, other technical challenges have limited the application of the technique. 2D-PAGE suffers from a limited dynamic range. Additionally, where proteins co-migrate, it is not possible to infer directly which species is responsible for the differential regulation of protein spot volume [18]. Furthermore, 2D-PAGE is not an effective method for the separation of membrane proteins, which represent approximately 50% of important drug targets [19]. Nonetheless, despite these challenges, 2D techniques are still used with success in some areas of cancer biology research [20].

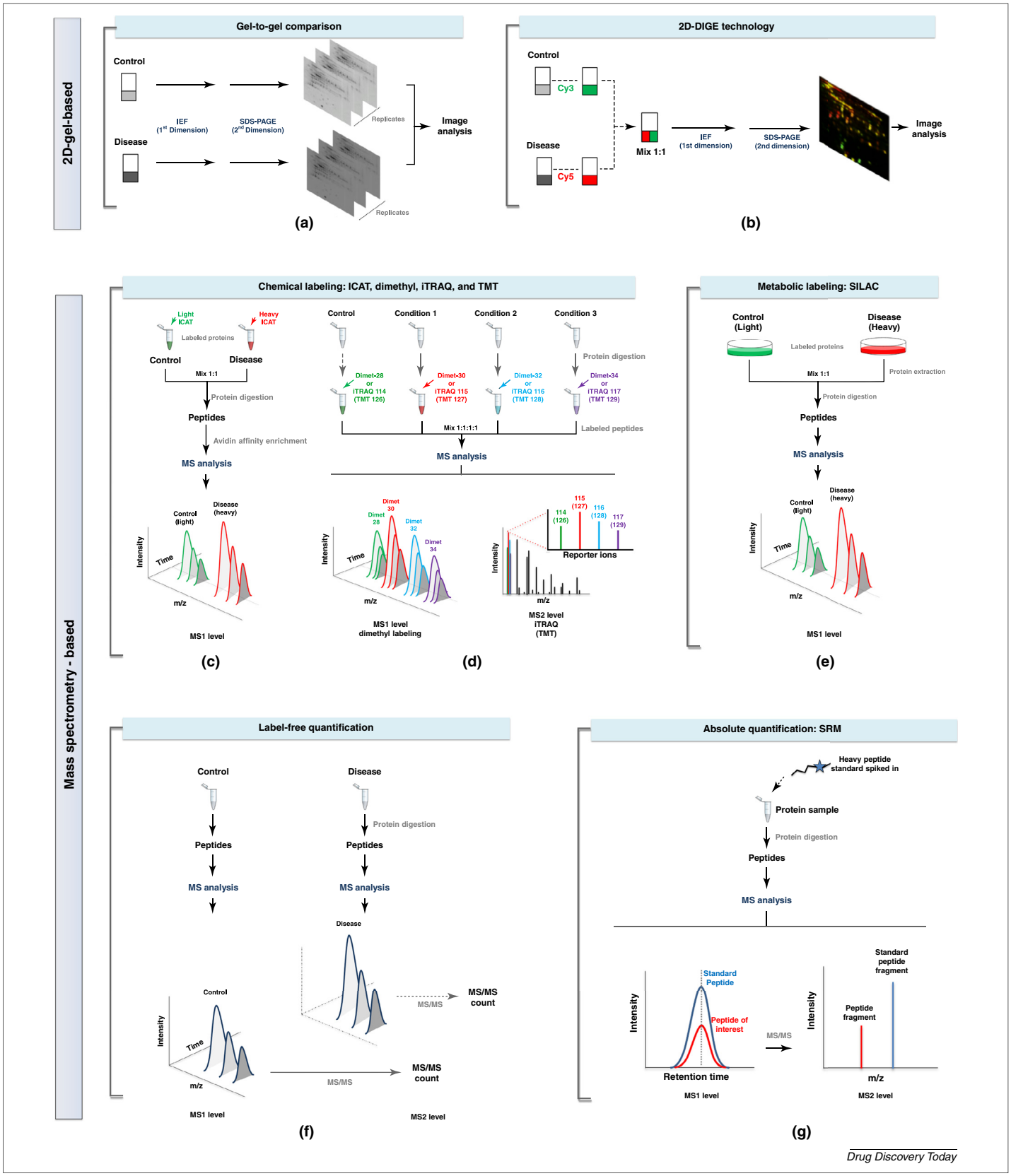

FIGURE 1

Overview of quantitative proteomics approaches used for cancer proteome studies. Relative or absolute quantification is achieved by either 2D-gel- or mass spectrometry (MS)-based proteomics. 2D gels allow relative protein quantification by comparing protein spot volume between different gels **(a)** or within the same gel using 2D-differential-in-gel electrophoresis (DIGE) technology **(b)**. For 2D-DIGE, each control or disease sample is labeled with either fluorescent Cy3 or Cy5 dyes before running the gel. Samples are combined and run on a single gel, and the resultant gel image is analyzed based on the fluorescent green or red intensities. In MS-based quantitative approaches, protein abundances can be estimated by stable isotope labeling or label-free approaches. Proteins can be

More recently, developments in high-performance liquid chromatography in combination with electrospray ionization and MS have led to so-called 'shotgun proteomics approaches' superseding 2D-PAGE, the method of choice for the analysis of complex proteomics in most areas of biology. In this approach, the entire set of proteins present in a target proteome (cell, tissue, fluids, or organism) is digested in solution with specific enzymes (most commonly trypsin). The mixture, comprising several thousand peptides from different proteins, is separated by liquid chromatography and analyzed in tandem by MS (LC–MS/MS), resulting in the identification and quantification of a large number of proteins in a single experiment. Furthermore, technological advances have seen successive generations of high-performance MS benefit from increased resolving power, mass accuracy, scan rate, and sensitivity. With the introduction of hybrid instruments, normally combining ion-trap and Orbitrap analyzers, MS-based proteomics has become the most effective tool for the investigation of complex proteomes, including the oncoproteome [21].

MS-based quantitative proteomics and drug discovery

Although there can be some utility in merely defining lists of proteins present in a sample, the ability to quantify these proteins is more informative if the underlying biological processes are to be understood. In relation to shotgun proteomics, the utility of such approaches owes much to the development of unique strategies and chemistries that have enabled the quantitation as well as identification of the proteins present in different samples. These innovations include stable isotope labeling strategies and label-free techniques.

Stable isotope labeling is based on the incorporation of heavy atoms on specific molecules that are introduced chemically or metabolically into the protein or peptide structures (reviewed in [22]). Methods based on the chemical labeling of proteins and/or peptides exhibit high quantification accuracy because systematic errors from sample handling can be minimized because different samples are combined in the initial steps of sample preparation. One of the first chemical labeling methods used for protein quantification by MS was the isotope-coded affinity tag (ICAT) approach [23]. This is based on the derivatization of cysteine residues with an ICAT reagent, comprising a thiol-reactive group, a biotin affinity tag, and a linker group containing 'light' or 'heavy' isotopes. Thiol-containing peptides are separated by biotin affinity chromatography and analyzed by LC–MS/MS. The presence of the ICAT tag results in separation of the peptides at the precursor (MS1) level (mass shift of +8 Da and +9 Da for the original and cleavable ICAT reagents, respectively) and allows for the relative quantification of peptides containing 'light' and 'heavy' tags (Fig. 1c). This approach has been successfully used to determine differential protein expression in prostate cancer cells exposed to androgens; in total, 77 proteins were found differentially regulated by the synthetic androgen R1881 in locally advanced carcinoma of the prostate (LACaP) cells, many of which had not previously been reported to be responsive to androgen stimulation [24].

Other common labeling chemistries used for quantitative shotgun proteomics are dimethyl [25], isobaric tags for relative and absolute quantification (iTRAQ) [26], and tandem mass tags (TMT) [27] chemical labels. Similar to ICAT, quantification using dimethyl label is performed at the MS1 level, whereas quantification using iTRAQ and TMT labels is carried out by the comparison of the relative intensities of MS/MS signature ions at the MS2 level, released during peptide fragmentation in tandem mass spectra (Fig. 1d). These methods allow for multiplexing several (usually up to eight) different samples in a single experiment and are based on the use of isobaric reagents that label primary amines in the peptides (N termini and the epsilon amino group of the lysine side chain). ICAT, dimethyl, iTRAQ, and TMT labeling can be readily applied to almost all types of sample, including tissue samples from animal models and patient biopsies. This constitutes a significant advantage over approaches based on the metabolic incorporation of isotopic labels, described elsewhere. With respect to sample preparation using dimethyl, iTRAQ, and TMT labels, protein samples from different treatments or conditions are digested, labeled separately with different tags, and then combined. The pooled peptide mixture is then analyzed by LC–MS/MS to provide both peptide identification and relative quantification. A large number of studies applying iTRAQ to oncoproteomics have been reported. De Souza and colleagues used iTRAQ in combination with cICAT labeling to define nine potential biomarkers associated with endometrial cancer [28]. Calderón-González and colleagues applied iTRAQ methodology to determine the protein expression profile of four different breast cancer cell lines (MCF7, MDA-MB-231, SK-BR-3, and T47D) in comparison to a MCF-10A nontumorigenic control cell line, with the aim of discovering novel biomarkers for the early detection of breast cancer [29]. All breast cancer cell lines shared 78 overexpressed proteins and 128 underexpressed proteins, mainly related to metabolic pathways and the generation of energy through anaerobic glycolysis instead of oxidative phosphorylation; all traits related to the Warburg effect [30]. Consequently, a set of six biomarkers was proposed to be of potential use for the diagnosis and treatment of breast cancer. In addition, the authors indicated panels of biomarkers found exclusively in each breast cancer cell line that can be used for the classification of different subtypes of the disease.

An alternative stable isotope labeling approach is metabolic labeling. Metabolic labeling has been shown to also provide precise quantification because the labels are incorporated into live

chemically tagged using an isotope-coded affinity tag (ICAT) reagent, where the ratio of 'light' and 'heavy' thiol-containing peptides is calculated between two different samples in the same MS run **(c)**. A similar labeling strategy can be performed at the peptide level using isobaric tags for relative and absolute quantification (iTRAQ), tandem mass tags (TMT), or dimethyl labeling. The relative protein abundances are measured in multiple samples by calculating the ratios of precursor ions at MS1 level (Dimet-28, -30, -32, and -34 for dimethyl labeling) or reporter ions at MS2 level (114-117 for iTRAQ and 126-129 for TMT) in the same MS run **(d)**. In the stable isotope labeling with amino acids in cell culture (SILAC) metabolic labeling method, proteins from two (or more) populations of cells are either unlabeled ('light') or metabolically labeled with 'heavy' SILAC amino acids. The relative abundances are determined by comparing 'light' and 'heavy' precursor ion intensities in the same MS run **(e)**. Label-free quantitative approaches are based on the comparison of precursor ion intensities or spectral counting in different MS runs **(f)**. In the targeted selected reaction monitoring (SRM) approach, a 'heavy' standard peptide is spiked into the protein sample at a known concentration. Peptide abundance is measured by comparing the native peptide to the spiked 'heavy' peptide. Quantification is performed on the basis of the signal generated when specific peptides are selected and fragmented, and predefined fragment ions are allowed to pass to the detector **(g).**

cells before any handling of the protein samples. One of the most popular commercial techniques characterized by its effectiveness and reproducibility is stable isotope labeling with amino acids in cell culture (SILAC) (Fig. 1e) [31]. In this method, cells are grown under different experimental conditions separately in culture media containing distinctive isotope-labeled lysine and arginine amino acids, resulting in 'light' (Arg-0 or Lys-0), 'medium' (Arg-6 or Lys-4), or 'heavy' (Arg-10 or Lys-8) forms of proteins in each experimental condition. The use of modified lysine and arginine ensures that peptides derived from trypsin digestion contain one labeled amino acid, which results in a predictable mass increase. Cells are typically cultured in isotopically defined media for several weeks to ensure that stable isotope-labeled amino acids are entirely incorporated into the proteome. Labeled proteins are extracted from cells, mixed, subjected to trypsin digestion, and finally those that are SILAC labeled are analyzed by LC–MS/MS. Relative quantification is accomplished by specific software that compares the intensities of the isotope clusters of the same peptide labeled with 'light', 'medium', and 'heavy' forms. SILAC is similar to ICAT in so far as the quantification is performed at the MS1 level.

Ren and colleagues used SILAC-based quantitative proteomics to study changes associated with hepatocellular carcinoma (HCC) by comparing the proteomes of the HepG2 liver cancer cell line and an immortalized normal hepatic cell line, L02 [32]. Among 63 differentially expressed proteins identified, phosphoglycerate mutase 1 (PGAM1) was identified as the most upregulated protein in HepG2 cells and was subsequently validated by reverse transcriptase (RT)-PCR and western blotting analyses, where the enzyme was also found overexpressed in approximately 67% (36/54) of HCC samples compared with normal liver tissue controls. The study showed that PGAM1 can have an important role in hepatocarcinogenesis and highlights the potential use of PGAM1 both as a diagnostic biomarker and therapeutic target.

Using the same quantitative approach, Zhou and colleagues performed a study that aimed to identify proteins regulated by the treatment of MDA-MB-231 human breast cancer cells with suberoylanilide hydroxamic acid (SAHA), a histone deacetylase (HDAC) inhibitor [33]. By combining parallel western-blot analysis and a SILAC-based quantitative MS approach, they discovered that the treatment of MDA-MB-231 with SAHA increased the lysine acetylation of 61 proteins, including both histone and nonhistone proteins. A number of these proteins had not previously been described as targets of HDAC inhibitors. Furthermore, the authors showed that lysine acetylation was sustained with prolonged treatment, indicating the potential effectiveness of SAHA treatment in cancer therapy.

Another recent study using SILAC-based quantitative proteomics demonstrated that sensitivity to the chemotherapeutic drug paclitaxel (PTX) in multiple cancer cell lines was related to levels of the tumor suppressor programmed cell death 4 (PDCD4); and levels of this protein in lung cancer tissues was positively correlated with a better prognosis in patients treated with PTX [34].

In recent years, advances in MS instrumentation and specific computational analysis tools have revealed the potential of label-free quantification of shotgun proteomic data as an alternative to chemical and metabolic labeling techniques [35]. Advantages of label-free quantification include the relative simplicity and low costs associated with the technique. Specifically, the use of label-free methods can avoid some limitations found in labeling-based methods, such as the additional steps required for sample preparation; the cost of labeling reagents; the limitation of sample numbers when multiplexing; and the amount of sample that can be analyzed. Two methods exist for estimating the relative abundance of a protein in multiple samples during label-free quantification. The first method, known as spectral counting, is based on the observation that more abundant peptides are sampled more frequently by MS than those of lower abundance. That is, the number of MS/MS spectra acquired for a given peptide is a function of its abundance in the sample [36]. The second method is based on the measure of ion intensities over a chromatographic elution profile observed in MS1 scans, where changes in protein abundance are estimated by comparing peptide intensities between different samples [37] (Fig. 1f).

Using a label-free approach, Winiewski and colleagues reported the quantification of 7576 proteins extracted from microdissected CRC samples and the identification of 1808 proteins that showed differential expression levels between normal and cancer tissues [38]. A label-free approach was also used in the investigation of protein markers associated with lymph node metastasis in CRC [39] and for the identification of candidate biomarkers present in the serum of patients with HCC [40].

Analyzing the subproteome of tumor cells, Wang and Hanash presented a method for analyzing the abundance of cell surface proteins based on the enrichment of surface membrane proteins by biotin labeling, followed by affinity chromatography, protein fractionation by reversed phase liquid chromatography (RPLC), and label-free absolute quantification [41]. This approach has provided valuable insights into how the cell surface proteome is regulated and how it changes in response to intracellular and extracellular stimuli.

Cancer PTM analysis: glycoproteomics and phosphoproteomics

By integrating data from the so-called 'omics' approaches, one might expect a comprehensive molecular profiling of oncogenic processes to be carried out in a feasible timescale [42]. However, even with improvements in LC–MS/MS instrumentation over the past decade, it remains a considerable challenge to detect and measure protein (micro) heterogeneities, such as protein PTMs that arise during and after protein synthesis, and which are, in many cases, of more functional significance than protein abundance per se. Given that PTMs are cellular events that occur at relatively low levels, their detection by MS involves specific methods for the enrichment of the modified protein and/or peptide population before MS analysis. Proteomics methods focused on the analysis of PTMs have benefited from improvements in not only MS instrumentation, but also protocols used to enrich for modified proteins and peptides. Below, we review approaches available to study glycosylation and phosphorylation, two key PTMs related to important signaling events in cancer.

Glycoproteomics and drug discovery

Protein glycosylation is an enzymatic process that is part of the secretory machinery of eukaryotic cells and occurs in the lumen of the endoplasmic reticulum and the Golgi apparatus, where glycan moieties are transferred by glycosyltransferases to nitrogen

(*N*-glycosylation) or oxygen (*O*-glycosylation) atoms of asparagine or serine/threonine amino acid side chains, respectively [43]. In the case of *N*-linked glycans, this process is highly site specific, occurring at a consensus motif formed by the triplet amino acid sequence Asn-Xaa-Ser/Thr (where Xaa could be any amino acid, except proline). Glycans have both important structural and biochemical roles, and are involved in many cellular processes, including direction of intra- and/or extracellular trafficking of glycoconjugates; modulation of cell adhesion and signaling; and regulation of cell proliferation and differentiation [43,44].

Most mammalian proteins are post-translationally modified and, even though certain modifications can be predicted by bioinformatic analysis of cDNA sequences, the extent and complexity of a given PTM, such as glycosylation, is almost unpredictable based solely on *in silico* analysis. Furthermore, the actual glycan diversity of a given glycoprotein can vary by several orders of magnitude depending on the utilization of glycosylation motifs, resulting in the generation of distinct glycoforms and/or the diversification of structural isomers by differences in constitutional isomers and/or linkage position among glycan monomers. These observations illustrate the inherent complexity associated with cancer-related (glyco) proteins as one of the main challenges in identifying reliable biomarkers.

Given that several fine-tuning mechanisms are altered upon oncogenesis, one might expect that significant changes in complex biological processes are propagated by changes affecting glycans and/or glycoproteins. Indeed, the increase in glycosyltransferase expression together with their Golgi localization is correlated with tumor transformation and progression in different cancers [45]. Moreover, specific glycan structures have been identified by MS as being associated with the degree of tumor malignancy [46–48]. Champattanachai and colleagues showed that aberrant protein *O*-GlcNAcylation is associated with malignancy in primary breast tumors, because *O*-GlcNAc transferase knockdown resulted in the inhibition of anchorage-independent growth *in vitro* [47]. In fact, more than half of cancer biomarkers discovered to date are glycosylated proteins, and the use of high-throughput approaches based on proteomics and MS have been successful in defining new biomarkers of this type [48,49]. Moreover, such approaches have also allowed researchers to explore qualitative and quantitative differences in the well-established cancer biomarkers, such as the *N*-glycans from the prostate-specific antigen (PSA) [46].

Given that sialylated glycoproteins are strongly correlated with tumor progression, Zhang and colleagues developed a chemical strategy for the determination of site-specific *N*-sialoglycan occupancy rates on the basis of specific oxidation of dihydroxyl groups from sialic acid [48]. The authors performed a high-throughput analysis of the *N*-sialoglycan occupancy rates in hepatocellular carcinoma and human liver tissues and found 76 *N*-sialoglycosites with occupancy rates higher than twofold compared with normal tissue.

Cell surface proteins are among the most variable protein groups within mammalian cells [50]. Given that several cell surface mammalian proteins are glycosylated, proteomic studies aiming to describe cancer-associated proteins have taken advantage of this feature by including an enrichment step in their experimental protocols to perform selective capture of glycosylated proteins in primary cultured tumors or cell lines. Among the experimental approaches that have been commonly used in glycoproteomics, chemical labeling of cell surface proteins and lectin capturing have been used with promising results [51,52].

Recently, Bausch-Fluck and colleagues reported a comprehensive MS-derived cell surface protein atlas, analyzing over 70 mammalian cell types, including cancer cells. Among the 1492 identified proteins, the Cell Surface Protein Atlas (CSPA) identified several *N*-glycosylated cell surface protein-specific markers, such as CD30 on the Hodgkin lymphoma cell lines and CD172a on glioblastoma cells [52].

Given that many clinically relevant proteins are of low abundance, the enrichment strategy has the main advantage of lowering the dynamic range of protein concentration, avoiding sampling of highly abundant or contaminant proteins. Tan and colleagues used a lectin-based enrichment strategy in combination with peptide labeling for relative quantitation to perform a large-scale analysis of core-fucosylated glycopeptides derived from serum samples from patients with pancreatic cancer [53]. It was reported that, out of the 322 identified proteins, eight exhibited significantly altered expression levels and, thus, are potential markers for pancreatic cancer.

Vakhrushev and colleagues developed a method for the precision mapping of the human *O*-GalNAc glycoproteome, which has been recently used to probe the *O*-glycoproteome of gastric cancer cell lines (AGS and MKN45) for potential biomarkers [54]. The authors successfully identified nearly 500 *O*-glycoproteins in gastric cancer cell lines as well as 26 exclusive *O*-glycoproteins in the serum of patients with gastric cancer. Comparison of the *O*-glycoproteome from the gastric cancer cell lines with a previous *O*-glycoproteome set, derived from 12 human cancer cell lines from different organs, enabled the identification of a new subset of 175 *O*-glycoproteins and *O*-glycosites that had not been reported previously.

Phosphoproteomics and drug discovery

Protein phosphorylation has a paramount role in modifying proteins in a reversible and highly dynamic, transient fashion, and modulates several aspects of protein structure and function [55]. The phosphorylation state of any given protein is a result of a dynamic interplay between protein kinase and phosphatase activities, which gives rise to the substoichiometric nature of protein phosphorylation, whereby specific sites can be phosphorylated from <1% to >90% [56]. Importantly, dysregulation of protein phosphorylation is a key driver of the cancer cell phenotype, and the desire to understand the aberrant global phosphorylation events observed in cancer has made phosphorylation one of the best-studied PTMs. Impairment in signaling networks resulting from overexpression of kinases, mutation of their corresponding genes, as well as altered negative regulatory mechanisms, have been recognized as ubiquitous features in several cancer types [57–59]. Moreover, multiple oncogenes and tumor suppressors driving dysregulated protein phosphorylation pathways have been thoroughly investigated as drug targets. Indeed, drugs targeting protein kinases are one of the most promising group of compounds currently available for cancer therapy [57,60,61], even though adaptive signaling changes frequently lead to activation of

alternative signaling networks, resulting in drug resistance and, ultimately, tumor relapse [62].

Analysis of phosphorylation is particularly dependent on an effective strategy for enriching phosphorylated peptides because of the substoichiometric nature of the modification, and the behavior of phosphopeptides within the MS. The negative charge of the phosphate group can be exploited to isolate phospho-Ser (pSer), phospho-Thr (pThr), and phospho-Tyrosine (pTyr)-containing peptides via selective binding to titanium-dioxide beads (TiO_2). Alternatively, phosphopeptides can be retained on immobilized metal affinity chromatography (IMAC) resin derivatized with iron (Fe^{3+}). However, pTyr sites are typically under-represented with this approach, because of the comparative infrequency of the modification compared with pSer and pThr. Therefore, for pTyr analysis, peptides are typically enriched using immunoaffinity-based methods employing panspecific anti-pTyr antibodies [56,63].

Sharma and colleagues developed a label-free approach to quantify phosphorylation peptides and determine their fractional occupancy in complex samples [64]. This strategy allowed for the identification of over 50 000 distinct phosphorylated peptides in HeLa S3 cell lysates. Interestingly, proteins phosphorylated on tyrosine residues were, on average, more abundant compared with the entire proteome. This observation led to the proposition that tyrosine phosphorylation acts as a separate functional regulatory post-translational modification in eukaryotic proteomes.

Using a 'multi-omics' approach, Smit and colleagues identified targets whose inhibition would increase the toxicity of the BRAF enzyme inhibitor vemurafenib toward melanoma cells [65]. As expected, treatment with vemurafenib led to the downregulation of phosphorylation within the mitogen-activated protein kinase (MAPK) pathway. Furthermore, out of approximately 5700 identified proteins and approximately 11 500 phosphosites, the authors found that the negative regulator of Rho-associated, coiled-coil containing protein kinase 1 (ROCK1) kinase, Rnd3, was downregulated, which pointed to ROCK1 as a potential combinatorial drug target for BRAF mutant melanoma. This was subsequently corroborated by gene-silencing experiments.

To gain a molecular understanding of the mechanisms by which EGFRvIII acts in glioblastoma multiforme, Huang and colleagues performed a large-scale analysis of EGFRvIII-activated phosphotyrosine-mediated signaling pathways by using iTRAQ labeling followed by pTyr and IMAC phospho-enrichments [66]. The authors identified and quantified 99 phosphorylation sites on 69 proteins and found that the activating phosphorylation site on the c-Met receptor was highly responsive to EGFRvIII levels, indicating cross-activation of the c-Met receptor tyrosine kinase by EGFRvIII. A combined treatment regimen using a c-Met kinase inhibitor and either an EGFR kinase inhibitor or cisplatin, resulted in enhanced cytotoxicity toward EGFRvIII-expressing cells compared with treatment with either compound alone.

More recently, Zhang and colleagues described a robust experimental framework and associated error model for iTRAQ-based quantification on an Orbitrap MS [67]. Their experimental model focused on the role of the Fms-like tyrosine kinase 3 (FLT3) receptor tyrosine kinase, an important receptor in normal hematopoietic development and leukemogenesis. Point mutations within the activation loop and in-frame tandem duplications of the juxtamembrane domain represent the most frequent molecular abnormalities observed in acute myeloid leukemia. Interestingly, these gain-of-function mutations correlated with different clinical outcomes, suggesting that signals from constitutive FLT3 mutants activate different downstream targets. Application of this error model to quantitative proteomics data for FLT3 signaling provided evidence that phosphorylation of tyrosine phosphatase SHP1 abrogates the transformative potential, but not overall kinase activity of FLT3-D835Y in acute myeloid leukemia.

Despite much technological and methodological progress in the field, the higher amount of starting material compared with the study of other PTMs, such as glycosylation, and the transient nature of phosphorylation are still among the main challenges of current phosphoproteomic approaches in cancer biology.

Secretomics and drug discovery

Proteins secreted from cells into the extracellular environment have an important role in many physiological and pathological processes. In cancer, it has been observed that the composition of secreted proteins is different compared with normal tissue, which makes them an important source for cancer biomarker and/or drug target discovery [68,69]. Thus, systemic experimental approaches aimed at characterizing cell secretomes provide important qualitative and quantitative evidence for understanding the process of tumor biology [70].

Stromal cells are often recruited by tumor cells to participate in tumorigenic development by inducing the production and release of molecular signals responsible for tumor progression, such as cellular growth factors, and by facilitating dispersion of tumor cells through the activity of proteases that degrade the extracellular matrix [71–73]. In this process, the secretion of specific molecules by tumor cells can act to drive cell signaling events contributing to the direct or indirect development and proliferation of cancer cells [73].

In addition, it is known that by secreting molecular factors, such as VEGF and proteases, tumor cells can mobilize noncancer bone marrow hematopoietic precursor cells to specific sites creating a suitable environment known as the premetastatic niche, into which tumor cells are able to locate and multiply [74].

Obenauf and colleagues showed that melanoma and adenocarcinoma cells submitted to BRAF, anaplastic lymphoma kinase (ALK), or EGFR kinase inhibitor therapy were able to induce a complex network of secreted factors promoting proliferation, migration, and metastasis of drug-resistant cancer cell clones. These factors also increased survival and suppressed the apoptotic activity of drug-sensitive tumor cells, contributing to rapid tumor progression [75].

Therefore, secretome analysis is of interest from the perspective of understanding the repertoire of factors potentially involved in the biochemical events associated with tumorigenesis [76,77] as well as for understanding the cellular and molecular complexity of the tumor microenvironment, and for the identification of factors contributing to metastasis.

Intracellular peptidomics

A promising area of investigation that has not yet been extensively explored in the area of cancer proteomics is the characterization of the intracellular peptidome. Uncontrolled cell growth and

near-constant proliferation and division of tumor cells require extensive protein degradation by proteasomes. In addition, it has been suggested that the proteasome has an extensive role in regulating the homeostasis of the intracellular peptidome and its deregulation has been suggested to be associated with cancer development [78,79]. In this regard, proteasome inhibitors, such as bortezomib and carfilzomib, have been used for the treatment of patients with multiple myeloma and mantle cell lymphoma, as antitumor agents that are able to regulate uncontrolled cell growth and induce apoptosis in several tumor cells [80,81]. Moreover, there is evidence that certain intracellular peptides have biological activities that could be exploited in the search for novel anticancer agents. For example, it was recently shown that the peptide WELVVLGKL (pep5) derived from G1/S-specific cyclin-D2 inhibited proteasome activity and induced cell death in several tumor cells, and reduced the volume of rat C6 glioblastoma when fused to a cell penetrating peptide (pep5-cpp) [82]. Therefore, intracellular peptidome mapping has the potential to identify not only potential biomarkers, but also novel bioactive peptides that can be exploited as new therapeutic targets.

Protein–protein interactomics and pathway analysis

In conventional MS-based proteomic workflows, the first step in sample preparation is the solubilization of proteins in the sample. Typically, highly denaturing conditions are used with the aim of solubilizing all of the proteins in the sample. However, protein function can be regulated independently of abundance, by association with other proteins and/or sequestration to specific subcellular localizations, and these events cannot be captured by conventional proteomic approaches. By using less stringent lysis conditions in combination with strategies to isolate specific proteins of interest (e.g., by immunoprecipitation, or epitope tagging and affinity purification) it is possible to study specific subproteomes with the aim of elucidating the regulation of discrete functional units underpinning cellular processes of interest. One example of this approach is the mapping of protein–protein interactions (PPI), as used in 'interactome' mapping studies, which have revealed important information on disease biology and, consequently, therapeutically relevant disease-associated proteins [83,84].

Interactome studies can provide insights into protein regulation and/or function in different disease states, which is particularly important for improving understanding of the molecular mechanisms underpinning events such as proliferation, migration, and angiogenesis. Targeting PPIs involved in regulating these processes represents an attractive avenue for the design of novel therapeutic strategies and less toxic anticancer drugs [85].

Similar to the genome and transcriptome databases, proteome expression database repositories (Table 1) have been devised in an attempt to make publicly available the massive amount of proteomic and PPI data from cell lines, biological fluids, different animal models, and clinical samples. Allied with bioinformatic tools, it is possible to mine the data in these repositories to identify and extract relations and patterns in the data that are not apparent in individual experiments, to formulate new, testable hypotheses to direct further research. The extensive data provided by the proteomic analysis of cancer cell lines and clinical samples encompassing different cancer types allow for the comparison of PPI from samples from both patients and healthy individuals, and from disease and normal cells. These analyses could lead to the identification of disruptions and abnormalities in specific PPIs of different signaling networks in different cancer types [86]. In addition, PPI analysis of clinical samples or cells from different classes of cancer treated with different drugs could also provide important information on the drug action, effectiveness, and resistance.

Cancer has been described as a disease of pathways [1,73], and it has been proposed that drug-discovery projects should shift from being protein-centric to being pathway-centric [87,88]. Indeed, it has been observed that patients with the same type of cancer typically show differences in protein expression and activation of specific oncogenic kinases [89–91] Conversely, It has been observed that, when compared with individual marker genes, protein subnetworks are more robust classifiers of cancer phenotype, and network-based classification achieves higher accuracy in prediction of therapeutic response [89,92]. Therefore, despite being challenging, there is clearly utility in mapping and identifying the pathways that lead to disease progression and metastasis.

With respect to targeting these aberrantly activated signaling networks, although most cancer drugs have been developed toward specific molecular targets, many of them target multiple proteins. For example, the multi-kinase inhibitors dasatinib, nilotinib, and imatinib, which were rationally designed to target BCR-ABL, have also been shown to inhibit other kinases, including platelet-derived growth factor receptor (PDGFR), discoidin domain receptor (DDRs), c-KIT, and SRC family kinases [22,93,94]. Interestingly, combination therapy with multiple drugs instead of a monotherapy targeting specific proteins has been more successful for the treatment of patients with cancer and also for diminishing the risk of drug resistance [73,95]. Proteomics has been useful in not only identifying new therapeutic targets, but also identifying network-level effects of different therapeutic agents.

The SILAC-aided proteomic and phosphoproteomic analysis of the effect of heat shock protein (HSP)-90 inhibitors 17-DMAG and geldanamycin on cancer cell lines showed multiple effects on several protein levels and multiple cellular processes, such as protein synthesis, protein degradation, cell cycle, and apoptosis [96,97]. Moreover, proteomic follow-up analysis of oncogenic signaling pathways in patients with cancer under specific drug therapy or cancer cell lines and tissues treated with different drugs has the potential to reveal unanticipated 'off-target' effects and new drug-specific pathway biomarkers that could guide clinical treatment decisions [22,94,98,99].

Targeted proteomic approaches

Targeted proteomic workflows are emerging as a complementary tool to discovery proteomics typified by the workflows described in the previous sections. Targeted proteomics differs from discovery proteomics in that it requires prior knowledge of the analytes to be studied. Hence, it is typically used for hypothesis-driven studies, often focusing on a particular set of peptides known to be derived from a given biological sample. Targeted proteomics typically uses an approach known as multiple reaction monitoring (MRM) or selected reaction monitoring (SRM), which is implemented on a triple-quadrupole MS (reviewed in [100]). A key advantage of MRM/SRM assays is that, by setting the MS to detect only a set of predefined analytes, it is possible to detect reproducibly analytes

TABLE 1

Database repositories containing proteomic and interactomic data

Database	Description	Website source
CCSB Interactome Database	An interactome database comprising human, virus, plant, bacteria, warm, and yeast PPIs	http://interactome.dfci.harvard.edu/
Database of Interacting Proteins (DIP)	Database combining both manually and automatic curated information creating a consistent set of PPIs	http://dip.doe-mbi.ucla.edu
Human Protein Reference Database (HPRD)	A centralized platform of curated proteomic data to depict and integrate information pertaining to domain architecture, PTMs, interaction networks, and disease association for each protein in the human proteome	http://www.hprd.org
InnateDB	Database with improved coverage of innate immunity interactome, integrating interactions and pathways from public databases with manually curated data into a centralized resource	http://www.innatedb.ca
IntAct	An open-source database system providing tools for analyses of molecular interaction data derived from literature curation or direct user submissions	http://www.ebi.ac.uk/intact
International Molecule Exchange (IMEx) consortium	An international collaboration between major interaction databases to combine curation efforts for increasing coverage and providing a nonredundant set of protein interactions in a single search interface	http://www.imexconsortium.org/
MatrixDB	A database focused on interactions established by extracellular proteins and polysaccharides; MatrixDB is a member of the IMEx consortium	http://matrixdb.ibcp.fr
Online predicted human interaction database – OPHID	Database integrating experimentally validated and predicted protein interactions for humans, *Saccharomyces cerevisiae*, *Caenorhabditis elegans*, *Drosophila melanogaster*, and *Mus musculus*	http://ophid.utoronto.ca
Proteopedia	A wiki encyclopedia of structural and functional information about protein, RNA, DNA, and other macromolecules, and their assemblies and interactions with small molecules	http://proteopedia.org/
The Biological General Repository for Interaction Databases (BioGRID)	A comprehensive repository database for protein and genetic interactions, chemical associations, and PTMs from major model organism species, compiled through literature curation	http://www.thebiogrid.org
The microbial protein interation database (MPIDB)	Database focused on microbial protein interactions curated from literature or imported from other databases	http://www.jcvi.org/mpidb
The PRoteome IDEntifications (PRIDE)	A public data repository for MS-based proteomics data, including protein and peptide identifications, as well as PTMs and supporting spectral evidence	http://www.ebi.ac.uk/pride
UniPep	A Swiss/American project to provide access to proteomics data from the Serum Biomarker group, offering a library of putative glycopeptides and theoretical proteotypic peptides	http://www.unipep.org

of low abundance in comparatively complex mixtures, where discovery proteomics approaches would either result in the analyte not being detected at all, or would detect it in only a subset of experiments. Additionally, because analytes are defined on the basis of their intact mass and multiple fragment ions, the assays are highly specific. Furthermore, known concentrations of isotopic-labeled heavy peptides (with either 13C- or 15N-containing amino acids) can be spiked into the sample to facilitate either relative or absolute quantification (Fig. 1g).

MRM/SRM assays have been successfully applied to samples ranging from cells and biopsy tissues to several kinds of biological fluid [100–104]. Recently, Sjöström and colleagues reported the use of combined strategies (shotgun and SRM) for breast cancer biomarker discovery [105]. After enriching breast tumor samples for *N*-glycopeptides, the authors compiled a list of proteins of interest and performed a multiplexed targeted analysis using SRM, resulting in the identification of ten proteins that were consistently differentially regulated between tumor samples. Importantly, SRM has been successfully applied to complex samples with protein concentrations that vary across many orders of magnitude, including human plasma. Indeed, Cima and colleagues measured a panel of candidate biomarkers for prostate cancer through a set of serum samples from over 100 individuals [104]. The identified protein signatures increased both sensitivity

and specificity of prostate cancer detection when compared with PSA measurements currently used as a diagnostic serum biomarker for prostate cancer.

Martínez-Aguilar and colleagues used SRM for the profiling of isoform-specific expression of the calcium-binding protein S100 in the three most common tumors of the thyroid gland (follicular adenoma, follicular thyroid carcinoma, and papillary thyroid carcinoma) in comparison with nine normal thyroid tissues [106]. Results from SRM analyses were also confirmed by metabolic (SILAC) labeling and western-blot analysis and allowed the identification of S100A31 as a novel candidate papillary thyroid carcinoma biomarker and the utility of S100A6, S100A4, and annexin A1 to discriminate follicular and papillary thyroid tumors.

A robust pipeline based on targeted proteomics for biomarker verification in plasma samples was recently developed and applied to the investigation of lung cancer biomarker candidates. The authors reported a total of 17 proteins as tumor markers for non-small cell lung cancer (NSCLC), including a novel plasma-based biomarker, the cell-adhesion protein zyxin [107].

Although they have clear utility in primary research, targeted proteomic strategies could also provide the platform that is able to take MS from the research laboratory into the clinical diagnostic setting. The ability to measure the abundance of analytes in complex biological samples such as serum, the high degree of specificity, the inherent reproducibility across technical replicates, and the relatively low cost of instrumentation, make targeted proteomics more suited to clinical diagnostics than are discovery proteomic approaches. Moreover, the ability to multiplex analyses, and the fact that assays do not rely on the development of specific antibodies, means that SRM/MRM-based assays have considerable potential as important diagnostic tools in the context of precision medicine.

TABLE 2

International large-scale omics initiatives

Initiative name	Description	Website
The Human Protein Atlas (HPA)	Focused on expression and localization of human proteins based on RNA and protein data. The cancer atlas, a subcategory of HPA, contains information on protein expression levels in tumor cells and provides useful information for identification of new potential cancer biomarkers	http://www.proteinatlas.org/
Clinical Proteomic Tumor Analysis Consortium (CPTAC)	Integrative effort joining genomic and proteomic data to detect proteins derived from alterations in the cancer genome	http://proteomics.cancer.gov/programs/cptacnetwork
Global Cancer Genomics Consortium (GCGC)	Focused on investigation of genomes and proteomes of cancer specimens to identify molecular drivers of cancer and collect cellular evidence of drug interaction, sensitivity, or resistance	https://smhs.gwu.edu/biochemistry/centers-institutes/gcgc
METAcancer Consortium	Aimed at identification of altered metabolites and metabolic pathways in breast cancer tissues through metabolomics; molecular classification of disease based on altered levels of metabolites, and identification of new prognostic and predictive biomarkers	http://www.metacancer-fp7.eu/
German Cancer Consortium (DKTK)	Focused on establishing interdisciplinary research centers to improve cancer prevention and diagnosis as well as development of more specific therapies	http://www.dkfz.de/en/index.html
The Biomarker Consortium	Aimed at acceleration of development of biomarker-based technologies to support drug development, preventive medicine, diagnosis and treatment	http://www.biomarkersconsortium.org/
Human Surfaceome Atlas	A repository for data representing the cell surface protein repertoire of 41 human and 31 mouse cell types generated by quantitative MS	http://wlab.ethz.ch/cspa/
Genome Medicine Database of Japan Proteomics (GeMDBJ)	Focused on identification of proteins associated with particular cancers as prognostic markers; proteome data of surgically resected tissues and tissue-cultured cells of various malignancies, as well as corresponding biological and clinicopathological data	https://gemdbj.nibio.go.jp/

Challenges for proteomics in cancer drug discovery

Current state-of-the-art proteomic technologies have made it possible to perform in-depth profiling of the cancer proteome with great reliability. With the application of specific workflows, it is possible to apply modern proteomics to interrogate cellular constituents, secreted proteins, PTMs, and PPIs. Mapping of altered signaling pathways and identification of alterations in protein expression that lead to tumor initiation, invasion, and metastasis have provided important clues to researchers for the discovery and design of more specific biomarker targets for diagnosis and treatment. However, despite these successes, more widespread adoption of the technology in cancer research is still hampered by several significant technical challenges. Despite improvements in instrumentation and sample fractionation procedures, the vast dynamic range of protein abundance coupled both protein isoform and disease heterogeneity still results in significant challenges. Moreover, different sample types can pose specific challenges to proteomic analyses, and sample preparation procedures can significantly alter the quality of data that can be obtained from analyses. For example, proteomic and phosphoproteomic analysis of breast cancer tumor samples has shown that levels of protein and phosphoprotein are affected by biospecimen type and pre-analytical sample manipulation procedures [108]. In addition, proteomics-based studies require specialized equipment and infrastructure, as well as trained personnel for sample preparation, processing, and analysis [109,110]. Moreover, a remaining bottleneck in most proteomic studies is the analysis of the large amounts of data generated to isolate biologically meaningful changes from background noise.

One approach to tackling these problems is the development of consortiums and initiatives formed by different groups of laboratories worldwide with different technical platforms to mine the cancer proteome in an integrative, large-scale, and collaborative manner (Table 2). By promoting and facilitating a culture of information sharing, these initiatives have enabled promising advances in the study and discovery of potential biomarkers for different types of cancer. Moreover, the multidisciplinary nature of these consortiums connects basic research with clinical and public health sciences, leading to a rational workflow focused on a common goal of reducing mortality by offering more precise and efficacious treatments for patients with cancer.

Concluding remarks

The application of proteomics to cancer research has provided invaluable insight into the biological processes that drive the hallmarks of cancer. Beyond merely cataloguing the proteome, advances in sample preparation, labeling, and instrumentation have made it possible to identify cancer-specific changes with more sensitivity than ever before. By selecting appropriate sample preparation and analysis approaches, it is possible to tailor proteomic technologies to study cell signaling networks, functionally important PTMs, PPIs, and protein expression changes. Although many technical challenges remain, the advent of

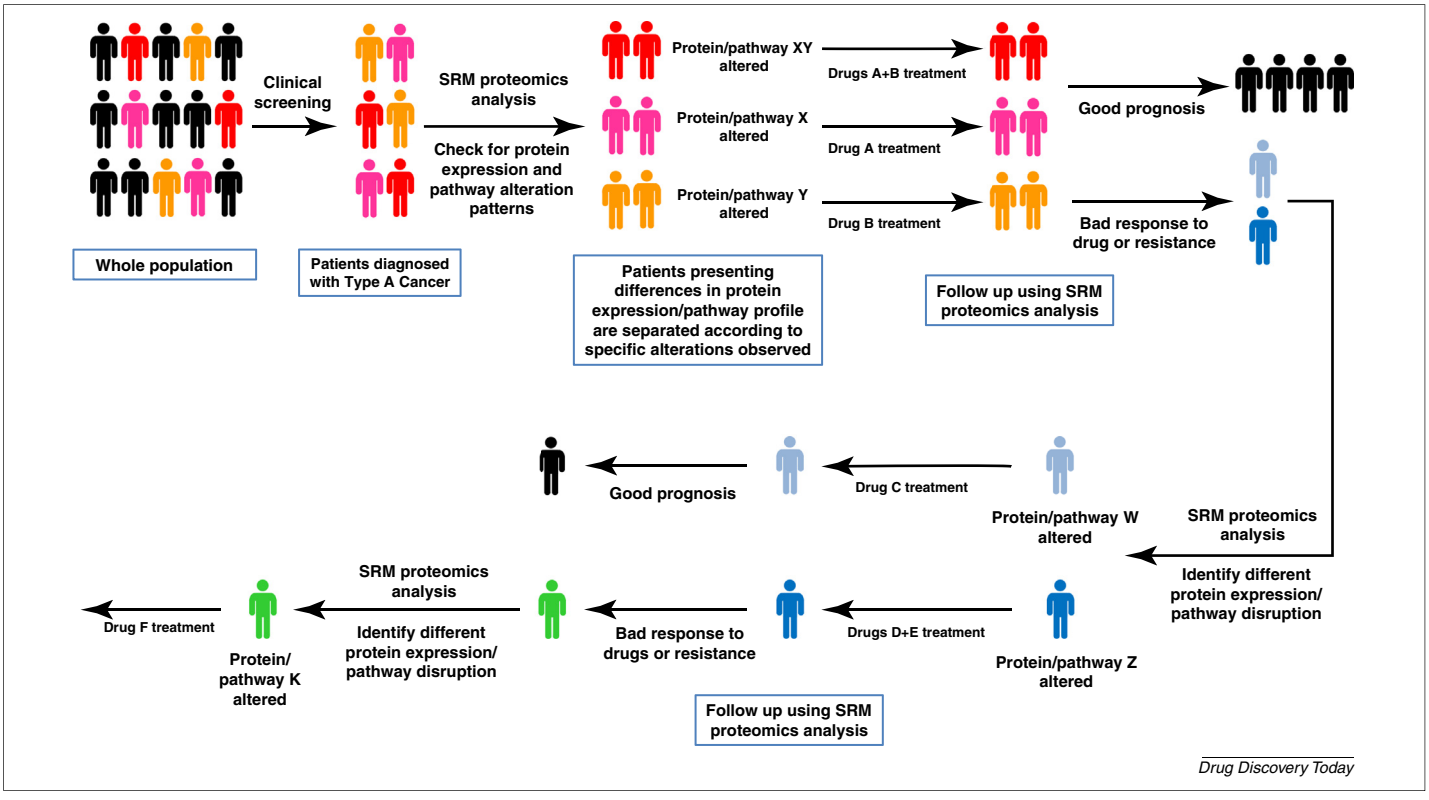

FIGURE 2

Precision medicine based on screening using a targeted selected reaction monitoring (SRM) approach. Clinically similar patients are screened for activation of specific signaling pathways through SRM-mass spectrometry (MS) analysis. Once the disrupted signaling pathway is identified, the patient undergoes mono- or combination chemotherapy. Follow-up testing allows for the identification of network reprogramming responsible for emergence of resistance and identifies new signaling pathways to direct second-round treatment.

consortium initiatives and proteomic data repositories, along with bioinformatic tools for mining these data, make it clear that proteomics will continue to have a key role in the discovery of both new biomarkers and new therapeutic targets.

However, an interesting question that remains is how this technology might be applied in the clinic to have a more direct role in patient management. It is conceivable that targeted proteomics could have an important role in the emerging precision medicine approach for cancer treatment, where individual variation in proteins as well as genes and metabolites are monitored to offer personalized diagnostics and treatment for each patient [111]. In this scheme, SRM analysis could be used to screen for the activation of key signaling networks in patient samples before treatment to direct treatment with mono- or combination-drug therapy (Fig. 2). Importantly, given that the emergence of resistance typically limits the durability of treatment response, follow-up screening could be used to identify network reprogramming events responsible for resistance and redirect treatment with compounds selective to newly activated bypass signaling pathways.

Acknowledgments

This work was supported by the Center of Toxins, Immune-response and Cell Signaling (CeTICS) grant 2013/07467-1 from the São Paulo Research Foundation (FAPESP). A.Z. is currently supported by the Young Investigator Grant 2014/06579-3 from FAPESP. E.S.K. is supported by PhD fellowship grant 2011/11308-0 and M.S.D. is supported by PD fellowship grant 2012/20186-9 from FAPESP. We thank Emer S. Ferro for insightful and valuable comments and suggestions on the manuscript.

References

1 Hanahan, D. and Weinberg, R.A. (2000) The hallmarks of cancer. *Cell* 100, 57–70
2 Burrell, R.A. *et al.* (2013) The causes and consequences of genetic heterogeneity in cancer evolution. *Nature* 501, 338–345
3 Poulikakos, P.I. and Rosen, N. (2011) Mutant BRAF melanomas: dependence and resistance. *Cancer Cell* 19, 11–15
4 Camidge, D.R. *et al.* (2014) Acquired resistance to TKIs in solid tumours: learning from lung cancer. *Nat. Rev. Clin. Oncol.* 11, 473–481
5 Ley, T.J. *et al.* (2008) DNA sequencing of a cytogenetically normal acute myeloid leukaemia genome. *Nature* 456, 66–72
6 Hudson, T.J. *et al.* (2010) International network of cancer genome projects. *Nature* 464, 993–998
7 Prahallad, A. *et al.* (2012) Unresponsiveness of colon cancer to BRAF (V600E) inhibition through feedback activation of EGFR. *Nature* 483, 100–103
8 De Gruttola, V.G. *et al.* (2001) Considerations in the evaluation of surrogate endpoints in clinical trials. Summary of a National Institutes of Health Workshop. *Control Clin. Trials* 22, 485–502
9 Domchek, S.M. *et al.* (2010) Association of risk-reducing surgery in BRCA1 or BRCA2 mutation carriers with cancer risk and mortality. *JAMA* 304, 967–975
10 Pao, W. *et al.* (2004) EGF receptor gene mutations are common in lung cancers from 'never smokers' and are associated with sensitivity of tumors to gefitinib and erlotinib. *Proc. Natl Acad. Sci. U. S. A.* 101, 13306–13311
11 Danila, D.C. *et al.* (2011) Circulating tumor cells as biomarkers in prostate cancer. *Clin. Cancer Res.* 17, 3903–3912
12 Ofarrell, P.H. (1975) High-resolution 2-dimensional electrophoresis of proteins. *J. Biol. Chem.* 250, 4007–4021
13 Fenn, J.B. *et al.* (1989) Electrospray ionization for mass-spectrometry of large biomolecules. *Science* 246, 64–71
14 Hillenkamp, F. *et al.* (1991) Matrix-assisted laser desorption ionization mass-spectrometry of biopolymers. *Anal. Chem.* 63, A1193–A1202
15 Hanash, S.M. *et al.* (1986) Lineage-related polypeptide markers in acute lymphoblastic-leukemia detected by two-dimensional gel-electrophoresis. *Proc. Natl. Acad. Sci. U. S. A.* 83, 807–811
16 Unlu, M. *et al.* (1997) Difference gel electrophoresis: a single gel method for detecting changes in protein extracts. *Electrophoresis* 18, 2071–2077
17 Zhou, G. *et al.* (2002) 2D differential in-gel electrophoresis for the identification of esophageal scans cell cancer-specific protein markers. *Mol. Cell. Proteomics* 1, 117–124
18 Campostrini, N. *et al.* (2005) Spot overlapping in two-dimensional maps: a serious problem ignored for much too long. *Proteomics* 5, 2385–2395
19 Drews, J. (2000) Drug discovery: a historical perspective. *Science* 287, 1960–1964
20 Poli, G. *et al.* (2015) 2D-DIGE proteomic analysis identifies new potential therapeutic targets for adrenocortical carcinoma. *Oncotarget* 6, 5695–5706
21 Makarov, A. *et al.* (2006) Performance evaluation of a hybrid linear ion trap/orbitrap mass spectrometer. *Anal. Chem.* 78, 2113–2120
22 Bantscheff, M. *et al.* (2007) Quantitative chemical proteomics reveals mechanisms of action of clinical ABL kinase inhibitors. *Nat. Biotechnol.* 25, 1035–1044
23 Gygi, S.P. *et al.* (1999) Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat. Biotechnol.* 17, 994–999
24 Meehan, K.L. and Sadar, M.D. (2004) Quantitative profiling of LNCaP prostate cancer cells using isotope-coded affinity tags and mass spectrometry. *Proteomics* 4, 1116–1134
25 Boersema, P.J. *et al.* (2009) Multiplex peptide stable isotope dimethyl labeling for quantitative proteomics. *Nat. Protoc.* 4, 484–494
26 Ross, P.L. *et al.* (2004) Multiplexed protein quantitation in *Saccharomyces cerevisiae* using amine-reactive isobaric tagging reagents. *Mol. Cell. Proteomics* 3, 1154–1169
27 Thompson, A. *et al.* (2003) Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. *Anal. Chem.* 75, 1895–1904
28 DeSouza, L. *et al.* (2005) Search for cancer markers from endometrial tissues using differentially labeled tags iTRAQ and cICAT with multidimensional liquid chromatography and tandem mass spectrometry. *J. Proteome Res.* 4, 377–386
29 Calderón-González, K.G. *et al.* (2015) Determination of the protein expression profiles of breast cancer cell lines by quantitative proteomics using iTRAQ labelling and tandem mass spectrometry. *J. Proteomics* 124, 50–78
30 Vander Heiden, M.G. *et al.* (2009) Understanding the Warburg effect: the metabolic requirements of cell proliferation. *Science* 324, 1029–1033
31 Ong, S.E. *et al.* (2002) Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol. Cell. Proteomics* 1, 376–386
32 Ren, F. *et al.* (2010) Quantitative proteomics identification of phosphoglycerate mutase 1 as a novel therapeutic target in hepatocellular carcinoma. *Mol. Cancer* 9, 81
33 Zhou, Q. *et al.* (2010) Screening for therapeutic targets of vorinostat by SILAC-based proteomic analysis in human breast cancer cells. *Proteomics* 10, 1029–1039
34 Xu, H. *et al.* (2015) Proteomic profiling of paclitaxel treated cells identifies a novel mechanism of drug resistance mediated by PDCD4. *J. Proteome Res.* 14, 80–91
35 Cox, J. *et al.* (2014) MaxLFQ allows accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction. *Mol. Cell. Proteomics* 3, 2513–2526
36 Liu, H. *et al.* (2004) A model for random sampling and estimation of relative protein abundance in shotgun proteomics. *Anal. Chem.* 76, 4193–4201
37 Bantscheff, M. *et al.* (2012) Quantitative mass spectrometry in proteomics: critical review update from 2007 to the present. *Anal. Bioanal. Chem.* 404, 939–965
38 Wiśniewski, J.R. *et al.* (2012) Extensive quantitative remodeling of the proteome between normal colon tissue and adenocarcinoma. *Mol. Syst. Biol.* 8, 611
39 Meding, S. *et al.* (2012) Tissue-based proteomics reveals FXYD3, S100A11 and GSTM3 as novel markers for regional lymph node metastasis in colon cancer. *J. Pathol.* 228, 459–470
40 Tsai, T. *et al.* (2015) LC–MS/MS-based serum proteomics for identification of candidate biomarkers for hepatocellular carcinoma. *Proteomics* 15, 2369–2381
41 Wang, H. and Hanash, S. (2015) Mass spectrometry based proteomics for absolute quantification of proteins from tumor cells. *Methods* 81, 34–40
42 The Cancer Genome Atlas research Network (2014) Comprehensive molecular profiling of lung adenocarcinoma. *Nature* 511, 543–550
43 Taylor, M.E. and Dricakamer, K., eds) (2006) *Introduction to Glycobiology*, Oxford University Press

44 Lau, K.S. *et al.* (2006) Complex N-glycan number and degree of branching cooperate to regulate cell proliferation and differentiation. *Cell* 129, 123–124
45 Brockhausen, I. (2006) Mucin-type O-glycans in human colon and breast cancer: glycodynamics and functions. *EMBO Rep.* 7, 599–604
46 Song, E. *et al.* (2014) Glycoproteomics: identifying the glycosylation of prostate specific antigen at normal and high isoeletric points by LC–MS/MS. *J. Proteome Res.* 13, 5570–5580
47 Champattanachai, V. *et al.* (2013) Proteomic analysis and abrogated expression of *O*-GlcNAcylted proteins associated with primary breast cancer. *Proteomics* 13, 2088–2099
48 Zhang, Y. *et al.* (2014) Mass spectrometry-based N-glycoproteomics for cancer biomarker discovery. *Clin. Proteomics* 11, 18
49 Kirmiz, C. *et al.* (2007) A serum glycomics approach to breast cancer biomarkers. *Mol. Cell. Proteomics* 6, 43–55
50 Schwanhausser, B. *et al.* (2011) Global quantification of mammalian gene expression control. *Nature* 473, 337–342
51 Woo, C.M. *et al.* (2015) Isotope-targeted glycoproteomics (IsoTaG): a mass-independent platform for intact N- and O-glycopeptide discovery and analysis. *Nat. Methods* 12, 561–567
52 Bausch-Fluck, D. *et al.* (2015) A mass spectrometric-derived cell surface protein atlas. *PLoS ONE* 10, e0121314
53 Tan, Z. *et al.* (2015) Large-scale identification of core-fucosylated glycopeptide sites in pancreatic cancer serum using mass spectrometry. *J. Proteome Res.* 14, 1968–1978
54 Vakhrushev, S.Y. *et al.* (2013) Enhanced mass spectrometric mapping of the human GalNAc-type O-glycoproteome with SimpleCells. *Mol. Cell. Proteomics* 12, 932–944
55 Cohen, P. (2002) The origins of protein phosphorylation. *Nat. Cell Biol.* 4, E127–E130
56 Macek, B. *et al.* (2009) Global and site-specific quantitative proteomics: principles and applications. *Annu. Rev. Pharmacol. Toxicol.* 49, 199–221
57 Daub, H. (2015) Quantitative proteomics of kinase inhibitor targets and mechanisms. *ACS Chem. Biol.* 10, 201–212
58 Morris, M.K. *et al.* (2015) Phosphoproteomics in drug discovery. *Drug Discov. Today* 19, 425–432
59 Harsha, H.C. and Pandey, A. (2010) Phosphoproteomics and cancer. *Mol. Oncol.* 4, 482–495
60 Cohen, P. (2002) Protein kinases: the major drug targets of the twenty-first century? *Nat. Rev. Drug. Discovery* 1, 309–315
61 Law, V. *et al.* (2014) DrugBank 4.0: shedding new light on drug metabolism. *Nucleic Acids Res.* 42, D1091–D1097
62 Rebecca, V.W. and Smalley, K.S.M. (2014) Change or die: targeting adaptive signaling to kinase inhibition in cancer cells. *Biochem. Pharmacol.* 91, 417–425
63 Curran, T.G. *et al.* (2015) MARQUIS: a multiplex method for absolute quantification of peptides and posttranslational modifications. *Nat. Commun.* 6, 5924
64 Sharma, K. *et al.* (2014) Ultradeep human phosphoproteome reveals a distinct regulatory nature of Tyr and Ser/Thr-based signaling. *Cell Rep.* 8, 1583–1594
65 Smit, M.A. *et al.* (2014) ROCK1 is a potential combinatorial drug target for BRAF mutant melanoma. *Mol. Syst. Biol.* 10, 772
66 Huang, P.H. *et al.* (2007) Quantitative analysis of EGFRvIII cellular signaling networks reveals a combinatorial therapeutic strategy for glioblastoma. *Proc. Natl. Acad. Sci. U. S. A.* 31, 12867–12872
67 Zhang, Y. *et al.* (2010) A robust error model for iTRAQ quantification reveals divergent signaling between oncogenic FLT3 mutants in acute myeloid leukemia. *Mol. Cell Proteomics* 9, 780–790
68 Makridakis, M. and Vlahou, A. (2010) Secretome proteomics for discovery of cancer biomarkers,. *J. Proteomics* 73, 2291–2305
69 Sepiashvili, L. *et al.* (2012) Potentially novel candidate biomarkers for head and neck squamous cell carcinoma identified using an integrated cell line-based discovery strategy. *Mol. Cell Proteomics* 11, 1404–1415
70 Schaaij-Visser, T.B.M. *et al.* (2013) The cancer secretome, current status and opportunities in the lung, breast and colorectal cancer context. *Biochim. Biophys. Acta* 1834, 2242–2258
71 Morrison, C.J. *et al.* (2009) Matrix metalloproteinase proteomics: substrates, targets and therapy. *Curr. Opin. Cell Biol.* 21, 645–653
72 Kessenbrock, K. *et al.* (2010) Matrix metalloproteinases: regulators of the tumor microenvironment. *Cell* 14, 52–67
73 Hanahan, D. and Weinberg, R.A. (2011) The hallmarks of cancer: the next generation. *Cell* 144, 57–70
74 Kaplan, R.N. *et al.* (2005) VEGFR1-positive haematopoietic bone marrow progenitors initiate the pre-metastatic niche. *Nature* 438, 820–827
75 Obenauf, A. *et al.* (2015) Therapy-induced tumour secretomes promote resistance and tumour progression. *Nature* 520, 368–372
76 Wu, C.C. *et al.* (2010) Candidate serological biomarkers for cancer identified from the secretomes of 23 cancer cell lines and the human protein atlas. *Mol. Cell Proteomics* 9, 1100–1117
77 Paltridge, J.L. *et al.* (2013) The secretome in cancer progression. *Biochim. Biophys. Acta* 1834, 2233–2241
78 Gelman, J.S. *et al.* (2013) Alterations of the intracellular peptidome in response to the proteasome inhibitor bortezomib. *PLoS ONE* 8, e53263
79 Dasgupta, S. *et al.* (2014) Proteasome inhibitors alter levels of intracellular peptides in HEK293T and SH-SY5Y cells. *PLoS ONE* 9, e103604
80 Kane, R.C. *et al.* (2003) Velcade: U.S. FDA approval for the treatment of multiple myeloma progressing on prior therapy. *Oncologist* 8, 508–513
81 Herndon, T.M. *et al.* (2013) U.S. Food and Drug Administration approval: carfilzomib for the treatment of multiple myeloma. *Clin. Cancer Res.* 19, 4559–4563
82 de Araujo, C.B. *et al.* (2014) A novel intracellular peptide derived from g1/s cyclin d2 induces cell death. *J. Biol. Chem.* 289, 16711–16726
83 Gulati, S. *et al.* (2013) Cancer networks and beyond: interpreting mutations using the human interactome and protein structure. *Semin. Cancer Biol.* 23, 219–226
84 Rolland, T. *et al.* (2014) A proteome-scale map of the human interactome network. *Cell* 159, 1212–1226
85 Kamdje, N.A.H. *et al.* (2014) New targeted therapies for breast cancer: a focus on tumor microenvironmental signals and chemoresistant breast cancers. *World J. Clin. Cases* 2, 769–786
86 Ivanov, A.A. (2013) Targeting protein-protein interactions as an anticancer strategy. *Trends Pharmacol. Sci.* 34, 393–400
87 Fishman, M.C. and Porter, J.A. (2005) Pharmaceuticals: a new grammar for drug discovery. *Nature* 437, 491–493
88 Ruffner, H. *et al.* (2007) Human protein-protein interaction networks and the value for drug discovery. *Drug Discov. Today* 12, 709–716
89 Chuang, H.Y. *et al.* (2007) Network-based classification of breast cancer metastasis. *Mol. Syst. Biol.* 3, 140
90 Johnson, H. and White, F.M. (2014) Quantitative analysis of signaling networks across differentially embedded tumors highlights interpatient heterogeneity in human glioblastoma. *J. Proteome Res.* 13, 4581–4593
91 Ein-Dor, L. *et al.* (2006) Thousands of samples are needed to generate a robust gene list for predicting outcome in cancer. *Proc. Natl. Acad. Sci. U. S. A.* 103, 5923–5928
92 Vidal, M. *et al.* (2011) Interactome networks and human disease. *Cell* 144, 986–998
93 Rix, U. *et al.* (2007) Chemical proteomic profiles of the BCR-ABL inhibitors imatinib, nilotinib, and dasatinib reveal novel kinase and nonkinase targets. *Blood* 110, 4055–4063
94 Médard, G. *et al.* (2015) Optimized chemical proteomics assay for kinase inhibitor profiling. *J. Proteome Res.* 14, 1574–1586
95 Duncan, J.S. *et al.* (2012) Dynamic reprogramming of the kinome in response to targeted MEK inhibition in triple-negative breast cancer. *Cell* 149, 307–321
96 Sharma, K. *et al.* (2012) Quantitative proteomics reveals that Hsp90 inhibition preferentially targets kinases and the DNA damage response. *Mol. Cell. Proteomics* 11 M111.014654
97 Wu, Z. *et al.* (2012) Systematic identification of the HSP90 candidate regulated proteome. *Mol. Cell. Proteomics* 11 M111.016675
98 Andersen, J.N. *et al.* (2010) Pathway-based identification of biomarkers for targeted therapeutics: personalized oncology with PI3K pathway inhibitors. *Sci. Transl. Med.* 2 43ra55
99 Kolch, W. and Pitt, A. (2010) Functional proteomics to dissect tyrosine kinase signalling pathways in cancer. *Nat. Rev. Cancer* 10, 618–629
100 Picotti, P. and Aebersold, R. (2012) Selected reaction monitoring-based proteomics: workflows, potential pitfalls and future directions. *Nat. Methods* 9, 555–566
101 Khirstenko, N. and Domon, B. (2015) Quantification of proteins in urine samples using targeted mass spectrometry methods. *Methods Mol. Biol.* 1243, 207–220
102 Iwai, L.K. *et al.* (2013) Phosphoproteomics of collagen receptor networks reveals SHP-2 phosphorylation downstream of wild-type DDR2 and its lung cancer mutants. *Biochem. J.* 454, 501–513
103 Hüttenhain, R. *et al.* (2012) Reproducible quantification of cancer-associated proteins in body fluids using targeted proteomics. *Sci. Transl. Med.* 4, 142ra94
104 Cima, I. *et al.* (2011) Cancer genetic-guided discovery of serum biomarker signatures for diagnosis and prognosis of prostate cancers. *Proc. Natl. Acad. Sci. U. S. A.* 108, 3342–3347
105 Sjöström, M. *et al.* (2015) A combined shotgun and targeted mass spectrometry strategy for breast cancer biomarker discovery. *J. Proteome Res.* 14, 2807–2818

106 Martínez-Aguilar, J. *et al.* (2015) A multiplexed, targeted mass spectrometry assay of the S100 protein family uncovers the isoform-specific expression in thyroid tumours. *BMC Cancer* 15, 199
107 Kim, Y.J. *et al.* (2015) Verification of the biomarker candidates for non-small-cell lung cancer using targeted proteomics approach. *J. Proteome Res.* 14, 1412–1419
108 Meric-Bernstam, F. *et al.* (2014) Influence of biospecimen variables on proteomic biomarkers in breast cancer. *Clin. Cancer Res.* 20, 3870–3883
109 Sallam, R.M. (2015) Proteomics in cancer biomarkers discovery: challenges and applications. *Dis. Markers* 321370, 1–12
110 Abbatiello, S.E. *et al.* (2015) Large-scale inter-laboratory study to develop, analytically validate and apply highly multiplexed, quantitative peptide assays to measure cancer-relevant proteins in plasma. *Mol. Cell Proteomics* Published online February 18, 2015. http://dx.doi.org/10.1074/mcp.M114.047050
111 Collins, F.S. and Varmus, H. (2015) A new initiative on precision medicine. *N. Engl. J. Med.* 372, 793–795