International Journal of
*Molecular Sciences*

MDPI

*Review*

# Bioinformatics Methods for Mass Spectrometry-Based Proteomics Data Analysis

**Chen Chen [1], Jie Hou [2,3], John J. Tanner [4] and Jianlin Cheng [1,*]**

1 Department of Electrical Engineering and Computer Science, University of Missouri, Columbia, MO 65211, USA; ccm3x@mail.missouri.edu
2 Department of Computer Science, Saint Louis University, St. Louis, MO 63103, USA; jie.hou@slu.edu
3 Program in Bioinformatics & Computational Biology, Saint Louis University, St. Louis, MO 63103, USA
4 Departments of Biochemistry and Chemistry, University of Missouri, Columbia, MO 65211, USA; tannerjj@missouri.edu
* Correspondence: chengji@missouri.edu

Received: 18 March 2020; Accepted: 18 April 2020; Published: 20 April 2020


**Abstract:** Recent advances in mass spectrometry (MS)-based proteomics have enabled tremendous progress in the understanding of cellular mechanisms, disease progression, and the relationship between genotype and phenotype. Though many popular bioinformatics methods in proteomics are derived from other omics studies, novel analysis strategies are required to deal with the unique characteristics of proteomics data. In this review, we discuss the current developments in the bioinformatics methods used in proteomics and how they facilitate the mechanistic understanding of biological processes. We first introduce bioinformatics software and tools designed for mass spectrometry-based protein identification and quantification, and then we review the different statistical and machine learning methods that have been developed to perform comprehensive analysis in proteomics studies. We conclude with a discussion of how quantitative protein data can be used to reconstruct protein interactions and signaling networks.

**Keywords:** bioinformatics analysis; computational proteomics; machine learning

## 1. Introduction

Proteins are essential molecules in nearly all biological processes. They provide the structural scaffolding in cells, and function in myriad processes, such as metabolism, biosignaling, gene regulation, protein synthesis, solute transport across membranes, immune function, and photosynthesis. Abnormal regulation of protein function is one of the most prominent factors in disease pathologies; thus, understanding how the proteome is perturbed by disease is an important goal of biomedical research. It is well known that transcriptome data, such as mRNA abundance, is insufficient to infer the protein abundance [1,2], and thus, direct measurements of protein activities are often necessary. Traditional approaches tend to focus on one or a few proteins; however, with the recent developments in the sample separation and mass spectrometry technologies, it is now possible to consider a complex biological system as an integrated unit. The rapid advancements in the experimental aspects of proteomics have inspired various downstream bioinformatics analysis methods that help to discover the relationship between molecular-level protein regulatory mechanisms and phenotypic behavior, such as disease development and progression [3,4].

The typical experiment strategy for MS-based proteomics can be divided into two broad categories based on the size of the protein analyzed by MS: bottom-up and top-down [5]. In the more common bottom-up approach, the protein samples are first proteolytically digested into peptides before analyzing in a mass spectrometer [6]. In top-down proteomics, intact proteins are directly analyzed by

MS [7,8]. In this review, we mainly focus on bioinformatics software and platforms designed for protein identification, quantification, and downstream analysis in bottom-up proteomics. The downstream analysis refers to various data analysis methods used to extract biological meaning from protein abundance data from MS experiments [9]. Similar to genomics, these bioinformatics methods are in rapid development and often require interdisciplinary efforts from mathematicians, statisticians, and computer scientists. Figure 1 shows the general workflow of bioinformatics analysis in mass spectrometry-based proteomics.

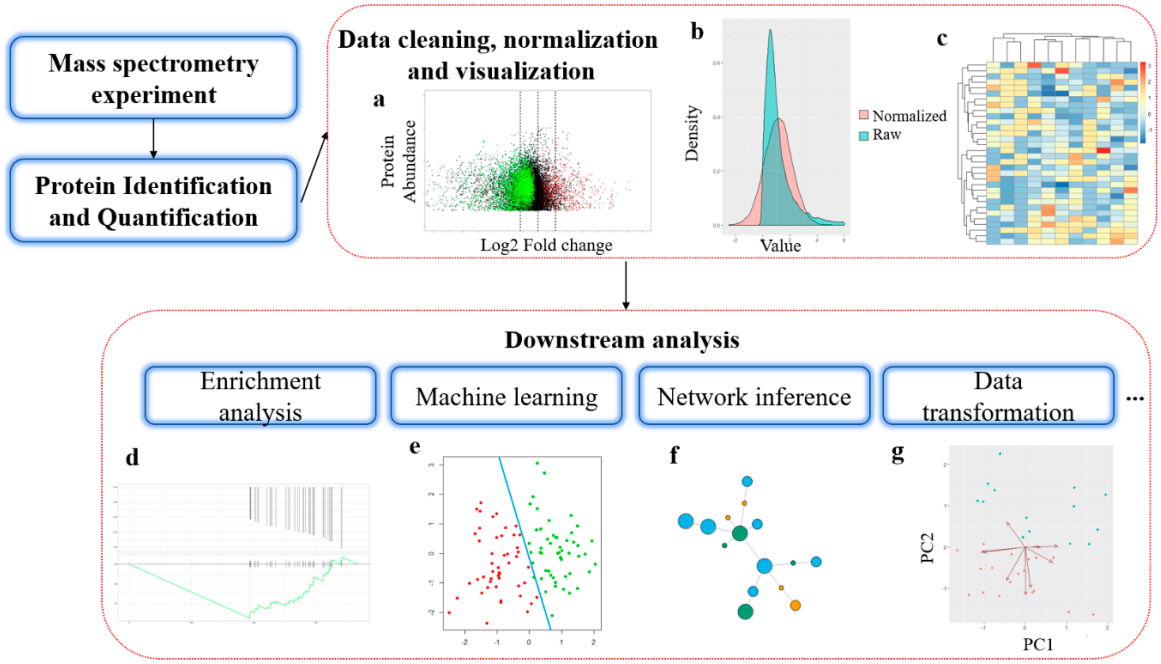


**Figure 1.** General workflow of bioinformatics analysis in mass spectrometry-based proteomics. (**a**) MA-plot from protein differential abundance analysis. X-axis is the log2 transformed fold change and Y-axis is the average protein abundance from replicates. (**b**) Distribution of protein abundance data before and after normalization. (**c**) Heatmap for protein abundance with clustering; (**d**) Protein set enrichment analysis, Y-axis in the above plot shows the ranked list metric, and in the bottom plot shows the running enrichment score. X-axis is the ranked position in protein list. (**e**) Machine learning-based sample clustering. (**f**) Illustration of a network inferred from proteomics data. (**g**) Dimensionality reduction of proteomics expression profile.

In this review, we first describe the tools and methods used to process the raw mass spectral data, including identification and quantification of peptides and proteins. In the following sections, we discuss various bioinformatics techniques used to process the proteomics data. Since the downstream analysis of proteomics does not have a standard workflow and can be highly specific to a particular research purpose, we first introduce the algorithms and tools used in three major applications: data preprocessing, statistical analysis, and enrichment analysis. Then we discuss popular machine learning methods and how they are applied to specific biomedical research topics in proteomics. Finally, we discuss how proteomics data can be used to reconstruct protein interaction and signaling networks. It is beyond the scope of this review to describe all the bioinformatics methods that have developed for proteomics analysis. Therefore, we will focus on the most popular software tools that are in use, as well as some novel methods that have been developed for emerging new experimental technologies.

## 2. Mass Spectrometry-Based Protein Identification and Quantification

### 2.1. Peptide and Protein Identification

After fragmentation spectra data from MS are acquired, the first procedure is to determine the sequence of the peptides. Two different strategies can be applied to this task: searching against the fragmentation spectra databases [10–12] and de novo peptide sequencing [13–15]. Table 1 lists a selection of software tools commonly used in peptide and protein identification reviewed in this study.

In the database matching approach, a target database is established from in silico digestion of all expressed or hypothetical protein sequences. Then a peptide spectrum match (PSM) score is calculated for each fragmentation spectra and all theoretical fragmentation spectra information from the target database. The peptide that has the highest PSM score can be used as a candidate for the query peptide. It is always a crucial task to choose appropriate searching algorithms that can yield high quality peptide spectrum matching results from databases. Traditional protein database search engines, such as SEQUEST [16], implement a scoring function based on normalized cross-correlation of the mass-to-charge ratio predicted from known amino acid sequences in databases and the fragment ions detected from the tandem mass spectrum. MASCOT [17] is another popular software developed later, which integrates mass measurements with protein sequence information, peptide molecular weights from protein digestion and tandem mass spectrometry data to generate a probability-based score for protein identification.

**Table 1.** Commonly used software packages for peptide and protein identification.

| Category | Name | Description |
| --- | --- | --- |
| Database search algorithms | Andromeda [18] | Probabilistic scoring-based peptide search engine integrated in MaxQuant. |
| | Mascot [17] | Probability-based database searching algorithm. |
| | MSPLIT-DIA [19] | Sensitive Peptide Identification for Data Independent Acquisition. |
| | MudPIT [20] | Multidimensional protein identification. |
| | PepArML [21] | An Unsupervised, Model-Free, Machine-Learning Combiner for Peptide Identifications from Tandem Mass Spectra. |
| | PepHMM [22] | A hidden Markov model-based scoring function for mass spectrometry database search. |
| | Protein Prospector [23] | An integrated framework of about twenty proteomic analysis tools. |
| | SEQUEST [24] | An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database |
| | TopPIC [25] | A software tool for top-down mass spectrometry-based complex proteoforms identification. |
| | X!Tandem/X!!Tandem [12,26] | An open source software that search tandem mass spectra with peptide sequences in database. |
| De novo peptide sequencing | DeepNovo-DIA [27] | De novo peptide sequencing by deep learning. |
| | EigenMS [28] | De novo Analysis of Peptide Tandem Mass Spectra. |
| | NovoHMM [29] | A hidden Markov model for de novo peptide sequencing. |
| | PEAKS [30] | A fast de novo sequencing tool. |
| | PECAN [31] | Library Free Peptide Detection for Data-Independent Acquisition Tandem Mass Spectrometry Data. |
| | PepNovo [14] | De novo peptide sequencing via probabilistic network modeling. |
| | pNovo 3 [32] | A software for precise de novo peptide sequencing using a learning-to-rank framework. |
| | SHERENGA [13] | De novo peptide sequencing via tandem mass spectrometry. |
| | SWPepNovo [33] | An Efficient De Novo Peptide Sequencing Tool for Large-scale MS/MS Spectra Analysis. |
| | UniNovo [34] | A universal de novo peptide sequencing algorithm with a modified offset frequency function. |

**Table 1.** *Cont.*

| Category | Name | Description |
| --- | --- | --- |
| Hybrid identification approach | ByOnic [35] | A hybrid of de novo sequencing and database search for protein identification by tandem mass spectrometry. |
| | DirecTag [36] | Accurate sequence tags from peptide MS/MS with statistical scoring method. |
| | InsPecT [37] | A software for identification of peptides posttranslational modification (PTM) from tandem mass spectra. |
| | JUMP [38] | A tag-based database search tool for peptide identification. |
| | PEAKS DB [39] | A hybrid de novo sequencing tool run in parallel with database search. |
| | ProteomeGenerator [40] | A hybrid framework for based on de novo transcriptome assembly and database matching. |
| Other software related to protein/PTM identification | DBParser [41] | Web-based software for shotgun proteomic data analyses. |
| | DIA-Umpire [42] | Comprehensive computational framework for data independent acquisition proteomics. |
| | MassSieve [43] | Panning MS/MS peptide data for proteins. |
| | MAYU [44] | A novel strategy that reliably estimates false discovery rates for protein identifications in large-scale datasets. |
| | ModifiComb [45] | Mapping substoichiometric post-translational modifications. |
| | Nokoi [46] | A decoy-free approach for improved peptide identification accuracy |
| | Param-Medic [47] | A strategy for inferring optimal search parameters for shotgun proteomics analysis. |
| | Perseus [48] | Platform for comprehensive analysis of proteomics data. |
| | PROVALT [49] | A heuristic method for computing false discovery rate (FDR) for protein identifications. |
| | MetaMorpheus [50] | Enhanced Global Post-translational Modification Discovery. |
| | PTMselect [51] | Optimization of protein modifications discovery by mass spectrometry. |

Choosing appropriate input parameters is key for database searching. The tolerances for precursor mass and fragment mass are two important parameters, which should be chosen wisely. The former controls the peptide candidates considered for each spectrum, while the fragment mass tolerance controls the upper limit of the absolute value of the difference between the detected and theoretical fragment masses for a match. For both parameters, setting values that are too narrow will exclude possible true PSMs, while allowing the values to be too wide will introduce a large amount of false PSMs. Several methods have been developed to optimize these database search parameters. They often rely on the observed m/z values of known ions in experimental data to infer instrument calibration, such as MSQuant [52], DtaRefinery [53], and nonparametric regression model [54]. In addition, Preview [55] can apply a fast database search to evaluate precursor and fragment mass error and nonspecific digestion during the experiment. Param-Medic [47] is a search parameter inference tool designed to determine the optimal search parameters by assembling and analyzing pairs of spectra that are likely from the same peptide ion to infer precursor and fragment mass error. Unlike Preview, Param-Medic is implemented in Python as a standalone tool and is convenient to be integrated into the stream pipeline for Linux users.

The scoring function of PSMs also plays an important role in the database searching approach. A well-calibrated scoring system is necessary for distinguishing among the different spectra. For example, Mascot uses probability-based scoring in which the total score is related to the probability that an observed match is a random event. SEQUEST [56] applies two scoring functions: the first one is used to select a limited range of peptide candidates for each spectrum (Sp), and the second uses cross-correlation between the observed and theoretical spectra (Xcorr). The recent version of MaxQuant software has incorporated the Andromeda search engine. Andromeda uses a score based on the probability of matching at least k out of the n theoretical masses by chance from the filtered MS/MS spectrum, where n is the total number of theoretical ions and k is the number of matching ions. According to the evaluation from Cox et al. [57], the score functions used by MaxQuant and Andromeda have very similar discriminatory power, while Andromeda achieves better accuracy in highly phosphorylated peptides. The support vector machines (SVM)-based peptide statistical scoring

method is an effective way to reduce the false discovery rate (FDR) in peptide identification [58]. The presence of the candidate peptide can then be determined by the SVM model trained with the vector representations of peptides from different database search metrics. Lin et al. [59] proposed the "residue evidence" score and demonstrated that this score function can result in performance improvements in a variety of datasets without introducing new trainable parameters. Another application, called PepHMM [22], introduced a new scoring function to improve the accuracy of peptide identification by integrating information on raw data accuracy, peak intensity, and correlation among known and detected peptides into a hidden Markov model to generate a confidence score based on statistical significance.

The database search approach is often followed by a second round of searching against a decoy database in order to reduce FDRs [60,61]. This procedure ensures that all remaining peptides have an FDR higher than the predefined cutoff for peptides identified from the decoy database. Decoy database searching estimates a threshold for removing identifications that have low statistical confidence, resulting in a higher percentage of true positive hits. Several commonly used protein sequence databases for spectrum matching are listed in Table 1. Many widely used analysis platforms, such as MASCOT [17] and MaxQuant [18], have integrated this strategy in their standard analysis pipelines. However, with the increasing size of the protein sequences database, the target-decoy search strategy is becoming computationally inefficient since its searching space is twice as large as the original target database search. To resolve this problem, many novel searching strategies have been reported in recent years. Gonnelli et al. [46] proposed Nokoi, which is a decoy-free approach for improved peptide identification accuracy. This approach is based on an L1-regularized logistic regression model trained on a large heterogeneous dataset using ranks supplied by the Mascot as true labels for PSMs. Kim et al. developed a target-small decoy search strategy that can handle cases where the sizes of target and decoy databases differ, which is done by reversing the target database and picking random proteins according to the ratio of small-decoy to target database sizes. Their evaluation shows that this approach reduces the database size and searching time while retaining the same level of accuracy as normal target-decoy search [62].

In de novo peptide sequencing, the peptide sequence is determined solely from fragmentation spectra information and fragmentation method properties. The analysis strategy for de novo peptide sequencing has been dominated by Graphical Probabilistic Model and Hidden Markov Model (HMM)-based methods, such as PepNovo [14] and NovoHMM [29,63]. However, many variants of the sequencing frameworks have been reported since then. In order to solve the performance deterioration in Electron Transfer Dissociation spectra and Higher-energy Collisional Dissociation spectra, Jeong et al. proposed a universal de novo peptide sequencing algorithm, called UniNovo [34]. UniNovo incorporates a new scoring criterion calculated from a modified offset frequency function to capture the dependencies between different types of ions. Low ion coverage is another well-known challenge in de novo peptide sequencing since the order of consecutive amino acids cannot be easily determined if all of their supporting fragment ions are missing. To address this problem, Yang et al. [32] recently proposed pNovo 3, which specifically designed to distinguish peptide candidates of each spectrum with a learning-to-rank framework. pNovo 3 can also generate different metrics measuring the similarity of the actual experimental spectrum and the theoretical spectra predicted by deep learning. Other deep learning-based methods [27,64] have been introduced to solve highly multiplexed spectra in recent years. These models often incorporate the highly customized architecture of convolutional and recurrent neural networks and can be trained with features attainable for de novo sequencing, such as spectra data, fragment ions information, and sequence patterns of amino acid.

It is also possible to achieve better performance by combining de novo peptide sequencing with the database matching approach. Such hybrid approaches first choose the most suitable protein database for searching based on peptide tag sequences and then perform error-tolerant searching against the selected database. This category of searching strategy includes InsPecT [37], DirecTag [36] and JUMP [38]. In contrast, PEAKS Studio [39] performs de novo sequencing of spectra before any

database searching. More recently, Cifani et al. [40] proposed ProteomeGenerator, which is a hybrid proteomics framework. This new approach calibrates the matching results from a target–decoy database with sample-specific controls, and can significantly improve the accuracy of isoform identification in non-canonical proteomes. Parallel PSM processing algorithms for large scale proteomics dataset have also been implemented [33].

Data-independent acquisition (DIA) mass spectrometry is now emerging as a new strategy for systematic analysis of peptide mixtures. Unlike traditional data-dependent acquisition (DDA), in which mixtures of precursor ions are selected based on co-selection and co-dissociation, DIA detects all fragmented ions at each chromatographic time point within a predefined m/z range, or uses the m/z ranges during isolation and fragmentation [65]. The strength of DIA is that all precursor ions are selected without bias, providing more reliable input for downstream systematic analysis. Various analysis strategies have been developed for DIA approach with different approaches. For example, MSPLIT-DIA [19] is a library matching method in which input from DIA spectrum are searched against library spectra and spectrum projections are evaluated based on normalized dot product. However, results from these library-based methods are limited to prior knowledge from data present in the library. To address this issue, many library-free methods have also been proposed. These tools typically reconstruct pseudo spectra from deconvolving the multiplexed spectra or highly correlated precursor ion groups, such as DIA-Umpire [42], or directly compute the confidence of existence for each query peptide from DIA data, such as FT-ARM [66] and PECAN [31]. Recently, Searle et al. [67] proposed a rapid, experiment-specific library generation workflow for DIA-MS for non-model organisms and non-canonical databases. In their system, libraries containing every peptide in a proteome are constructed first, followed by a refinement process using empirical data built directly from protein sequence databases.

Posttranslational modifications (PTMs) in proteins can also be identified with MS. However, simply searching the database with all possible modifications positions can be extremely time-consuming since the large number of all possible position/modification combinations. Several methods exist to help decrease the computational requirements or increase the accuracy of results, such as ModifiComb [45], PTMselect [51], and G-PTM-D [68]. PEAKS PTM [69] is the PTM identification method integrated in the PEAKS Studio software. In PEAKS PTM, unassigned spectra with high de novo scores are searched against the identified proteins. Sequence data from genomic or transcriptomic databases can also aid peptide identification by providing possible protein sequences. This strategy is widely used in proteogenomics, which is an emerging area of research that uses information from both genomic and transcriptomic databases to facilitate the identification of novel peptides from MS-based proteomics data [70]. MetaMorpheus [50] is another novel tool, which incorporating multinotch searches in global PTM discovery. Compared with the G-PTM-D approach, it achieves a higher number of identified PTMs with a significant increase in search speed.

Once the peptide identification is complete, the next step is to reconstruct the peptide sequences into original proteins. This procedure is called protein inference. Longer peptides are more informative in this step due to their uniqueness. In comparison, it is usually not straightforward to construct a reliable list of proteins from shorter peptides, since some of the peptides may be shared by two or more proteins. These "degenerate peptides" usually have multiple optimal solutions for protein assignment. Many models built for peptide assembly adapt the parsimonious rule, in which the smallest set of proteins that account for the detected peptides is reported [41,43,71]. Probabilistic models are also widely used in protein inference and was first introduced by ProteinProphet [72]. More statistical modeling frameworks, such as Hierarchical Statistical Model [73] and Bayesian Inference Model [74], have been proposed since then. The Bayesian Inference Model applies Bayesian models built based on posterior probabilities and has improved the performance of the original ProteinProphet by approximately 6%. The performance of protein inference is often affected by the strength of evidence, such as the PSM score from the observed peptides [75]. Typically, for large proteomics data sets, a stringent cutoff for protein identification FDRs is set to reduce the number of incorrectly reported proteins. Novel FDR estimation frameworks for protein inference have also been developed.

Reiter et al. proposed MAYU [44], which extends the target-decoy strategy for FDR estimation of PSM to the protein level. Validation in various large proteomics datasets shows that the size of the dataset impacts the reliability of protein identifications results. Recently, Wu et al. [76] proposed a new FDR estimation approach in which the null distribution is generated from a combination of logistic regression model with the Permutation + BH (Benjamini–Hochberg) method. Moreover, the authors showed that this approach achieves consistently better performance than MAYU.

### *2.2. Protein Abundance Quantification*

The abundance level of proteins across the entire proteome in a sample can be acquired from quantitative proteomics analysis. The increasing volume and extent of quantitative proteomics data inspired the development of numerous subsequent bioinformatics analysis methods. Currently, there is no standard analysis pipeline to construct the global protein expression profiles from samples in a straightforward fashion, since many software and algorithms have been developed for dedicated upstream experimental techniques. Nevertheless, most experimental quantification methods fall into two categories: labeled methods and label-free methods. In this section, we first review the latest software tools designed for both techniques, and then introduce several software tools dedicated for an emerging technique in clinical studies, called imaging mass spectrometry. Table 2 lists software tools for protein abundance quantification reviewed in this section.

**Table 2.** Commonly used software package for quantitative proteomics.

| Category | Name | Description |
| --- | --- | --- |
| Label-based | IsobariQ [77] | A relative quantification software that can be used for both iTRAQ (Isobaric tags for relative and absolute quantitation) and TMT (Tandem mass tag) labeling. |
| | iTracker [78] | Allows quantitative information gained using the iTRAQ protocol to be linked with peptide identifications from popular tandem MS identification tools |
| | Libra [79] | The iTRAQ quantification module of the TPP (Trans-Proteomic Pipeline). |
| | MaxQuant [18] | One of the most frequently used platforms for mass-spectrometry (MS)-based proteomics data analysis. |
| | ProteinPilot (ABSciex) | Full solution for protein identification and label-based protein expression experiments. |
| | Proteome Discoverer (Thermo Scientific) | Proteomics workflows for a wide range of applications. |
| | PVIEW [80] | LC-MS/MS Data Viewer and Analyzer developed by Princeton. |
| | XPRESS [81] | Quantitative profiling of differentiation-induced microsomal proteins using isotope-coded affinity tags and mass spectrometry. |
| Label-free | emPAI [82] | Exponentially modified protein abundance index. |
| | Mascot Distiller (Matrix Science) | A single, intuitive interface to a wide range of native (binary) mass spectrometry data files. |
| | MaxLFQ [83] | Label free quantification module integrated in MaxQuant. |
| | VIPER [84] | Visual Inspection of Peak/Elution Relationships. |
| Integrated platform | OpenMS [85] | A cross-platform software for the flexible analysis of MS data. |
| | Peaks Studio X + [86] | A peptide/protein identification & quantification software platform that offers complete solutions, including PTM and sequence variants. |
| | Skyline [87] | An open-source Windows client application for accelerating targeted proteomics experimentation. |

Currently, there are two popular approaches for label-based quantification: MS1 (first stage MS) based labeling [88,89], and MS2 (second stage MS) based labeling. In MS1 based labeling, different samples will have different isotope patterns in MS1 spectra depending on their labeling. This can be done for both in vitro and in vivo samples. Isotope coded affinity tags (ICATs) [90], dimethyl-based labeling [91] and isotope coded protein labeling (ICPL) [91] are common strategies for in vitro experiments. Stable isotope labeling by amino acids in cell culture (SILAC) [92] and $^{15}N$ metabolic labeling [93] are the most popular strategies for in vivo labeling. Various bioinformatics

methods have been developed for MS1 labeling experiments, including MaxQuant [18], PVIEW [80] and XPRESS [81]. Most software tools can handle samples from multiple labeling methods. For example, XPRESS is able to analyze ICAT, SILAC, and ICPL labeled samples, and it can also calculate the relative abundance of proteins based on the elution profiles of the labeled peptide pairs. PVIEW can handle SILAC, ICPL and ICAT labeled samples, and it can even perform nonlinear alignment label-free quantification and label free XIC-based quantification. On the contrary, MaxQuant is designed specifically for high-resolution data with SILAC labeling from Thermo Orbitrap and FT mass spectrometers. With the Perseus framework [48], it is convenient to perform downstream statistical analysis for raw quantification results from MaxQuant.

In MS2 based labeling, the quantification signals can be detected at the low mass range of MS2. Isobaric labeling is a common strategy in MS2 based labeling [94]. In isobaric labeling, samples are labeled with certain chemical groups (such as TMT [95] and iTRAQ [96]) that have identical mass but a different distribution of heavy isotopes. It can achieve relatively large multiplexing capacity during an MS run but has the problem of ratio compression due to cofragmentation [97]. Most instrument vendors provide commercial software for the analysis of raw data from spectrometers. For example, the ProteinPilot (from ABSciex) and the Proteome Discoverer (from Thermo Scientific) can handle raw data from their corresponding instrument vendors. There is also a wide range of free software available, including iTracker [78], IsobariQ [77] and Libra [98]. These free software tools often use public peak file formats, such as mzXML, to import and process data. In addition, the Trans-Proteomic Pipeline (TPP) [79], provides a full workflow from raw data processing to statistical analysis. TPP integrated a collection of software used in quantification analysis, such as XPRESS and Libra, and can process isotopically or isobarically labeled samples as a single software system. Figure 2 illustrates a typical workflow of protein abundance quantification using TPP. Skyline [87] and OpenMS [85] are similar integrated platforms for convenient and flexible analysis of MS data.

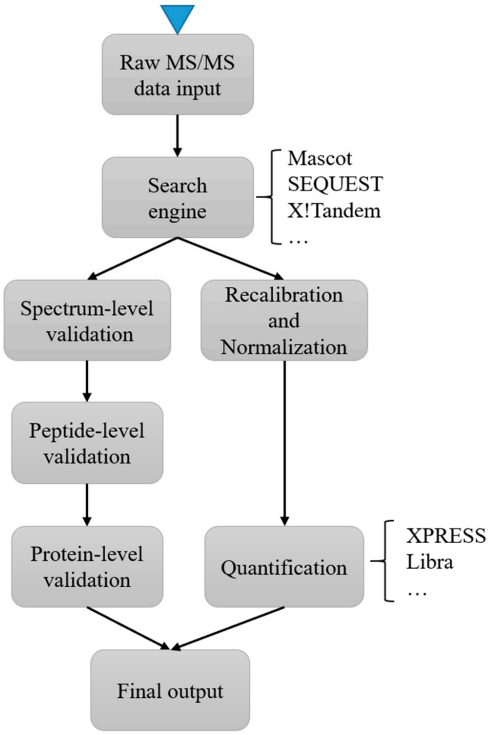


**Figure 2.** A typical workflow for protein abundance quantification using Trans-Proteomic Pipeline (TPP). After the raw data files are converted to one of the supported open XML formats, the data are processed with a designated search engine and the search results are validated at spectrum, peptide, and protein level. Meanwhile, the quantification tools will perform the protein quantification with the normalized data. The final output are ready for various downstream analysis.

In label-free quantification, spectra for different samples are acquired from separate LC-MS/MS experiments [99]. However, this may introduce unwanted variations, such as the inconsistency of chromatographic measurements from multiple experiments. To address this issue, different intensity normalization methods have been implemented, such as total ion count (TIC), MaxLFQ (integrated into MaxQuant) [83]. Usually, the label-free quantification is performed based on analysis of the signal intensity of peptide precursor ions of the fragmented proteins, and most software are designed using this approach, such as Mascot Distiller [100], MaxQuant [18], and VIPER [84]. In addition, tandem mass spectra counting can be used as an alternative quantification approach. The exponentially modified protein abundance index (emPAI) [82] integrated into Mascot is one of the tools using this approach. The observed peptide count can also be corrected by the peptide detection probability based on characteristic physicochemical peptide properties using machine learning models [101].

Models such as iBAQ [102] and LCMSE [103] are built to measure the global absolute quantifications of proteins. For example, iBAQ values are generated from dividing total intensities by the identified peptides for one protein. Nonetheless, it is infeasible to directly compare the abundance levels of different proteins due to their distinct efficiencies of ionization and detectability. Spike-in standards [104] or the "proteomic ruler" approach [105] can be applied to improve the quality of absolute abundance estimation.

Imaging mass spectrometry (IMS) is emerging as a popular protein quantification technique in clinical studies, particularly in neuroscience [106]. IMS has the capability to systemically detect localization patterns of proteins and their abundance in various types of cells and biological tissues. In IMS, proteins from different biological samples are desorbed and ionized with different probes. Popular techniques used in IMS include laser desorption/ionization (LDI), matrix assisted-LDI (MALDI) [107], time-of-flight -secondary ion mass spectrometry (TOF-SIMS) [108] and electrospray based desorption (DESI) [109]. These techniques are designed for various purposes with respect to their different capability of spatial resolution and detection power. In order to improve the quality of raw images from IMS, image processing algorithms such as edge-preserving image denoising/clustering [110] and spatially aware clustering [111] have been proposed to generate more informative and precise segmentation maps. Software suites that can perform baseline corrections, subtractions, denoising, smoothing, recalibration, and normalization of IMS data in an integrated pipeline have also been developed [112].

## 3. Data Preparation and Downstream Bioinformatics Analysis in Proteomics

### *3.1. Data Preprocessing and Normalization*

In peptide-based MS quantitation, the abundance of a protein is inferred from a limited number of peptides, sometimes only one. Typically, researchers need to carefully apply cutoffs for minimum peptide numbers to increase the reliability of the inferences from only a small number of peptides. For instance, Peaks Studio Quantification uses only the top three peptides for the quantification of a protein, while it should be using as many as possible as long as they are reliably detected. In addition, to generate comparable and reliable results for downstream analysis, it is always necessary to perform preprocessing and normalization of the raw quantitative data.

Normalization refers to a data correction process that removes any non-biological related variations and makes the results reliable and aligned across samples for downstream analysis. Selecting an appropriate normalization method is crucial in proteomics analysis, especially considering the fact that inadequate normalization is usually being performed prior to LC/MS [113]. Several types of normalization methods are developed based on different statistical assumptions. The distribution of the protein abundance is often extremely skewed towards zero, so it is common practice to perform logarithm transformation on the intensity values before further normalization. For linear regression-based normalization such as RlrMA and LinRegMA [114], they assume that bias in the measurements and the scale of the observed protein abundance are linearly dependent. For local

regression normalization [115], the assumption is that bias and protein abundance have a nonlinear dependency and can be explained with nonlinear models such as local polynomial regression. Variance stabilization normalization (VSN) [116] is another statistical model, which aims to eliminate the dependency between variances and mean abundances and scaling data from different samples into the same level through parametric transformations and maximum likelihood estimation. Quantile, median and EigenMS [117] normalizations are also used in proteomics in various studies. Välikangas et al. [114] systematically evaluated eleven commonly used normalization techniques in four separate proteomics datasets, and the results indicate that these normalization methods generally have the same level of performance on most of the proteomics data, while VSN generally performed consistently well in the differential expression analysis.

Despite the various normalization methods mentioned in this section, the optimum normalization approach is dependent upon the experiment environment in different MS-based proteomics studies and needs to be evaluated individually with MA plots. Moreover, hierarchical clustering and heatmaps are useful visualization techniques to help to determine the effectiveness of the preprocessing steps.

The raw quantitative data may also contain missing values, especially for those with relatively low concentration in the cellular context. Missing values are common in most high throughput technologies due to the stochasticity in sampling during the experiment [118]. These missing values can be removed or included based on the distribution of all proteins detected. Various machine learning models have been developed to impute missing values for omics data, including empirical distribution sampling [119], k-nearest neighbors (kNN) imputation [120] and singular value decomposition (SVD) imputation [121]. However, it has been shown that the performance of these imputation methods will decrease drastically when the data contains more missing values [122]. Novel missing value approaches have been developed with higher imputation accuracy and statistical sensitivity. For example, Wei et al. [123] proposed GSimp, which adapts an iterative Gibbs sampler to infer the missing values for MS data. More recently, Berg et al. [119] developed an efficient imputation model based on sampling from normal distributions with better performance at high false positive rates level.

### 3.2. Statistical Analysis of Quantitative Protein Data

Once the protein abundance data has gone through data cleaning, filtering, and normalization process, it is ready for further statistical analysis and further downstream studies. The most straightforward statistical analysis is to examine if there are any significant changes in protein levels between two different conditions. This is typically done by performing a t-test between the observed protein abundances from two different groups [124]. In case there are two or more factors required to estimate, ANOVA (analysis of variance) [125] can be performed instead. However, the sample size in proteomics data is relatively small due to the limited multiplexity, and this impairs the statistical power of the t-test and results in insignificant $p$-values for proteins with a large fold change, but also a relatively large variance. To address this issue, several statistical models have been proposed. For example, Kammers et al. [126] demonstrated better results can be achieved by applying moderated t-statistics from the empirical Bayes procedure Linear Models for Microarray Data (LIMMA) in the identification of differentially expressed proteins. The LIMMA model, as is suggested by its name, is originated from significant change detection in microarray data. LIMMA is designed to reduce the variances of the measurement to a pooled estimate based on all sample data and can achieve more robust and accurate results than traditional t-test, especially on relatively small proteomic datasets. Linear mixed-effects models [127], mean/median sweeps [128], and "masterpool" normalization [128] are other commonly used methods in relative protein abundance estimation and statistical analysis of proteomic data.

For all statistical tests, it is necessary to set up an FDR threshold since multiple hypotheses are being tested at the same time. The Benjamini–Hochberg procedure [129] and FDR estimation from permutation [130] are widely used approaches in proteomics. It has been demonstrated that

introducing FDR-controlling procedures can effectively reduce the number of false positives in the statistical analysis of proteomic data [131].

### 3.3. Enrichment Analysis in Proteomics

Enrichment analysis is a method that helps to identify genes or proteins that are overrepresented in the predefined gene set of interest. The gene set usually consists of a list of genes that share common functions or in the same pathway or network. The advantage of performing enrichment analysis on proteomics data is that we can test hypotheses on the systemic measurements of the proteome instead of the transcriptome. Moreover, the input of such analysis can incorporate additional information after the transcription process, such as differential translation rate and PTM. Publicly available online databases, such as DAVID [132] and STRING [133], have provided gene sets built from prior knowledge and automatic tools to perform automatic enrichment analysis online. For PTMs, databases such as PhosphoSitePlus [134] and Signor [135], have curated additional information about modification position/type and their related diseases from literature mining. In such workflows, the users provide a list of gene identifiers or modifications and then select a gene set database based on their research interest. While databases such as Uniprot and Ensembl can accept protein names as input, many other databases require converting protein names into gene names before further steps. However, it is still a challenging task for identifier conversions as protein names and corresponding genes are in a many-to-many relationship, and such conversions are often labor-intensive and will result in information loss. Currently, several web cross-reference services are developed to solve the conversion tasks, such as PICR [136] or CRONOS [137], but the users should be cautious since these tools may not represent up-to-date information and may need manual correction. It is also worth mentioning that biological databases are rapidly growing in size and number, whereas accurate curations of these new data are often lagging (the "Data avalanche") [138]. Inadequate database curation and a lack of common data formats for cross-database referencing often results in labor-intensive work before enrichment studies.

Gene Ontology (GO) Enrichment [139] is the most widely used technique in enrichment analysis. The GO terms can be considered as a set of predefined groups to which different genes are assigned based on their functional characteristics, thus helping to reduce the redundancy in terminologies. The GO terms are hierarchically clustered with the top nodes describe the three main categories of the terms: "biological process", "molecular function", and "cellular component". Each term has a unique identifier and is connected to other related terms. The Amigo database [140] provides GO term annotation for many species, but not all proteins have a complete and precise annotation. For those proteins that do not have complete annotations, informative GO terms from proteins with similar sequences can be used instead. GO term prediction algorithms such as ProLoc-GO [141], PFP [142] and IGNA [143] are designed for this problem. The most common statistical tests used in GO enrichment are Fisher's exact test and the hypergeometric test. Statistically significant GO terms are those that appear more frequently in the input protein list than would be expected by chance, and they may indicate interesting biological processes for further studies. However, since GO terms usually represent ORF products, rather than mature proteoforms, researchers should carefully examine the GO terms in the enrichment results to ensure that they have appropriate relations between the proteoforms and corresponding genes.

Similar to GO terms, prior knowledge about regulatory pathway networks and diseases can also be used to perform enrichment analysis [144]. A biological pathway describes biological actions and chemical reactions among molecules within a cell that result in certain biological processes. Databases such as PANTHER [145] and Reactome [146] curated the interaction maps for different pathways as well as a set of enrichment tools to perform enrichment directly on the database webpage. Many independent tools can also use the public interface to run enrichment analysis based on the data curated in those only databases directly. For example, Pathview [147] is an R package for KEGG pathway [148]-based data integration and visualization. Protein set enrichment analysis (PSEA) is

another popular enrichment approach and is derived from gene set enrichment analysis (GSEA). In PSEA, the enrichment score is computed from a weighted running sum statistic and proteins without significant changes in abundance may negatively affect the enrichment score. While many GSEA software can also perform PSEA, tools developed specifically for protein quantification data, such as PSEA-Quant [149], may provide a more convenient and reliable workflow for proteomics. Figure 3 shows the common enrichment analysis visualization tools in proteomics.

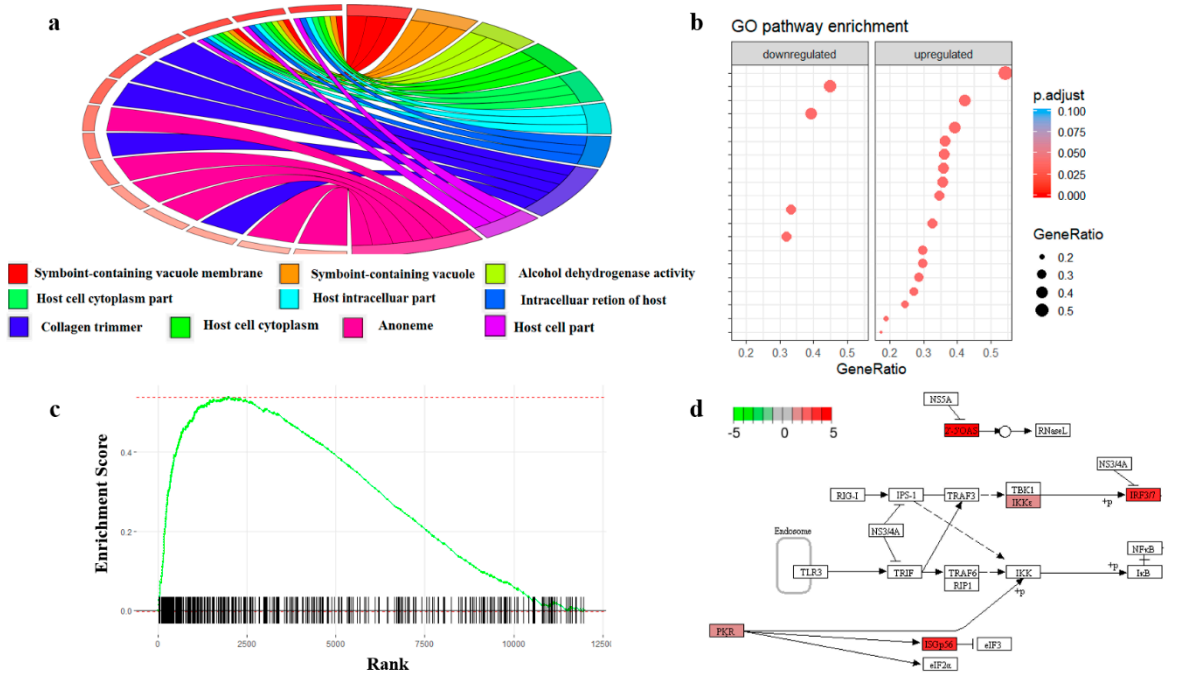


**Figure 3.** Illustration of enrichment analysis with proteomics data. (**a**) Gene Ontology (GO) enrichment with Circos plot. The left part of the circle shows the differentially expressed proteins and their significant levels. The right part shows the enriched GO Terms. (**b**) Pathway enrichment in dot plot. The colors of the dots represent the adjusted *p*-value and the sizes of the nodes represents the ratio of the proteins that are differentially expressed to the total proteins in the pathway. (**c**) The running sum score of the gene set enrichment analysis (GSEA). Significant proteins will increase the enrichment score and proteins that are not in the list will decrease the score. (**d**) Protein enrichment with KEGG pathway database with Pathview. The color of the boxes represents the log2 fold change of the protein abundances.

### *3.4. Machine Learning Approaches*

Machine learning is powerful in downstream bioinformatics analysis. It can extract informative features from a huge amount of proteomics data and construct models that can be validated from separate datasets rather than providing simple descriptive information. There are two main types of tasks of machine learning: supervised learning and unsupervised learning. Common applications with supervised learning often involve using the quantitative proteomics data to build models that can predict the annotations for new samples. The format of the labels in supervised learning can be either discrete categories (classification) or continuous numeric values (regression). For example, clinical outcomes (such as survival time) can be predicted from quantitative proteomics data with models trained from samples collected from patients with known clinical information [150]. Applications of supervised learning in proteomics usually involve taking the quantitative results from MS as input and manually labeled or experimentally verified annotations as the target output. In unsupervised learning, the goal is to infer the natural structure and dependency existed within the data given. Both supervised learning and unsupervised learning are powerful methods in the downstream analysis for proteomics, and enormous tools have been implemented for various research purposes.

Common models in supervised learning include Bayesian classifiers, Logistic Regression, Decision trees, Random Forest, Support Vector Machines (SVM), and Artificial Neural Networks. A variety of applications in proteomics is reported in several works. Deeb et al. [151] used protein expression profiles to perform classification for patients with diffuse large B-cell lymphoma. According to their study, highly ranked proteins from the trained models can be recognized as core signaling molecules in pathobiology for different subtypes. Dan et al. [152] used an SVM classifier to identify diagnostic markers for tuberculosis by proteomic fingerprinting of serum. Their model also helped to identify several potential biomarkers for new diagnostic methods. Tyanova et al. [153] used proteomic profiles to identify functional differences between breast cancer subtypes. They further identified several proteins with different expression signatures across the subtypes without the involvement of gene copy number variations. These findings provide novel insight into future subtype-specific therapeutics development. Protein subcellular localization is another fruitful field for supervised learning methods [154]. Due to the nature of high dimensionality for proteomics data, dimensionality reduction techniques such as principal component analysis (PCA), Linear Discriminant Analysis (LDA) and t-Distributed Stochastic Neighbor Embedding (t-SNE) are often applied prior to building the models [155]. Regularization is another widely used technique in supervised learning. It reduces the model complexity and number of features required for prediction by adding larger penalties to models that have more parameters. Deep Learning also becomes a popular approach in proteomics since it is good at extracting useful information from large numbers of features. For example, Ding et al. constructed a deep autoencoder model to identify informative features from genomics and proteomics data, which was then used to predict the effectiveness of drugs in cancer cell lines [156]. They further showed that the features derived from deep learning models contain relevant information of the cellular signaling system and contribute to the improvements of model accuracy. Cross-validation is widely used for performance assessment in supervised learning [157]. It also helps to avoid the overfitting problem, since the performances of the models are always evaluated from separated validation datasets rather than the samples used in model training.

The simplest form of unsupervised learning is hierarchical clustering, according to protein abundance values. Clustering can be used as a quality assessment for MS experiment results as we can compare the blindly grouped samples with the prior knowledge about sample similarities. Several more complicated unsupervised learning approaches have been reported: Peptide identification Arbiter by Machine Learning (PepArML) [158] is an unsupervised learning method built to determine the peptide associated with the tandem MS spectrum. ProtVec [159] is an unsupervised distributed representation for biological sequences and can be applied to many common problems in proteomics, such as protein family classification and structure prediction. Compared with supervised learning, unsupervised learning is less frequently used in proteomics. However, it does not require ground truth for model training and is particularly useful in heavy data-driven tasks such as feature extraction.

## 4. Protein Networks Reconstruction

In biomedical studies, a new trend is to explore the complex contextual relationships between molecules of interest within the cellular environment. This paradigm created a new branch in life sciences, called systems biology or networks biology [160]. Unlike the traditional mechanism studies, which follow the "one gene, one protein, one function" principle, the systems biology views the interactions between genes and their functions as a large network. In this network, the nodes represent functional molecules within cells, such as genes or proteins. The edges between nodes indicate the functional relationship between the molecules. The connection can be either direct interactions such as chemical reactions between kinases and their substrates, or indirect relationships such as the transcriptional regulation between transcription factors and their targets. Analyzing the biological processes with the networks graphical model provides a novel perspective between genotypes and phenotypes and can also efficiently utilize the huge volume of omics data.

Currently, there are two main categories in proteomics-based network biology: protein-protein interaction (PPI) networks and signaling networks. The PPI network describes the direct interactions between two proteins and can be validated through the analysis of protein complexes using affinity purification-mass spectrometry (AP-MS) [161]. In AP-MS, certain proteins of interest are tagged and then detected from copurified protein components by MS. However, false-positive interactions can also be introduced during the experiment. Several computational frameworks have been established to compensate for limitations in experimental design and a lack of proper controls in AP-MS. Rinner et al. [162] proposed the MasterMap system that facilitates the quantitative protein complexes analysis with label-free MS. Glatter et al. [163] proposed the PP2A integrated workflow that significantly improved the data throughput, sensitivity, and robustness in human protein-based global AP-MS analysis. Many graph-based clustering methods, such as superparamagnetic clustering [164] and hyperclique pattern discovery [165] also proved to be effective approaches for detecting actual functional protein complexes.

An objective of studying signaling networks is to interpret the relationships between enzymes and their substrates. The interactions between kinases and their targets are particularly interesting since they are essential cellular signaling molecules and frequently associated with diseases such as cancers and endocrine disorders [166,167]. MS-based phosphoproteome analysis of kinase signaling also helps reveal the regulatory modification pattern of proteins in bacteria [168–170]. Moreover, proteomics can help identify the acetylation modification, which has emerged as a major interest in epigenetics and cellular signaling recently [171–174]. With the advancements in MS-based techniques, it is now possible to identify a wide range of PTM events, such as phosphorylation in a single MS run with high coverage and quality. Furthermore, the PTM information can help to predict the status of the nodes and edges in signaling networks. One important application in this field is kinase activity inferences based on abundances of phosphorylated proteins. Several machine learning methods, such as IKAP [175], KSEA [176], and kinact [177], have been implemented for this task. Reconstruction of the signaling network is another important topic. In the HPN-DREAM challenge held by Dialogue for Reverse Engineering Assessment and Methods (DREAM), one of the main categories is to infer causal signaling networks from time-course Breast cancer proteomic data. A systematic assessment [178] has been published with a comprehensive evaluation of all submitted models. Since then, many new network inference algorithms have been developed, such as PerseusNet [48], GNET2 [179], Neglog [180] and HIPPIE 2.0 [181]. Dynamic network reconstruction models for time-series proteomics data have also been reported. Usually, they are designed to quantify changes in network in response to different physiological conditions or when stimuli factors are introduced. COVAIN [182] is an integrated toolbox using time-series and correlation network analysis to investigate responses on different hierarchies of cellular contexts. Considering the popularity of network studies in biology, most network inference tools have been packaged as R and Python libraries, or integrated web services for better accessibility among the academic community.

Tremendous effort has been made to elucidate disease pathogenesis in a network view. Wang et al. [183] showed that Se-and Zn-related proteins play significant roles in Endemic Dilated Cardiomyopathy Keshan Disease from networks constructed from protein expression profile. Pirhaji et al. [184] designed a random forest-based algorithm called PIUMet, to study abnormal signaling pathways in metabolisms of sphingolipids, fatty acids, and steroids within Huntington's disease. Many biological network databases are freely available to the public, such as KEGG [148], Pathway Commons [185] and BioGRID [186]. Usually, these databases will also provide programmatic interfaces that allow bioinformaticians to run comprehensive analysis without the need to understand the structure and format of data stored. However, it is still a challenging task to systematically infer the protein networks since most species have a large pool of proteins. Thus, network inference problems usually have an extremely high computation cost and are often limited by available computational resources.

## 5. Discussion and Future Perspectives

MS-based proteomics has greatly improved our understanding of the complex biological mechanisms that underlie human health and disease. Currently, the upstream of MS-based proteomic analysis, including protein identification, characterization, and quantification, is becoming increasingly convenient and reliable as automated pipelines have been provided in most of the experimental platforms. New multiplexing technologies have made it possible to analyze hundreds of samples at a higher throughput. With the latest isobaric tag-based multiplexing, up to 11 samples can be analyzed in a single MS run [187]. Despite recent advancements, several major issues still exist for proteomics. Similar to next-generation sequencing, it is still challenging to precisely quantify proteins at low abundance level. Moreover, reducing the number of missing values is not always possible. Novel techniques that can improve the effectiveness of ion sources, spectra resolution, and dynamic detectors with broader range may become the trend for future upstream proteomics development.

Combining the proteome data with other omics technologies is emerging as a new paradigm in bioinformatics. For example, pointwise comparisons between proteomics and transcriptomics can be established since the identifiers of genes and proteins can be mapped between the two omics spaces. Proteins/genes with discordant trends may indicate the involvement of significant transcriptional and post-translational regulation mechanisms. Quantitative proteomics using in vivo SILAC mouse technology has been successfully applied in studying the post-transcriptional mechanisms in circadian regulation of the liver [188]. The reverse engineering of protein signaling networks can also benefit from other omics data. Virtual Inference of Protein-activity by Enriched Regulon (VIPER) algorithm [189] can perform computational analysis of protein activity based on the relationship between transcription factors and their potential targets identified from transcriptome data. Given the fact that different types of omics data are often complementary to each other, the MS-based proteomics will become much more powerful when analyzed from a multi-omics perspective.

As MS-based proteomics technology continues to evolve, the associated bioinformatics tools need to be updated accordingly. So far, most of the analysis methods mentioned in this review have been provided with a convenient and user-friendly interface. Table 3 lists all downstream bioinformatics tools and databases for proteomics analysis mentioned in this review. We expect more bioinformatics methods from interdisciplinary studies will emerge in the future and enhance the current understandings of complex systems biology.

**Table 3.** Downstream bioinformatics analysis software tools.

| Category | Name | Description |
| --- | --- | --- |
| Downstream bioinformatics analysis tools | COVAIN [182] | A software for statistics, time series, and correlation network analysis. |
| | CRONOS [137] | A cross-reference navigation server. |
| | GNET2 [179] | An R package for module inference of biological network. |
| | GSimp [123] | A Gibbs sampler-based left-censored missing value imputation approach for metabolomics studies. |
| | HIPPIE v2.0 [181] | A tool for enhancing meaningfulness and reliability of protein–protein interaction networks. |
| | IKAP [175] | A heuristic framework for inference of kinase activities from phosphoproteomics data. |
| | INGA [143] | A tool for protein function prediction combining interaction networks, domain assignments, and sequence similarity. |
| | KSEA [176] | A web-based tool for kinase activity inference from quantitative phosphoproteomics. |
| | Neglog [180] | Reconstruction of Human Protein–Protein Interaction Networks. |
| | Pathview [147] | An R package for pathway-based data integration and visualization. |
| | PP2A [163] | An integrated workflow for charting the human interaction proteome. |
| | ProLoc-GO [141] | Utilizing informative Gene Ontology terms for sequence-based prediction of protein subcellular localization. |
| | PSEA-Quant [149] | A Protein Set Enrichment Analysis on Label-Free and Label-Based Protein Quantification Data. |
| | viper [189] | Virtual Inference of Protein-activity by Enriched Regulon analysis. |

**Table 3.** *Cont.*

| Category | Name | Description |
| --- | --- | --- |
| Databases and web services for downstream analysis | STRING [190] | A database for quality-controlled protein-protein association networks. |
| | SIGNOR [135] | A database of causal relationships between biological entities. |
| | KEGG [148] | A disease and pathway database. |
| | PANTHER Pathway [145] | An Ontology-Based Pathway Database Coupled with Data Analysis Tools. |
| | Pathway Commons [185] | A web resource for biological pathway data. |
| | PhosphoSitePlus [134] | A comprehensive web services for post-translational modifications. |
| | PICR [136] | Reconciling protein identifiers across multiple source databases. |
| | Reactome [146] | A database of reactions, pathways, and biological processes. |

**Author Contributions:** C.C. and J.C. were involved in the concept of the study; C.C., J.H., J.J.T., and J.C. were involved in the drafting and revising of the manuscript. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by an NIH grant (R01GM093123) and two NSF grants (DBI1759934, IIS1763246).

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

## Abbreviations

| | |
| --- | --- |
| MS | Mass spectrometry |
| FDR | False discovery rate |
| AP-MS | Affinity purification-mass spectrometry |
| DIA | Data-independent acquisition |
| PTM | Posttranslational modifications |
| LC-MS/MS | Liquid-chromatography coupled with tandem mass spectrometry |
| MALDI IMS | Matrix-assisted laser desorption/ionization imaging mass spectrometry |
| PSEA | Protein set enrichment analysis |
| GSEA | Gene set enrichment analysis |
| VSN | Variance stabilization normalization |
| HMM | Hidden Markov model |
| SVM | Support vector machines |
| LIMMA | Linear Models for Microarray Data |
| PSM | peptide spectrum match |
| ICAT | Isotope coded affinity tag |
| ICPL | isotope coded protein labeling |
| SILAC | stable isotope labeling by amino acids in cell culture |
| PPI | Protein-protein interaction |

## References

1. Anderson, L.; Seilhamer, J. A comparison of selected mRNA and protein abundances in human liver. *Electrophoresis* **1997**, *18*, 533–537. [CrossRef] [PubMed]
2. Gygi, S.P.; Rochon, Y.; Franza, B.R.; Aebersold, R. Correlation between Protein and mRNA Abundance in Yeast. *Mol. Cell. Biol.* **1999**, *19*, 1720. [CrossRef] [PubMed]
3. Sharaf, A.; Mensching, L.; Keller, C.; Rading, S.; Scheffold, M.; Palkowitsch, L.; Djogo, N.; Rezgaoui, M.; Kestler, H.A.; Moepps, B.; et al. Systematic Affinity Purification Coupled to Mass Spectrometry Identified p62 as Part of the Cannabinoid Receptor CB2 Interactome. *Front Mol. Neurosci.* **2019**, *12*, 224. [CrossRef] [PubMed]
4. Strasser, S.D.; Ghazi, P.C.; Starchenko, A.; Boukhali, M.; Edwards, A.; Suarez-Lopez, L.; Lyons, J.; Changelian, P.S.; Monahan, J.B.; Jacobsen, J.; et al. Substrate-based kinase activity inference identifies MK2 as driver of colitis. *Integr. Biol.* **2019**, *11*, 301–314.
5. Kar, U.K.; Simonian, M.; Whitelegge, J.P. Integral membrane proteins: Bottom-up, top-down and structural proteomics. *Expert Rev. Proteomics* **2017**, *14*, 715–723. [CrossRef]

6. Gillet, L.C.; Leitner, A.; Aebersold, R. Mass Spectrometry Applied to Bottom-Up Proteomics: Entering the High-Throughput Era for Hypothesis Testing. *Annu. Rev. Anal. Chem.* **2016**, *9*, 449–472. [CrossRef]
7. Toby, T.K.; Fornelli, L.; Kelleher, N.L. Progress in Top-Down Proteomics and the Analysis of Proteoforms. *Annu. Rev. Anal. Chem.* **2016**, *9*, 499–519. [CrossRef]
8. Donnelly, D.P.; Rawlins, C.M.; DeHart, C.J.; Fornelli, L.; Schachner, L.F.; Lin, Z.; Lippens, J.L.; Aluri, K.C.; Sarin, R.; Chen, B.; et al. Best practices and benchmarks for intact protein analysis for top-down mass spectrometry. *Nat. Methods* **2019**, *16*, 587–594. [CrossRef]
9. Domon, B.; Aebersold, R. Challenges and Opportunities in Proteomics Data Analysis. *Mol. Cell. Proteom.* **2006**, *5*, 1921. [CrossRef]
10. Eng, J.K.; McCormack, A.L.; Yates, J.R. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J. Am. Soc. Mass Spectrom.* **1994**, *5*, 976–989. [CrossRef]
11. Geer, L.Y.; Markey, S.P.; Kowalak, J.A.; Wagner, L.; Xu, M.; Maynard, D.M.; Yang, X.; Shi, W.; Bryant, S.H. Open mass spectrometry search algorithm. *J. Proteome Res.* **2004**, *3*, 958–964. [CrossRef] [PubMed]
12. Craig, R.; Beavis, R.C. TANDEM: Matching proteins with tandem mass spectra. *Bioinformatics* **2004**, *20*, 1466–1467. [CrossRef]
13. Dančík, V.; Addona, T.A.; Clauser, K.R.; Vath, J.E.; Pevzner, P.A. De novo peptide sequencing via tandem mass spectrometry. *J. Comput. Biol.* **1999**, *6*, 327–342. [CrossRef] [PubMed]
14. Frank, A.; Pevzner, P. PepNovo: De novo peptide sequencing via probabilistic network modeling. *Anal. Chem.* **2005**, *77*, 964–973. [CrossRef]
15. Shevchenko, A.; Chernushevich, I.; Ens, W.; Standing, K.G.; Thomson, B.; Wilm, M.; Mann, M. Rapid 'de novo'peptide sequencing by a combination of nanoelectrospray, isotopic labeling and a quadrupole/time-of-flight mass spectrometer. *Rapid Commun. Mass Spectrom.* **1997**, *11*, 1015–1024. [CrossRef]
16. Eng, J.K.; Fischer, B.; Grossmann, J.; Maccoss, M.J. A fast SEQUEST cross correlation algorithm. *J. Proteome Res.* **2008**, *7*, 4598–4602. [CrossRef]
17. Perkins, D.N.; Pappin, D.J.; Creasy, D.M.; Cottrell, J.S. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* **1999**, *20*, 3551–3567. [CrossRef]
18. Tyanova, S.; Temu, T.; Cox, J. The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. *Nat. Protoc.* **2016**, *11*, 2301–2319. [CrossRef]
19. Wang, J.; Tucholska, M.; Knight, J.D.; Lambert, J.P.; Tate, S.; Larsen, B.; Gingras, A.C.; Bandeira, N. MSPLIT-DIA: Sensitive peptide identification for data-independent acquisition. *Nat. Methods* **2015**, *12*, 1106–1108. [CrossRef]
20. Schirmer, E.C.; Yates, J.R., 3rd; Gerace, L. MudPIT: A powerful proteomics tool for discovery. *Discov. Med.* **2003**, *3*, 38–39.
21. Edwards, N.J. PepArML: A Meta-Search Peptide Identification Platform for Tandem Mass Spectra. *Curr. Protoc. Bioinforma.* **2013**, *44*, 11–23. [CrossRef] [PubMed]
22. Wan, Y.; Yang, A.; Chen, T. PepHMM: A hidden Markov model based scoring function for mass spectrometry database search. *Anal. Chem.* **2006**, *78*, 432–437. [CrossRef] [PubMed]
23. Chalkley, R.J.; Baker, P.R.; Huang, L.; Hansen, K.C.; Allen, N.P.; Rexach, M.; Burlingame, A.L. Comprehensive analysis of a multidimensional liquid chromatography mass spectrometry dataset acquired on a quadrupole selecting, quadrupole collision cell, time-of-flight mass spectrometer: II. New developments in Protein Prospector allow for reliable and comprehensive automatic analysis of large datasets. *Mol. Cell. Proteomics* **2005**, *4*, 1194–1204. [PubMed]
24. Brodbelt, J.S.; Russell, D.H. Focus on the 20-year anniversary of SEQUEST. *J. Am. Soc. Mass Spectrom.* **2015**, *26*, 1797–1798. [CrossRef] [PubMed]
25. Kou, Q.; Xun, L.; Liu, X. TopPIC: A software tool for top-down mass spectrometry-based proteoform identification and characterization. *Bioinformatics* **2016**, *32*, 3495–3497. [CrossRef] [PubMed]
26. Bjornson, R.D.; Carriero, N.J.; Colangelo, C.; Shifman, M.; Cheung, K.H.; Miller, P.L.; Williams, K. X!!Tandem, an improved method for running X!tandem in parallel on collections of commodity computers. *J. Proteome Res.* **2008**, *7*, 293–299. [CrossRef] [PubMed]
27. 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]

28. Karpievitch, Y.V.; Nikolic, S.B.; Wilson, R.; Sharman, J.E.; Edwards, L.M. Metabolomics data normalization with EigenMS. *PLoS One* **2014**, *9*, e116221. [CrossRef]
29. Fischer, B.; Roth, V.; Roos, F.; Grossmann, J.; Baginsky, S.; Widmayer, P.; Gruissem, W.; Buhmann, J.M. NovoHMM: A hidden Markov model for de novo peptide sequencing. *Anal. Chem.* **2005**, *77*, 7265–7273. [CrossRef]
30. Ma, B.; Zhang, K.; Hendrie, C.; Liang, C.; Li, M.; Doherty-Kirby, A.; Lajoie, G. PEAKS: Powerful software for peptide de novo sequencing by tandem mass spectrometry. *Rapid Commun. Mass Spectrom.* **2003**, *17*, 2337–2342. [CrossRef]
31. Ting, Y.S.; Egertson, J.D.; Bollinger, J.G.; Searle, B.C.; Payne, S.H.; Noble, W.S.; MacCoss, M.J. PECAN: Library-free peptide detection for data-independent acquisition tandem mass spectrometry data. *Nat. Methods* **2017**, *14*, 903–908. [CrossRef] [PubMed]
32. Yang, H.; Chi, H.; Zeng, W.F.; Zhou, W.J.; He, S.M. pNovo 3: Precise de novo peptide sequencing using a learning-to-rank framework. *Bioinformatics* **2019**, *35*, i183–i190. [CrossRef] [PubMed]
33. Li, C.; Li, K.; Li, K.; Xie, X.; Lin, F. SWPepNovo: An Efficient De Novo Peptide Sequencing Tool for Large-scale MS/MS Spectra Analysis. *Int. J. Biol. Sci.* **2019**, *15*, 1787–1801. [CrossRef] [PubMed]
34. Jeong, K.; Kim, S.; Pevzner, P.A. UniNovo: A universal tool for de novo peptide sequencing. *Bioinformatics* **2013**, *29*, 1953–1962. [CrossRef]
35. Bern, M.; Kil, Y.J.; Becker, C. Byonic: Advanced peptide and protein identification software. *Curr. Protoc. Bioinform.* **2012**, *40*, 13.20.1–13.20.14. [CrossRef]
36. Tabb, D.L.; Ma, Z.Q.; Martin, D.B.; Ham, A.J.; Chambers, M.C. DirecTag: Accurate sequence tags from peptide MS/MS through statistical scoring. *J. Proteome Res.* **2008**, *7*, 3838–3846. [CrossRef]
37. Tanner, S.; Shu, H.; Frank, A.; Wang, L.C.; Zandi, E.; Mumby, M.; Pevzner, P.A.; Bafna, V. InsPecT: Identification of posttranslationally modified peptides from tandem mass spectra. *Anal. Chem.* **2005**, *77*, 4626–4639. [CrossRef]
38. Wang, X.; Li, Y.; Wu, Z.; Wang, H.; Tan, H.; Peng, J. JUMP: A tag-based database search tool for peptide identification with high sensitivity and accuracy. *Mol. Cel.l Proteomics* **2014**, *13*, 3663–3673. [CrossRef]
39. Zhang, J.; Xin, L.; Shan, B.; Chen, W.; Xie, M.; Yuen, D.; Zhang, W.; Zhang, Z.; Lajoie, G.A.; Ma, B. PEAKS DB: De novo sequencing assisted database search for sensitive and accurate peptide identification. *Mol. Cell. Proteomics* **2012**, *11*, M111.010587. [CrossRef]
40. Cifani, P.; Dhabaria, A.; Chen, Z.; Yoshimi, A.; Kawaler, E.; Abdel-Wahab, O.; Poirier, J.T.; Kentsis, A. ProteomeGenerator: A Framework for Comprehensive Proteomics Based on de Novo Transcriptome Assembly and High-Accuracy Peptide Mass Spectral Matching. *J. Proteome Res.* **2018**, *17*, 3681–3692. [CrossRef]
41. Yang, X.; Dondeti, V.; Dezube, R.; Maynard, D.M.; Geer, L.Y.; Epstein, J.; Chen, X.; Markey, S.P.; Kowalak, J.A. DBParser: Web-based software for shotgun proteomic data analyses. *J. Proteome Res.* **2004**, *3*, 1002–1008. [CrossRef] [PubMed]
42. Tsou, C.C.; Avtonomov, D.; Larsen, B.; Tucholska, M.; Choi, H.; Gingras, A.C.; Nesvizhskii, A.I. DIA-Umpire: Comprehensive computational framework for data-independent acquisition proteomics. *Nat. Methods* **2015**, *12*, 258–264. [CrossRef]
43. Slotta, D.J.; McFarland, M.A.; Markey, S.P. MassSieve: Panning MS/MS peptide data for proteins. *Proteomics* **2010**, *10*, 3035–3039. [CrossRef] [PubMed]
44. Reiter, L.; Claassen, M.; Schrimpf, S.P.; Jovanovic, M.; Schmidt, A.; Buhmann, J.M.; Hengartner, M.O.; Aebersold, R. Protein identification false discovery rates for very large proteomics data sets generated by tandem mass spectrometry. *Mol. Cell. Proteomics* **2009**, *8*, 2405–2417. [CrossRef] [PubMed]
45. Savitski, M.M.; Nielsen, M.L.; Zubarev, R.A. ModifiComb, a new proteomic tool for mapping substoichiometric post-translational modifications, finding novel types of modifications, and fingerprinting complex protein mixtures. *Mol. Cell. Proteomics* **2006**, *5*, 935–948. [CrossRef] [PubMed]
46. Gonnelli, G.; Stock, M.; Verwaeren, J.; Maddelein, D.; De Baets, B.; Martens, L.; Degroeve, S. A decoy-free approach to the identification of peptides. *J. Proteome Res.* **2015**, *14*, 1792–1798. [CrossRef]
47. May, D.H.; Tamura, K.; Noble, W.S. Param-Medic: A Tool for Improving MS/MS Database Search Yield by Optimizing Parameter Settings. *J. Proteome Res.* **2017**, *16*, 1817–1824. [CrossRef]
48. 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]

49. Weatherly, D.B.; Atwood, J.A.; Minning, T.A.; Cavola, C.; Tarleton, R.L.; Orlando, R. A Heuristic Method for Assigning a False-discovery Rate for Protein Identifications from Mascot Database Search Results. *Mol. Cell. Proteomics* **2005**, *4*, 762. [CrossRef]
50. Solntsev, S.K.; Shortreed, M.R.; Frey, B.L.; Smith, L.M. Enhanced Global Post-translational Modification Discovery with MetaMorpheus. *J. Proteome Res.* **2018**, *17*, 1844–1851. [CrossRef]
51. Perchey, R.T.; Tonini, L.; Tosolini, M.; Fournié, J.-J.; Lopez, F.; Besson, A.; Pont, F. PTMselect: Optimization of protein modifications discovery by mass spectrometry. *Sci. Rep.* **2019**, *9*, 4181. [CrossRef]
52. Mortensen, P.; Gouw, J.W.; Olsen, J.V.; Ong, S.E.; Rigbolt, K.T.; Bunkenborg, J.; Cox, J.; Foster, L.J.; Heck, A.J.; Blagoev, B.; et al. MSQuant, an open source platform for mass spectrometry-based quantitative proteomics. *J. Proteome Res.* **2010**, *9*, 393–403. [CrossRef] [PubMed]
53. Petyuk, V.A.; Mayampurath, A.M.; Monroe, M.E.; Polpitiya, A.D.; Purvine, S.O.; Anderson, G.A.; Camp, D.G., 2nd; Smith, R.D. DtaRefinery, a software tool for elimination of systematic errors from parent ion mass measurements in tandem mass spectra data sets. *Mol. Cell. Proteomics* **2010**, *9*, 486–496. [CrossRef] [PubMed]
54. Petyuk, V.A.; Jaitly, N.; Moore, R.J.; Ding, J.; Metz, T.O.; Tang, K.; Monroe, M.E.; Tolmachev, A.V.; Adkins, J.N.; Belov, M.E.; et al. Elimination of systematic mass measurement errors in liquid chromatography-mass spectrometry based proteomics using regression models and a priori partial knowledge of the sample content. *Anal. Chem.* **2008**, *80*, 693–706. [CrossRef] [PubMed]
55. Kil, Y.J.; Becker, C.; Sandoval, W.; Goldberg, D.; Bern, M. Preview: A program for surveying shotgun proteomics tandem mass spectrometry data. *Anal. Chem.* **2011**, *83*, 5259–5267. [CrossRef]
56. Tabb, D.L. The SEQUEST family tree. *J. Am. Soc. Mass Spectrom.* **2015**, *26*, 1814–1819. [CrossRef]
57. Cox, J.; Neuhauser, N.; Michalski, A.; Scheltema, R.A.; Olsen, J.V.; Mann, M. Andromeda: A peptide search engine integrated into the MaxQuant environment. *J. Proteome Res.* **2011**, *10*, 1794–1805. [CrossRef]
58. Webb-Robertson, B.-J.M. Support vector machines for improved peptide identification from tandem mass spectrometry database search. In *Mass Spectrometry of Proteins and Peptides*; Humana Press: Totowa, NJ, USA, 2009; pp. 453–460.
59. Lin, A.; Howbert, J.J.; Noble, W.S. Combining High-Resolution and Exact Calibration To Boost Statistical Power: A Well-Calibrated Score Function for High-Resolution MS2 Data. *J. Proteome Res.* **2018**, *17*, 3644–3656. [CrossRef]
60. Elias, J.E.; Gygi, S.P. Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nat. Methods* **2007**, *4*, 207–214. [CrossRef]
61. Elias, J.E.; Gygi, S.P. Target-decoy search strategy for mass spectrometry-based proteomics. *Methods Mol. Biol.* **2010**, *604*, 55–71.
62. Kim, H.; Lee, S.; Park, H. Target-small decoy search strategy for false discovery rate estimation. *BMC Bioinforma.* **2019**, *20*, 438. [CrossRef]
63. Fischer, B.; Roth, V.; Grossmann, J.; Baginsky, S.; Gruissem, W.; Roos, F.; Widmayer, P.; Buhmann, J.M. A hidden markov model for de novo peptide sequencing. In *Advances in Neural Information Processing Systems*; MIT Press: Cambridge, MA, USA, 2005; pp. 457–464.
64. Tran, N.H.; Zhang, X.; Xin, L.; Shan, B.; Li, M. De novo peptide sequencing by deep learning. *Proce. Nat. Acad. Sci.* **2017**, *114*, 8247–8252. [CrossRef]
65. Chapman, J.D.; Goodlett, D.R.; Masselon, C.D. Multiplexed and data-independent tandem mass spectrometry for global proteome profiling. *Mass Spectrom. Rev.* **2014**, *33*, 452–470. [CrossRef] [PubMed]
66. Weisbrod, C.R.; Eng, J.K.; Hoopmann, M.R.; Baker, T.; Bruce, J.E. Accurate peptide fragment mass analysis: Multiplexed peptide identification and quantification. *J. Proteome Res.* **2012**, *11*, 1621–1632. [CrossRef]
67. Searle, B.C.; Swearingen, K.E.; Barnes, C.A.; Schmidt, T.; Gessulat, S.; Kuster, B.; Wilhelm, M. Generating high quality libraries for DIA MS with empirically corrected peptide predictions. *Nat. Commun.* **2020**, *11*, 1548. [CrossRef]
68. Li, Q.; Shortreed, M.R.; Wenger, C.D.; Frey, B.L.; Schaffer, L.V.; Scalf, M.; Smith, L.M. Global Post-Translational Modification Discovery. *J. Proteome Res.* **2017**, *16*, 1383–1390. [CrossRef] [PubMed]
69. Han, X.; He, L.; Xin, L.; Shan, B.; Ma, B. PeaksPTM: Mass spectrometry-based identification of peptides with unspecified modifications. *J. Proteome Res.* **2011**, *10*, 2930–2936. [CrossRef]
70. Nesvizhskii, A.I. Proteogenomics: Concepts, applications and computational strategies. *Nat. Methods* **2014**, *11*, 1114. [CrossRef] [PubMed]

71. Alves, P.; Arnold, R.J.; Novotny, M.V.; Radivojac, P.; Reilly, J.P.; Tang, H. Advancement in protein inference from shotgun proteomics using peptide detectability. In *Biocomputing 2007*; World Scientific: Singapore, 2007; pp. 409–420.
72. Nesvizhskii, A.I.; Keller, A.; Kolker, E.; Aebersold, R. A statistical model for identifying proteins by tandem mass spectrometry. *Anal. Chem.* **2003**, *75*, 4646–4658. [CrossRef] [PubMed]
73. Shen, C.; Wang, Z.; Shankar, G.; Zhang, X.; Li, L. A hierarchical statistical model to assess the confidence of peptides and proteins inferred from tandem mass spectrometry. *Bioinformatics* **2008**, *24*, 202–208. [CrossRef]
74. Li, Y.F.; Arnold, R.J.; Li, Y.; Radivojac, P.; Sheng, Q.; Tang, H. A Bayesian approach to protein inference problem in shotgun proteomics. *J. Comput. Biol.* **2009**, *16*, 1183–1193. [CrossRef] [PubMed]
75. Serang, O.; MacCoss, M.J.; Noble, W.S. Efficient marginalization to compute protein posterior probabilities from shotgun mass spectrometry data. *J. Proteome Res.* **2010**, *9*, 5346–5357. [CrossRef] [PubMed]
76. Wu, G.; Wan, X.; Xu, B. A new estimation of protein-level false discovery rate. *BMC Genomics* **2018**, *19*, 567. [CrossRef] [PubMed]
77. Arntzen, M.O.; Koehler, C.J.; Barsnes, H.; Berven, F.S.; Treumann, A.; Thiede, B. IsobariQ: Software for isobaric quantitative proteomics using IPTL, iTRAQ, and TMT. *J. Proteome Res.* **2011**, *10*, 913–920. [CrossRef]
78. Shadforth, I.P.; Dunkley, T.P.; Lilley, K.S.; Bessant, C. i-Tracker: For quantitative proteomics using iTRAQ. *BMC Genomics* **2005**, *6*, 145. [CrossRef]
79. Deutsch, E.W.; Mendoza, L.; Shteynberg, D.; Farrah, T.; Lam, H.; Tasman, N.; Sun, Z.; Nilsson, E.; Pratt, B.; Prazen, B.; et al. A guided tour of the Trans-Proteomic Pipeline. *Proteomics* **2010**, *10*, 1150–1159. [CrossRef]
80. Khan, Z.; Bloom, J.S.; Garcia, B.A.; Singh, M.; Kruglyak, L. Protein quantification across hundreds of experimental conditions. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 15544–15548. [CrossRef]
81. Han, D.K.; Eng, J.; Zhou, H.; Aebersold, R. Quantitative profiling of differentiation-induced microsomal proteins using isotope-coded affinity tags and mass spectrometry. *Nat. Biotechnol.* **2001**, *19*, 946–951. [CrossRef]
82. Ishihama, Y.; Oda, Y.; Tabata, T.; Sato, T.; Nagasu, T.; Rappsilber, J.; Mann, M. 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.* **2005**, *4*, 1265–1272. [CrossRef]
83. 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. Proteomics* **2014**, *13*, 2513–2526. [CrossRef]
84. Monroe, M.E.; Tolic, N.; Jaitly, N.; Shaw, J.L.; Adkins, J.N.; Smith, R.D. VIPER: An advanced software package to support high-throughput LC-MS peptide identification. *Bioinformatics* **2007**, *23*, 2021–2023. [CrossRef] [PubMed]
85. Sturm, M.; Bertsch, A.; Gropl, C.; Hildebrandt, A.; Hussong, R.; Lange, E.; Pfeifer, N.; Schulz-Trieglaff, O.; Zerck, A.; Reinert, K.; et al. OpenMS - an open-source software framework for mass spectrometry. *BMC Bioinforma.* **2008**, *9*, 163. [CrossRef] [PubMed]
86. Tran, N.H.; Rahman, M.Z.; He, L.; Xin, L.; Shan, B.; Li, M. Complete De Novo Assembly of Monoclonal Antibody Sequences. *Sci. Rep.* **2016**, *6*, 31730. [CrossRef]
87. MacLean, B.; Tomazela, D.M.; Shulman, N.; Chambers, M.; Finney, G.L.; Frewen, B.; Kern, R.; Tabb, D.L.; Liebler, D.C.; MacCoss, M.J. Skyline: An open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics* **2010**, *26*, 966–968. [CrossRef] [PubMed]
88. Cao, R.; Chen, K.; Song, Q.; Zang, Y.; Li, J.; Wang, X.; Chen, P.; Liang, S. Quantitative proteomic analysis of membrane proteins involved in astroglial differentiation of neural stem cells by SILAC labeling coupled with LC–MS/MS. *J. Proteome Res.* **2012**, *11*, 829–838. [CrossRef]
89. Merrill, A.E.; Hebert, A.S.; MacGilvray, M.E.; Rose, C.M.; Bailey, D.J.; Bradley, J.C.; Wood, W.W.; El Masri, M.; Westphall, M.S.; Gasch, A.P.; et al. NeuCode labels for relative protein quantification. *Mol. Cell. Proteomics* **2014**, *13*, 2503–2512. [CrossRef]
90. 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]
91. Schmidt, A.; Kellermann, J.; Lottspeich, F. A novel strategy for quantitative proteomics using isotope-coded protein labels. *Proteomics* **2005**, *5*, 4–15. [CrossRef]

92. Ong, S.E.; Blagoev, B.; Kratchmarova, I.; Kristensen, D.B.; Steen, H.; Pandey, A.; Mann, M. Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics. *Mol. Cell. Proteomics.* **2002**, *1*, 376–386. [CrossRef]
93. Oda, Y.; Huang, K.; Cross, F.R.; Cowburn, D.; Chait, B.T. Accurate quantitation of protein expression and site-specific phosphorylation. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 6591–6596. [CrossRef]
94. Rauniyar, N.; Yates, J.R., 3rd. Isobaric labeling-based relative quantification in shotgun proteomics. *J. Proteome Res.* **2014**, *13*, 5293–5309. [CrossRef] [PubMed]
95. Zecha, J.; Satpathy, S.; Kanashova, T.; Avanessian, 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. Proteomics* **2019**, *18*, 1468–1478. [CrossRef] [PubMed]
96. Wiese, S.; Reidegeld, K.A.; Meyer, H.E.; Warscheid, B. Protein labeling by iTRAQ: A new tool for quantitative mass spectrometry in proteome research. *Proteomics* **2007**, *7*, 340–350. [CrossRef] [PubMed]
97. Li, H.; Hwang, K.B.; Mun, D.G.; Kim, H.; Lee, H.; Lee, S.W.; Paek, E. Estimating influence of cofragmentation on peptide quantification and identification in iTRAQ experiments by simulating multiplexed spectra. *J. Proteome Res.* **2014**, *13*, 3488–3497. [CrossRef] [PubMed]
98. Pedrioli, P.G.; Raught, B.; Zhang, X.D.; Rogers, R.; Aitchison, J.; Matunis, M.; Aebersold, R. Automated identification of SUMOylation sites using mass spectrometry and SUMmOn pattern recognition software. *Nat. Methods* **2006**, *3*, 533–539. [CrossRef] [PubMed]
99. Nahnsen, S.; Bielow, C.; Reinert, K.; Kohlbacher, O. Tools for label-free peptide quantification. *Mol. Cell. Proteomics* **2013**, *12*, 549–556. [CrossRef]
100. Leung, K.Y.; Lescuyer, P.; Campbell, J.; Byers, H.L.; Allard, L.; Sanchez, J.C.; Ward, M.A. A novel strategy using MASCOT Distiller for analysis of cleavable isotope-coded affinity tag data to quantify protein changes in plasma. *Proteomics* **2005**, *5*, 3040–3044. [CrossRef]
101. Mallick, P.; Schirle, M.; Chen, S.S.; Flory, M.R.; Lee, H.; Martin, D.; Ranish, J.; Raught, B.; Schmitt, R.; Werner, T.; et al. Computational prediction of proteotypic peptides for quantitative proteomics. *Nat. Biotechnol.* **2007**, *25*, 125–131. [CrossRef]
102. Wilhelm, M.; Schlegl, J.; Hahne, H.; Gholami, A.M.; Lieberenz, M.; Savitski, M.M.; Ziegler, E.; Butzmann, L.; Gessulat, S.; Marx, H.; et al. Mass-spectrometry-based draft of the human proteome. *Nature* **2014**, *509*, 582–587. [CrossRef]
103. Silva, J.C.; Gorenstein, M.V.; Li, G.Z.; Vissers, J.P.; Geromanos, S.J. Absolute quantification of proteins by LCMSE: A virtue of parallel MS acquisition. *Mol. Cell. Proteomics* **2006**, *5*, 144–156. [CrossRef]
104. 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] [PubMed]
105. Wisniewski, J.R.; Hein, M.Y.; Cox, J.; Mann, M. A "proteomic ruler" for protein copy number and concentration estimation without spike-in standards. *Mol. Cell. Proteomics* **2014**, *13*, 3497–3506. [CrossRef] [PubMed]
106. Hanrieder, J.; Phan, N.T.; Kurczy, M.E.; Ewing, A.G. Imaging mass spectrometry in neuroscience. *ACS Chem. Neurosci.* **2013**, *4*, 666–679. [CrossRef] [PubMed]
107. Baker, T.C.; Han, J.; Borchers, C.H. Recent advancements in matrix-assisted laser desorption/ionization mass spectrometry imaging. *Curr. Opin. Biotechnol.* **2017**, *43*, 62–69. [CrossRef] [PubMed]
108. Jungnickel, H.; Laux, P.; Luch, A. Time-of-Flight Secondary Ion Mass Spectrometry (ToF-SIMS): A New Tool for the Analysis of Toxicological Effects on Single Cell Level. *Toxics* **2016**, *4*, 5. [CrossRef]
109. Girod, M.; Shi, Y.; Cheng, J.X.; Cooks, R.G. Desorption electrospray ionization imaging mass spectrometry of lipids in rat spinal cord. *J. Am. Soc. Mass Spectrom.* **2010**, *21*, 1177–1189. [CrossRef]
110. Alexandrov, T.; Becker, M.; Deininger, S.O.; Ernst, G.; Wehder, L.; Grasmair, M.; von Eggeling, F.; Thiele, H.; Maass, P. Spatial segmentation of imaging mass spectrometry data with edge-preserving image denoising and clustering. *J. Proteome Res.* **2010**, *9*, 6535–6546. [CrossRef]
111. Alexandrov, T.; Kobarg, J.H. Efficient spatial segmentation of large imaging mass spectrometry datasets with spatially aware clustering. *Bioinformatics* **2011**, *27*, i230–i238. [CrossRef]
112. Kallback, P.; Shariatgorji, M.; Nilsson, A.; Andren, P.E. Novel mass spectrometry imaging software assisting labeled normalization and quantitation of drugs and neuropeptides directly in tissue sections. *J. Proteomics* **2012**, *75*, 4941–4951. [CrossRef]

113. Wisniewski, J.R.; Mann, M. A Proteomics Approach to the Protein Normalization Problem: Selection of Unvarying Proteins for MS-Based Proteomics and Western Blotting. *J. Proteome Res.* **2016**, *15*, 2321–2326. [CrossRef]
114. Välikangas, T.; Suomi, T.; Elo, L.L. A systematic evaluation of normalization methods in quantitative label-free proteomics. *Brief. Bioinforma.* **2016**, *19*, 1–11. [CrossRef] [PubMed]
115. Berger, J.A.; Hautaniemi, S.; Järvinen, A.-K.; Edgren, H.; Mitra, S.K.; Astola, J. Optimized LOWESS normalization parameter selection for DNA microarray data. *BMC Bioinforma.* **2004**, *5*, 194. [CrossRef] [PubMed]
116. Huber, W.; von Heydebreck, A.; Sultmann, H.; Poustka, A.; Vingron, M. Variance stabilization applied to microarray data calibration and to the quantification of differential expression. *Bioinformatics* **2002**, *18* (Suppl. 1), S96–S104. [CrossRef]
117. Bern, M.; Goldberg, D. De novo analysis of peptide tandem mass spectra by spectral graph partitioning. *J. Comput. Biol.* **2006**, *13*, 364–378. [CrossRef] [PubMed]
118. Wei, R.; Wang, J.; Su, M.; Jia, E.; Chen, S.; Chen, T.; Ni, Y. Missing Value Imputation Approach for Mass Spectrometry-based Metabolomics Data. *Sci. Rep.* **2018**, *8*, 663. [CrossRef] [PubMed]
119. Berg, P.; McConnell, E.W.; Hicks, L.M.; Popescu, S.C.; Popescu, G.V. Evaluation of linear models and missing value imputation for the analysis of peptide-centric proteomics. *BMC Bioinforma.* **2019**, *20*, 102. [CrossRef] [PubMed]
120. Ling, W.; Dong-Mei, F. Estimation of Missing Values Using a Weighted K-Nearest Neighbors Algorithm. In Proceedings of the 2009 International Conference on Environmental Science and Information Application Technology, Wuhan, China, 4–5 July 2009; pp. 660–663.
121. Bergamo, G.C.; Dias, C.T.d.S.; Krzanowski, W.J. Distribution-free multiple imputation in an interaction matrix through singular value decomposition. *Sci. Agric.* **2008**, *65*, 422–427. [CrossRef]
122. Webb-Robertson, B.-J.M.; Wiberg, H.K.; Matzke, M.M.; Brown, J.N.; Wang, J.; McDermott, J.E.; Smith, R.D.; Rodland, K.D.; Metz, T.O.; Pounds, J.G.; et al. Review, Evaluation, and Discussion of the Challenges of Missing Value Imputation for Mass Spectrometry-Based Label-Free Global Proteomics. *J. Proteome Res.* **2015**, *14*, 1993–2001. [CrossRef]
123. Wei, R.; Wang, J.; Jia, E.; Chen, T.; Ni, Y.; Jia, W. GSimp: A Gibbs sampler based left-censored missing value imputation approach for metabolomics studies. *PLoS Comput. Biol.* **2018**, *14*, e1005973. [CrossRef]
124. Krzywinski, M.; Altman, N. Significance, P values and t-tests. *Nat. Methods* **2013**, *10*, 1041. [CrossRef]
125. McHugh, M.L. Multiple comparison analysis testing in ANOVA. *Biochem. Med. (Zagreb)* **2011**, *21*, 203–209. [CrossRef]
126. Kammers, K.; Cole, R.N.; Tiengwe, C.; Ruczinski, I. Detecting significant changes in protein abundance. *EuPA Open Proteomics* **2015**, *7*, 11–19. [CrossRef]
127. Hill, E.G.; Schwacke, J.H.; Comte-Walters, S.; Slate, E.H.; Oberg, A.L.; Eckel-Passow, J.E.; Therneau, T.M.; Schey, K.L. A statistical model for iTRAQ data analysis. *J. Proteome Res.* **2008**, *7*, 3091–3101. [CrossRef]
128. Herbrich, S.M.; Cole, R.N.; West, K.P., Jr.; Schulze, K.; Yager, J.D.; Groopman, J.D.; Christian, P.; Wu, L.; O'Meally, R.N.; May, D.H.; et al. Statistical inference from multiple iTRAQ experiments without using common reference standards. *J. Proteome Res.* **2013**, *12*, 594–604. [CrossRef]
129. van Iterson, M.; Boer, J.M.; Menezes, R.X. Filtering, FDR and power. *BMC Bioinforma.* **2010**, *11*, 450. [CrossRef]
130. Xie, Y.; Pan, W.; Khodursky, A.B. A note on using permutation-based false discovery rate estimates to compare different analysis methods for microarray data. *Bioinformatics* **2005**, *21*, 4280–4288. [CrossRef]
131. Choi, H.; Nesvizhskii, A.I. False Discovery Rates and Related Statistical Concepts in Mass Spectrometry-Based Proteomics. *J. Proteome Res.* **2008**, *7*, 47–50. [CrossRef]
132. Dennis, G.; Sherman, B.T.; Hosack, D.A.; Yang, J.; Gao, W.; Lane, H.C.; Lempicki, R.A. DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biol.* **2003**, *4*, R60. [CrossRef]
133. Szklarczyk, D.; Morris, J.H.; Cook, H.; Kuhn, M.; Wyder, S.; Simonovic, M.; Santos, A.; Doncheva, N.T.; Roth, A.; Bork, P.; et al. The STRING database in 2017: Quality-controlled protein-protein association networks, made broadly accessible. *Nucleic. Acids. Res.* **2017**, *45*, D362–D368. [CrossRef]
134. Hornbeck, P.V.; Kornhauser, J.M.; Tkachev, S.; Zhang, B.; Skrzypek, E.; Murray, B.; Latham, V.; Sullivan, M. PhosphoSitePlus: A comprehensive resource for investigating the structure and function of experimentally determined post-translational modifications in man and mouse. *Nucleic. Acids. Res.* **2012**, *40*, D261–D270. [CrossRef]

135. Perfetto, L.; Briganti, L.; Calderone, A.; Cerquone Perpetuini, A.; Iannuccelli, M.; Langone, F.; Licata, L.; Marinkovic, M.; Mattioni, A.; Pavlidou, T.; et al. SIGNOR: A database of causal relationships between biological entities. *Nucleic. Acids. Res.* **2015**, *44*, D548–D554. [CrossRef]
136. Côté, R.G.; Jones, P.; Martens, L.; Kerrien, S.; Reisinger, F.; Lin, Q.; Leinonen, R.; Apweiler, R.; Hermjakob, H. The Protein Identifier Cross-Referencing (PICR) service: Reconciling protein identifiers across multiple source databases. *BMC Bioinforma.* **2007**, *8*, 401. [CrossRef]
137. Waegele, B.; Dunger-Kaltenbach, I.; Fobo, G.; Montrone, C.; Mewes, H.W.; Ruepp, A. CRONOS: The cross-reference navigation server. *Bioinformatics* **2008**, *25*, 141–143. [CrossRef]
138. Howe, D.; Costanzo, M.; Fey, P.; Gojobori, T.; Hannick, L.; Hide, W.; Hill, D.P.; Kania, R.; Schaeffer, M.; St Pierre, S.; et al. Big data: The future of biocuration. *Nature* **2008**, *455*, 47–50. [CrossRef]
139. Gene Ontology, C. The Gene Ontology (GO) database and informatics resource. *Nucleic. Acids. Res.* **2004**, *32*, D258–D261. [CrossRef]
140. Carbon, S.; Ireland, A.; Mungall, C.J.; Shu, S.; Marshall, B.; Lewis, S.; the AmiGO Hub; the Web Presence Working Group. AmiGO: Online access to ontology and annotation data. *Bioinformatics* **2008**, *25*, 288–289. [CrossRef]
141. Huang, W.-L.; Tung, C.-W.; Ho, S.-W.; Hwang, S.-F.; Ho, S.-Y. ProLoc-GO: Utilizing informative Gene Ontology terms for sequence-based prediction of protein subcellular localization. *BMC Bioinforma.* **2008**, *9*, 80. [CrossRef]
142. Hawkins, T.; Chitale, M.; Luban, S.; Kihara, D. PFP: Automated prediction of gene ontology functional annotations with confidence scores using protein sequence data. *Proteins: Struct. Funct. Bioinforma.* **2009**, *74*, 566–582. [CrossRef]
143. Piovesan, D.; Giollo, M.; Leonardi, E.; Ferrari, C.; Tosatto, S.C.E. INGA: Protein function prediction combining interaction networks, domain assignments and sequence similarity. *Nucleic Acids Res.* **2015**, *43*, W134–W140. [CrossRef]
144. Welzenbach, J.; Neuhoff, C.; Heidt, H.; Cinar, U.M.; Looft, C.; Schellander, K.; Tholen, E.; Große-Brinkhaus, C. Integrative Analysis of Metabolomic, Proteomic and Genomic Data to Reveal Functional Pathways and Candidate Genes for Drip Loss in Pigs. *Int. J. Mol. Sci.* **2016**, *17*, 1426. [CrossRef]
145. Mi, H.; Thomas, P. PANTHER Pathway: An Ontology-Based Pathway Database Coupled with Data Analysis Tools. In *Protein Networks and Pathway Analysis*; Nikolsky, Y., Bryant, J., Eds.; Humana Press: Totowa, NJ, USA, 2009; pp. 123–140. [CrossRef]
146. Croft, D.; O'Kelly, G.; Wu, G.; Haw, R.; Gillespie, M.; Matthews, L.; Caudy, M.; Garapati, P.; Gopinath, G.; Jassal, B.; et al. Reactome: A database of reactions, pathways and biological processes. *Nucleic Acids Res.* **2010**, *39*, D691–D697. [CrossRef]
147. Luo, W.; Brouwer, C. Pathview: An R/Bioconductor package for pathway-based data integration and visualization. *Bioinformatics* **2013**, *29*, 1830–1831. [CrossRef] [PubMed]
148. Kanehisa, M.; Furumichi, M.; Tanabe, M.; Sato, Y.; Morishima, K. KEGG: New perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res.* **2017**, *45*, D353–D361. [CrossRef] [PubMed]
149. Lavallée-Adam, M.; Rauniyar, N.; McClatchy, D.B.; Yates, J.R. PSEA-Quant: A Protein Set Enrichment Analysis on Label-Free and Label-Based Protein Quantification Data. *J. Proteome Res.* **2014**, *13*, 5496–5509. [CrossRef]
150. Isik, Z.; Ercan, M.E. Integration of RNA-Seq and RPPA data for survival time prediction in cancer patients. *Comput. Biol. Med.* **2017**, *89*, 397–404. [CrossRef]
151. Deeb, S.J.; Tyanova, S.; Hummel, M.; Schmidt-Supprian, M.; Cox, J.; Mann, M. Machine Learning-based Classification of Diffuse Large B-cell Lymphoma Patients by Their Protein Expression Profiles. *Mol. Cell. Proteomics* **2015**, *14*, 2947. [CrossRef]
152. Agranoff, D.; Fernandez-Reyes, D.; Papadopoulos, M.C.; Rojas, S.A.; Herbster, M.; Loosemore, A.; Tarelli, E.; Sheldon, J.; Schwenk, A.; Pollok, R.; et al. Identification of diagnostic markers for tuberculosis by proteomic fingerprinting of serum. *Lancet* **2006**, *368*, 1012–1021. [CrossRef]
153. Tyanova, S.; Albrechtsen, R.; Kronqvist, P.; Cox, J.; Mann, M.; Geiger, T. Proteomic maps of breast cancer subtypes. *Nat. Commun.* **2016**, *7*, 10259. [CrossRef]
154. Itzhak, D.N.; Tyanova, S.; Cox, J.; Borner, G.H. Global, quantitative and dynamic mapping of protein subcellular localization. *Elife* **2016**, *5*, e16950. [CrossRef]

155. Su, Y.; Shi, Q.; Wei, W. Single cell proteomics in biomedicine: High-dimensional data acquisition, visualization, and analysis. *PROTEOMICS* **2017**, *17*, 1600267. [CrossRef]
156. Ding, M.Q.; Chen, L.; Cooper, G.F.; Young, J.D.; Lu, X. Precision Oncology beyond Targeted Therapy: Combining Omics Data with Machine Learning Matches the Majority of Cancer Cells to Effective Therapeutics. *Mol. Cancer Res.* **2018**, *16*, 269. [CrossRef] [PubMed]
157. Wolpert, D.H. The Supervised Learning No-Free-Lunch Theorems. In *Soft Computing and Industry: Recent Applications*; Roy, R., Köppen, M., Ovaska, S., Furuhashi, T., Hoffmann, F., Eds.; Springer London: London, UK, 2002; pp. 25–42. [CrossRef]
158. Edwards, N.; Wu, X.; Tseng, C.-W. An Unsupervised, Model-Free, Machine-Learning Combiner for Peptide Identifications from Tandem Mass Spectra. *Clin. Proteomics* **2009**, *5*, 23–36. [CrossRef]
159. Asgari, E.; Mofrad, M.R.K. Continuous Distributed Representation of Biological Sequences for Deep Proteomics and Genomics. *PLOS ONE* **2015**, *10*, e0141287. [CrossRef] [PubMed]
160. Palsson, B.Ø. *Systems Biology: Properties of Reconstructed Networks*; Cambridge University Press: Cambridge, UK, 2006.
161. Bauer, A.; Kuster, B. Affinity purification-mass spectrometry. *Eur. J. Biochem.* **2003**, *270*, 570–578. [CrossRef] [PubMed]
162. Rinner, O.; Mueller, L.N.; Hubálek, M.; Müller, M.; Gstaiger, M.; Aebersold, R. An integrated mass spectrometric and computational framework for the analysis of protein interaction networks. *Nat. Biotechnol.* **2007**, *25*, 345–352. [CrossRef] [PubMed]
163. Glatter, T.; Wepf, A.; Aebersold, R.; Gstaiger, M. An integrated workflow for charting the human interaction proteome: Insights into the PP2A system. *Mol. Systems Biol.* **2009**, *5*, 237. [CrossRef] [PubMed]
164. Tornow, S.; Mewes, H.W. Functional modules by relating protein interaction networks and gene expression. *Nucleic Acids Rese.* **2003**, *31*, 6283–6289. [CrossRef]
165. Xiong, H.U.I.; He, X.; Ding, C.; Zhang, Y.A.; Kumar, V.; Holbrook, S.R. Identification of functional modules in protein complexes via hyperclique pattern discovery. In *Biocomputing 2005*; World Scientific: Singapore, 2004; pp. 221–232. [CrossRef]
166. Kozina, N.; Mihaljević, Z.; Lončar, B.M.; Mihalj, M.; Mišir, M.; Radmilović, D.M.; Justić, H.; Gajović, S.; Šešelja, K.; Bazina, I.; et al. Impact of High Salt Diet on Cerebral Vascular Function and Stroke in Tff3−/−/C57BL/6N Knockout and WT (C57BL/6N) Control Mice. *Int. J. Mol. Sci.* **2019**, *20*, 5188. [CrossRef]
167. Benabdelkamel, H.; Masood, A.; Okla, M.; Al-Naami, Y.M.; Alfadda, A.A. A Proteomics-Based Approach Reveals Differential Regulation of Urine Proteins between Metabolically Healthy and Unhealthy Obese Patients. *Int. J. Mol. Sci.* **2019**, *20*, 4905. [CrossRef]
168. Schmidl, S.R.; Gronau, K.; Pietack, N.; Hecker, M.; Becher, D.; Stulke, J. The phosphoproteome of the minimal bacterium Mycoplasma pneumoniae: Analysis of the complete known Ser/Thr kinome suggests the existence of novel kinases. *Mol. Cell. Proteomics* **2010**, *9*, 1228–1242. [CrossRef]
169. Arora, G.; Sajid, A.; Arulanandh, M.D.; Singhal, A.; Mattoo, A.R.; Pomerantsev, A.P.; Leppla, S.H.; Maiti, S.; Singh, Y. Unveiling the novel dual specificity protein kinases in Bacillus anthracis: Identification of the first prokaryotic dual specificity tyrosine phosphorylation-regulated kinase (DYRK)-like kinase. *J. Biol. Chem.* **2012**, *287*, 26749–26763. [CrossRef]
170. Ravikumar, V.; Shi, L.; Krug, K.; Derouiche, A.; Jers, C.; Cousin, C.; Kobir, A.; Mijakovic, I.; Macek, B. Quantitative phosphoproteome analysis of Bacillus subtilis reveals novel substrates of the kinase PrkC and phosphatase PrpC. *Mol. Cell. Proteomics* **2014**, *13*, 1965–1978. [CrossRef]
171. Singhal, A.; Arora, G.; Virmani, R.; Kundu, P.; Khanna, T.; Sajid, A.; Misra, R.; Joshi, J.; Yadav, V.; Samanta, S.; et al. Systematic Analysis of Mycobacterial Acylation Reveals First Example of Acylation-mediated Regulation of Enzyme Activity of a Bacterial Phosphatase. *J. Biol. Chem.* **2015**, *290*, 26218–26234. [CrossRef]
172. Birhanu, A.G.; Yimer, S.A.; Holm-Hansen, C.; Norheim, G.; Aseffa, A.; Abebe, M.; Tonjum, T. Nepsilon- and O-Acetylation in Mycobacterium tuberculosis Lineage 7 and Lineage 4 Strains: Proteins Involved in Bioenergetics, Virulence, and Antimicrobial Resistance Are Acetylated. *J. Proteome Res.* **2017**, *16*, 4045–4059. [CrossRef]
173. 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.* **2019**. [CrossRef]

174. Pang, H.; Li, W.; Zhang, W.; Zhou, S.; Hoare, R.; Monaghan, S.J.; Jian, J.; Lin, X. Acetylome profiling of Vibrio alginolyticus reveals its role in bacterial virulence. *J. Proteomics* **2020**, *211*, 103543. [CrossRef]
175. Mischnik, M.; Sacco, F.; Cox, J.; Schneider, H.-C.; Schäfer, M.; Hendlich, M.; Crowther, D.; Mann, M.; Klabunde, T. IKAP: A heuristic framework for inference of kinase activities from Phosphoproteomics data. *Bioinformatics* **2015**, *32*, 424–431. [CrossRef]
176. Wiredja, D.D.; Koyutürk, M.; Chance, M.R. The KSEA App: A web-based tool for kinase activity inference from quantitative phosphoproteomics. *Bioinformatics* **2017**, *33*, 3489–3491. [CrossRef]
177. Wirbel, J.; Cutillas, P.; Saez-Rodriguez, J. Phosphoproteomics-Based Profiling of Kinase Activities in Cancer Cells. In *Cancer Systems Biology: Methods and Protocols*; von Stechow, L., Ed.; Springer New York: New York, NY, USA, 2018; pp. 103–132. [CrossRef]
178. Hill, S.M.; Heiser, L.M.; Cokelaer, T.; Unger, M.; Nesser, N.K.; Carlin, D.E.; Zhang, Y.; Sokolov, A.; Paull, E.O.; Wong, C.K.; et al. Inferring causal molecular networks: Empirical assessment through a community-based effort. *Nat. Methods* **2016**, *13*, 310. [CrossRef]
179. Chen, C.; Hou, J.; Cheng, J. GNET2: Constructing gene regulatory networks from expression data through functional module inference. *Bioconductor* **2019**.
180. Mei, S.; Zhang, K. Neglog: Homology-Based Negative Data Sampling Method for Genome-Scale Reconstruction of Human Protein–Protein Interaction Networks. *Int. J. Mol. Sci.* **2019**, *20*, 5075. [CrossRef] [PubMed]
181. Alanis-Lobato, G.; Andrade-Navarro, M.A.; Schaefer, M.H. HIPPIE v2.0: Enhancing meaningfulness and reliability of protein–protein interaction networks. *Nucleic Acids Res.* **2016**, *45*, D408–D414. [CrossRef] [PubMed]
182. Sun, X.; Weckwerth, W. COVAIN: A toolbox for uni- and multivariate statistics, time-series and correlation network analysis and inverse estimation of the differential Jacobian from metabolomics covariance data. *Metabolomics* **2012**, *8*, 81–93. [CrossRef]
183. Wang, S.; Lv, Y.; Wang, Y.; Du, P.; Tan, W.; Lammi, M.J.; Guo, X. Network Analysis of Se-and Zn-related Proteins in the Serum Proteomics Expression Profile of the Endemic Dilated Cardiomyopathy Keshan Disease. *Biol. Trace Element Res.* **2018**, *183*, 40–48. [CrossRef]
184. Pirhaji, L.; Milani, P.; Leidl, M.; Curran, T.; Avila-Pacheco, J.; Clish, C.B.; White, F.M.; Saghatelian, A.; Fraenkel, E. Revealing disease-associated pathways by network integration of untargeted metabolomics. *Nat. Methods* **2016**, *13*, 770. [CrossRef]
185. Cerami, E.G.; Gross, B.E.; Demir, E.; Rodchenkov, I.; Babur, O.; Anwar, N.; Schultz, N.; Bader, G.D.; Sander, C. Pathway Commons, a web resource for biological pathway data. *Nucleic Acids Res.* **2011**, *39*, D685–D690. [CrossRef]
186. Cheerathodi, M.R.; Meckes, D.G., Jr. BioID Combined with Mass Spectrometry to Study Herpesvirus Protein-Protein Interaction Networks. *Methods Mol. Biol.* **2020**, *2060*, 327–341.
187. Pappireddi, N.; Martin, L.; Wuhr, M. A Review on Quantitative Multiplexed Proteomics. *Chembiochem* **2019**, *20*, 1210–1224. [CrossRef]
188. Robles, M.S.; Cox, J.; Mann, M. In-vivo quantitative proteomics reveals a key contribution of post-transcriptional mechanisms to the circadian regulation of liver metabolism. *PLoS Genet.* **2014**, *10*, e1004047. [CrossRef]
189. Alvarez, M.J.; Giorgi, F.; Califano, A. Using viper, a package for Virtual Inference of Protein-activity by Enriched Regulon analysis. *Bioconductor* **2014**, 1–14.
190. Szklarczyk, D.; Gable, A.L.; Lyon, D.; Junge, A.; Wyder, S.; Huerta-Cepas, J.; Simonovic, M.; Doncheva, N.T.; Morris, J.H.; Bork, P.; et al. STRING v11: Protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* **2019**, *47*, D607–D613. [CrossRef]