



Review

Mass Spectrometry-Based Proteomics Workflows in Cancer Research: The Relevance of Choosing the Right Steps

Paula Carrillo-Rodriguez ^{1,2}, Frode Selheim ^{1,*} and Maria Hernandez-Valladares ^{1,3,4,*}

¹ Proteomics Unit of University of Bergen (PROBE), University of Bergen, Jonas Lies vei 91, 5009 Bergen, Norway

² Vall d'Hebron Institute of Oncology (VHIO), 08035 Barcelona, Spain

³ Department of Physical Chemistry, University of Granada, Avenida de la Fuente Nueva S/N, 18071 Granada, Spain

⁴ Instituto de Investigación Biosanitaria ibs.GRANADA, 18012 Granada, Spain

* Correspondence: frode.selheim@uib.no (F.S.); mariahv@ugr.es (M.H.-V.); Tel.: +47-55586378 (F.S.); +34-958241000 (ext. 20284) (M.H.-V.)

Simple Summary: Liquid chromatography–mass spectrometry (LC-MS)-based proteomics is a powerful technology for discovering new cancer biomarkers. In addition to last generation instrumentation, it uses experimental designs of different complexity that describe key steps from sample selection to data analysis and interpretation. All aspects must be optimized to obtain the most satisfactory results. However, planning proteomics procedures can be challenging unless their advantages and drawbacks are known. This review aims to highlight the methodological features that cancer researchers must consider before executing an LC-MS-based proteomics project. Based on these features, we suggest straightforward and complex workflows whereby researchers can discover new molecules or therapeutic pathways to defeat or significantly decrease the impact of oncological diseases.

Abstract: The qualitative and quantitative evaluation of proteome changes that condition cancer development can be achieved with liquid chromatography–mass spectrometry (LC-MS). LC-MS-based proteomics strategies are carried out according to predesigned workflows that comprise several steps such as sample selection, sample processing including labeling, MS acquisition methods, statistical treatment, and bioinformatics to understand the biological meaning of the findings and set predictive classifiers. As the choice of best options might not be straightforward, we herein review and assess past and current proteomics approaches for the discovery of new cancer biomarkers. Moreover, we review major bioinformatics tools for interpreting and visualizing proteomics results and suggest the most popular machine learning techniques for the selection of predictive biomarkers. Finally, we consider the approximation of proteomics strategies for clinical diagnosis and prognosis by discussing current barriers and proposals to circumvent them.

Keywords: mass spectrometry; proteomics; sample preparation; data-dependent acquisition (DDA); data-independent acquisition (DIA); workflows; data analysis; bioinformatics



Citation: Carrillo-Rodriguez, P.; Selheim, F.; Hernandez-Valladares, M. Mass Spectrometry-Based Proteomics Workflows in Cancer Research: The Relevance of Choosing the Right Steps. *Cancers* **2023**, *15*, 555. <https://doi.org/10.3390/cancers15020555>

Academic Editors: Hiroyuki Katayama and Maxim V. Berezovski

Received: 4 January 2023

Accepted: 12 January 2023

Published: 16 January 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Liquid chromatography–mass spectrometry (LC-MS) proteomics is the current technology of choice for describing and quantifying the proteome of cells (as well as a single cell or subcellular fractions of cells), tissue, plasma, or other biological fluids and exosomes to understand the gene and cellular functions of particular conditions. Protein functions are usually identified by studying protein expression regulation, their posttranslational modifications (PTMs), and their protein–protein interaction networks. Thus, LC-MS-based proteomics analyses can provide a comprehensive picture of intra- and extra-cellular signaling [1].

Several LC-MS-based proteomics strategies are widely used to study the proteome of biological systems in medical research. They include top-down proteomics or the analysis of intact proteins (e.g., KRAS proteoforms in colorectal cancer cases); targeted proteomics used to verify, validate, and absolutely quantify candidate cancer biomarkers; and bottom-up or shotgun proteomics used to study whole proteomes [2–7]. The latter approach is widely utilized in the study of patient cohorts suffering cancer and other diseases. Shotgun proteomics workflows comprise several steps: selection of sample type, assessment of sample size, sample processing, data acquisition from the mass spectrometer, data cleaning and statistics, data interpretation and visualization, and machine learning. Unlike other omics technologies, there is little protocol standardization in LC-MS-based proteomics workflows, and therefore each project is carefully carried out according to a previously discussed experimental design of varied complexity depending on the number of samples and their nature, quantification method, enrichment of PTMs, and bioinformatics analyses.

LC-MS-based proteomics has been increasingly used in the search for disease biomarkers in the last decade. This has been accompanied by the continuous upgrading of mass spectrometers and the development of faster and more sensitive acquisition methods. In the recent past the identification of more than 2500 unique peptides per LC gradient minute and the quantification of approximately 5000 proteins with 21 min LC gradients have been reported for a quadrupole-Orbitrap mass spectrometer equipped with a differential ion mobility device [8]. Herein, we highlight several recent successes in which LC-MS-based proteomics has enabled the discovery of a classifier of five proteins (WAP four-disulfide core domain protein 2, WFDC2; prothymosin alpha, PTMA; nectin-4, PVRL4; fibrinogen alpha chain, FIBA; and nectin-2, PVRL2) for distinguishing between benign and malignant ovarian tumors [9], and a panel of six proteins (alpha-1-antichymotrypsin, AACT; trombospondin-4, TSP4; malate dehydrogenase mitochondrial, MDHM; calreticulin, CALR; protein LEG1 homolog, LEG1; and alpha-2-HS-glycoprotein, AHSG) for the detection of gliomas [10]. However, more studies are required to improve many of the current diagnostic assays as well as for the discovery of new prognostic biomarkers that will help us understand disease development and patient responses to treatments [11].

In cancer research, oncologists are aware of how LC-MS-based proteomics can significantly contribute to preclinical drug discovery. Nonetheless, multi-steps workflows employed in LC-MS-based proteomics and the multiple methodological choices available for each of them may overwhelm researchers that have not tried this technology before. For this reason, we present and describe these workflow steps, suggesting straightforward and branched roadmaps leading to the discovery of cancer biomarkers. Finally, we will discuss how proteomics strategies can be part of clinical diagnosis and prognosis assays, contemplating current barriers and encouraging proposals to circumvent them.

2. LC-MS-Based Proteomics Strategies from Sample Selection to Data Acquisition in Cancer Research: Steps and Main Considerations

Each step of an LC-MS-based proteomics workflow represents an opportunity to maximize proteome coverage and obtain the most successful findings. Therefore, all possible approaches at each step must be carefully considered in order to create the most productive workflow. Here, we describe the core steps and propose simple and efficient tools to augment the quality and quantity of MS-based data (Figure 1). We will expand these workflows by adding steps covering PTM enrichment and machine learning for more experienced cancer researchers.

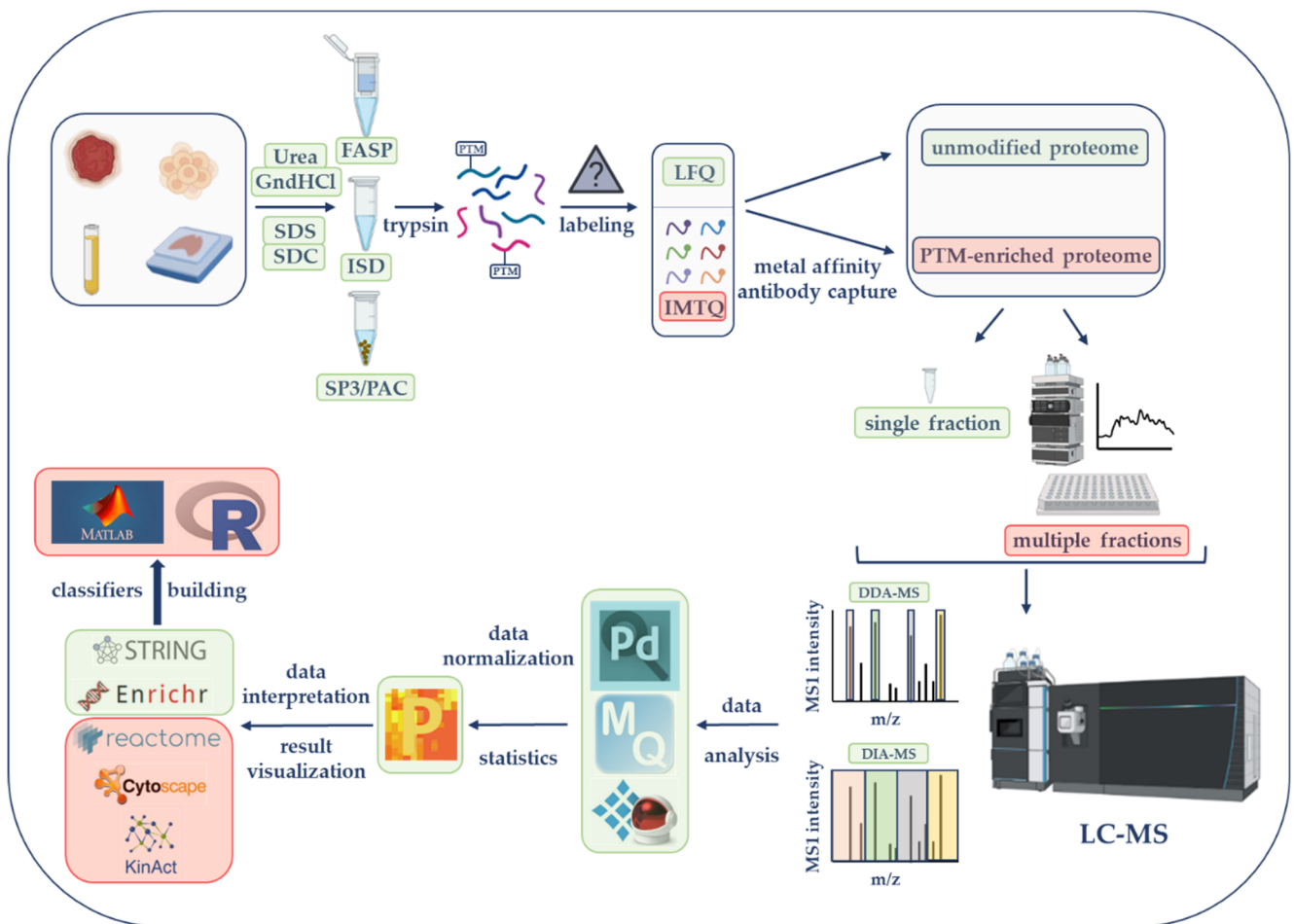


Figure 1. Liquid chromatography–mass spectrometry (LC-MS)-based proteomics workflow. Initial steps consist of sample lysis and solubilization in the presence of chaotropic agents (urea or guanidine hydrochloride, GndHCl) or detergents (e.g., sodium dodecyl sulfate, SDS; sodium deoxycholate, SDC). Samples are further processed by the filter-aided sample preparation (FASP), the in-solution digestion (ISD) procedure, single-pot, solid-phase-enhanced sample preparation (SP3), or protein aggregation capture (PAC) before trypsin digestion. According to the selected quantification approach, peptides will be kept unlabeled (for label-free quantification, LFQ) or will be labeled with tandem mass tags (TMTs) for isobaric mass tag quantification (IMTQ). A small portion will be utilized for the characterization of the unmodified or so-called global proteome while the rest of the sample will be used for posttranslational modifications (PTMs). Unmodified and modified samples can be analyzed as single fractions or as multiple fractions after being chromatographically fractionated. Peptides will be run on a mass spectrometer with data-dependent acquisition (DDA) or data-independent acquisition (DIA) methods. MS data will be analyzed by commercially or publicly available software followed by the utilization of several bioinformatics tools to perform gene ontology enrichment, protein network, and PTM characterization studies. These and more software and online applications are described in Table 1. The final step will involve the use of several artificial intelligence tools for classification modeling. The steps illustrated in this figure represent the most efficient strategies according to our experience. Steps framed with a green rectangle correspond to basic global proteomics workflows, while those framed with a light-red rectangle are utilized by experienced researchers or by those that seek PTM information. The figure was created with features obtained from BioRender (<https://biorender.com/> (accessed on 25 May 2022)).

2.1. Sample Type's Selection and Cohort Size

LC-MS-based proteomics can analyze any type of oncological samples from which proteins can be extracted. These include freshly frozen tissue or cells, formalin-fixed paraffin

embedded (FFPE) tissue, blood fractions plasma or serum, feces, and other biological fluids such as urine, saliva, buccal swabs, and cerebral spinal fluid. While it might not be possible to select a sample's type in some retrospective projects because of material availability, it is becoming easier to find different sample types (i.e., tissue and plasma) from the same patient thanks to the standard operating procedures (SOPs) that are being established in prospective studies by new biobanking policies [12,13]. In fact, the development and compliance of SOPs that include detailed criteria for proper sample collection (e.g., reagents and chemicals added; duration of the procedure) and storage (e.g., addition of cryoprotectants; storage temperature and acceptable duration) have become essential to guarantee sample quality and reduce variability of the project data. However, more efforts are required toward the elaboration of global SOPs that can facilitate sample sharing among different research groups and hospital biobanks.

Besides sample availability, three main factors determine the choice of sample type for proteomics research. The first factor is the tumor type and location. Biofluids in closer contact with tumors are probably a better source for potential biomarkers.

The second factor is the researcher's skills along with equipment availability in the laboratory. The sample must be optimally processed in order to obtain the highest number of identified proteins and accurate quantitative values. Thus, while sample preparation for LC-MS analysis of leukemic blasts can involve uncomplicated procedures [14–17], FFPE tissue and plasma (key sample types in cancer proteomics) require more complex protocols that include reversal of chemical crosslinking, removal of reagents and protein extraction, and effective depletion of most abundant proteins, respectively [18–20]. Recently, a protocol that combines tissue disruption by ultrasonication, heat-induced antigen retrieval, and two alternative methods for efficient detergent removal has enabled satisfactory quantitative proteomic analysis of limited amounts of FFPE material [21]. Currently, plasma researchers are mainly using columns to selectively deplete the most abundant plasma proteins [22]. However, issues of reproducibility and indirect removal of relevant proteins have already been reported [23,24]. As an alternative, the use of nanoparticles with different surface chemistry was proven to identify ~4000 plasma proteins [25,26]. Nonetheless, the cost of this procedure, which is only available in a robotic system, becomes especially high in discovery studies.

The last factor to consider is the number of study subjects and their samples needed to achieve an acceptable study power, typically 80%. Although Levin demonstrated that for a study to be powered at 80% with a detectable fold change of 1.5 comparing two sample groups for all proteins, the minimum sample size was 60 per group [27], Nakayasu et al., found that the number of required biological replicates in a study of that power depends on the variability [11]. The variability in a study is the sum of the biological and the technical variability. Moreover, the study design (i.e., number of biological replicates and number of groups) depends on heterogeneity and homogeneity in a group or between groups. Therefore, it is important to identify samples that are homogenous in a group during the study design, and it is desirable that the groups to compare are as different as possible. Furthermore, the biological variability in a study is highly dependent on the sample origin; i.e., cancer cells are expected to have less variability compared to tumor tissue. The lower the variability in a study the higher the power of analysis, and as a result, a higher number of statistically changed proteins with smaller differences will be found.

The power of previously published LC-MS-based proteomics studies was rarely described. However, current journal practices and policies promote the inclusion of detailed descriptions of the experimental design that provide the necessary power of the study.

2.2. Sample Preparation Strategies

The choice of a sample preparation methodology is a key step of any LC-MS-based proteomics workflow [28,29]. Only the use of unbiased preparation approaches that produce a high number of identified and quantified proteins can provide satisfactory descriptions of the proteomes under study. Most sample processing for LC-MS analysis can be mainly cat-

egorized into in-solution (ISD), filter-based, and bead-based methods (Figure 1). While ISD protocols extract proteomes by the addition of concentrated solutions of chaotropic agents or detergents such as urea and guanidine hydrochloride (GndHCl) or sodium deoxycholate (SDC), respectively [30,31], filter-based or bead-based workflows allow protein extraction with detergents such as sodium dodecyl sulfate (SDS) or SDC and digestion in the presence of ammonium bicarbonate buffer after detergent removal. Other buffers such as HEPES and triethylammonium bicarbonate (TEAB) are used during digestion and are compatible with tandem mass tag (TMT)-labeling for relative quantification (see Section 2.3).

Although classical ISD protocols are less frequently used, recent attractive ISD solutions such as microreactor tips with on-column TMT labeling [32] and SDC-based ISD with TMT labeling in a 96-well plate format, SimPLIT [33], have been presented as efficient, fast, and low-cost approaches for the digestion of fluorescence-activated cell sorting (FACS)-sorted samples and global proteomics samples, respectively.

The first sample preparation and digestion methodology for MS-based proteomics using spin filters with a ≥ 3000 molecular weight cutoff membrane was introduced nearly two decades ago [34]. However, this method did not become popular in the proteomics community until it was presented as filter-aided sample preparation (FASP), which incorporated urea in a high concentration to successfully remove SDS [35]. Since then, FASP in combination with StageTip-based fractionation and multi-enzyme digestion FASP protocols has been extensively used for in-depth analysis of proteomes [36–38]. Magnetic bead-based sample preparation approaches for proteomics experiments were introduced, such as single-pot, solid-phase-enhanced sample preparation (SP3), and the protein aggregation capture (PAC). SP3 uses carboxylate-modified hydrophilic beads that bind proteins in a nonselective fashion through the use of ethanol-driven solvation capture. It is compatible with most of the common chemical agents used to facilitate cell or tissue lysis such as detergents, chaotropes, salts, and organic solvents [39]. As the entire SP3 procedure occurs in a single sample tube and takes little time when compared to other procedures [40], it is not surprising that the SP3 technology is becoming more and more popular among new and experienced MS-based proteomics researchers [41,42]. PAC, which employs the inherent instability of denatured proteins for non-specific immobilization on microparticles by aggregation capture, was shown to be more efficient than ISD and FASP procedures in the preparation of phosphopeptides and peptides from tissue and secretome samples [43]. Both protocols were also reported to be successful on automated devices [8,44–46].

To secure optimal sample preparation protocols for LC-MS-based proteomic studies aiming at the discovery of acute myeloid leukemia (AML) biomarkers, our research group has been testing novel techniques over the past few years. We started evaluating ISD and FASP proteomic workflows with leukemic blast samples isolated from peripheral blood [17]. Using two different quantitative approaches, label-free (LF) and stable isotopes labeling with amino acids in cell culture (SILAC), FASP workflows were selected to produce the highest number of quantified proteins with reduced number of missed cleavages. However, the use of fractionation methods such as the mixed mode with styrene-divinylbenzene-reverse phase sulfonate plugs in both FASP and ISD workflows, employing one (trypsin) and two proteases (Lys-C and trypsin) at the digestion step, respectively, quantified approximately 2200 proteins with an Orbitrap Elite mass spectrometer (Thermo Scientific, Waltham, MA, USA).

Because of the long processing time in the FASP procedure, we recently compared the performance of the ISD method using GndHCl in the lysis buffer and two proteases and the SP3 strategy using lysis buffers containing SDS or GndHCl and one protease with HeLa cell and human plasma samples [20]. Our results showed that the SP3 protocol, using either buffer, achieved the highest number of LF-quantified proteins in HeLa cells (5895–6131 without peptide fractionation; 7817–8136 with high pH reversed-phase LC fractionation) and plasma samples (397–411 without depletion and fractionation steps; 1397 after Top12 abundant protein depletion and high pH reversed-phase LC fractionation). Therefore, we

have recently used the SP3 protocol with SDS-based lysis buffer for the proteomic analysis of AML samples [15,16].

Thus, we and other authors recommend the use of the SP3 procedure which represents a very robust and efficient processing tool for both concentrated and diluted protein materials [41,47–49]. To facilitate large studies, the use of automation (e.g., KingFisher™ Flex, Thermo Fisher Scientific, Waltham, MA, USA) in a 96-well format was proved to have a great impact on the reproducibility of bead-based sample preparation protocols [8,50].

2.3. Quantification Strategies

Quantitative LC-MS-based proteomics experiments involve the use (or not, as in the LF quantification (LFQ) approach) of specific mass tags that are recognized by the instrument and are usually introduced into proteins or peptides metabolically or by chemical means, respectively [51]. SILAC utilizes the cell's own metabolism to incorporate isotopically labeled amino acids into its proteome, which can be mixed with the proteome of unlabeled cells [52,53]. Thus, differences in protein expression can be analyzed by comparing the abundance of the labeled versus unlabeled proteins. The chemical derivatization processes include methodologies such as isotope-coded affinity tags (ICATs), dimethyl labeling, and isobaric mass tags among others [54–58]. Isobaric tags for relative and absolute quantification (iTRAQ), which consist of a reporter group, a balance group, and a peptide reactive group, are used to quantify up to eight peptide samples [59]. When the samples are pooled and analyzed simultaneously, the same peptide from the different samples will appear at the same mass in the MS1 scan. However, when the peptides are fragmented at the MS2 level, the peptide fragments provide amino acid sequence information and tag fragments, i.e., reporter ions. The ratios of these reporter ions are representative of the proportions of that peptide in each of the eight samples [59]. Herein, we will describe popular quantitative approaches with the use of TMT (another isobaric tag technique) and LFQ in current LC-MS-based proteomics.

LFQ was introduced early in the past decade as an alternative procedure to expensive and time-consuming stable isotope-based labeling methods. LFQ quantification is based on the intensities obtained from the extracted ion chromatogram (XIC) of MS1 signals or on spectral counting of the precursors, whereas peptide identification is carried out, as described for isobaric tag quantification, with peptidic features from fragment ions at MS2 [60]. It requires initial measurement of the sample concentration under consistent conditions and a strict adherence to the sample preparation workflow, including fractionation to resolve peptides with a consequent increase in the coverage of complex proteomes. LFQ has become highly employed in global proteomics and phosphoproteomics thanks to algorithms such as MaxLFQ, which handles fraction-dependent normalization information, calculation of pair-wise sample protein ratios from the peptide XIC ratios, and transfer of peptide identifications in one run to unidentified peptides in the subsequent run by matching their mass and retention times (i.e., the “match-between-runs, MBR” feature) [61]. Therefore, MBR can significantly increase the number of annotated identifications and provide more data for downstream quantification of proteins [62]. Recently, MBR has also been applied to TMT quantification using the three-dimensional MS1 features to transfer identifications from identified to unidentified MS2 spectra between LC-MS runs in order to utilize reporter ion intensities in unidentified spectra for quantification [63].

The TMT labeling system is used at the peptide level and consists of mass tagging reagents of the same nominal mass. Similar to iTRAQ labels, these tags are composed of an amine-reactive group, a spacer arm, and a mass reporter that are used for MS2 quantification. Commercial TMT kits (Thermo Fisher Scientific) contain 6, 10, 11, 16, or 18 labels (also called channels) that can be used in different experiment sets when a reference channel comprising a small aliquot from each sample serves as a normalization bridge among the different sets. This allows accurate quantification of large sample cohorts. Despite the tag cost, more and more proteomics researchers are using the TMT labeling approach since several optimized TMT labeling protocols covering important issues such

as the peptide:tag ratio and reaction buffer have been recently published in addition to simplified commercial and free software workflows [64–66].

2.4. PTM Enrichments

The study of protein regulation by covalent modifications, PTMs, becomes necessary to understand the complexity and functionality of proteomes in cancer development [67]. PTMs that involve a mass increase in a peptide sequence can be identified and quantified with the LC-MS technology. Because of the substoichiometric abundance of many PTMs, their study involves enrichment procedures in order to remove unmodified peptides. The description and enrichment procedures of the most frequent PTMs are beyond the scope of this manuscript. However, we will herein focus on peptide phosphorylation as one of the major cellular signaling events, and we recommend several recent reviews regarding other PTMs and strategies to characterize them [1,68].

Phosphopeptide enrichment has been classically performed using metal oxide affinity chromatography (MOAC) with titanium dioxide beads, immobilized metal affinity chromatography (IMAC) with iron affinity gel, and sequential elution from IMAC (SIMAC) with a combination of both reagents [69–71]. Our group successfully constructed a dataset comprising more than 12,000 quantified class I (i.e., probability of site localization ≥ 0.75) phosphorylation sites from approximately 3000 proteins in an AML cohort with 41 patients using the IMAC protocol [72]. Nonetheless, the enrichment procedure has been remarkably eased by the use of magnetic material (e.g., MagReSyn Ti-IMAC HP beads from Resyn Biosciences) in the last few years [73].

2.5. Peptide Fractionation to Increase Proteome Coverage

Peptide fractionation is a necessary step before LC-MS analysis in order to achieve maximal proteome coverage in samples from complex organisms. Most popular fractionation techniques are based on peptide properties such as charge, polarity, and hydrophobicity [74]. Strong cation exchange, strong anion exchange, and mixed mode methodologies have been widely used as stuck disks on pipette tips or in the in-StageTip format [17,38,75]. However, in order to produce more fractions and take advantage of the increasing sensitivity of last generation mass spectrometers, offline high pH reversed-phase chromatography using C18 sorbents proved to be an excellent strategy to quantify up to 8434 mouse protein groups and 16,152 localized class I phosphosites when 46 and 12 TMT-labeled peptidic fractions were analyzed during a 30 min and a 60 min elution gradient, respectively [44]. Using the same number of peptidic fractions and length of LC gradients, 11,292 protein groups and 30,304 localized class I phosphosites were identified in HeLa lysates in an LFQ strategy [76].

Alternatively, high-resolution isoelectric point focusing (HiRIEF) applied at the (iTRAQ-labeled) peptide level in the 3.7–5.0 pH range identified 13,078 human and 10,637 mouse proteins when the 72 fractions obtained from the strip were analyzed during a 50 min gradient [77]. In a recent study, the analysis of TMT-labeled peptides from 141 non-small-cell lung cancer tumor samples that were fractionated on two strips (pH 3.7–4.9 and pH 3–10) and analyzed during a 60 min elution gradient quantified 13,975 proteins [78]. However, HiRIEF with two pH-range strips (2.5–3.7; 3–10) did not appear to efficiently perform in a cell-cycle arrest study that identified 19,075 localized class I phosphosites from a total of 132 TMT-labeled fractions analyzed during a 50 min elution gradient [79].

All things considered, the choice of peptide fractionation method is subject to the number of fractions that can be affordably analyzed, i.e., the MS time and the proteome depth sought.

2.6. MS Methods for Data Acquisition

LC-MS-based proteomics basically employs two MS data acquisition strategies, data-dependent acquisition (DDA) and data-independent acquisition (DIA), for global proteomics studies.

In DDA mode, the MS alternates between full-scan spectral acquisition at the MS1 level and MS2 sequential analysis of MS1 precursors selected according to their charge state (i.e., ≥ 2) and relative high intensity. Although this acquisition mode can be used for LF- or TMT-labeled samples, it introduces an abundance bias into the sampling and variability when running both biological and technical replicates. In order to alleviate these inherent DDA effects, the MS dynamic exclusion technology that adds masses with the highest intensity to a temporary exclusion list for a period of typically 30–60 s while peptides of lower abundance are sequenced and the already-mentioned software MBR tool have been widely used [61,80].

However, the development of new publicly available and commercial software solutions has encouraged the introduction and establishment of the DIA strategy in many proteomics platforms. In DIA mode, all MS1 precursors within a m/z range of interest are sequentially selected and fragmented at the MS2 level using isolation windows of different widths. It thus offers potentially deeper coverage of the data, decreasing the need for offline fractionation. As DIA does not suffer from the stochastic identifications of peptides that DDA suffers from, cross-sample comparisons in large cohorts are thus made much easier. Because of the complex deconvoluting processes of the fragmentation spectra, DIA is currently used for LF- and SILAC-spiked samples only. Originally, experimentally derived DDA run-spectral libraries were necessary to facilitate DIA spectral deconvolution. However, some current DIA applications that are discussed below (see Table 1) allow spectral analysis without their use.

Recent reports have shown that TMT-DDA methodology provides an excellent workflow to study proteomes and phosphoproteomes in depth. A TMT-based quantitative proteomic profiling of human monocyte-derived macrophages and foam cells identified 5146 proteins, among which 1515 and 182 were differentially expressed in macrophages/monocytes and foam cells/macrophages, respectively [81]. A three TMT 11-plex quantitative proteomic and phosphoproteomic analysis of human post-mortem cortex across asymptomatic phase Alzheimer's disease, symptomatic Alzheimer's disease, and healthy individuals identified 11,378 protein groups and 51,736 phosphopeptides [82]. However, DIA-based approaches that do not require expensive labels and time-consuming fractionation steps have become a powerful alternative for both proteomic and phosphoproteomics characterization [8,73,83]. A recent DIA with parallel accumulation-serial fragmentation (PASEF, a mass spectrometry technique that enables hundreds of MS/MS events per second at full sensitivity) study identified over 7700 proteins in HeLa cells in 44 min with quadruplicate single-shot injections and over 35,000 phosphosites after stimulation with epidermal growth factor in triplicate 31 min runs [84].

When TMT quantification is preferred, the synchronous precursor selection (SPS) MS3 technology in Orbitrap Tribrid mass spectrometers can be used to obtain a higher accuracy than the one provided by MS2 acquisition. Moreover, a real-time search (RTS) step between the MS2 and MS3 scans, which allows an MS3 scan acquisition only if the MS2 spectrum provides a positive peptide identification, can be selected in order to increase the scan rate of data acquisition and match the number of peptide identifications usually observed in MS2 acquisition [85–87].

3. LC-MS-Based Proteomics Data Analysis and Bioinformatics

Raw files with LC-MS data are complex and contain technical and biological information that must be properly visualized to interpret the data and communicate the results in a clear manner. We use visualization tools to examine the quality of the LC-MS data, to analyze the data at the peptide and protein level, and finally to show protein networks. Herein we describe our recommended LC-MS data and bioinformatics software and online tools that do not require any programming skills. However, R and Python scripts for data analysis and visualization are increasingly being used in the LC-MS-based proteomics community because of their capabilities, diversity, graphical quality, and the utility of their libraries. Therefore, programmer participation in proteomics projects is advantageous.

Data analysis and visualization approaches with Python and R programming language can be found in other studies [88,89].

Initially in the data analysis workflow, LC-MS raw files can be analyzed with different software, based on the previous use of the FAIMS interface in the mass spectrometer. Presently, Proteome Discoverer (PD), PEAKS Xpro (Bioinformatics Solutions Inc., Ontario, Canada), and Spectronaut are the only software solutions that can fully handle ion mobility information from DDA and DIA datasets. However, DDA and DIA FAIMS-free LC-MS datasets can be analyzed with publicly available software such as MaxQuant [90,91], MSFragger [92], and DIA-NN [93] (Table 1). These software programs are supported by several online tutorials and workshops that can help beginners analyze their first LC-MS datasets (<https://www.youtube.com/c/MaxQuantChannel> (accessed on 8 August 2022); https://www.maxquant.org/summer_school/ (accessed on 12 August 2022); <https://msfragger.nesvilab.org/>; <https://github.com/vdemichev/DiaNN> (accessed on 22 August 2022)). DIA data analysis without the need for experimentally derived spectral libraries can be performed with DIA-NN, Spectronaut, and PEAKS Xpro.

There are several available platforms to clean and process LC-MS data. However, we recommend Perseus because of its diverse features including normalization, statistical testing, protein interaction, gene ontology (GO) enrichment, PTM analysis, machine learning, data visualization, and many more, thanks to an increasing number of plugins that are being continuously created by Perseus developers and others [94–96]. GO enrichment, pathways enrichment, and protein network analyses can be performed with tools other than the Perseus platform (Table 1). We recommend the sequential use of Enrichr-STRING-Cytoscape to investigate the effects of regulated proteins and visualize their interactions [97–99].

More experienced researchers can use the multiple LC-MS data management tools at OpenMS [100] and full analysis workflows at the Trans-Proteomics Pipeline (TPP; [101]). PTM studies using LC-MS-based phosphoproteomics can benefit from the use of WebLogo and IceLogo to create a graphical representation of the alignment of multiple amino acid sequences [102,103], KSEA and Kinact for kinase prediction [104–106], and Omnipath for kinase–substrate relationships [107].

Table 1. Recommended software and online applications * to analyze LC-MS-based proteomics data.

| LC-MS Data | Data Processing | Gene Ontology (GO)/Pathway Analysis | Protein Networks | Visualization Tools |
|--|--|--|--|--|
| MaxQuant [90,91] MSFragger [92] DIA-NN [93] Proteome Discoverer (F) ¹ Mascot Distiller/Server ¹ Spectronaut (F) ¹ PEAKS Xpro (F) ¹ MSStats [117] Progenesis QI for Proteomics ¹ | Perseus [94,95] Prostar [108] Proteome Discoverer ¹ Qlucore ¹ | Enrichr [97] A GO tool [109] Reactome [111] DAVID [114] InnateDB [115] FunRich [116] QIAGEN IPA ¹ | STRING [98] Omnipath [107] PINA [112,113] Perseus | Cytoscape [99] OpenPIP [110] Perseus |

* Recommended software and applications are based on our own experience and the practice of renowned proteomics laboratories; ¹ Software commercially available; F represents software that can analyze high-field asymmetric waveform ion mobility spectrometry (FAIMS) data.

4. Artificial Intelligence Strategies on Proteomics Data

Artificial intelligence (AI) is an area of computer science that can predict or classify objects or events driven by available data [118]. Through its algorithms, machine learning (ML), a discipline of AI, learns from data and makes predictions without being explicitly programmed.

LC-MS-based discovery proteomics studies that usually result in the identification and quantification of thousands of proteins must overcome the challenge of determining criteria or a strategy to prioritize biomarker candidates at the end of the workflow. Although statistical significance and fold change are the most frequent criteria when comparing groups, supervised ML is becoming recognized as a powerful approach to prioritizing biomarker candidates according to their performance in predicting the phenotype outcome [119,120]. LC-MS-based proteomics researchers are learning ML techniques such as logistic regression (LR), random forest (RF), K-Nearest Neighbors (KNN), and support vector machine (SVM). LR is a technique borrowed by ML from statistics, and it is used to predict the probability of a binary event occurring (e.g., acquiring ovarian cancer or not; sensitive or resistant to treatment). It has successfully associated proteomic subtypes with the risk of gastric lesion progression and identified salivary proteomic biomarkers for oral cancer screening [121,122]. RF models comprise multiple decision trees. A decision tree algorithm (typically the classification and regression tree, CART) is characterized by decision tree nodes to question the data until the leaf node is reached and the best split to subset the data is achieved. This modeling strategy has been successfully applied in the detection of biomarkers for prostate cancer progression and prediction of lung cancer from control cases and other tumors [123,124]. KNN is one of the simplest ML algorithms. It stores all the available data and classifies a new incoming data point based on the similarity with a well-suited category. KNN algorithms in combination with resampling and feature dimensionality reduction methods have recently contributed to cancer prediction using entropy data and improved the classification performance of lung cancer subtypes [125,126]. SVM aims to find a hyperplane that has the maximum margin (i.e., the maximum distance between data points of both classes) and distinctly classifies the data points. SVM works with kernel functions in order to determine the shape of the hyperplane and decision boundaries. This ML approach has been employed in the differentiation of breast cancer subtypes and for early detection of ovarian cancer [127,128]. These ML strategies can be run on the Perseus platform, in R or Python languages or with commercially available software such as MATLAB (MathWorks, Natick, MA, USA), Qlucore (New York, NY, USA), or JMP (SAS, Cary, NC, USA).

However, biomarker prioritization is not the only area of the LC-MS-based proteomics field where AI offers significant benefits. Despite the ongoing development of faster and more sensitive mass spectrometers by several companies, proteomics data never reach the completeness achieved by sequencing-based methodologies [129]. Therefore, it is expected that the AI contributions to several steps of the LC-MS-based proteomics workflow can substantially improve data quality and data interpretation. The impact of AI strategies on proteomics processes and data integration has been described in other publications [129,130]. In this section, we describe current approaches for MS2 prediction and peptide identification. To identify peptides at MS2, experimentally measured peptides are matched against ones calculated *in silico* using a sequence-reversed database to control the FDR. A widely used ML algorithm at this stage is the one of Percolator's (available from PD software), which optimizes the number of true matches at a specified FDR by working with multiple peptide sequence features and experimental peptide data [131]. Another great ML tool is MS2PIP, which has achieved excellent correlations (0.9–0.95 Pearson correlation coefficient) between experimental and predicted spectra on different project datasets by considering the chemical properties of amino acids for spectral prediction [132,133].

Nonetheless, given the great number of peptides that can be obtained, especially from complex proteomes and of MS2 spectra available in public repositories such as PRIDE, Peptide Atlas, MassIVE, JPOST, iPROX, and Panorama within the ProteomeXchange Consortium (<http://www.proteomexchange.org/> (accessed on 1 September 2022)) [134–140], the deep learning (DL) methodology is becoming more and more popular for MS2 prediction in our research community. As the fragmentation of a peptide bond depends not only on the adjacent amino acids but also on those far away from the bond, deep neural networks represent an ideal strategy to map the long-term sequential dependencies [129].

Popular neural-network-based DL software includes Prosit, whose models include different collision energies for peptide fragmentation [141], and DeepMass:Prism, which additionally takes into account the fragmentation energy and works with DDA and DIA data [142].

On the other hand, DL models are successful at detecting LC-MS features, at assessing spectral quality for identification, and at predicting the likeliest peptides and their corresponding proteins [129,130,143–146]. Moreover, DL strategies are becoming useful in the de novo peptide sequencing approach that aims to determine amino acid sequences using fragmentation data and without prior database knowledge. DeepNovo software is based on “automatically generating a description for an image”, where “image” represents MS2 spectra with intensity and mass/charge data. While convolutional neural networks (CNNs) are used to encode the “image”, long short-term memory (LSTM) recurrent neural networks (RNNs) are employed to describe the content of the “image” acquired in the DDA and DIA mode [147–151]. Additionally, DL algorithms are used to build classifiers that discriminate decoys and targets during the peptide identification process with DIA data (e.g., DIA-NN; [93]) and for the intensity-based rescoring of Sequest HT search engine results with the INFERYS workflow on the commercial PD platform [152]. At the last stage of the MS data analysis, CNN models can be run in protein inference strategies such as DeepPep, which quantifies the change in probabilistic score of peptide-spectrum matches (PSMs) in the presence or absence of a specific protein, hence selecting candidate proteins with the largest impact on the peptide profile [153].

Regarding the expertise requirements for including AI techniques in proteomics workflows, while ML approaches can be used by experienced proteomics researchers, the incorporation of DL algorithms requires a close collaboration with AI experts. This reinforces the complexity of MS-created data and the necessity to encourage the greater involvement of bioinformaticians and biomathematicians in LC-MS-based proteomics projects.

5. Fever of Single-Cell LC-MS-Based Proteomics

The study of the proteome of single cells has become an attractive solution to understanding the molecular basis of cell-specific functions and how cell types might respond to different stimuli. However, single-cell LC-MS-based proteomics must face several challenges associated with the low amount of protein material found in single cells (~150 pg) and with the proteome dynamic range expanding over several orders of magnitude [76,154]. It is indeed a technological scenario that requires much more expertise than that needed to perform bulk proteomics analyses.

There are several robotized and miniaturized sample preparation strategies that have been applied to the small-cell population or single-cell level. These include nanodroplet processing in one pot for trace samples (nanoPOTS) of glass chips with hydrophilic pedestals surrounded by a hydrophobic surface to serve as nanodroplet reaction vessels [155], nanoliter-scale oil-air-droplet (OAD) chips consisting of a stationary nanoliter microreactor with an oil-air-droplet sandwich structure [156], and commercial solutions such as cellenONE and proteoCHIP equipment (Cellenion, Lyon, France) that provide single cell sorting, isolation, and nanoliter acoustic dispensing for further sample processing including TMT labeling [157]. The latter approach identified approximately 2000 protein groups across 158 multiplexed single-cell samples from HeLa and HEK293T cultures. However, quantification of single-cell proteomes using the TMT approach was complicated by isotope cross talk between single-cell sample channels and the carrier channel, which contains peptides from one to several hundred cells and is employed to reduce the assay sensitivity from 10- to 200-fold [158,159]. Therefore, although MS sensitivity has significantly improved of late, allowing the use of a 25-cell sample in the carrier channel [160], LF procedures, and alternative methods for data acquisition such as DIA, which has fundamentally higher data completeness than DDA, are becoming more and more prevalent. The combination of PASEF and DIA has recently led to the quantification of up to 2000 proteins per single HeLa cell in a cell-cycle arrest experiment [161].

The proteomics strategies using isolated single cells described above become less attractive when tissues are under study. Procedures that analyze isolated cells from tissues will miss data regarding cell context, which is essential to fully understand cellular functions, cell-to-cell interactions (especially those between normal and cancer cells), and tissue heterogeneity. In order to retain their natural neighborhood, cells have been isolated by automated laser microdissection within a spatial region using a new approach called deep visual proteomics (DVP), which combines AI-driven image-based segmentation and classification for the analysis of cells showing a specific antibody-based bioimage [162,163].

6. Conclusions and Prospects

In cancer research, oncologists are aware of how LC-MS-based proteomics can significantly contribute to preclinical drug discovery. Nonetheless, multi-step workflows employed in LC-MS-based proteomics and the multiple methodological choices available for each of them might be overwhelming for researchers new to this technology. We here recommend the use of bead-based sample preparation followed by DIA with LFQ methodologies or DDA with TMT quantification for broad proteome and phosphoproteome characterizations. LC-MS-based applications in the clinic require fast and robust workflows. Therefore, bead-based sample preparation, phosphopeptide enrichment by the Ti-IMAC HP technique, and DIA represent an attractive protocol package for the characterization of cancer samples. These workflows have recently been applied to the analysis of different types of cancer samples providing, in some examples, 7000 phosphosites from FFPE lung biopsies with limited tissue amounts and 2103 proteins from plasma samples of breast cancer patients [164,165].

The study of patient samples from large cancer cohorts will benefit from the growing availability of efficient automated sample preparation methods that also enable PTM enrichment and improve the reproducibility of the data. Moreover, the introduction of new, robust, and fast chromatographic technology (Evosep, Odense, Denmark) greatly contributes to the feasibility of large cancer studies. Although sample preparation and data acquisition could be sequentially implemented in hospital laboratories, we must still put more effort into the development and engagement of bioinformatics solutions that provide a fast and easy interpretation of the proteomic data. Proteomics pipelines based on efficient workflows will be crucial in precision medicine decisions.

Author Contributions: Conceptualization, M.H.-V.; methodology, M.H.-V.; software, M.H.-V.; formal analysis, P.C.-R.; investigation, M.H.-V. and P.C.-R.; resources, M.H.-V.; data curation, M.H.-V. and P.C.-R.; writing—original draft preparation, M.H.-V. and P.C.-R.; writing—review and editing, M.H.-V., F.S. and P.C.-R.; visualization, M.H.-V. and P.C.-R.; supervision, M.H.-V.; project administration, M.H.-V.; funding acquisition, F.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Research Council of Norway INFRASTRUKTUR-program (project number: 295910).

Acknowledgments: We thank Stacey Dmello for providing us with the HeLa cells in this study. Mass-spectrometry-based proteomic analyses were performed by the Proteomics Unit at the University of Bergen (PROBE). We thank Olav Mjaavatten for excellent technical support.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Martinez-Val, A.; Guzman, U.H.; Olsen, J.V. Obtaining Complete Human Proteomes. *Annu. Rev. Genom. Hum. Genet.* **2022**, *23*, 99–121. [[CrossRef](#)] [[PubMed](#)]
2. Toby, T.K.; Fornelli, L.; Srzentic, K.; DeHart, C.J.; Levitsky, J.; Friedewald, J.; Kelleher, N.L. A comprehensive pipeline for translational top-down proteomics from a single blood draw. *Nat. Protoc.* **2019**, *14*, 119–152. [[CrossRef](#)] [[PubMed](#)]
3. van Bentum, M.; Selbach, M. An Introduction to Advanced Targeted Acquisition Methods. *Mol. Cell. Proteom.* **2021**, *20*, 100165. [[CrossRef](#)]

4. Lee, H.; Kim, S.I. Review of Liquid Chromatography-Mass Spectrometry-Based Proteomic Analyses of Body Fluids to Diagnose Infectious Diseases. *Int. J. Mol. Sci.* **2022**, *23*, 2187. [[CrossRef](#)]
5. Martelli, C.; Iavarone, F.; D'Angelo, L.; Arba, M.; Vincenzoni, F.; Inserra, I.; Delfino, D.; Rossetti, D.V.; Caretto, M.; Massimi, L.; et al. Integrated proteomic platforms for the comparative characterization of medulloblastoma and pilocytic astrocytoma pediatric brain tumors: A preliminary study. *Mol. Biosyst.* **2015**, *11*, 1668–1683. [[CrossRef](#)] [[PubMed](#)]
6. Borrás, E.; Sabido, E. What is targeted proteomics? A concise revision of targeted acquisition and targeted data analysis in mass spectrometry. *Proteomics* **2017**, *17*, 17–18. [[CrossRef](#)]
7. Ntai, I.; Fornelli, L.; DeHart, C.J.; Hutton, J.E.; Doubleday, P.F.; LeDuc, R.D.; van Nispen, A.J.; Fellers, R.T.; Whiteley, G.; Boja, E.S.; et al. Precise characterization of KRAS4b proteoforms in human colorectal cells and tumors reveals mutation/modification cross-talk. *Proc. Natl. Acad. Sci. USA* **2018**, *115*, 4140–4145. [[CrossRef](#)]
8. Bekker-Jensen, D.B.; Martinez-Val, A.; Steigerwald, S.; Ruther, P.; Fort, K.L.; Arrey, T.N.; Harder, A.; Makarov, A.; Olsen, J.V. A Compact Quadrupole-Orbitrap Mass Spectrometer with FAIMS Interface Improves Proteome Coverage in Short LC Gradients. *Mol. Cell. Proteom.* **2020**, *19*, 716–729. [[CrossRef](#)]
9. Ni, M.; Zhou, J.; Zhu, Z.; Yuan, J.; Gong, W.; Zhu, J.; Zheng, Z.; Zhao, H. A Novel Classifier Based on Urinary Proteomics for Distinguishing Between Benign and Malignant Ovarian Tumors. *Front. Cell Dev. Biol.* **2021**, *9*, 712196. [[CrossRef](#)]
10. Wu, J.; Zhang, J.; Wei, J.; Zhao, Y.; Gao, Y. Urinary biomarker discovery in gliomas using mass spectrometry-based clinical proteomics. *Chin. Neurosurg. J.* **2020**, *6*, 11. [[CrossRef](#)]
11. Nakayasu, E.S.; Gritsenko, M.; Piehowski, P.D.; Gao, Y.; Orton, D.J.; Schepmoes, A.A.; Fillmore, T.L.; Frohnert, B.I.; Rewers, M.; Krischer, J.P.; et al. Tutorial: Best practices and considerations for mass-spectrometry-based protein biomarker discovery and validation. *Nat. Protoc.* **2021**, *16*, 3737–3760. [[CrossRef](#)] [[PubMed](#)]
12. Bonizzi, G.Z.L.; Capra, M.; Cassi, C.; Taliento, G.; Ivanova, M.; Guerini-Rocco, E.; Fumagali, M.; Monturano, M.; Albini, A.; Viale, G.; et al. Standard operating procedures for biobank in oncology. *Front. Mol. Biosci.* **2022**, *9*. [[CrossRef](#)]
13. Greco, V.; Piras, C.; Pieroni, L.; Urbani, A. Direct Assessment of Plasma/Serum Sample Quality for Proteomics Biomarker Investigation. *Methods Mol. Biol.* **2017**, *1619*, 3–21. [[CrossRef](#)] [[PubMed](#)]
14. Hernandez-Valladares, M.; Aasebo, E.; Mjaavatten, O.; Vaudel, M.; Bruserud, O.; Berven, F.; Selheim, F. Reliable FASP-based procedures for optimal quantitative proteomic and phosphoproteomic analysis on samples from acute myeloid leukemia patients. *Biol. Proced. Online* **2016**, *18*, 13. [[CrossRef](#)]
15. Aasebo, E.; Brenner, A.K.; Hernandez-Valladares, M.; Birkeland, E.; Berven, F.S.; Selheim, F.; Bruserud, O. Proteomic Comparison of Bone Marrow Derived Osteoblasts and Mesenchymal Stem Cells. *Int. J. Mol. Sci.* **2021**, *22*, 5665. [[CrossRef](#)] [[PubMed](#)]
16. Aasebo, E.; Brenner, A.K.; Hernandez-Valladares, M.; Birkeland, E.; Mjaavatten, O.; Reikvam, H.; Selheim, F.; Berven, F.S.; Bruserud, O. Patient Heterogeneity in Acute Myeloid Leukemia: Leukemic Cell Communication by Release of Soluble Mediators and Its Effects on Mesenchymal Stem Cells. *Diseases* **2021**, *9*, 74. [[CrossRef](#)]
17. Aasebo, E.; Mjaavatten, O.; Vaudel, M.; Farag, Y.; Selheim, F.; Berven, F.; Bruserud, O.; Hernandez-Valladares, M. Freezing effects on the acute myeloid leukemia cell proteome and phosphoproteome revealed using optimal quantitative workflows. *J. Proteom.* **2016**, *145*, 214–225. [[CrossRef](#)]
18. Dapic, I.; Uwugiaren, N.; Kers, J.; Mohammed, Y.; Goodlett, D.R.; Corthals, G. Evaluation of Fast and Sensitive Proteome Profiling of FF and FFPE Kidney Patient Tissues. *Molecules* **2022**, *27*, 1137. [[CrossRef](#)]
19. Dressler, F.F.; Schoenfeld, J.; Revyakina, O.; Vogele, D.; Kiefer, S.; Kirfel, J.; Gemoll, T.; Perner, S. Systematic evaluation and optimization of protein extraction parameters in diagnostic FFPE specimens. *Clin. Proteom.* **2022**, *19*, 10. [[CrossRef](#)]
20. Neset, L.; Takayidza, G.; Berven, F.S.; Hernandez-Valladares, M. Comparing Efficiency of Lysis Buffer Solutions and Sample Preparation Methods for Liquid Chromatography-Mass Spectrometry Analysis of Human Cells and Plasma. *Molecules* **2022**, *27*, 3390. [[CrossRef](#)]
21. Buczak, K.; Kirkpatrick, J.M.; Truckenmueller, F.; Santinha, D.; Ferreira, L.; Roessler, S.; Singer, S.; Beck, M.; Ori, A. Spatially resolved analysis of FFPE tissue proteomes by quantitative mass spectrometry. *Nat. Protoc.* **2020**, *15*, 2956–2979. [[CrossRef](#)] [[PubMed](#)]
22. Cao, X.; Sandberg, A.; Araujo, J.E.; Cvetkovski, F.; Berglund, E.; Eriksson, L.E.; Pernemalm, M. Evaluation of Spin Columns for Human Plasma Depletion to Facilitate MS-Based Proteomics Analysis of Plasma. *J. Proteome Res.* **2021**, *20*, 4610–4620. [[CrossRef](#)] [[PubMed](#)]
23. Keshishian, H.; Burgess, M.W.; Specht, H.; Wallace, L.; Clauser, K.R.; Gillette, M.A.; Carr, S.A. Quantitative, multiplexed workflow for deep analysis of human blood plasma and biomarker discovery by mass spectrometry. *Nat. Protoc.* **2017**, *12*, 1683–1701. [[CrossRef](#)] [[PubMed](#)]
24. Kverneland, A.H.; Ostergaard, O.; Emdal, K.B.; Svane, I.M.; Olsen, J.V. Differential ultracentrifugation enables deep plasma proteomics through enrichment of extracellular vesicles. *Proteomics* **2022**, e2200039. [[CrossRef](#)] [[PubMed](#)]
25. Blume, J.E.; Manning, W.C.; Troiano, G.; Hornburg, D.; Figa, M.; Hesterberg, L.; Platt, T.L.; Zhao, X.; Cuaresma, R.A.; Everley, P.A.; et al. Rapid, deep and precise profiling of the plasma proteome with multi-nanoparticle protein corona. *Nat. Commun.* **2020**, *11*, 3662. [[CrossRef](#)]
26. Ferdosi, S.; Stukalov, A.; Hasan, M.; Tangeysh, B.; Brown, T.R.; Wang, T.; Elgierari, E.M.; Zhao, X.; Huang, Y.; Alavi, A.; et al. Enhanced Competition at the Nano-Bio Interface Enables Comprehensive Characterization of Protein Corona Dynamics and Deep Coverage of Proteomes. *Adv. Mater.* **2022**, *34*, e2206008. [[CrossRef](#)]

27. Levin, Y. The role of statistical power analysis in quantitative proteomics. *Proteomics* **2011**, *11*, 2565–2567. [[CrossRef](#)]
28. Alexovic, M.; Sabo, J.; Longuespee, R. Microproteomic sample preparation. *Proteomics* **2021**, *21*, e2000318. [[CrossRef](#)]
29. Varnavides, G.; Mader, M.; Anrather, D.; Hartl, N.; Reiter, W.; Hartl, M. In Search of a Universal Method: A Comparative Survey of Bottom-Up Proteomics Sample Preparation Methods. *J. Proteome Res.* **2022**, *21*, 2397–2411. [[CrossRef](#)]
30. Foster, L.J.; De Hoog, C.L.; Mann, M. Unbiased quantitative proteomics of lipid rafts reveals high specificity for signaling factors. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 5813–5818. [[CrossRef](#)]
31. Kelstrup, C.D.; Jersie-Christensen, R.R.; Batth, T.S.; Arrey, T.N.; Kuehn, A.; Kellmann, M.; Olsen, J.V. Rapid and deep proteomes by faster sequencing on a benchtop quadrupole ultra-high-field Orbitrap mass spectrometer. *J. Proteome Res.* **2014**, *13*, 6187–6195. [[CrossRef](#)] [[PubMed](#)]
32. Myers, S.A.; Rhoads, A.; Cocco, A.R.; Peckner, R.; Haber, A.L.; Schweitzer, L.D.; Krug, K.; Mani, D.R.; Clauser, K.R.; Rozenblatt-Rosen, O.; et al. Streamlined Protocol for Deep Proteomic Profiling of FAC-sorted Cells and Its Application to Freshly Isolated Murine Immune Cells. *Mol. Cell. Proteom.* **2019**, *18*, 995–1009. [[CrossRef](#)] [[PubMed](#)]
33. Sialana, F.J.; Roumeliotis, T.I.; Bouguenina, H.; Chan Wah Hak, L.; Wang, H.; Caldwell, J.; Collins, I.; Chopra, R.; Choudhary, J.S. SimPLIT: Simplified Sample Preparation for Large-Scale Isobaric Tagging Proteomics. *J. Proteome Res.* **2022**, *21*, 1842–1856. [[CrossRef](#)] [[PubMed](#)]
34. Manza, L.L.; Stamer, S.L.; Ham, A.J.; Codreanu, S.G.; Liebler, D.C. Sample preparation and digestion for proteomic analyses using spin filters. *Proteomics* **2005**, *5*, 1742–1745. [[CrossRef](#)] [[PubMed](#)]
35. Wisniewski, J.R.; Zougman, A.; Nagaraj, N.; Mann, M. Universal sample preparation method for proteome analysis. *Nat. Methods* **2009**, *6*, 359–362. [[CrossRef](#)] [[PubMed](#)]
36. Wisniewski, J.R.; Mann, M. Consecutive proteolytic digestion in an enzyme reactor increases depth of proteomic and phospho-proteomic analysis. *Anal. Chem.* **2012**, *84*, 2631–2637. [[CrossRef](#)] [[PubMed](#)]
37. Wisniewski, J.R.; Rakus, D. Multi-enzyme digestion FASP and the ‘Total Protein Approach’-based absolute quantification of the *Escherichia coli* proteome. *J. Proteom.* **2014**, *109*, 322–331. [[CrossRef](#)]
38. Wisniewski, J.R.; Zougman, A.; Mann, M. Combination of FASP and StageTip-based fractionation allows in-depth analysis of the hippocampal membrane proteome. *J. Proteome Res.* **2009**, *8*, 5674–5678. [[CrossRef](#)]
39. Hughes, C.S.; Moggridge, S.; Muller, T.; Sorensen, P.H.; Morin, G.B.; Krijgsveld, J. Single-pot, solid-phase-enhanced sample preparation for proteomics experiments. *Nat. Protoc.* **2019**, *14*, 68–85. [[CrossRef](#)]
40. Hayoun, K.; Gouveia, D.; Grenga, L.; Pible, O.; Armengaud, J.; Alpha-Bazin, B. Evaluation of Sample Preparation Methods for Fast Proteotyping of Microorganisms by Tandem Mass Spectrometry. *Front. Microbiol.* **2019**, *10*, 1985. [[CrossRef](#)]
41. Sielaff, M.; Kuharev, J.; Bohn, T.; Hahlbrock, J.; Bopp, T.; Tenzer, S.; Distler, U. Evaluation of FASP, SP3, and iST Protocols for Proteomic Sample Preparation in the Low Microgram Range. *J. Proteome Res.* **2017**, *16*, 4060–4072. [[CrossRef](#)] [[PubMed](#)]
42. Waas, M.; Pereckas, M.; Jones Lipinski, R.A.; Ashwood, C.; Gundry, R.L. SP2: Rapid and Automatable Contaminant Removal from Peptide Samples for Proteomic Analyses. *J. Proteome Res.* **2019**, *18*, 1644–1656. [[CrossRef](#)] [[PubMed](#)]
43. Batth, T.S.; Tollenaere, M.X.; Ruther, P.; Gonzalez-Franquesa, A.; Prabhakar, B.S.; Bekker-Jensen, S.; Deshmukh, A.S.; Olsen, J.V. Protein Aggregation Capture on Microparticles Enables Multipurpose Proteomics Sample Preparation. *Mol. Cell. Proteom.* **2019**, *18*, 1027–1035. [[CrossRef](#)] [[PubMed](#)]
44. Franciosa, G.; Smits, J.G.A.; Minuzzo, S.; Martinez-Val, A.; Indraccolo, S.; Olsen, J.V. Proteomics of resistance to Notch1 inhibition in acute lymphoblastic leukemia reveals targetable kinase signatures. *Nat. Commun.* **2021**, *12*, 2507. [[CrossRef](#)]
45. Ruther, P.L.; Husic, I.M.; Bangsgaard, P.; Gregersen, K.M.; Pantmann, P.; Carvalho, M.; Godinho, R.M.; Friedl, L.; Cascalheira, J.; Taurozzi, A.J.; et al. SPIN enables high throughput species identification of archaeological bone by proteomics. *Nat. Commun.* **2022**, *13*, 2458. [[CrossRef](#)] [[PubMed](#)]
46. Muller, T.; Kalxdorf, M.; Longuespee, R.; Kazdal, D.N.; Stenzinger, A.; Krijgsveld, J. Automated sample preparation with SP3 for low-input clinical proteomics. *Mol. Syst. Biol.* **2020**, *16*, e9111. [[CrossRef](#)]
47. Moggridge, S.; Sorensen, P.H.; Morin, G.B.; Hughes, C.S. Extending the Compatibility of the SP3 Paramagnetic Bead Processing Approach for Proteomics. *J. Proteome Res.* **2018**, *17*, 1730–1740. [[CrossRef](#)]
48. Dagley, L.F.; Infusini, G.; Larsen, R.H.; Sandow, J.J.; Webb, A.I. Universal Solid-Phase Protein Preparation (USP(3)) for Bottom-up and Top-down Proteomics. *J. Proteome Res.* **2019**, *18*, 2915–2924. [[CrossRef](#)]
49. van der Pan, K.; Kassem, S.; Khatri, I.; de Ru, A.H.; Janssen, G.M.C.; Tjokrodrijo, R.T.N.; Al Makindji, F.; Stavrakaki, E.; de Jager, A.L.; Naber, B.A.E.; et al. Quantitative proteomics of small numbers of closely-related cells: Selection of the optimal method for a clinical setting. *Front. Med.* **2022**, *9*, 997305. [[CrossRef](#)]
50. Tape, C.J.; Worboys, J.D.; Sinclair, J.; Gourlay, R.; Vogt, J.; McMahan, K.M.; Trost, M.; Lauffenburger, D.A.; Lamont, D.J.; Jorgensen, C. Reproducible automated phosphopeptide enrichment using magnetic TiO₂ and Ti-IMAC. *Anal. Chem.* **2014**, *86*, 10296–10302. [[CrossRef](#)]
51. Bantscheff, M.; Schirle, M.; Sweetman, G.; Rick, J.; Kuster, B. Quantitative mass spectrometry in proteomics: A critical review. *Anal. Bioanal. Chem.* **2007**, *389*, 1017–1031. [[CrossRef](#)]
52. Geiger, T.; Wisniewski, J.R.; Cox, J.; Zanivan, S.; Kruger, M.; Ishihama, Y.; Mann, M. Use of stable isotope labeling by amino acids in cell culture as a spike-in standard in quantitative proteomics. *Nat. Protoc.* **2011**, *6*, 147–157. [[CrossRef](#)]
53. Rigbolt, K.T.; Blagoev, B. Proteome-wide quantitation by SILAC. *Methods Mol. Biol.* **2010**, *658*, 187–204. [[CrossRef](#)]

54. Gygi, S.P.; Rist, B.; Gerber, S.A.; Turecek, F.; Gelb, M.H.; Aebersold, R. Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat. Biotechnol.* **1999**, *17*, 994–999. [[CrossRef](#)]
55. Hsu, J.L.; Huang, S.Y.; Chow, N.H.; Chen, S.H. Stable-isotope dimethyl labeling for quantitative proteomics. *Anal. Chem.* **2003**, *75*, 6843–6852. [[CrossRef](#)]
56. Ross, P.L.; Huang, Y.N.; Marchese, J.N.; Williamson, B.; Parker, K.; Hattan, S.; Khainovski, N.; Pillai, S.; Dey, S.; Daniels, S.; et al. Multiplexed protein quantitation in *Saccharomyces cerevisiae* using amine-reactive isobaric tagging reagents. *Mol. Cell. Proteom.* **2004**, *3*, 1154–1169. [[CrossRef](#)]
57. Thompson, A.; Schafer, J.; Kuhn, K.; Kienle, S.; Schwarz, J.; Schmidt, G.; Neumann, T.; Johnstone, R.; Mohammed, A.K.; Hamon, C. Tandem mass tags: A novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS. *Anal. Chem.* **2003**, *75*, 1895–1904. [[CrossRef](#)]
58. Yao, X.; Freas, A.; Ramirez, J.; Demirev, P.A.; Fenselau, C. Proteolytic ¹⁸O labeling for comparative proteomics: Model studies with two serotypes of adenovirus. *Anal. Chem.* **2001**, *73*, 2836–2842. [[CrossRef](#)]
59. Unwin, R.D. Quantification of proteins by iTRAQ. *Methods Mol. Biol.* **2010**, *658*, 205–215. [[CrossRef](#)]
60. Wong, J.W.; Cagney, G. An overview of label-free quantitation methods in proteomics by mass spectrometry. *Methods Mol. Biol.* **2010**, *604*, 273–283. [[CrossRef](#)]
61. Cox, J.; Hein, M.Y.; Luber, C.A.; Paron, I.; Nagaraj, N.; Mann, M. Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. *Mol. Cell. Proteom.* **2014**, *13*, 2513–2526. [[CrossRef](#)]
62. Bielow, C.; Mastrobuoni, G.; Kempa, S. Proteomics Quality Control: Quality Control Software for MaxQuant Results. *J. Proteome Res.* **2016**, *15*, 777–787. [[CrossRef](#)]
63. Yu, S.H.; Kyriakidou, P.; Cox, J. Isobaric Matching between Runs and Novel PSM-Level Normalization in MaxQuant Strongly Improve Reporter Ion-Based Quantification. *J. Proteome Res.* **2020**, *19*, 3945–3954. [[CrossRef](#)]
64. Hutchinson-Bunch, C.; Sanford, J.A.; Hansen, J.R.; Gritsenko, M.A.; Rodland, K.D.; Piehowski, P.D.; Qian, W.J.; Adkins, J.N. Assessment of TMT Labeling Efficiency in Large-Scale Quantitative Proteomics: The Critical Effect of Sample pH. *ACS Omega* **2021**, *6*, 12660–12666. [[CrossRef](#)]
65. Yu, K.; Wang, Z.; Wu, Z.; Tan, H.; Mishra, A.; Peng, J. High-Throughput Profiling of Proteome and Posttranslational Modifications by 16-Plex TMT Labeling and Mass Spectrometry. *Methods Mol. Biol.* **2021**, *2228*, 205–224. [[CrossRef](#)]
66. Zecha, J.; Satpathy, S.; Kanashova, T.; Avanesian, S.C.; Kane, M.H.; Clauser, K.R.; Mertins, P.; Carr, S.A.; Kuster, B. TMT Labeling for the Masses: A Robust and Cost-efficient, In-solution Labeling Approach. *Mol. Cell. Proteom.* **2019**, *18*, 1468–1478. [[CrossRef](#)]
67. Pieroni, L.; Iavarone, F.; Olianas, A.; Greco, V.; Desiderio, C.; Martelli, C.; Manconi, B.; Sanna, M.T.; Messana, I.; Castagnola, M.; et al. Enrichments of post-translational modifications in proteomic studies. *J. Sep. Sci.* **2020**, *43*, 313–336. [[CrossRef](#)]
68. Hernandez-Valladares, M.; Wangen, R.; Berven, F.S.; Guldbrandsen, A. Protein Post-Translational Modification Crosstalk in Acute Myeloid Leukemia Calls for Action. *Curr. Med. Chem.* **2019**, *26*, 5317–5337. [[CrossRef](#)]
69. Thingholm, T.E.; Jensen, O.N.; Robinson, P.J.; Larsen, M.R. SIMAC (sequential elution from IMAC), a phosphoproteomics strategy for the rapid separation of monophosphorylated from multiply phosphorylated peptides. *Mol. Cell. Proteom.* **2008**, *7*, 661–671. [[CrossRef](#)]
70. Thingholm, T.E.; Jorgensen, T.J.; Jensen, O.N.; Larsen, M.R. Highly selective enrichment of phosphorylated peptides using titanium dioxide. *Nat. Protoc.* **2006**, *1*, 1929–1935. [[CrossRef](#)]
71. Thingholm, T.E.; Larsen, M.R. Phosphopeptide Enrichment by Immobilized Metal Affinity Chromatography. *Methods Mol. Biol.* **2016**, *1355*, 123–133. [[CrossRef](#)] [[PubMed](#)]
72. Aasebo, E.; Berven, F.S.; Bartaula-Brevik, S.; Stokowy, T.; Hovland, R.; Vaudel, M.; Doskeland, S.O.; McCormack, E.; Batth, T.S.; Olsen, J.V.; et al. Proteome and Phosphoproteome Changes Associated with Prognosis in Acute Myeloid Leukemia. *Cancers* **2020**, *12*, 709. [[CrossRef](#)]
73. Bekker-Jensen, D.B.; Bernhardt, O.M.; Högrefe, A.; Martinez-Val, A.; Verbeke, L.; Gandhi, T.; Kelstrup, C.D.; Reiter, L.; Olsen, J.V. Rapid and site-specific deep phosphoproteome profiling by data-independent acquisition without the need for spectral libraries. *Nat. Commun.* **2020**, *11*, 787. [[CrossRef](#)]
74. Manadas, B.; Mendes, V.M.; English, J.; Dunn, M.J. Peptide fractionation in proteomics approaches. *Expert Rev. Proteom.* **2010**, *7*, 655–663. [[CrossRef](#)] [[PubMed](#)]
75. Kulak, N.A.; Pichler, G.; Paron, I.; Nagaraj, N.; Mann, M. Minimal, encapsulated proteomic-sample processing applied to copy-number estimation in eukaryotic cells. *Nat. Methods* **2014**, *11*, 319–324. [[CrossRef](#)] [[PubMed](#)]
76. Bekker-Jensen, D.B.; Kelstrup, C.D.; Batth, T.S.; Larsen, S.C.; Haldrup, C.; Bramsen, J.B.; Sorensen, K.D.; Hoyer, S.; Orntoft, T.F.; Andersen, C.L.; et al. An Optimized Shotgun Strategy for the Rapid Generation of Comprehensive Human Proteomes. *Cell Syst.* **2017**, *4*, 587–599. [[CrossRef](#)]
77. Branca, R.M.; Orre, L.M.; Johansson, H.J.; Granholm, V.; Huss, M.; Perez-Bercoff, A.; Forshed, J.; Kall, L.; Lehtio, J. HiRIEF LC-MS enables deep proteome coverage and unbiased proteogenomics. *Nat. Methods* **2014**, *11*, 59–62. [[CrossRef](#)]
78. Lehtio, J.; Arslan, T.; Siavelis, I.; Pan, Y.; Socciaelli, F.; Berkovska, O.; Umer, H.M.; Mermelekas, G.; Pirmoradian, M.; Jonsson, M.; et al. Proteogenomics of non-small cell lung cancer reveals molecular subtypes associated with specific therapeutic targets and immune evasion mechanisms. *Nat. Cancer* **2021**, *2*, 1224–1242. [[CrossRef](#)]
79. Panizza, E.; Branca, R.M.M.; Oliviusson, P.; Orre, L.M.; Lehtio, J. Isoelectric point-based fractionation by HiRIEF coupled to LC-MS allows for in-depth quantitative analysis of the phosphoproteome. *Sci. Rep.* **2017**, *7*, 4513. [[CrossRef](#)]

80. Hodge, K.; Have, S.T.; Hutton, L.; Lamond, A.I. Cleaning up the masses: Exclusion lists to reduce contamination with HPLC-MS/MS. *J. Proteom.* **2013**, *88*, 92–103. [[CrossRef](#)]
81. Zhang, Y.; Fu, Y.; Jia, L.; Zhang, C.; Cao, W.; Alam, N.; Wang, R.; Wang, W.; Bai, L.; Zhao, S.; et al. TMT-based quantitative proteomic profiling of human monocyte-derived macrophages and foam cells. *Proteome Sci.* **2022**, *20*, 1. [[CrossRef](#)] [[PubMed](#)]
82. Ping, L.; Kundinger, S.R.; Duong, D.M.; Yin, L.; Gearing, M.; Lah, J.J.; Levey, A.I.; Seyfried, N.T. Global quantitative analysis of the human brain proteome and phosphoproteome in Alzheimer's disease. *Sci. Data* **2020**, *7*, 315. [[CrossRef](#)] [[PubMed](#)]
83. Frohlich, K.; Brombacher, E.; Fahrner, M.; Vogeled, D.; Kook, L.; Pinter, N.; Bronsert, P.; Timme-Bronsert, S.; Schmidt, A.; Barenfaller, K.; et al. Benchmarking of analysis strategies for data-independent acquisition proteomics using a large-scale dataset comprising inter-patient heterogeneity. *Nat. Commun.* **2022**, *13*, 2622. [[CrossRef](#)] [[PubMed](#)]
84. Skowronek, P.; Thielert, M.; Voytik, E.; Tanzer, M.C.; Hansen, F.M.; Willems, S.; Karayel, O.; Brunner, A.D.; Meier, F.; Mann, M. Rapid and In-Depth Coverage of the (Phospho-)Proteome With Deep Libraries and Optimal Window Design for dia-PASEF. *Mol. Cell. Proteom.* **2022**, *21*, 100279. [[CrossRef](#)] [[PubMed](#)]
85. Erickson, B.K.; Mintseris, J.; Schweppe, D.K.; Navarrete-Perea, J.; Erickson, A.R.; Nusinow, D.P.; Paulo, J.A.; Gygi, S.P. Active Instrument Engagement Combined with a Real-Time Database Search for Improved Performance of Sample Multiplexing Workflows. *J. Proteome Res.* **2019**, *18*, 1299–1306. [[CrossRef](#)]
86. McAlister, G.C.; Nusinow, D.P.; Jedrychowski, M.P.; Wuhr, M.; Huttlin, E.L.; Erickson, B.K.; Rad, R.; Haas, W.; Gygi, S.P. MultiNotch MS3 enables accurate, sensitive, and multiplexed detection of differential expression across cancer cell line proteomes. *Anal. Chem.* **2014**, *86*, 7150–7158. [[CrossRef](#)]
87. Schweppe, D.K.; Eng, J.K.; Yu, Q.; Bailey, D.; Rad, R.; Navarrete-Perea, J.; Huttlin, E.L.; Erickson, B.K.; Paulo, J.A.; Gygi, S.P. Full-Featured, Real-Time Database Searching Platform Enables Fast and Accurate Multiplexed Quantitative Proteomics. *J. Proteome Res.* **2020**, *19*, 2026–2034. [[CrossRef](#)]
88. Gatto, L.; Breckels, L.M.; Naake, T.; Gibb, S. Visualization of proteomics data using R and bioconductor. *Proteomics* **2015**, *15*, 1375–1389. [[CrossRef](#)]
89. Schessner, J.P.; Voytik, E.; Bludau, I. A practical guide to interpreting and generating bottom-up proteomics data visualizations. *Proteomics* **2022**, *22*, e2100103. [[CrossRef](#)]
90. Cox, J.; Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* **2008**, *26*, 1367–1372. [[CrossRef](#)]
91. Sinitcyn, P.; Hamzeiy, H.; Salinas Soto, F.; Itzhak, D.; McCarthy, F.; Wichmann, C.; Steger, M.; Ohmayer, U.; Distler, U.; Kaspar-Schoenefeld, S.; et al. MaxDIA enables library-based and library-free data-independent acquisition proteomics. *Nat. Biotechnol.* **2021**, *39*, 1563–1573. [[CrossRef](#)] [[PubMed](#)]
92. Kong, A.T.; Leprevost, F.V.; Avtonomov, D.M.; Mellacheruvu, D.; Nesvizhskii, A.I. MSFragger: Ultrafast and comprehensive peptide identification in mass spectrometry-based proteomics. *Nat. Methods* **2017**, *14*, 513–520. [[CrossRef](#)] [[PubMed](#)]
93. Demichev, V.; Messner, C.B.; Vernardis, S.I.; Lilley, K.S.; Ralser, M. DIA-NN: Neural networks and interference correction enable deep proteome coverage in high throughput. *Nat. Methods* **2020**, *17*, 41–44. [[CrossRef](#)]
94. Rudolph, J.D.; Cox, J. A Network Module for the Perseus Software for Computational Proteomics Facilitates Proteome Interaction Graph Analysis. *J. Proteome Res.* **2019**, *18*, 2052–2064. [[CrossRef](#)]
95. Tyanova, S.; Temu, T.; Sinitcyn, P.; Carlson, A.; Hein, M.Y.; Geiger, T.; Mann, M.; Cox, J. The Perseus computational platform for comprehensive analysis of (pro)teomics data. *Nat. Methods* **2016**, *13*, 731–740. [[CrossRef](#)] [[PubMed](#)]
96. Yu, S.H.; Ferretti, D.; Schessner, J.P.; Rudolph, J.D.; Borner, G.H.H.; Cox, J. Expanding the Perseus Software for Omics Data Analysis With Custom Plugins. *Curr. Protoc. Bioinform.* **2020**, *71*, e105. [[CrossRef](#)] [[PubMed](#)]
97. Chen, E.Y.; Tan, C.M.; Kou, Y.; Duan, Q.; Wang, Z.; Meirelles, G.V.; Clark, N.R.; Ma'ayan, A. Enrichr: Interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinform.* **2013**, *14*, 128. [[CrossRef](#)]
98. von Mering, C.; Jensen, L.J.; Snel, B.; Hooper, S.D.; Krupp, M.; Foglierini, M.; Jouffre, N.; Huynen, M.A.; Bork, P. STRING: Known and predicted protein-protein associations, integrated and transferred across organisms. *Nucleic Acids Res.* **2005**, *33*, D433–D437. [[CrossRef](#)]
99. Shannon, P.; Markiel, A.; Ozier, O.; Baliga, N.S.; Wang, J.T.; Ramage, D.; Amin, N.; Schwikowski, B.; Ideker, T. Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res.* **2003**, *13*, 2498–2504. [[CrossRef](#)]
100. Rost, H.L.; Sachsenberg, T.; Aiche, S.; Bielow, C.; Weissner, H.; Aicheler, F.; Andreotti, S.; Ehrlich, H.C.; Gutenbrunner, P.; Kenar, E.; et al. OpenMS: A flexible open-source software platform for mass spectrometry data analysis. *Nat. Methods* **2016**, *13*, 741–748. [[CrossRef](#)]
101. Deutsch, E.W.; Mendoza, L.; Shteynberg, D.; Slagel, J.; Sun, Z.; Moritz, R.L. Trans-Proteomic Pipeline, a standardized data processing pipeline for large-scale reproducible proteomics informatics. *Proteom. Clin. Appl.* **2015**, *9*, 745–754. [[CrossRef](#)] [[PubMed](#)]
102. Colaert, N.; Helsens, K.; Martens, L.; Vandekerckhove, J.; Gevaert, K. Improved visualization of protein consensus sequences by iceLogo. *Nat. Methods* **2009**, *6*, 786–787. [[CrossRef](#)] [[PubMed](#)]
103. Crooks, G.E.; Hon, G.; Chandonia, J.M.; Brenner, S.E. WebLogo: A sequence logo generator. *Genome Res.* **2004**, *14*, 1188–1190. [[CrossRef](#)] [[PubMed](#)]

104. Casado, P.; Rodriguez-Prados, J.C.; Cosulich, S.C.; Guichard, S.; Vanhaesebroeck, B.; Joel, S.; Cutillas, P.R. Kinase-substrate enrichment analysis provides insights into the heterogeneity of signaling pathway activation in leukemia cells. *Sci. Signal.* **2013**, *6*, rs6. [[CrossRef](#)] [[PubMed](#)]
105. Rodrigues, C.H.; Ascher, D.B.; Pires, D.E. Kinact: A computational approach for predicting activating missense mutations in protein kinases. *Nucleic Acids Res.* **2018**, *46*, W127–W132. [[CrossRef](#)]
106. Wiredja, D.D.; Koyuturk, M.; Chance, M.R. The KSEA App: A web-based tool for kinase activity inference from quantitative phosphoproteomics. *Bioinformatics* **2017**, *33*, 3489–3491. [[CrossRef](#)]
107. Turei, D.; Valdeolivas, A.; Gul, L.; Palacio-Escat, N.; Klein, M.; Ivanova, O.; Olbei, M.; Gabor, A.; Theis, F.; Modos, D.; et al. Integrated intra- and intercellular signaling knowledge for multicellular omics analysis. *Mol. Syst. Biol.* **2021**, *17*, e9923. [[CrossRef](#)]
108. Wieczorek, S.; Combes, F.; Lazar, C.; Giai Gianetto, Q.; Gatto, L.; Dorffer, A.; Hesse, A.M.; Coute, Y.; Ferro, M.; Bruley, C.; et al. DAPAR & ProStaR: Software to perform statistical analyses in quantitative discovery proteomics. *Bioinformatics* **2017**, *33*, 135–136. [[CrossRef](#)]
109. Scholz, C.; Lyon, D.; Refsgaard, J.C.; Jensen, L.J.; Choudhary, C.; Weinert, B.T. Avoiding abundance bias in the functional annotation of post-translationally modified proteins. *Nat. Methods* **2015**, *12*, 1003–1004. [[CrossRef](#)]
110. Helmy, M.; Mee, M.; Ranjan, A.; Hao, T.; Vidal, M.; Calderwood, M.A.; Luck, K.; Bader, G.D. OpenPIP: An Open-source Platform for Hosting, Visualizing and Analyzing Protein Interaction Data. *J. Mol. Biol.* **2022**, *434*, 167603. [[CrossRef](#)]
111. Gillespie, M.; Jassal, B.; Stephan, R.; Milacic, M.; Rothfels, K.; Senff-Ribeiro, A.; Griss, J.; Sevilla, C.; Matthews, L.; Gong, C.; et al. The reactome pathway knowledgebase 2022. *Nucleic Acids Res.* **2022**, *50*, D687–D692. [[CrossRef](#)] [[PubMed](#)]
112. Du, Y.; Cai, M.; Xing, X.; Ji, J.; Yang, E.; Wu, J. PINA 3.0: Mining cancer interactome. *Nucleic Acids Res.* **2020**, *49*, D1351–D1357. [[CrossRef](#)] [[PubMed](#)]
113. Cowley, M.J.; Pinese, M.; Kassahn, K.S.; Waddell, N.; Pearson, J.V.; Grimmond, S.M.; Biankin, A.V.; Hautaniemi, S.; Wu, J. PINA v2.0: Mining interactome modules. *Nucleic Acids Res.* **2012**, *40*, D862–D865. [[CrossRef](#)] [[PubMed](#)]
114. Sherman, B.T.; Hao, M.; Qiu, J.; Jiao, X.; Baseler, M.W.; Lane, H.C.; Imamichi, T.; Chang, W. DAVID: A web server for functional enrichment analysis and functional annotation of gene lists (2021 update). *Nucleic Acids Res.* **2022**, *50*, W216–W221. [[CrossRef](#)] [[PubMed](#)]
115. Breuer, K.; Foroushani, A.K.; Laird, M.R.; Chen, C.; Sribnaia, A.; Lo, R.; Winsor, G.L.; Hancock, R.E.; Brinkman, F.S.; Lynn, D.J. InnateDB: Systems biology of innate immunity and beyond—recent updates and continuing curation. *Nucleic Acids Res.* **2013**, *41*, D1228–D1233. [[CrossRef](#)]
116. Benito-Martin, A.; Peinado, H. FunRich proteomics software analysis, let the fun begin! *Proteomics* **2015**, *15*, 2555–2556. [[CrossRef](#)] [[PubMed](#)]
117. Choi, M.; Chang, C.Y.; Clough, T.; Broudy, D.; Killeen, T.; MacLean, B.; Vitek, O. MSstats: An R package for statistical analysis of quantitative mass spectrometry-based proteomic experiments. *Bioinformatics* **2014**, *30*, 2524–2526. [[CrossRef](#)]
118. Yin, X.; Liao, H.; Yun, H.; Lin, N.; Li, S.; Xiang, Y.; Ma, X. Artificial intelligence-based prediction of clinical outcome in immunotherapy and targeted therapy of lung cancer. *Semin. Cancer Biol.* **2022**, *86*, 146–159. [[CrossRef](#)]
119. Kawahara, R.; Meirelles, G.V.; Heberle, H.; Domingues, R.R.; Granato, D.C.; Yokoo, S.; Canevarolo, R.R.; Winck, F.V.; Ribeiro, A.C.; Brandao, T.B.; et al. Integrative analysis to select cancer candidate biomarkers to targeted validation. *Oncotarget* **2015**, *6*, 43635–43652. [[CrossRef](#)]
120. Swan, A.L.; Mobasher, A.; Allaway, D.; Liddell, S.; Bacardit, J. Application of machine learning to proteomics data: Classification and biomarker identification in postgenomics biology. *OMICS* **2013**, *17*, 595–610. [[CrossRef](#)]
121. Li, X.; Zheng, N.R.; Wang, L.H.; Li, Z.W.; Liu, Z.C.; Fan, H.; Wang, Y.; Dai, J.; Ni, X.T.; Wei, X.; et al. Proteomic profiling identifies signatures associated with progression of precancerous gastric lesions and risk of early gastric cancer. *EBioMedicine* **2021**, *74*, 103714. [[CrossRef](#)] [[PubMed](#)]
122. Ishikawa, S.; Ishizawa, K.; Tanaka, A.; Kimura, H.; Kitabatake, K.; Sugano, A.; Edamatsu, K.; Ueda, S.; Iino, M. Identification of Salivary Proteomic Biomarkers for Oral Cancer Screening. *Vivo* **2021**, *35*, 541–547. [[CrossRef](#)] [[PubMed](#)]
123. Toth, R.; Schiffmann, H.; Hube-Magg, C.; Buscheck, F.; Hoflmayer, D.; Weidemann, S.; Lebok, P.; Fraune, C.; Minner, S.; Schlomm, T.; et al. Random forest-based modelling to detect biomarkers for prostate cancer progression. *Clin. Epigenetics* **2019**, *11*, 148. [[CrossRef](#)] [[PubMed](#)]
124. Zhang, C.; Leng, W.; Sun, C.; Lu, T.; Chen, Z.; Men, X.; Wang, Y.; Wang, G.; Zhen, B.; Qin, J. Urine Proteome Profiling Predicts Lung Cancer from Control Cases and Other Tumors. *EBioMedicine* **2018**, *30*, 120–128. [[CrossRef](#)]
125. Song, C.; Li, X. Cost-Sensitive KNN Algorithm for Cancer Prediction Based on Entropy Analysis. *Entropy* **2022**, *24*, 253. [[CrossRef](#)] [[PubMed](#)]
126. Wang, C.; Long, Y.; Li, W.; Dai, W.; Xie, S.; Liu, Y.; Zhang, Y.; Liu, M.; Tian, Y.; Li, Q.; et al. Exploratory study on classification of lung cancer subtypes through a combined K-nearest neighbor classifier in breathomics. *Sci. Rep.* **2020**, *10*, 5880. [[CrossRef](#)]
127. Tyanova, S.; Albrechtsen, R.; Kronqvist, P.; Cox, J.; Mann, M.; Geiger, T. Proteomic maps of breast cancer subtypes. *Nat. Commun.* **2016**, *7*, 10259. [[CrossRef](#)]
128. Wu, J.; Ji, Y.; Zhao, L.; Ji, M.; Ye, Z.; Li, S. A Mass Spectrometric Analysis Method Based on PPCA and SVM for Early Detection of Ovarian Cancer. *Comput. Math. Methods Med.* **2016**, *2016*, 6169249. [[CrossRef](#)]
129. Mann, M.; Kumar, C.; Zeng, W.F.; Strauss, M.T. Artificial intelligence for proteomics and biomarker discovery. *Cell Syst.* **2021**, *12*, 759–770. [[CrossRef](#)]

130. Meyer, J.G. Deep learning neural network tools for proteomics. *Cell Rep. Methods* **2021**, *1*, 100003. [[CrossRef](#)]
131. Kall, L.; Canterbury, J.D.; Weston, J.; Noble, W.S.; MacCoss, M.J. Semi-supervised learning for peptide identification from shotgun proteomics datasets. *Nat. Methods* **2007**, *4*, 923–925. [[CrossRef](#)] [[PubMed](#)]
132. Degroeve, S.; Martens, L. MS2PIP: A tool for MS/MS peak intensity prediction. *Bioinformatics* **2013**, *29*, 3199–3203. [[CrossRef](#)] [[PubMed](#)]
133. Xu, R.; Sheng, J.; Bai, M.; Shu, K.; Zhu, Y.; Chang, C. A Comprehensive Evaluation of MS/MS Spectrum Prediction Tools for Shotgun Proteomics. *Proteomics* **2020**, *20*, e1900345. [[CrossRef](#)] [[PubMed](#)]
134. Desiere, F.; Deutsch, E.W.; King, N.L.; Nesvizhskii, A.I.; Mallick, P.; Eng, J.; Chen, S.; Eddes, J.; Loevenich, S.N.; Aebersold, R. The PeptideAtlas project. *Nucleic Acids Res.* **2006**, *34*, D655–D658. [[CrossRef](#)]
135. Perez-Riverol, Y.; Bai, J.; Bandla, C.; Garcia-Seisdedos, D.; Hewapathirana, S.; Kamatchinathan, S.; Kundu, D.J.; Prakash, A.; Frericks-Zipper, A.; Eisenacher, M.; et al. The PRIDE database resources in 2022: A hub for mass spectrometry-based proteomics evidences. *Nucleic Acids Res.* **2022**, *50*, D543–D552. [[CrossRef](#)]
136. Wang, M.; Wang, J.; Carver, J.; Pullman, B.S.; Cha, S.W.; Bandeira, N. Assembling the Community-Scale Discoverable Human Proteome. *Cell Syst.* **2018**, *7*, 412–421. [[CrossRef](#)]
137. Ma, J.; Chen, T.; Wu, S.; Yang, C.; Bai, M.; Shu, K.; Li, K.; Zhang, G.; Jin, Z.; He, F.; et al. iProX: An integrated proteome resource. *Nucleic Acids Res.* **2019**, *47*, D1211–D1217. [[CrossRef](#)]
138. Okuda, S.; Watanabe, Y.; Moriya, Y.; Kawano, S.; Yamamoto, T.; Matsumoto, M.; Takami, T.; Kobayashi, D.; Araki, N.; Yoshizawa, A.C.; et al. jPOSTrepo: An international standard data repository for proteomes. *Nucleic Acids Res.* **2017**, *45*, D1107–D1111. [[CrossRef](#)]
139. Sharma, V.; Eckels, J.; Schilling, B.; Ludwig, C.; Jaffe, J.D.; MacCoss, M.J.; MacLean, B. Panorama Public: A Public Repository for Quantitative Data Sets Processed in Skyline. *Mol. Cell. Proteom.* **2018**, *17*, 1239–1244. [[CrossRef](#)]
140. Vizcaino, J.A.; Deutsch, E.W.; Wang, R.; Csordas, A.; Reisinger, F.; Rios, D.; Dianes, J.A.; Sun, Z.; Farrah, T.; Bandeira, N.; et al. ProteomeXchange provides globally coordinated proteomics data submission and dissemination. *Nat. Biotechnol.* **2014**, *32*, 223–226. [[CrossRef](#)]
141. Gessulat, S.; Schmidt, T.; Zolg, D.P.; Samaras, P.; Schnatbaum, K.; Zerweck, J.; Knaute, T.; Rechenberger, J.; Delanghe, B.; Huhmer, A.; et al. ProSIT: Proteome-wide prediction of peptide tandem mass spectra by deep learning. *Nat. Methods* **2019**, *16*, 509–518. [[CrossRef](#)] [[PubMed](#)]
142. Tiwary, S.; Levy, R.; Gutenbrunner, P.; Salinas Soto, F.; Palaniappan, K.K.; Deming, L.; Berndl, M.; Brant, A.; Cimermanic, P.; Cox, J. High-quality MS/MS spectrum prediction for data-dependent and data-independent acquisition data analysis. *Nat. Methods* **2019**, *16*, 519–525. [[CrossRef](#)] [[PubMed](#)]
143. Ma, C.; Zhu, Z.; Ye, J.; Yang JPei, J.; Xu, S.; Zhou, R.; Yu, C.; Mo, F.; Wen, B.; Liu, S. DeepRT: Deep learning for peptide retention time prediction in proteomics. *arXiv* **2017**, arXiv:1705.05368.
144. Ma, C.; Ren, Y.; Yang, J.; Ren, Z.; Yang, H.; Liu, S. Improved Peptide Retention Time Prediction in Liquid Chromatography through Deep Learning. *Anal. Chem.* **2018**, *90*, 10881–10888. [[CrossRef](#)] [[PubMed](#)]
145. Serrano, G.; Guruceaga, E.; Segura, V. DeepMSPeptide: Peptide detectability prediction using deep learning. *Bioinformatics* **2019**, *36*, 1279–1280. [[CrossRef](#)] [[PubMed](#)]
146. Zohora, F.T.; Rahman, M.Z.; Tran, N.H.; Xin, L.; Shan, B.; Li, M. DeepIso: A Deep Learning Model for Peptide Feature Detection from LC-MS map. *Sci. Rep.* **2019**, *9*, 17168. [[CrossRef](#)]
147. Tran, N.H.; Zhang, X.; Xin, L.; Shan, B.; Li, M. De novo peptide sequencing by deep learning. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, 8247–8252. [[CrossRef](#)]
148. Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural Comput.* **1997**, *9*, 1735–1780. [[CrossRef](#)]
149. Karpathy, A.; Fei-Fei, L. Deep visual-semantic alignments for generating image descriptions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 3128–3137.
150. Vinyals, O.; Toshev, A.; Bengio, S.; Erhan, D. Show and tell: A neural image caption generator. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 3156–3164.
151. Tran, N.H.; Qiao, R.; Xin, L.; Chen, X.; Liu, C.; Zhang, X.; Shan, B.; Ghodsi, A.; Li, M. Deep learning enables de novo peptide sequencing from data-independent-acquisition mass spectrometry. *Nat. Methods* **2019**, *16*, 63–66. [[CrossRef](#)]
152. Zolg, D.P.; Gessulat, S.; Paschke, C.; Graber, M.; Rathke-Kuhnert, M.; Seefried, F.; Fitzemeier, K.; Berg, F.; Lopez-Ferrer, D.; Horn, D.; et al. INFERYS rescoring: Boosting peptide identifications and scoring confidence of database search results. *Rapid Commun. Mass Spectrom.* **2021**, e9128. [[CrossRef](#)]
153. Kim, M.; Eetemadi, A.; Tagkopoulos, I. DeepPep: Deep proteome inference from peptide profiles. *PLoS Comput. Biol.* **2017**, *13*, e1005661. [[CrossRef](#)] [[PubMed](#)]
154. Muntel, J.; Gandhi, T.; Verbeke, L.; Bernhardt, O.M.; Treiber, T.; Bruderer, R.; Reiter, L. Surpassing 10 000 identified and quantified proteins in a single run by optimizing current LC-MS instrumentation and data analysis strategy. *Mol. Omics* **2019**, *15*, 348–360. [[CrossRef](#)]
155. Zhu, Y.; Piehowski, P.D.; Zhao, R.; Chen, J.; Shen, Y.; Moore, R.J.; Shukla, A.K.; Petyuk, V.A.; Campbell-Thompson, M.; Mathews, C.E.; et al. Nanodroplet processing platform for deep and quantitative proteome profiling of 10–100 mammalian cells. *Nat. Commun.* **2018**, *9*, 882. [[CrossRef](#)] [[PubMed](#)]

156. Li, Z.Y.; Huang, M.; Wang, X.K.; Zhu, Y.; Li, J.S.; Wong, C.C.L.; Fang, Q. Nanoliter-Scale Oil-Air-Droplet Chip-Based Single Cell Proteomic Analysis. *Anal. Chem.* **2018**, *90*, 5430–5438. [[CrossRef](#)] [[PubMed](#)]
157. Ctordecka, C.H.D.; Seth, A.; Mendjan, S.; Tourniaire, G.; Mechtler, K. An automated workflow for multiplexed single-cell proteomics sample preparation at unprecedented sensitivity. *bioRxiv* **2022**. [[CrossRef](#)]
158. Budnik, B.; Levy, E.; Harmange, G.; Slavov, N. SCoPE-MS: Mass spectrometry of single mammalian cells quantifies proteome heterogeneity during cell differentiation. *Genome Biol.* **2018**, *19*, 161. [[CrossRef](#)]
159. Schoof, E.M.; Furtwangler, B.; Uresin, N.; Rapin, N.; Savickas, S.; Gentil, C.; Lechman, E.; Keller, U.A.D.; Dick, J.E.; Porse, B.T. Quantitative single-cell proteomics as a tool to characterize cellular hierarchies. *Nat. Commun.* **2021**, *12*, 3341. [[CrossRef](#)]
160. Cheung, T.K.; Lee, C.Y.; Bayer, F.P.; McCoy, A.; Kuster, B.; Rose, C.M. Defining the carrier proteome limit for single-cell proteomics. *Nat. Methods* **2021**, *18*, 76–83. [[CrossRef](#)]
161. Brunner, A.D.; Thielert, M.; Vasilopoulou, C.; Ammar, C.; Coscia, F.; Mund, A.; Hoerning, O.B.; Bache, N.; Apalategui, A.; Lubeck, M.; et al. Ultra-high sensitivity mass spectrometry quantifies single-cell proteome changes upon perturbation. *Mol. Syst. Biol.* **2022**, *18*, e10798. [[CrossRef](#)]
162. Mund, A.; Coscia, F.; Kriston, A.; Hollandi, R.; Kovacs, F.; Brunner, A.D.; Migh, E.; Schweizer, L.; Santos, A.; Bzorek, M.; et al. Deep Visual Proteomics defines single-cell identity and heterogeneity. *Nat. Biotechnol.* **2022**, *40*, 1231–1240. [[CrossRef](#)]
163. Mund, A.; Brunner, A.D.; Mann, M. Unbiased spatial proteomics with single-cell resolution in tissues. *Mol. Cell* **2022**, *82*, 2335–2349. [[CrossRef](#)] [[PubMed](#)]
164. An, R.; Yu, H.; Wang, Y.; Lu, J.; Gao, Y.; Xie, X.; Zhang, J. Integrative analysis of plasma metabolomics and proteomics reveals the metabolic landscape of breast cancer. *Cancer Metab.* **2022**, *10*, 13. [[CrossRef](#)] [[PubMed](#)]
165. Friedrich, C.; Schallenberg, S.; Kirchner, M.; Ziehm, M.; Niquet, S.; Haji, M.; Beier, C.; Neudecker, J.; Klauschen, F.; Mertins, P. Comprehensive micro-scaled proteome and phosphoproteome characterization of archived retrospective cancer repositories. *Nat. Commun.* **2021**, *12*, 3576. [[CrossRef](#)] [[PubMed](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.