Advances in Mass Spectrometry-Based Proteomics and Its Application in Cancer Research

4

Markus Hardt

Abstract

With the advent of high-resolution/high mass accuracy instrumentation, sophisticated informatic approaches, and advances in liquid chromatography, mass spectrometry-based proteomics has emerged as an indispensable and widely used tool for the identification, characterization, and quantification of proteins on a large scale. Deep proteome analyses can now sequence over 14,000 protein isoforms for a single human cell line rivaling the depth of next-generation RNA sequencing technology. Without additional enrichment steps, highly sensitive MS-based proteomic studies yield comprehensive identification of major post-translational modifications (PTMs). Isotopic labeling techniques enable the comparison of multiple samples in a single mass spectrometry experiment, while data-independent acquisition strategies provide comprehensive protein coverage and quantification against complex backgrounds.

Keywords

Mass spectrometry · Quantitative proteomics · Post-translational modifications · Cancer pathways · Isotopic labeling

Proteins are the essential mediators of cellular function: Their biological activities and interactions catalyze biochemical reactions and thereby facilitate physiological and pathological processes. Characterization of the functional state of proteins as well as measuring changes in protein abundances reveals fundamental insights into

M. Hardt (✉)
Center for Salivary Diagnostics, The Forsyth Institute, Cambridge, MA, USA

Harvard School of Dental Medicine, Boston, MA, USA
e-mail: mhardt@forsyth.org

these processes and ultimately deepens our understanding of the underlying biochemistry and molecular biology. While gene expression analysis by real-time PCR or via transcriptome sequencing provides valuable insights into biological pathways, it can only infer protein abundance information. There is a growing consensus that correlation between mRNA and protein levels is in general modest (Gygi et al. 1999b; Schwanhäusser et al. 2011; Skelly et al. 2013; Lundberg et al. 2010). The protein phenotype appears to be buffered against transcriptional variation (Fu et al. 2009) Correlations of transcripts and proteins depend on cellular location and biological function (Conrads et al. 2005) and are controlled by tissue-specific posttranscriptional regulation (Franks et al. 2017). Therefore, direct measurements of proteins are preferable since they will more accurately reflect cellular status and provide insights into the molecular mechanisms that underlie physiological and pathological processes. Mass spectrometry-based proteomics has emerged as the method of choice for the identification, characterization, and quantification of proteins (Picotti et al. 2013; Aebersold and Mann 2016). Protein identification and characterization is critical to identify alternatively spliced proteins, proteolytic processing, and post-translational modifications that alter the composition and functional status of proteins at the post-transcriptional level. It is estimated that the diversity of the roughly 20,300 protein-coding genes is increased to over 500,000 proteoforms by alternative splicing and post-translational modifications (phosphorylation, glycosylation, proteolytic truncations) (Smith et al. 2013).

The ability to identify proteins at a large scale has been primarily driven by the advances in mass spectrometric instrumentation, informatic workflows, and separation of complex protein mixtures. Liquid chromatography (LC) and two-dimensional polyacrylamide gel electrophoresis (2D-PAGE) are two of the most commonly applied separation techniques prior to mass spectrometric analysis. Quantitative proteomics has developed into an indispensable tool for cancer research to analyze disease-related tissues and body fluids in order to identify proteins, protein post-translational modifications, or protein complexes that can be used to detect the disease early, prognose disease outcome, and monitor response to therapeutic intervention and for the elucidation of molecular mechanisms for the development of novel therapeutics. Oncoproteomics has been extensively reviewed, from proteomic studies of tumor tissue and cancer cell lines to profiling of plasma and other body fluids for cancer biomarkers (Huang et al. 2017; Belczacka et al. 2018; Tan et al. 2012; Cantor et al. 2015; Veenstra 2013; Faria et al. 2017). Here, we highlight the most promising quantitative proteomics approaches in the context of studying cancer signaling pathways.

4.1   Differential Analysis by 2D-PAGE

In 2D-PAGE, proteins are initially resolved by isoelectric focusing followed by a separation based on molecular mass. After protein staining, specialized image analysis software is used to identify differentially expressed protein spots. Spots

of interest are excised, and proteins are in-gel digested with exogenous proteases (i.e., trypsin). The resulting peptides are recovered and their molecular masses measured by mass spectrometry. In the early stages of proteomics, subsequent peptide identification was performed by peptide mass fingerprinting (PMF) using matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS), in which peptide masses were matched against theoretically predicted peptide masses in a database of candidate proteins (Monteoliva and Albar 2004). Nowadays, with the ubiquity of LC-MS/MS instrumentation, protein identification is typically performed in higher throughput and with higher accuracy through the matching of peptide fragmentation data obtained from tandem mass spectrometry experiments (MS/MS) to theoretically predicted peptide fragment masses. 2D-PAGE is a classical proteomics workflow that provides a straightforward visual and quantitative comparison of differences in protein composition. It has been extensively employed in cancer research, including for the detection of tumor-associated proteins in colorectal cancer tissue samples (Wang et al. 2007; Stulík et al. 1999; Xing et al. 2006) and in the combination with laser capture microdissection (LCM) (Shi et al. 2011). However, the application of 2D-PAGE has been declining in recent years due to its limitations in throughput and reproducibility, sensitivity, dynamic range, and its laborious nature (Belczacka et al. 2018). The detection limits of the most commonly used stains range from 500 ng/mm^2 (colloidal Coomassie Brilliant Blue) to 0.1 ng/mm^2 (silver stain and fluorescent dyes). Some of these drawbacks can be overcome by the usage of narrow-range IPG strips to increase the resolving power in the initial isoelectric focusing dimension. To increase reproducibility and improve comparative analyses, difference gel electrophoresis (DIGE) was developed, a form of multiplexed 2D-PAGE where up to three different protein samples are fluorescently labeled prior to gel separation. DIGE-based differential proteomics analysis has been successfully used in the discovery phase of cancer biomarker studies (colorectal, prostate cancer) when the proximal tissue samples are being analyzed at greater depth before validation of potential markers by ELISA in serum (Hamelin et al. 2011; Pang et al. 2010).

A particular strength of 2D-PAGE is the ability to detect and visualize proteoforms – the different molecular structures that the protein products of a single gen can assume due to genetic variations, alternatively spliced RNA transcripts, and post-translational modifications (PTMs) (Smith et al. 2013). PTMs including proteolytic processing, deamidation, glycosylation, acetylation, alkylation, cysteine oxidation, tyrosine nitration, and phosphorylation regulate many cellular signaling pathways. PTMs alter the molecular mass and/or the isoelectric point of the protein. For example, different phosphorylation states of a protein are observable as horizontal spot trains in a 2D gel, whereas glycosylation can alter both the pI and the molecular weight of proteins resulting in clusters shifted both horizontally and vertically (Löster and Kannicht 2008). ProMoST (http://proteomics.mcw.edu/promost.html) is a webtool that can be used to calculate gel shifts introduced by PTMs to facilitate more detailed analyses (Halligan et al. 2004).

4.2 Top-Down Proteomics

Ideally, intact proteins would be analyzed directly by mass spectrometry without the need for proteolytic digestion; protein identification, in turn, would be achieved by MS/MS fragmentation of the whole protein. As a liquid phase-alternative to 2D-PAGE, "top-down" proteomics has made substantial progress in the last decade, and it is now feasible to measure over 3000 proteoforms using a four-dimensional LC separation system that is integrated with high-resolution electrospray mass spectrometry analysis (Tran et al. 2011). However, similar to 2D-PAGE, the high complexity and dynamic range of protein concentrations encountered in proteome research currently limit the applicability of the top-down approach to large-scale discovery analyses. Nonetheless, "native mass spectrometry" experiments in which biological analytes are ionized by electrospray from nondenaturing solvents to preserve noncovalent interactions in the gas phase, have been used to analyze specific macromolecular assemblies including protein-protein and protein-ligand complexes (Hernández and Robinson 2007; Zhou et al. 2011; Leney and Heck 2017).

4.3 Bottom-Up (Shotgun) Proteomics

Instead of top-down intact protein analysis, "bottom-up" proteomics has been the more practical approach and has been widely adapted in the field (Aebersold and Mann 2016). In the bottom-up strategy, peptides are generated by enzymatic digestion of proteins with sequence-specific, exogenous proteases such as trypsin. The resulting peptides are separated by reversed-phase liquid chromatography (LC) and injected into the hyphenated tandem mass spectrometer. Peptides are isolated in the gas phase and subjected to fragmentation, thereby generating tandem mass spectra (MS/MS; MS²). Collision-induced dissociation (CID) and higher-energy collisional dissociation (HCD) are two of the most commonly used fragmentation techniques to generate sequence information for peptide identification. Electron-transfer dissociation (ETD) and electron-capture dissociation (ECD) can be useful alternative strategies for the identification of larger and post-translationally modified peptides. Post-translational modifications (PTMs) such as phosphorylation and glycosylation are labile and readily lost over peptide backbone fragmentation (Mikesh et al. 2006; Syka et al. 2004; Zubarev et al. 2000). The resulting MS/MS fragmentation data are submitted to database search engines (i.e., MASCOT (Perkins et al. 1999), SEQUEST (Eng et al. 1994), X!Tandem (Craig and Beavis 2004), MyriMatch (Tabb et al. 2007), and OMSSA (Geer et al. 2004)) for protein/peptide identification. These search engines match and score the empirically acquired spectra against theoretically predicted fragmentation patterns of peptides derived from in silico digestions of proteins stored in protein sequence databases (Nesvizhskii 2010; Eng et al. 2011). Alternatively, MS/MS spectra can be matched via correlation analysis to previously observed and identified spectra using spectral library search engines such as SpectraST (Lam et al. 2007), X!Hunter (Craig et al. 2006), and BiblioSpec (Frewen et al. 2006). Though spectral library searching is typically considered to be

a more sensitive approach than sequence database searches, its adaption in the field has been fairly limited (Deutsch et al. 2018). PeptideAtlas (Desiere et al. 2004), the Global Proteome Machine Database (Craig et al. 2006), and the MassIVE Knowledge Base (Wang et al. 2018) are efforts to leverage the large number of peptide identifications contained in public proteomics datasets to create spectral library resources that can support future proteomics experiments.

Peptide sequences can also be derived from MS/MS fragmentation data by de novo sequencing approaches using algorithms including PEAKS (Ma et al. 2003), PepNovo (Frank and Pevzner 2005), Novor (Ma 2015), and Lutefisk (Taylor and Johnson 2001) that do not rely on reference databases (Allmer 2011). De novo sequencing frameworks designed for top-down proteomics can be advantageous in the analysis of high-resolution bottom-up MS/MS datasets (Vyatkina et al. 2017).

Combining the results of multiple search engines with tools such as iProphet (Shteynberg et al. 2011) can improve the confidence of peptide-spectrum matches and increase the overall number of distinct peptides and proteins identified since each search engine has its own specific strengths which can be complementary to others (Shteynberg et al. 2013).

Currently, there are two major data acquisition strategies used in bottom-up proteomics: The preferred method for proteome discovery is data-dependent acquisition (DDA), which aims to maximize the number of protein and peptide identifications per experiment to achieve comprehensive proteome coverage. Hallmarks of this approach include the 1-h yeast proteome (Hebert et al. 2013b) and draft maps of the human proteome with coverages of up to 92% of the protein-coding sequences (Wilhelm et al. 2014; Kim et al. 2014). To achieve this level of proteome coverage, additional fractionation techniques (strong anion exchange; off-gel electrophoresis) were employed to distribute sample complexity across additional data acquisitions. Applied on single cell lines (i.e., HeLa human cervical carcinoma), over 10,255 proteoforms stemming from 9205 genes can be identified by deep proteomics analysis (Nagaraj et al. 2011). Proteomics analyses of a panel of 11 commonly studied cell lines (Geiger et al. 2012) and the NCI-60 panel of 59 cancer cell lines (Gholami et al. 2013) suggests that at least ~10,000 proteins are about the average proteome coverage of a human cell line. A more recent study showed that adding an off-line high pH peptide fractionation step prior to low pH LC-MS/MS analysis can deepen the protein coverage even further to over 12,000 proteins for HeLa cells (Bekker-Jensen et al. 2017). A key strength of the described DDA methods is the fact that no a priori knowledge about the identity of the expected proteins is required and therefore unanticipated proteins and PTMs can be discovered, potentially providing new biological understanding.

Data-independent acquisition (DIA) also referred to as SWATH (Sequential Windowed data-independent Acquisition of Total High-resolution) is a more recently developed methodology that aims to obtain complete fragment ion coverage across samples (Ludwig et al. 2018). In DDA experiments, a full precursor ion spectrum of all co-eluting peptides is acquired at the MS^1 level, after which as many as possible precursor peptides are isolated, fragmented, and MS^2 spectra acquired within the instrument cycle time. In DIA experiments by contrast, predetermined

windows of m/z values are sequentially isolated for fragmentation (Gillet et al. 2012). In each instrument cycle, the entire precursor ion m/z range gets fragmented, resulting in highly multiplexed fragment ion spectra. Precursor-fragment ion relationships can be reconstructed with bioinformatic tools such as DIA-Umpire (Tsou et al. 2015, 2016) to create "pseudo"-spectra that are conventionally searched against protein databases to create internal spectral libraries that contain peptide identifications. These internal spectral libraries or external spectral libraries built from DDA data are then used to perform targeted extraction (Röst et al. 2014). Key advantage of the DIA approach is its unbiased nature: All precursor and all fragment ions are acquired all the time without losing low abundant ions; the identities of quantified peptides do not need to be specified a priori, which is ideal when the data is acquired over the course of a multi-year study. DIA measurements comprise an archival record of the sample content that can be re-interrogated when new proteins, proteoforms, or post-translational modifications sites of interest emerge.

4.4   Relative Quantitation in Bottom-Up Proteomics

In bottom-up proteomics, quantitation is achieved by either label-free or stable isotope labeling methods (Bantscheff et al. 2012). Stable isotope-based methods are the gold standard for quantification; however they require metabolic labeling or an additional chemical labeling step during sample preparation. Label-free approaches are simpler and more economical, providing relative quantitation for an unlimited number of samples (including clinical specimens) and can be based on either DDA or DIA datasets (Nahnsen et al. 2013). State-of-the-art mass spectrometers provide the necessary high mass resolution and high mass accuracy that are required for the accurate extraction of ion chromatograms (XICs; elution profiles) of precursor ions at the MS^1 level that are used to determine peptide quantities. In the past, when bottom-up proteomics was mostly performed on low-resolution ion trap instruments, the number of identified MS/MS spectra for a given peptide (spectral counts) was used as a surrogate measurement for peptide abundance (Ishihama et al. 2005). While the spectral count approach has been used to create one of the drafts of the human proteome (Kim et al. 2014), XIC-based approaches are now the most commonly employed label-free methodology due to their superior sensitivity. By aligning the retention times of XIC areas and propagating MS/MS-based peptide identifications across data acquisitions ("matching between runs"), the overall number of detectable peptides between samples can be boosted which leads to more comprehensive comparative analyses (Bateman et al. 2013). Numerous academic and commercial proteomics data analysis packages including PEAKS (Ma et al. 2003) and Scaffold (Searle 2010) offer label-free quantitative workflows in addition to their identification pipelines (Nahnsen et al. 2013; Mueller et al. 2008). Particularly noteworthy is the continuously expanding proteomics software tool suite under the MaxQuant umbrella which is freely available and has become one of the most widely used proteomics data analysis platforms. MaxQuant incorporates the peptide database search engine Andromeda (Cox et al. 2011) and the MaxLFQ

workflow for label-free quantitation (Cox et al. 2014) and supports as well other MS[1]- and MS[2]-level (isobaric) labeling approaches (Tyanova et al. 2016).

In contrast to the stochastic precursor ion selection in DDA, DIA systemically parallelizes the fragmentation of all detectable ions, thereby minimizing selection bias, which in turn results in improved dynamic range and sensitivity. Specific peptides can be identified and quantified by applying targeted extraction of either MS[1] precursor or MS[2] fragment ion intensities using spectral library-based OpenSWATH (Röst et al. 2014), Skyline (Maclean et al. 2010), or commercial software (PeakView SWATH 2.0, SCIEX; Spectronaut, Biognosys). The performance of these "peptide-centric" query tools in terms of identification precision, robustness, and specificity has been benchmarked against reference datasets and compared to the "data-centric" DIA-Umpire approach (Tsou et al. 2015) that does not rely on existing assay libraries (Navarro et al. 2016). Targeted extraction relies on the generation of sample-specific assay libraries that contain precursor and fragment ion m/z values, normalized retention times, and relative ion intensities of targeted peptides. Retention times are typically normalized using a set of reference peptides (Escher et al. 2012). DIA studies often rely on sample-specific libraries that are acquired on the same instrument in DDA mode prior to the DIA analysis (Gillet et al. 2012; Röst et al. 2014; Hüttenhain et al. 2013). Alternatively, repositories of assay libraries for human proteins have been created that are optimized for specific MS instruments. These resources contribute to simplified and reproducible targeted SWATH/DIA analysis across laboratories (Rosenberger et al. 2014). A multi-laboratory evaluation study across 11 sites demonstrated that SWATH acquisitions are capable of reproducibly detecting and quantifying a large-scale protein set (Collins et al. 2017).

4.5 Multiplexed Quantitation Using Stable Isotope Labeling Methods

The analysis of cancer signaling networks requires the ability to quantify proteins across multiple conditions so that temporal dynamics can be captured. A broad variety of chemical and metabolic stable isotope labeling methods have been developed that allow for multiplexing (Gevaert et al. 2008). Stable isotope labeling strategies can provide relative and absolute quantitation; however, the specifics of the labeling reactions can limit the number of samples that can be interrogated in contrast to label-free approaches. Isotope-coded affinity tags (ICAT) are one of the first stable isotope chemical labeling reagents that became widely adapted in proteomics (Gygi et al. 1999a). ICAT reagents are comprised of a reactive group specific toward cysteinyl residues, a stable isotope label (heavy/light), and a biotin affinity tag for selective enrichment to reduce sample complexity. ICAT allows for the duplex analysis for comparison of protein levels across two biological states. The exclusive reliance of ICAT on cysteine-containing peptides limits its general applicability as quantitation approach, and it has been mostly replaced by a new generation of isobaric labeling strategies based on *N*-hydroxysuccinimide (NHS) chemistry. The TMT (tandem mass tag) (Thompson et al. 2003) and iTRAQ (isobaric tags for relative and absolute

quantitation) (Ross et al. 2004) labels share isobaric stable isotope moieties as design features, which render differentially labeled samples "silent" – indistinguishable during chromatographic separation and in precursor MS[1] acquisition. Only upon MS/MS fragmentation the low molecular weight reporter ions are released, and their relative ion abundances are used for quantitation. Currently, there are up to eight reporter ions available for iTRAQ (Choe et al. 2007) and up to ten for TMT (Erickson et al. 2017), each allowing for multiplexed analysis in single LC-MS/MS experiments. For projects entailing larger sample numbers, one of the isotope channels is typically used for a control reference mixture.

The dynamic range of isobaric multiplex quantitation methodologies can be limited by isotopic contamination, background interference, low signal-to-noise ratio, and ratio compression (Ow et al. 2009; Karp et al. 2010). Applying an additional isolation and fragmentation event (MS[3] scan) (Ting et al. 2011) and gas-phase purification through proton transfer ion-ion reactions (Wenger et al. 2011) has been shown to eliminate interferences. Co-isolating and co-fragmenting of multiple MS[2] fragments (MultiNotch MS[3]) can boost sensitivity and improve the dynamic range of the isobaric tagging approach (Mcalister et al. 2014).

Dimethyl labeling using different isotopomers of formaldehyde provides a more economical triplex stable isotope quantitation method at the peptide level (Boersema et al. 2008). Chemical isotope labels are typically introduced late in the sample preparation process, which makes these labeling strategies broadly applicable; however, at the same time, they are more susceptible to variability introduced during processing.

SILAC (stable isotope labeling by amino acids) is a metabolic labeling method alternative to chemical isotope tags (Mann 2006). SILAC relies on the in vitro incorporation of essential amino acids that feature substituted stable isotope nuclei (e.g., Arg or Lys labeled with ^{13}C or ^{15}N). SILAC labeling is insensitive to variability introduced at the sample processing and analysis stage since all sample handling issues affect all proteins and peptides equally. SILAC and ^{15}N metabolic labeling has been used for comparative proteomics analysis in cell culture systems (Ong et al. 2002; Everley et al. 2004, 2006) and model organisms including yeast (de Godoy et al. 2008), *C. elegans* and *D. melanogaster* (Sury et al. 2010), and rodents (Kruger et al. 2008; Wu et al. 2004). Full incorporation into the entire organisms requires feeding more than one generation exclusively with the essential, stable isotopically labeled lysine amino acids. A comprehensive analysis employing triple SILAC-based proteomics (using Arg0, Lys0; Arg6-L-$^{13}C_6$ and Lys4-L-$^2H_4$; Arg10-L-$^{13}C_6^{15}N_4$ and Lys8-L-$^{13}C_6^{15}N_2$), RNA-seq-based transcriptomic profiling, and antibody-based confocal microscopy revealed that three functionally different human cancer cell lines shared expression levels for more than half of their expressed genes, while close to 20% were substantially altered (Lundberg et al. 2010).

In the super-SILAC method, lysates from multiple SILAC-labeled cancer cell lines are combined to serve as internal, isotopically labeled peptide standards to measure fold change ratios between human tumor proteomes (Geiger et al. 2010). By combining SILAC and TMT labeling in the same experiment, a strategy termed

"hyperplexing," it is possible to extend the number of samples that can be quantified in the same LC-MS run (Dephoure and Gygi 2012).

The advent of mass spectrometers capable of ultra-high mass resolution (>200,000) made it possible to reveal the small mass differences (milliDaltons) introduced by the differences in the neuron-binding energetics of isotopes such as ^{2}H (+ 1.0062), ^{13}C (+ 1.0034), and ^{15}N (+ 0.997). The neuron encoding (NeuCode) method (available as amine-reactive labels and SILAC reagents) takes advantage of the ability to embed these mass defect-based neutron signatures into isotopologues. At standard resolution, these isotopologues are concealed during MS^1 and MS/MS analysis and therefore do not increase spectral complexity (Hebert et al. 2013a). The multiplexed quantitative information is only revealed at high-resolution scans. NeuCode is applicable to DDA (Overmyer et al. 2018) and DIA approaches (Minogue et al. 2015) as well as targeted proteomics (Potts et al. 2016) and top-down applications (Rhoads et al. 2014; Shortreed et al. 2016).

4.6   Quantitation by Targeted Proteomics

Targeted proteomics provides accurate and quantitative measurements of protein abundances and thereby enables hypothesis-driven research using mass spectrometry (Picotti et al. 2013). In contrast to DDA- and DIA-based proteomics analyses, the identities of the proteins of interest are known a priori in targeted proteomics experiments. For any given protein, peptides are selected that are "proteotypic," meaning that each peptide has a unique sequence, is readily detected by MS, and has been repeatedly and consistently identified in previous studies (Mallick et al. 2007). By selectively subjecting these proteotypic peptides to precursor ion isolation and continuous fragmentation, characteristic fragment (product) ion abundances for the most intense transitions can be recorded over the chromatographic elution profile, and this information is then used to estimate relative protein abundances. These types of experiments are typically performed on triple quadrupole instruments operating in multiple reaction monitoring (MRM) mode, which is also referred to as selected reaction monitoring (SRM). To increase specificity, typically multiple product ions are measured. Absolute protein abundances can be determined by using spike-in, isotopically labeled reference peptides (Gerber et al. 2003) or mTRAQ chemically labeled standards (Desouza et al. 2008) or in label-free format when anchor proteins are used to create a quantitation model (Ludwig et al. 2011). An efficient method to define custom MRM assay conditions in high-throughput format is through the usage of crude synthetic peptide libraries (Picotti et al. 2010). To achieve proteome-wide coverage for absolute protein quantification, an in vitro protein expression system has been used to synthesize over 18,000 recombinant proteins from full-length human cDNA libraries, which were then digested and labeled with mTRAQ (Matsumoto et al. 2017). Alternatively, ProteomeTools is a brute force project to create a resource comprised of the

comprehensive LC-MS analysis of over 1.4 synthetic million peptides that cover tryptic and non-tryptic peptides representative of the canonical human proteome, as well as additional peptides covering splicing variants, post-translational modifications, and other sequences representing interesting biology such as disease-associated mutations (Zolg et al. 2017).

Compared to shotgun proteomics approaches, MRM assays provide higher sensitivity, specificity, and a broad dynamic range. Once established, individual MRM assays can be multiplexed at the peptide level (Picotti and Aebersold 2012). Measurements have been shown to be highly reproducible across laboratory sites (Addona et al. 2009). SRMAtlas (www.srmatlas.org) and PASSEL (www.peptideatlas.org/passel) both host freely accessible proteome-wide assay libraries along with empirical performance data that facilitate the design of targeted MRM assays (Farrah et al. 2012; Kusebauch et al. 2014, 2016). MRM assays for 1157 cancer-associated proteins have been developed, of which 182 were detected in depleted plasma and 408 in urine across a cohort of cancer patients and healthy controls using a label-free MRM strategy (Hüttenhain et al. 2012).

By combining peptide immunoaffinity enrichment with stable isotope-labeled standards and MRM-MS, it is possible to create automated, multiplexed assays with sufficient sensitivity to quantify low-abundance target proteins in plasma as an alternative to traditional enzyme-linked immunosorbent assay (ELISA)-based testing (Whiteaker et al. 2010).

The advent of high-resolution/accurate mass (HRAM) instrumentation has enabled the development of the parallel reaction monitoring method (PRM), in which the monitoring of a single product ion in an MRM assay is substituted with the parallel detection of all target product ions in a high-resolution MS/MS analysis (Peterson et al. 2012; Bourmaud et al. 2016). While MRM and PRM provide the best quantitation performance, both are throughput limited in terms of how many proteins can be quantified in a single MS experiment. SWATH/DIA provides a compelling alternative for reproducible quantitation in which a targeted data analysis strategy is employed to extract specific fragment ion abundances out of the comprehensive fragment ion map provided by the DIA dataset. Similar to MRM/PRM, reference libraries containing SWATH assay conditions can be built (Schubert et al. 2015) and shared via repositories (Rosenberger et al. 2014). SWATH/DIA assays have been shown to perform well across multiple laboratory sites (Collins et al. 2017). Additional throughput can be achieved for targeted proteomics assays when multiplexing is extended to the sample level by utilizing isobaric labels. In the TOMAHAQ method, synthetic TMT0-labeled spiked-in peptides trigger the MultiNotch MS^3 acquisition of co-eluting TMT10-labeled endogenous peptides, which allowed for the quantitation of 69 target proteins across 180 cancer cells within 48 h (Erickson et al. 2017). The setup and data analysis for this approach have been simplified by the recent development of the TomahaqCompanion tool (Rose et al. 2018).

By carefully selecting protein targets based on their involvement in particular biochemical pathways, it is possible to quantitatively investigate the response of cellular systems to external stimulation (Matsumoto and Nakayama 2018). The

multiplex MRM approach has been used to study the protein expression in major metabolic energy pathways of breast cancer cells in response to hypoxia, glucose deprivation, and estradiol stimulation (Drabovich et al. 2012; Murphy and Pinto 2010). Leveraging their in vitro proteome-assisted MRM assay library (iMPAQT) that covers over 18,000 proteins, Matsumoto et al. were able to explore the global impact of oncogenic transformation on fibroblasts (2017). Alternatively, by integrating detailed information about biological processes on the basis of literature evidence and computational predictions, it is possible to carefully select protein quantitation targets that can serve as sentinels or proxies for system responses (Soste et al. 2014).

4.7   Characterization of Post-translational Modifications

With continued improvements in mass accuracy, resolution, and sensitivity of mass spectrometry instruments, proteomic expression analyses feature deeper proteome and higher protein sequence coverages that enable more exhaustive characterizations of post-translational modifications (PTMs). PTMs including phosphorylation, glycosylation, and ubiquitination are important modulators of protein function: For example, most proteolytic enzymes are activated from their inactive precursor (zymogen) state by proteolytic cleavage (Klein et al. 2017). Many phosphorylations lead to protein conformational changes that modulate protein activity, i.e., protein binding. Ubiquitination marks proteins for degradation. Glycosylation often regulates protein function and enzymatic activities, alters protein-protein interactions, and changes the subcellular localization of numerous proteins. In mass spectrometric analyses, most PTMs lead to characteristic mass shifts in MS[1] spectra, and their location on specific amino acid residues can be determined by fragmentation analysis. However, the combinatorial nature of post-translational modifications creates a heterogeneity that constitutes a formidable analytical challenge as the vast structural diversity that can be generated via oligomerization and branching of glycans (complex carbohydrates) illustrates (Laine 1994). Hundreds of protein modification kinds (biological and artificial) have been reported in the Unimod (Creasy and Cottrell 2004) and RESID (Garavelli 2004) databases. The most actively studied post-translational modifications include phosphorylation, methylation, ubiquitination, methylation, acetylation, and O-GlcNAcylation (Doll and Burlingame 2015). Together, over 260,000 PTM sites have been identified in the human proteome so far (Doll and Burlingame 2015). Comprehensive information on empirically observed in vivo and in vitro post-translational modifications can be found in online bioinformatic resources including PhosphoSitePlus (PSP) (www.phosphosite.org), iPTMnet (https://research.bioinformatics.udel.edu/iptmnet/), and Phospho.ELM (http://phospho.elm.eu.org/) along with additional tools useful for PTM analysis (Hornbeck et al. 2012; Huang et al. 2018; Dinkel et al. 2011).

4.8   Phosphorylation

Protein phosphorylation is one of the central means by which cells transiently modulate protein function as exemplified by signal transduction pathways. The localization, the extent of phosphorylation, and the site-specific occupancy or stoichiometry are important determinants of protein functional modulation. Phosphorylation states are mediated by a network of kinases that phosphorylate serine, threonine, and tyrosine residues and phosphatases that remove phosphorylations. Deregulated kinase activities have been associated with the ability of cancer cells to circumvent physiological constraints on cell proliferation. Kinase inhibition (i.e., of the serine/threonine kinase mammalian target of rapamycin (mTOR)) has emerged as one of the most heavily pursued classes of drug targets in oncology (Dowling et al. 2010). With over 518 genes identified, protein kinases are one of the largest protein families in eukaryotes (Manning et al. 2002). It is estimated that a typical eukaryotic cell harbors between 700,000 and 1000,000 potential phosphorylation sites (Ubersax and Ferrell 2007; Boersema et al. 2010). Analysis of 50,000 phosphopeptides in HeLa S3 cancer cells revealed that at least three-quarters of the 11,000 identified proteins were phosphorylated (Sharma et al. 2014). Interestingly, the 150 most abundant phosphopeptides accounted for 20% of the cumulative phosphopeptide signal (Sharma et al. 2014). Phosphoproteomics analysis of nine mouse tissues (12,000 proteins; ~36,000 phosphorylation sites) revealed that most phosphoproteins are widely expressed but display tissue-specific phosphorylation to adapt to tissue function (Huttlin et al. 2010).

Phosphotyrosine accounts for only 1% of phosphorylations, owing to its primary regulatory and not structural role in proteins and a short half-life due to the presence of highly active phosphotyrosine phosphatases (Sharma et al. 2014). Many phosphoproteins such as transcription factors and protein kinases have low copy numbers. Combined with the substoichiometric levels observed for many regulatory protein phosphorylations, enrichment strategies are necessary to comprehensively profile protein phosphorylations (Macek et al. 2009). Enrichment can be performed at the phosphoprotein level prior to digestion using immobilized metal affinity chromatography (IMAC) (Collins et al. 2005) or after digestion using phosphopeptide enrichment by metal oxide affinity chromatography (e.g., using titanium dioxide (TiO2)) or IMAC. In the case of phosphotyrosine, immunoaffinity purification using phosphotyrosine-specific antibodies is preferred (Boersema et al. 2010; Kettenbach and Gerber 2011; Rush et al. 2005; Breitkopf and Asara 2012). Mass spectrometric characterization of phosphopeptides is challenging due to their overall low abundance, susceptibility to ion suppression, and limited fragmentation patterns (Dreier et al. 2018). Phosphopeptide-selective mass spectrometric detection methods include precursor ion and neutral loss scanning based on the diagnostic PO_3^- and H_3PO_4 ion losses that are caused by the lability of the O-phosphate bond in collision-induced dissociation (Le Blanc et al. 2003; Carr et al. 2005). Compared to pSer and pThr, phosphorylations of tyrosine (pTyr) are relatively stable and remain attached to MS/MS fragments, which facilitates their analysis. Also, pTyr yields characteristic immonium ions that can be used as an alternative means to

identify phosphorylation sites (Steen et al. 2003). In ion trap instruments, detection of neutral losses can be used to trigger the acquisition of MS³ spectra in which the neutral loss precursor ion undergoes an additional round of isolation and fragmentation to yield better fragmentation coverage (Gruhler et al. 2005). Peptide fragmentation by ETD or ECD yields more extensive peptide backbone cleavages without shedding the labile phosphate groups first which, in turn, also facilitates phosphopeptide identification (Chi et al. 2007; Stensballe et al. 2000). Large-scale, quantitative phosphoproteomics has been used to define the downstream signaling networks of mTOR, identifying Grb10 as a potential mTORC1-regulated tumor suppressor (Hsu et al. 2011; Yu et al. 2011). The dynamic nature of the phosphoproteome mandates the acquisition of temporal profiles of the in vivo phosphoproteome to capture the cellular response upon stimulation (Olsen et al. 2006). By streamlining conventional multi-step phosphoproteomics workflows into a simplified parallel 96-well plate format protocol, sufficient sample throughput is now achievable to perform global profiling of phosphorylation in a time-resolved fashion (Humphrey et al. 2015). The NCI Clinical Proteomics Tumor Analysis Consortium (CPTAC) recently provided an optimized, highly reproducible workflow for proteome/phosphoproteome analysis that utilizes TMT-10 for multiplexed quantitation of over 10,000 proteins in a breast cancer xenograft model (Mertins et al. 2018).

An inherent challenge in large-scale phosphoproteomics analyses is the fact that changes in phosphoprotein expression levels can interfere with the interpretation of site-specific phosphorylation stoichiometries (Wu et al. 2011). Measuring the degree of phosphorylation requires the quantification of the cognate phosphorylated and non-phosphorylated peptides. This can be accomplished by splitting samples into two and forcing dephosphorylation in one fraction by phosphatase treatment and leaving the other fraction untreated. After differential stable isotope labeling, the two fractions are combined, and the degree of phosphorylation can be estimated by comparing the intensities of the differentially labeled unphosphorylated peptides (Zhang et al. 2002; Hegeman et al. 2004). Alternatively, spike-ins of synthetic isotopologues of the phosphorylated/non-phosphorylated peptides in conjunction with targeted mass spectrometry (MRM or PRM) can be used for absolute quantification of site-specific phosphorylation stoichiometry (Dekker et al. 2018; Jin et al. 2010). By normalizing for total phosphoprotein amount using multiple unmodified peptides, it is possible to estimate the degree of phosphorylation by calculating the ratios of phosphorylated/unphosphorylated peptide intensities for phosphoproteins of interest without stable isotope labeling (Steen et al. 2005). For large-scale phosphoproteomics studies that rely on phosphopeptide enrichment, parallel proteomics analyses can provide the necessary information on total phosphoprotein abundances to determine phosphorylation site stoichiometries (Wu et al. 2011; Olsen et al. 2010). Typical signaling pathway analysis is performed by collapsing discrete site measurements to the protein level. The curated PTMsigDB database aims to leverage site-specific post-translational modification information to capture signaling events more accurately as demonstrated in the phosphoproteome analysis of PI3K-inhibited breast cancer cells (Krug et al. 2018).

4.9   Ubiquitination

The ubiquitin-proteasome pathway controls the degradation of 80–90% of intracellular proteins. Ubiquitination is a process by which one or multiple ubiquitin monomers are covalently attached to the amino group at the protein N-terminus or at lysine side chains of substrate proteins, thereby forming branched proteins. Eukaryotic ubiquitin consists of 76 amino acids and is evolutionary conserved. Ubiquitination is catalyzed by a ubiquitin-activating enzyme (E1), a ubiquitin-conjugating enzyme (E2), and a ubiquitin ligase (E3), which confers substrate specificity. De-ubiquitinating enzyme can reverse the ubiquitin conjugation, creating a steady state with poly-ubiquitinated proteins (n > 4) targeted for degradation by the 26S proteasome. As an important regulator of cell proliferation, differentiation, and survival, alterations of the ubiquitin ligase pathways have been linked to cancer (Ding et al. 2014; Mani and Gelmann 2005). Characterization of ubiquitination sites by mass spectrometry is commonly performed after antibody enrichment of peptides containing the Lys-GlyGly sequence that is formed during tryptic digestions of ubiquitinated proteins (Xu et al. 2010). More recently, an immunoaffinity strategy based on the recognition of the C-terminal 13 amino acids of ubiquitin has allowed for the identification of over 63,000 unique ubiquitination sites, including N-terminal ubiquitination, across 9200 proteins in 2 human cell lines (Akimov et al. 2018).

4.10   Proteogenomics

In an effort to elucidate how somatic gene mutations impact the cancer proteome and the post-translational modification landscape, CPTAC used quantitative MS and phosphoproteomics to characterize hundreds of ovarian, breast, and colon/rectal tumors whose genome and transcriptome were previously defined by The Cancer Genome Atlas (TCGA) (Mertins et al. 2016; Zhang et al. 2014, 2016). Integrating genomic and proteomics/phosphoproteomics measurements allowed to explore the effect of copy number alterations on protein abundance and test whether transcriptome-derived subtypes are reflected in protein expression patterns. Proteogenomics promises to deepen our understanding of cancer biology and identify alterations in cancer signaling pathways and potential therapeutic targets with higher levels of confidence. The human cancer proteome variation cancer database (CanProVar) provides a bridge between genomic and proteomics data by compiling protein sequence alterations in different types of cancers (Zhang et al. 2017; Li et al. 2010) along with extensive annotation, which can be used for the detection of variant peptides in shotgun and targeted proteomics experiments (Li et al. 2011).

4.11   Ultrasensitive Proteomics via Cellular Pre-fractionation

Given the microheterogeneity of the cancer microenvironment, it can be of advantage to analyze specific cell types individually in order to more accurately reveal their biochemical potentials. Cellular populations can be specifically purified by antibody-based methods such as fluorescence-activated cell sorting, CyTOF mass cytometry, or immune magnetic separation. CyTOF mass cytometry uses rare earth metals as unique antibody reporters that are monitored by inductively coupled plasma mass spectrometry (ICP-MS) in multiplex format to reveal marker expression in individual cells (Bandura et al. 2009). ICP-MS offers an extraordinary level of sensitivity which enables the detection of metal-labeled antibodies at levels corresponding to single cells. Alternatively, cellular subpopulations can be dissected from tissue using laser capture microdissection (LCM) prior to MS-based proteomics analysis (Altelaar and Heck 2012). In-depth LC-MS analysis of approximately 3000 LCM-derived tumor cells can yield the identification of 1000–2000 proteins (Umar et al. 2007; Wiśniewski et al. 2011), a number that can be boosted to over 4000 protein identifications from microdissected cells from formalin-fixed and paraffin-embedded human tissue specimens with the incorporation of additional off-line fractionation steps (Wiśniewski et al. 2011).

4.12   Imaging Mass Spectrometry

MALDI and secondary ion mass spectrometry (SIMS) imaging mass spectrometry (IMS) combine the parallel molecular detection by mass spectrometry with microscopic imaging to visualize the spatial distribution of proteins and metabolites (Cornett et al. 2007; Schwamborn and Caprioli 2010). MALDI-IMS yields 2D molecular maps that provide the localization and relative abundance of thousands of analytes in thin tissue sections with typical pixel size in the range of 50–200 μm in an untargeted manner (McDonnell and Heeren 2007; Schober et al. 2012). The discovery nature of MALDI imaging can be complemented by imaging mass cytometry, which utilizes the multiplexing capability of CyTOF mass cytometry for the targeted multiplexed localization of up to 32 proteins with subcellular resolution. This approach was pioneered to characterize tumor cell subpopulations and highlight the heterogeneity of human breast cancer microenvironments (Giesen et al. 2014).

4.13   Outlook

The field of mass spectrometry-based proteomics continues to rapidly evolve and mature. Each new generation of mass spectrometers pushes the limits of performance in terms of resolving power, mass accuracy, and sensitivity. Many of these improvements continue to trickle down into mainstream instrumentation available to the average user. How do these technological innovations impact the field?

Ultra-high resolution opens the window to investigate the fine structure of isotopologues. This advancement has already led to the development of novel stable isotope labeling strategies that take advantage of mass defect-based neutron encoding for multiplexed quantitation (Hebert et al. 2013a). The resolved isotopologues structures could also be harnessed by a next generation of informatic pipelines that capitalize on the encoded elemental composition information in an effort to improve peptide/protein identification rates.

In terms of sensitivity, one promising approach entails a switch from serial to parallel accumulation of MS precursor and subsequent release and fragmentation based on ion mobility. The speed and sensitivity of MS/MS experiments can be increased by parallel accumulation and serial fragmentation (PASEF) that is employed on trapped ion mobility-mass spectrometry (TIMS)-mass spectrometers (Meier et al. 2015). Other opportunities exist to increase sensitivity by improving and better integrating sample preparation and data acquisition workflows (Specht and Slavov 2018). Increased sensitivity will open up the transformative potential of single-cell proteomics, in which the contribution of each cell type to complex microenvironments such as cancers can be determined.

Integration with other omics approaches and resolving the spatial distribution of proteins are key aspects to reveal protein function and elucidate their role in physiology and pathology. The Human Protein Atlas project (www.proteinatlas.org) is a pioneering resource to study spatial proteomics across the major tissues and organs of the human body (Uhlen et al. 2015) and at the subcellular level (Thul et al. 2017) based on immunohistochemistry and complemented by RNA sequencing and mass spectrometry. The Human Pathology Atlas companion extends this groundbreaking system-level analysis to the transcriptome of the 17 major cancer types (Uhlén et al. 2017).

Finally, live monitoring of data acquisition will provide opportunities to fine-tune workflows in real time so that qualitative and quantitative performance can be optimized. The MaxQuant.Live framework is a first example of how real-time monitoring can be used for on-the-fly recalibration of mass and retention times which increases the efficiency of LC-MS experiments (Wichmann et al. 2018). In the future, further integration of entire workflows from automated sample preparation, data measurements, and data analysis will make the development of adaptive and smart data acquisitions a reality.

References

Addona TA et al (2009) Multi-site assessment of the precision and reproducibility of multiple reaction monitoring-based measurements of proteins in plasma. Nat Biotechnol 27(7):633–641

Aebersold R, Mann M (2016) Mass-spectrometric exploration of proteome structure and function. Nature 537(7620):347–355

Akimov V et al (2018) UbiSite approach for comprehensive mapping of lysine and N-terminal ubiquitination sites. Nat Struct Mol Biol 25(7):631–640

Allmer J (2011) Algorithms for the de novo sequencing of peptides from tandem mass spectra. Expert Rev Proteomics 8(5):645–657

Altelaar AM, Heck AJ (2012) Trends in ultrasensitive proteomics. Curr Opin Chem Biol 16(1–2):206–213

Bandura DR et al (2009) Mass cytometry: technique for real time single cell multitarget immunoassay based on inductively coupled plasma time-of-flight mass spectrometry. Anal Chem 81(16):6813–6822

Bantscheff M et al (2012) Quantitative mass spectrometry in proteomics: critical review update from 2007 to the present. Anal Bioanal Chem 404(4):939–965

Bateman NW et al (2013) Maximizing peptide identification events in proteomic workflows utilizing data-dependent acquisition. Mol Cell Proteomics 13(1):329–338

Bekker-Jensen DB et al (2017) An optimized shotgun strategy for the rapid generation of comprehensive human proteomes. Cell Syst 4(6):587–599.e4

Belczacka I et al (2018) Proteomics biomarkers for solid tumors: current status and future prospects. Mass Spectrom Rev 136:E359

Boersema PJ et al (2008) Triplex protein quantification based on stable isotope labeling by peptide dimethylation applied to cell and tissue lysates. Proteomics 8(22):4624–4632

Boersema PJ et al (2010) In-depth qualitative and quantitative profiling of tyrosine phosphorylation using a combination of phosphopeptide immunoaffinity purification and stable isotope dimethyl labeling. Mol Cell Proteomics 9(1):84–99

Bourmaud A, Gallien S, Domon B (2016) Parallel reaction monitoring using quadrupole-Orbitrap mass spectrometer: principle and applications. Proteomics 16(15–16):2146–2159

Breitkopf SB, Asara JM (2012) Determining in vivo phosphorylation sites using mass spectrometry. In: Ausubel FM et al (eds) Current protocols in molecular biology, Chapter 18(1), pp Unit18.19.1–27

Cantor DI, Nice EC, Baker MS (2015) Recent findings from the human proteome project: opening the mass spectrometry toolbox to advance cancer diagnosis, surveillance and treatment. Expert Rev Proteomics 12(3):279–293

Carr SA, Annan RS, Huddleston MJ (2005) Mapping posttranslational modifications of proteins by MS-based selective detection: application to phosphoproteomics. Methods Enzymol 405:82–115

Chi A et al (2007) Analysis of phosphorylation sites on proteins from Saccharomyces cerevisiae by electron transfer dissociation (ETD) mass spectrometry. Proc Natl Acad Sci U S A 104(7):2193–2198

Choe L et al (2007) 8-plex quantitation of changes in cerebrospinal fluid protein expression in subjects undergoing intravenous immunoglobulin treatment for Alzheimer's disease. Proteomics 7(20):3651–3660

Collins MO et al (2005) Proteomic analysis of in vivo phosphorylated synaptic proteins. J Biol Chem 280(7):5972–5982

Collins BC et al (2017) Multi-laboratory assessment of reproducibility, qualitative and quantitative performance of SWATH-mass spectrometry. Nat Commun 8(1):291

Conrads KA et al (2005) A combined proteome and microarray investigation of inorganic phosphate-induced pre-osteoblast cells. Mol Cell Proteomics 4(9):1284–1296

Cornett DS et al (2007) MALDI imaging mass spectrometry: molecular snapshots of biochemical systems. Nat Methods 4(10):828–833

Cox J et al (2011) Andromeda: a peptide search engine integrated into the MaxQuant environment. J Proteome Res 10(4):1794–1805

Cox J et al (2014) Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. Mol Cell Proteomics 13(9):2513–2526

Craig R, Beavis RC (2004) TANDEM: matching proteins with tandem mass spectra. Bioinformatics 20(9):1466–1467

Craig R et al (2006) Using annotated peptide mass spectrum libraries for protein identification. J Proteome Res 5(8):1843–1849

Creasy DM, Cottrell JS (2004) Unimod: protein modifications for mass spectrometry. Proteomics 4(6):1534–1536

de Godoy LMF et al (2008) Comprehensive mass-spectrometry-based proteome quantification of haploid versus diploid yeast. Nature 455(7217):1251–1254

Dekker LJM et al (2018) Determination of site-specific phosphorylation ratios in proteins with targeted mass spectrometry. J Proteome Res 17(4):1654–1663

Dephoure N, Gygi SP (2012) Hyperplexing: a method for higher-order multiplexed quantitative proteomics provides a map of the dynamic response to rapamycin in yeast. Sci Signal 5(217):rs2–rs2

Desiere F et al (2004) Integration with the human genome of peptide sequences obtained by high-throughput mass spectrometry. Genome Biol 6(1):R9

Desouza LV et al (2008) Multiple reaction monitoring of mTRAQ-labeled peptides enables absolute quantification of endogenous levels of a potential cancer marker in cancerous and normal endometrial tissues. J Proteome Res 7(8):3525–3534

Deutsch EW et al (2018) Expanding the use of spectral libraries in proteomics. J Proteome Res 17(12):4051–4060

Ding F et al (2014) The role of the ubiquitin-proteasome pathway in cancer development and treatment. Front Biosci 19:886–895

Dinkel H et al (2011) Phospho.ELM: a database of phosphorylation sites–update 2011. Nucleic Acids Res 39(Database issue):D261–D267

Doll S, Burlingame AL (2015) Mass spectrometry-based detection and assignment of protein post-translational modifications. ACS Chem Biol 10(1):63–71

Dowling RJO et al (2010) Dissecting the role of mTOR: lessons from mTOR inhibitors. Biochim Biophys Acta 1804(3):433–439

Drabovich AP et al (2012) Quantitative analysis of energy metabolic pathways in MCF-7 breast cancer cells by selected reaction monitoring assay. Mol Cell Proteomics 11(8):422–434

Dreier RF et al (2018) Global ion suppression limits the potential of mass spectrometry based phosphoproteomics. J Proteome Res 18(1):493–507. https://doi.org/10.1021/acs.jproteome.8b00812

Eng JK, McCormack AL, Yates JR (1994) An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. J Am Soc Mass Spectrom 5(11):976–989

Eng JK et al (2011) A face in the crowd: recognizing peptides through database search. Mol Cell Proteomics 10(11):R111.009522

Erickson BK et al (2017) A strategy to combine sample multiplexing with targeted proteomics assays for high-throughput protein signature characterization. Mol Cell 65(2):361–370

Escher C et al (2012) Using iRT, a normalized retention time for more targeted measurement of peptides. Proteomics 12(8):1111–1121

Everley P et al (2004) Quantitative cancer proteomics: stable isotope labeling with amino acids in cell culture (SILAC) as a tool for prostate cancer research. Mol Cell Proteomics 3(7):729–735

Everley PA et al (2006) Enhanced analysis of metastatic prostate cancer using stable isotopes and high mass accuracy instrumentation. J Proteome Res 5(5):1224–1231

Faria SS et al (2017) A timely shift from shotgun to targeted proteomics and how it can be groundbreaking for cancer research. Front Oncol 7(10):13

Farrah T et al (2012) PASSEL: the PeptideAtlas SRM experiment library. Proteomics 12(8):1170–1175

Frank A, Pevzner P (2005) PepNovo: de novo peptide sequencing via probabilistic network modeling. Anal Chem 77(4):964–973

Franks A, Airoldi E, Slavov N (2017) Post-transcriptional regulation across human tissues (Vogel C (ed)). PLoS Comput Biol 13(5):e1005535

Frewen BE et al (2006) Analysis of peptide MS/MS spectra from large-scale proteomics experiments using spectrum libraries. Anal Chem 78(16):5678–5684

Fu J et al (2009) System-wide molecular evidence for phenotypic buffering in Arabidopsis. Nat Genet 41(2):166–167

Garavelli JS (2004) The RESID database of protein modifications as a resource and annotation tool. Proteomics 4(6):1527–1533

Geer LY et al (2004) Open mass spectrometry search algorithm. J Proteome Res 3(5):958–964

Geiger T et al (2010) Super-SILAC mix for quantitative proteomics of human tumor tissue. Nat Methods 7(5):383–385

Geiger T et al (2012) Comparative proteomic analysis of eleven common cell lines reveals ubiquitous but varying expression of most proteins. Mol Cell Proteomics 11(3):M111.014050–M111.014050

Gerber SA et al (2003) Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS. Proc Natl Acad Sci U S A 100(12):6940–6945

Gevaert K et al (2008) Stable isotopic labeling in proteomics (Dunn MJ (ed)). Proteomics 8(23–24):4873–4885

Gholami AM et al (2013) Global proteome analysis of the NCI-60 cell line panel. Cell Rep 4(3):609–620

Giesen C et al (2014) Highly multiplexed imaging of tumor tissues with subcellular resolution by mass cytometry. Nat Methods 11:417–422

Gillet LC et al (2012) 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(6):O111.016717–O111.016717

Gruhler A et al (2005) Quantitative phosphoproteomics applied to the yeast pheromone signaling pathway. Mol Cell Proteomics 4(3):310–327

Gygi SP, Rist B et al (1999a) Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. Nat Biotechnol 17(10):994–999

Gygi SP, Rochon Y et al (1999b) Correlation between protein and mRNA abundance in yeast. Mol Cell Biol 19(3):1720–1730

Halligan BD et al (2004) ProMoST (Protein Modification Screening Tool): a web-based tool for mapping protein modifications on two-dimensional gels. Nucleic Acids Res 32(Web Server issue):W638–W644

Hamelin C et al (2011) Identification and verification of heat shock protein 60 as a potential serum marker for colorectal cancer. FEBS J 278(24):4845–4859

Hebert AS, Merrill AE et al (2013a) Amine-reactive neutron-encoded labels for highly plexed proteomic quantitation. Mol Cell Proteomics 12(11):3360–3369

Hebert AS, Richards AL et al (2013b) The one hour yeast proteome. Mol Cell Proteomics 13(1):339–347

Hegeman AD et al (2004) An isotope labeling strategy for quantifying the degree of phosphorylation at multiple sites in proteins. J Am Soc Mass Spectrom 15(5):647–653

Hernández H, Robinson CV (2007) Determining the stoichiometry and interactions of macromolecular assemblies from mass spectrometry. Nat Protoc 2(3):715–726

Hornbeck PV et al (2012) PhosphoSitePlus: a comprehensive resource for investigating the structure and function of experimentally determined post-translational modifications in man and mouse. Nucleic Acids Res 40(Database issue):D261–D270

Hsu PP et al (2011) The mTOR-regulated phosphoproteome reveals a mechanism of mTORC1-mediated inhibition of growth factor signaling. Science 332(6035):1317–1322

Huang Z et al (2017) Proteomic profiling of human plasma for cancer biomarker discovery (Pandey A (ed)). Proteomics 17(6):1600240

Huang H et al (2018) iPTMnet: an integrated resource for protein post-translational modification network discovery. Nucleic Acids Res 46(D1):D542–D550

Humphrey SJ, Azimifar SB, Mann M (2015) High-throughput phosphoproteomics reveals in vivo insulin signaling dynamics. Nat Biotechnol 33(9):990–995

Hüttenhain R et al (2012) Reproducible quantification of cancer-associated proteins in body fluids using targeted proteomics. Sci Transl Med 4(142):142ra94–142ra94

Hüttenhain R et al (2013) Quantitative measurements of N-linked glycoproteins in human plasma by SWATH-MS (Figeys D (ed)). Proteomics 13(8):1247–1256

Huttlin EL et al (2010) A tissue-specific atlas of mouse protein phosphorylation and expression. Cell 143(7):1174–1189

Ishihama Y et al (2005) Exponentially modified protein abundance index (emPAI) for estimation of absolute protein amount in proteomics by the number of sequenced peptides per protein. Mol Cell Proteomics 4(9):1265–1272

Jin LL et al (2010) Measurement of protein phosphorylation stoichiometry by selected reaction monitoring mass spectrometry. J Proteome Res 9(5):2752–2761

Karp NA et al (2010) Addressing accuracy and precision issues in iTRAQ quantitation. Mol Cell Proteomics 9(9):1885–1897

Kettenbach AN, Gerber SA (2011) Rapid and reproducible single-stage phosphopeptide enrichment of complex peptide mixtures: application to general and phosphotyrosine-specific phosphoproteomics experiments. Anal Chem 83(20):7635–7644

Kim M-S et al (2014) A draft map of the human proteome. Nature 509(7502):575–581

Klein T et al (2017) Proteolytic cleavage-mechanisms, function, and "omic" approaches for a near-ubiquitous posttranslational modification. Chem Rev 118(3):1137–1168. https://doi.org/10.1021/acs.chemrev.7b00120

Krug K et al (2018) A curated resource for phosphosite-specific signature analysis. Mol Cell Proteomics 18(3):576–593. https://doi.org/10.1074/mcp.TIR118.000943

Kruger M et al (2008) SILAC mouse for quantitative proteomics uncovers kindlin-3 as an essential factor for red blood cell function. Cell 134(2):353–364

Kusebauch U et al (2014) Using PeptideAtlas, SRMAtlas, and PASSEL: comprehensive resources for discovery and targeted proteomics (Baxevanis AD et al (ed)). Curr Protoc Bioinformatics 46(1):13.25.1–28

Kusebauch U et al (2016) Human SRMAtlas: a resource of targeted assays to quantify the complete human proteome. Cell 166(3):766–778

Laine RA (1994) A calculation of all possible oligosaccharide isomers both branched and linear yields $1.05 \times 10(12)$ structures for a reducing hexasaccharide: the Isomer Barrier to development of single-method saccharide sequencing or synthesis systems. Glycobiology 4(6):759–767

Lam H et al (2007) Development and validation of a spectral library searching method for peptide identification from MS/MS. Proteomics 7(5):655–667

Le Blanc JCY et al (2003) Unique scanning capabilities of a new hybrid linear ion trap mass spectrometer (Q TRAP) used for high sensitivity proteomics applications. Proteomics 3(6):859–869

Leney AC, Heck AJR (2017 Jan) Native mass spectrometry: what is in the name? J Am Soc Mass Spectrom 28(1):5–13. PMCID: PMC5174146

Li J, Duncan DT, Zhang B (2010) CanProVar: a human cancer proteome variation database. Hum Mutat 31(3):219–228

Li J et al (2011) A bioinformatics workflow for variant peptide detection in shotgun proteomics. Mol Cell Prot 10(5):M110.006536

Löster K, Kannicht C (2008) 2-dimensional electrophoresis: detection of glycosylation and influence on spot pattern. Methods Mol Biol 446(Chapter 14):199–214

Ludwig C et al (2011) Estimation of absolute protein quantities of unlabeled samples by selected reaction monitoring mass spectrometry. Mol Cell Proteomics 11(3):M111.013987–M111.013987

Ludwig C et al (2018) Data-independent acquisition-based SWATH-MS for quantitative proteomics: a tutorial. Mol Syst Biol 14(8):e8126

Lundberg E et al (2010) Defining the transcriptome and proteome in three functionally different human cell lines. Mol Syst Biol 6:450

Ma B (2015) Novor: real-time peptide de novo sequencing software. J Am Soc Mass Spectrom 26(11):1885–1894

Ma B et al (2003) PEAKS: powerful software for peptide de novo sequencing by tandem mass spectrometry. Rapid Commun Mass Spectrom 17(20):2337–2342

Macek B, Mann M, Olsen JV (2009) Global and site-specific quantitative phosphoproteomics: principles and applications. Annu Rev Pharmacol Toxicol 49(1):199–221

Maclean B et al (2010) Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. Bioinformatics 26(7):966–968

Mallick P et al (2007) Computational prediction of proteotypic peptides for quantitative proteomics. Nat Biotechnol 25(1):125–131

Mani A, Gelmann EP (2005) The ubiquitin-proteasome pathway and its role in cancer. J Clin Oncol Off J Am Soc Clin Oncol 23(21):4776–4789

Mann M (2006) Functional and quantitative proteomics using SILAC. Nat Rev Mol Cell Biol 7(12):952–958

Manning G et al (2002) The protein kinase complement of the human genome. Science 298(5600):1912–1934

Matsumoto M, Nakayama KI (2018) The promise of targeted proteomics for quantitative network biology. Curr Opin Biotechnol 54:88–97

Matsumoto M et al (2017) A large-scale targeted proteomics assay resource based on an in vitro human proteome. Nat Methods 14(3):251–258

Mcalister GC et al (2014) MultiNotch MS3 enables accurate, sensitive, and multiplexed detection of differential expression across cancer cell line proteomes. Anal Chem 86(14):7150–7158

McDonnell LA, Heeren RMA (2007) Imaging mass spectrometry. Mass Spectrom Rev 26(4):606–643

Meier F et al (2015) Parallel accumulation-serial fragmentation (PASEF): multiplying sequencing speed and sensitivity by synchronized scans in a trapped ion mobility device. J Proteome Res 14(12):5378–5387

Mertins P et al (2016) Proteogenomics connects somatic mutations to signalling in breast cancer. Nature 534(7605):55–62

Mertins P et al (2018) Reproducible workflow for multiplexed deep-scale proteome and phosphoproteome analysis of tumor tissues by liquid chromatography-mass spectrometry. Nat Protoc 13(7):1632–1661

Mikesh LM et al (2006) The utility of ETD mass spectrometry in proteomic analysis. Biochim Biophys Acta 1764(12):1811–1822

Minogue CE et al (2015) Multiplexed quantification for data-independent acquisition. Anal Chem 87(5):2570–2575

Monteoliva L, Albar JP (2004) Differential proteomics: an overview of gel and non-gel based approaches. Brief Funct Genomic Proteomic 3(3):220–239

Mueller LN et al (2008) An assessment of software solutions for the analysis of mass spectrometry based quantitative proteomics data. J Proteome Res 7(1):51–61

Murphy JP, Pinto DM (2010) Targeted proteomic analysis of glycolysis in cancer cells. J Proteome Res 10(2):604–613

Nagaraj N et al (2011) Deep proteome and transcriptome mapping of a human cancer cell line. Mol Syst Biol 7(1):548–548

Nahnsen S et al (2013) Tools for label-free peptide quantification. Mol Cell Proteomics 12(3):549–556

Navarro P et al (2016) A multicenter study benchmarks software tools for label-free proteome quantification. Nat Biotechnol 34:1130–1136

Nesvizhskii AI (2010) A survey of computational methods and error rate estimation procedures for peptide and protein identification in shotgun proteomics. J Proteome 73(11):2092–2123

Olsen JV et al (2006) Global, in vivo, and site-specific phosphorylation dynamics in signaling networks. Cell 127(3):635–648

Olsen JV et al (2010) Quantitative phosphoproteomics reveals widespread full phosphorylation site occupancy during mitosis. Sci Signal 3(104):ra3–ra3

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(5):376–386

Overmyer KA et al (2018) Multiplexed proteome analysis with neutron-encoded stable isotope labeling in cells and mice. Nat Protoc 13(1):293–306

Ow SY et al (2009) iTRAQ underestimation in simple and complex mixtures: "the good, the bad and the ugly". J Proteome Res 8(11):5347–5355

Pang J et al (2010) Profiling protein markers associated with lymph node metastasis in prostate cancer by DIGE-based proteomics analysis. J Proteome Res 9(1):216–226

Perkins DN et al (1999) Probability-based protein identification by searching sequence databases using mass spectrometry data. Electrophoresis 20(18):3551–3567

Peterson AC, Russell JD, Bailey DJ, Westphall MS, Coon JJ (2012) Parallel reaction monitoring for high resolution and high mass accuracy quantitative, targeted proteomics. Mol Cell Proteomics 11(11):1475–1488

Picotti P, Aebersold R (2012) Selected reaction monitoring-based proteomics: workflows, potential, pitfalls and future directions. Nat Methods 9(6):555–566

Picotti P et al (2010) High-throughput generation of selected reaction-monitoring assays for proteins and proteomes. Nat Methods 7(1):43–46

Picotti P, Bodenmiller B, Aebersold R (2013) Proteomics meets the scientific method. Nat Methods 10(1):24–27

Potts GK et al (2016) Neucode labels for multiplexed, absolute protein quantification. Anal Chem 88(6):3295–3303

Rhoads TW et al (2014) Neutron-encoded mass signatures for quantitative top-down proteomics. Anal Chem 86(5):2314–2319

Rose CM et al (2018) TomahaqCompanion: a tool for the creation and analysis of isobaric label based multiplexed targeted assays. J Proteome Res 18(2):594–605

Rosenberger G et al (2014) A repository of assays to quantify 10,000 human proteins by SWATH-MS. Sci Data 1:140031

Ross PL et al (2004) Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents. Mol Cell Proteomics 3(12):1154–1169

Röst HL et al (2014) OpenSWATH enables automated, targeted analysis of data-independent acquisition MS data. Nat Biotechnol 32(3):219–223

Rush J et al (2005) Immunoaffinity profiling of tyrosine phosphorylation in cancer cells. Nat Biotechnol 23(1):94–101

Schober Y et al (2012) Single cell matrix-assisted laser desorption/ionization mass spectrometry imaging. Anal Chem 84(15):6293–6297

Schubert OT et al (2015) Building high-quality assay libraries for targeted analysis of SWATH MS data. Nat Protoc 10(3):426–441

Schwamborn K, Caprioli RM (2010) Molecular imaging by mass spectrometry--looking beyond classical histology. Nat Rev Cancer 10(9):639–646

Schwanhäusser B et al (2011) Global quantification of mammalian gene expression control. Nature 473(7347):337–342

Searle BC (2010) Scaffold: a bioinformatic tool for validating MS/MS-based proteomic studies (Martens L, Hermjakob H (eds)). Proteomics 10(6):1265–1269

Sharma K et al (2014) Ultradeep human phosphoproteome reveals a distinct regulatory nature of Tyr and Ser/Thr-based signaling. Cell Rep 8(5):1583–1594

Shi H et al (2011) Proteomic analysis of advanced colorectal cancer by laser capture microdissection and two-dimensional difference gel electrophoresis. J Proteome 75(2):339–351

Shortreed MR et al (2016) Elucidating proteoform families from proteoform intact-mass and lysine-count measurements. J Proteome Res 15(4):1213–1221

Shteynberg D et al (2011) iProphet: multi-level integrative analysis of shotgun proteomic data improves peptide and protein identification rates and error estimates. Mol Cell Proteomics 10(12):M111.007690–M111.007690

Shteynberg D et al (2013) Combining results of multiple search engines in proteomics. Mol Cell Proteomics 12(9):2383–2393

Skelly DA et al (2013) Integrative phenomics reveals insight into the structure of phenotypic diversity in budding yeast. Genome Res 23(9):1496–1504

Smith LM, Kelleher NL, Consortium for Top Down Proteomics (2013) Proteoform: a single term describing protein complexity. Nat Methods 10(3):186–187

Soste M et al (2014) A sentinel protein assay for simultaneously quantifying cellular processes. Nat Methods 11(10):1045–1048

Specht H, Slavov N (2018) Transformative opportunities for single-cell proteomics. J Proteome Res 17(8):2565–2571

Steen H et al (2003) Phosphotyrosine mapping in Bcr/Abl oncoprotein using phosphotyrosine-specific immonium ion scanning. Mol Cell Proteomics 2(3):138–145

Steen H et al (2005) Stable isotope-free relative and absolute quantitation of protein phosphorylation stoichiometry by MS. Proc Natl Acad Sci U S A 102(11):3948–3953

Stensballe A et al (2000) Electron capture dissociation of singly and multiply phosphorylated peptides. Rapid Commun Mass Spectrom 14(19):1793–1800

Stulík J et al (1999) Protein abundance alterations in matched sets of macroscopically normal colon mucosa and colorectal carcinoma. Electrophoresis 20(18):3638–3646

Sury MD, Chen J-X, Selbach M (2010) The SILAC fly allows for accurate protein quantification in vivo. Mol Cell Proteomics 9(10):2173–2183

Syka JEP et al (2004) Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. Proc Natl Acad Sci U S A 101(26):9528–9533

Tabb DL, Fernando CG, Chambers MC (2007) MyriMatch: highly accurate tandem mass spectral peptide identification by multivariate hypergeometric analysis. J Proteome Res 6(2):654–661

Tan HT, Lee YH, Chung MCM (2012) Cancer proteomics. Mass Spectrom Rev 31(5):583–605

Taylor J, Johnson R (2001) Implementation and uses of automated de novo peptide sequencing by tandem mass spectrometry. Anal Chem 73(11):2594–2604

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(8):1895–1904

Thul PJ et al (2017) A subcellular map of the human proteome. Science 356(6340):eaal3321

Ting L et al (2011) MS3 eliminates ratio distortion in isobaric multiplexed quantitative proteomics. Nat Methods 8(11):937–940

Tran JC et al (2011) Mapping intact protein isoforms in discovery mode using top-down proteomics. Nature 480(7376):254–258

Tsou C-C et al (2015) DIA-Umpire: comprehensive computational framework for data-independent acquisition proteomics. Nat Methods 12(3):258–64– 7 p following 264

Tsou C-C et al (2016) Untargeted, spectral library-free analysis of data-independent acquisition proteomics data generated using Orbitrap mass spectrometers (Aebersold R et al (eds)). Proteomics 16(15–16):2257–2271

Tyanova S, Temu T, Cox J (2016) The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. Nat Protoc 11(12):2301–2319

Ubersax JA, Ferrell JE (2007) Mechanisms of specificity in protein phosphorylation. Nat Rev Mol Cell Biol 8(7):530–541

Uhlen M et al (2015) Tissue-based map of the human proteome. Science 347(6220):1260419–1260419

Uhlén M et al (2017) A pathology atlas of the human cancer transcriptome. Science 357(6352):eaan2507

Umar A et al (2007) NanoLC-FT-ICR MS improves proteome coverage attainable for approximately 3000 laser-microdissected breast carcinoma cells. Proteomics 7(2):323–329

Veenstra TD (2013) Proteomic applications in cancer detection and discovery. Wiley, Hoboken

Vyatkina K et al (2017) De novo sequencing of peptides from high-resolution bottom-up tandem mass spectra using top-down intended methods (Mathivanan S (ed)). Proteomics 17(23–24):1600321

Wang Y et al (2007) Differential expression of mimecan and thioredoxin domain-containing protein 5 in colorectal adenoma and cancer: a proteomic study. Exp Biol Med (Maywood) 232(9):1152–1159

Wang M et al (2018) Assembling the community-scale discoverable human proteome. Cell Syst 7(4):412–421.e5

Wenger CD et al (2011) Gas-phase purification enables accurate, multiplexed proteome quantification with isobaric tagging. Nat Methods 8(11):933–935

Whiteaker JR et al (2010) An automated and multiplexed method for high throughput peptide immunoaffinity enrichment and multiple reaction monitoring mass spectrometry-based quantification of protein biomarkers. Mol Cell Proteomics 9(1):184–196

Wichmann C et al (2018) MaxQuant.Live enables global targeting of more than 25,000 peptides. bioRxiv:1–15

Wilhelm M et al (2014) Mass-spectrometry-based draft of the human proteome. Nature 509(7502):582–587

Wiśniewski JR, Ostasiewicz P, Mann M (2011) 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(7):3040–3049

Wu CC et al (2004) Metabolic labeling of mammalian organisms with stable isotopes for quantitative proteomic analysis. Anal Chem 76(17):4951–4959

Wu R et al (2011) Correct interpretation of comprehensive phosphorylation dynamics requires normalization by protein expression changes. Mol Cell Proteomics 10(8):M111.009654

Xing X et al (2006) Identification of differentially expressed proteins in colorectal cancer by proteomics: down-regulation of secretagogin. Proteomics 6(9):2916–2923

Xu G, Paige JS, Jaffrey SR (2010) Global analysis of lysine ubiquitination by ubiquitin remnant immunoaffinity profiling. Nat Biotechnol 28(8):868–873

Yu Y et al (2011) Phosphoproteomic analysis identifies Grb10 as an mTORC1 substrate that negatively regulates insulin signaling. Science 332(6035):1322–1326

Zhang X et al (2002) N-Terminal peptide labeling strategy for incorporation of isotopic tags: a method for the determination of site-specific absolute phosphorylation stoichiometry. Rapid Commun Mass Spectrom 16(24):2325–2332

Zhang B et al (2014) Proteogenomic characterization of human colon and rectal cancer. Nature 513(7518):382–387

Zhang H et al (2016) Integrated proteogenomic characterization of human high-grade serous ovarian cancer. Cell 166(3):755–765

Zhang M et al (2017) CanProVar 2.0: an updated database of human cancer proteome variation. J Proteome Res 16(2):421–432

Zhou M, Morgner N, Barrera NP, Politis A, Isaacson SC, Matak-Vinković D et al (2011 Oct 21) Mass spectrometry of intact V-type ATPases reveals bound lipids and the effects of nucleotide binding. Science 334(6054):380–385. PMCID: PMC3927129

Zolg DP et al (2017) Building proteometools based on a complete synthetic human proteome. Nat Methods 14(3):259–262

Zubarev RA et al (2000) Electron capture dissociation for structural characterization of multiply charged protein cations. Anal Chem 72(3):563–573