



Review

# Review of Liquid Chromatography-Mass Spectrometry-Based Proteomic Analyses of Body Fluids to Diagnose Infectious Diseases

Hayoung Lee <sup>1,2</sup>  and Seung Il Kim <sup>1,2,\*</sup>

<sup>1</sup> Research Center for Bioconvergence Analysis, Korea Basic Science Institute (KBSI), Ochang 28119, Korea; lhy3221@kbsi.re.kr

<sup>2</sup> Bio-Analytical Science Division, University of Science and Technology (UST), Daejeon 34113, Korea

\* Correspondence: ksi@kbsi.re.kr

**Abstract:** Rapid and precise diagnostic methods are required to control emerging infectious diseases effectively. Human body fluids are attractive clinical samples for discovering diagnostic targets because they reflect the clinical statuses of patients and most of them can be obtained with minimally invasive sampling processes. Body fluids are good reservoirs for infectious parasites, bacteria, and viruses. Therefore, recent clinical proteomics methods have focused on body fluids when aiming to discover human- or pathogen-originated diagnostic markers. Cutting-edge liquid chromatography-mass spectrometry (LC-MS)-based proteomics has been applied in this regard; it is considered one of the most sensitive and specific proteomics approaches. Here, the clinical characteristics of each body fluid, recent tandem mass spectroscopy (MS/MS) data-acquisition methods, and applications of body fluids for proteomics regarding infectious diseases (including the coronavirus disease of 2019 [COVID-19]), are summarized and discussed.

**Keywords:** bodily fluid; biomarker discovery; COVID-19; infectious disease; mass spectrometry; pathogen; proteomics



**Citation:** Lee, H.; Kim, S.I. Review of Liquid Chromatography-Mass Spectrometry-Based Proteomic Analyses of Body Fluids to Diagnose Infectious Diseases. *Int. J. Mol. Sci.* **2022**, *23*, 2187. <https://doi.org/10.3390/ijms23042187>

Academic Editor: Karel Doležal

Received: 13 December 2021

Accepted: 14 February 2022

Published: 16 February 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

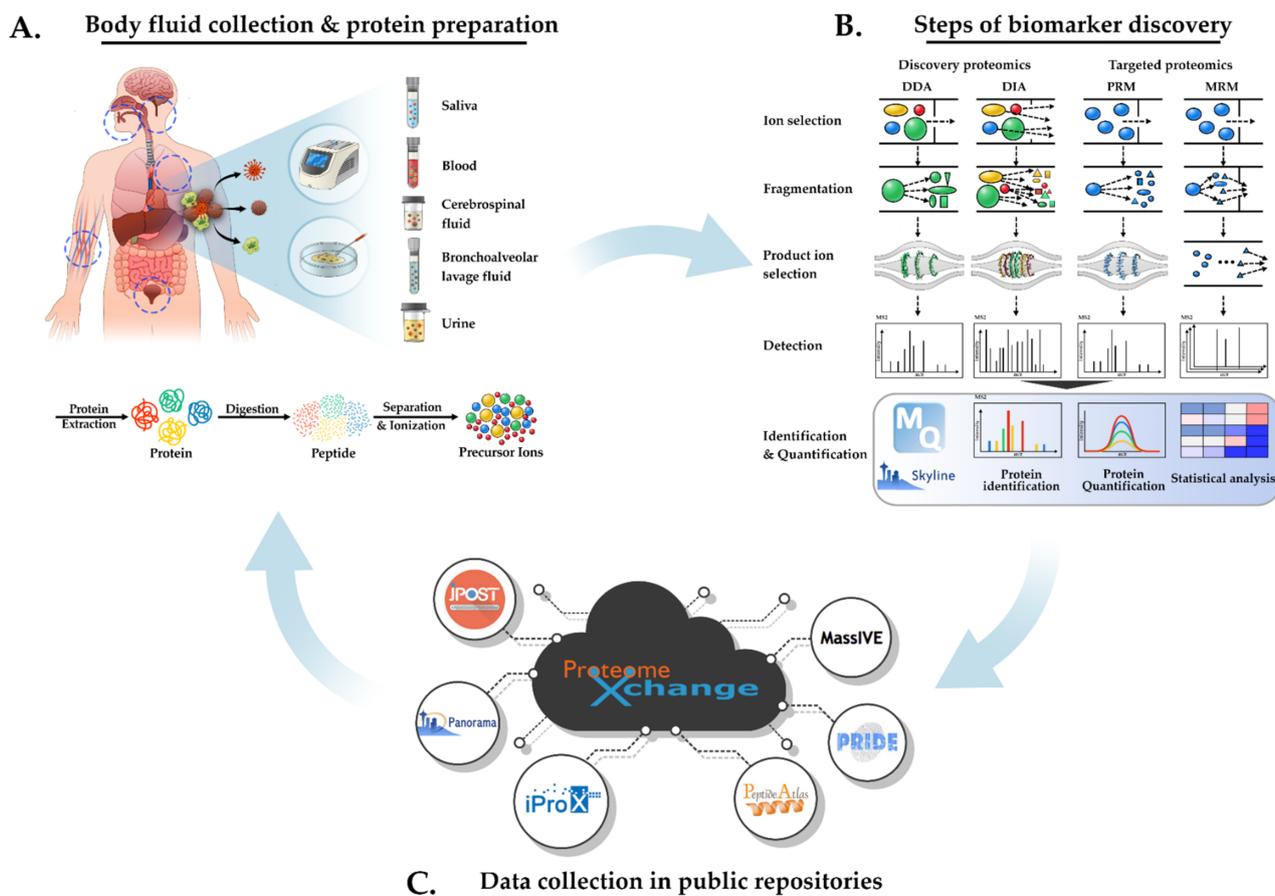
Infectious diseases are major threats to global health, as evidenced by the current coronavirus disease of 2019 (COVID-19) pandemic. Even though many treatments for infectious diseases are available, they remain one of the leading causes of death in the world [1–4]. Therefore, it is crucial to better understand the physiology of pathogens (bacteria, viruses, fungi, and parasites) at the molecular level [5]. The first step when studying bacterial pathogens involves identifying molecular features that contribute to pathogenicity. These features can be potential therapeutic targets, and their inhibition can eradicate or counteract bacterial infections.

Proteomics is a useful tool for studying infectious diseases because it can provide large-scale protein information involved in the pathogenesis, infection mechanisms, and pathological symptoms of hosts. Modern proteomic methods have evolved from gel-based techniques to gel-free mass spectrometry (MS) approaches known as “shotgun” proteomics. Shotgun proteomics utilizes nano-high precision liquid chromatography (HPLC) systems coupled to high-resolution mass spectrometers; it has revolutionized the proteomic research field by allowing large-scale protein characterization with high throughput [6]. For the discovery of biomarker candidates or pathogen detection in the infected host, MS-based proteomics have several advantages. MS-based proteomics have high performance regarding its detection limits (<1 nM), repeatability, and reproducibility, in comparison to immunoassays. The advantage of this technology is that it can be conducted in a multiplexed manner, without the use of antibodies or comparable binders [7]. In addition, MS-based analyses are now generally integrated into many hospital laboratories for the

routine identification of bacterial pathogens in clinical samples, as well as for antibiotic resistance testing. Proteomic studies related to infectious diseases can be categorized into three groups according to target samples as follows: (1) pure-cultured pathogen, (2) infected host proteome, and (3) pathogens in the infected host.

- (1) Pure-cultured pathogens: the proteomic analysis of pathogens grown under pure culture conditions has several advantages regarding the characterization of pathogens. First, this approach provides accurate translational information of each gene at the genome-wide level. Second, it is easy to control the cell culture conditions of pathogens and acquire their proteomic responses. However, the pure-culture conditions of pathogens are different from real infection conditions. In general, host systems infected with pathogens provide more severe and diverse culture conditions. Additionally, many pathogens related to human diseases cannot be cultured in laboratory environments [8].
- (2) Infected host proteome: the host proteome infected with pathogens is another important target for proteomic analysis. This approach can provide valuable information on host–pathogen interactions, the infection mechanisms of pathogens (pathogenic bacteria or viruses), and the pathological symptoms of hosts. The study of the interactions between microbial pathogens and their hosts is called “infectomics”; it constitutes a growing area of interest in proteomics. Infection sites within a host are also diverse, including the respiratory system, digestive system, nerve systems, skin, and body fluid. Therefore, many clinical samples are available. However, though host-derived biomarkers are useful for monitoring disease status, they are limited for discerning between similar diseases [9].
- (3) Pathogens in the infected host: detecting pathogens (pathogenic bacteria or viruses) from an infected host is the most direct method for the diagnosis, prognosis, treatment, and clinical characterization of infectious diseases. Body fluids can be useful samples for this analysis.

Thus, MS-based proteomics are expected to be used in the future as a tool to rapidly identify pathogens from human biological specimens. Figure 1 shows an overview of LC-MS based proteomics to discover biomarkers for infectious diseases. However, to date, the direct detection of pathogens from the host is still difficult to routinely implement, owing to several technical hurdles. At this point, this review is about the detection of pathogens in the body fluids of an infected host. At first, it covers the characteristics of each body fluid as samples for clinical proteomics, in Section 2 (Figure 1A). It also briefly introduces current proteomics approaches (data-dependent acquisition/data-independent acquisition (DDA/DIA), and targeted proteomics) that have been applied to identify the pathogen in the infected host, in addition to reviewing a coronavirus disease-19 (COVID-19) case to showcase the most up-to-date technology applied, in Section 3 (Figure 1B). Afterwards, future challenges are discussed in Section 4 (Figure 1C).



**Figure 1.** Schematic depictions of liquid chromatography-mass spectrometry (LC-MS) to discover biomarkers for infectious diseases. These workflows are designed to discover pathogen-originated biomarkers. (A) Body fluid collection and protein preparation. Body fluids are useful for diagnosing infectious diseases. The body fluids of patients with infectious diseases are screened using polymerase chain reaction (PCR) or culture tests and then collected. Proteins are extracted from body fluids and enzymatically digested into tryptic peptides. The resulting peptides are applied to LC/MS for separation and ionization. (B) The steps of biomarker discovery using various acquisition methods. Protein identification, quantification, and statistical analysis methods are used to identify useful biomarkers. Mass spectrometry (MS) analysis is categorized by discovery proteomics (data-dependent acquisition [DDA] and data-independent acquisition [DIA]) and targeted proteomics (parallel reaction monitoring [PRM] and multiple reaction monitoring [MRM]). The principles of each mass technology are described in detail in Section 2.2. The advantages and disadvantages of each approach are summarized in Table 3. Acquired fragmented spectra are translated into peptide sequences and then inferred to identify proteins using proteomics software such as MaxQuant or Skyline. The intensities or peak areas of the acquired peptides are used for comparative analysis of the corresponding proteins between clinical samples. As a next step, various statistical analyses of MS data can help to discover potential biomarkers indicative of specific infectious diseases. (C) Data collection in public repositories for further applications. Many MS data produced by previous studies can be deposited in public repositories. These data should be further curated to be housed in an open database, which can be used for discovery or validation studies.

## 2. Using Body Fluids for the Proteomic Analysis of Infectious Diseases

### 2.1. Body Fluids as Valuable Clinical Samples for Proteomic Analysis

As described in the introduction, proteomic targets for infectious diseases can be categorized into three groups. Among them, two targets (host-infected with pathogens and pathogens in the infected host) are important for screening clinically available biomarkers

for diagnosis or treatment [10]. Body fluids can be used as the main source of these targets [5,11]. Body fluids are defined as “liquids within the human body”; they can be classified as being either systematic or proximal (Figure 1A). Systematic fluids represent the host’s overall physiological state. However, proximal fluids are limited to specific tissue, but represent the status of adjacent tissue.

The advantages of body fluids include the fact that they are easier to collect with less invasive methods, with less demanding sample preparation methods than those with the tissue before proteomic analysis [12]. Several body fluids can be analyzed, but we chose five commonly studied body fluids. The characteristics of the representative body fluids are summarized in Table 1.

**Table 1.** Characteristics of body fluids.

Sample Type	Blood	BALF <sup>a</sup>	CSF <sup>b</sup>	Urine	Saliva	Ref.
Non-invasive collection	Moderate	Moderate	No	Yes	Yes	[13]
Expert required	Moderate	Yes (local anesthesia)	Yes (local anesthesia)	No	No	[14]
Protein concentration (mg/mL)	60–80	0.05–0.2	0.2–0.8	0.08 <sup>c</sup>	0.5–2	[14–17]
Complexity	Highest	High	High	Moderate	High	[18]
Proposed * SOPs for collection	Yes	Yes	Yes	Yes	Yes	[14,19–22]

\* SOPs, standard operating procedures; BALF<sup>a</sup>, bronchoalveolar lavage fluid; CSF<sup>b</sup>, cerebrospinal fluid; <sup>c</sup> variable depends on hydration.

Blood is the most popular body fluid for host proteomic analysis because it represents the proteome of the whole organism, as it contacts every tissue in the body. It also contains many hidden or unknown proteins, such as cytokines, hormones, and antibodies, thus providing insights on the clinical or physiological conditions of the host [23]. Both these characteristics are valuable aspects that other body fluids do not have. Therefore, blood is a useful clinical sample to guide the treatment of various diseases. Serum or plasma is the blood fraction used for this purpose. However, fewer than 15 of the most abundant but least informative proteins account for more than 90% of the total protein components of these fractions. The high complexity of protein contents and additional modifications make the analysis more complicated [24]. Strategies for removing these abundant proteins without losing informative low-copy proteins are essential to ensure the effective screening of most indicative biomarkers. There are still technical challenges in identifying low-abundance proteins in the blood.

Bronchoalveolar lavage fluid (BALF) is a biofluid obtained using fiber-optic bronchoscopy. This fluid reflects the protein composition of the pulmonary airway. Therefore, proteomic analysis of the BALF can provide information on airway diseases. BALF can be a more sensitive clinical sample than nasal swabs regarding respiratory molecular diagnostics, but its sampling method is ineffective and invasive and requires considerable time and costs. The BALF proteome is dominated by plasma-derived proteins such as albumin and immunoglobulins (65–80%), which makes its analysis difficult [25]. However, BALF can be a valuable clinical sample for the screening of lung-specific disease biomarkers [14].

Cerebrospinal fluid (CSF) is a body fluid that surrounds the ventricular system of the central nervous system, spinal cord, and brain. CSF transports waste products and nutrients, mediating molecular exchange with blood plasma. As it connects with the blood system, most of its protein contents are similar, but the protein concentration is lower than that of plasma [15,26]. CSF also acts as a mechanical support for the spinal cord and brain. It is considered to be an ideal clinical sample for the detection of neurological disorders or diseases, such as multiple sclerosis, meningitis, and spinal cord injuries. However, the disadvantage of CSF is its associated invasive sample collection method, which requires lumbar puncture. Blood contamination and the presence of blood plasma proteins in the CSF are further disadvantages regarding the identification of biomarkers by CSF proteomic analysis [15].

Urine is a liquid secreted by the kidneys; it is the result of glomerular filtration of the plasma to eliminate waste products, such as urea and metabolites. It also includes proteins secreted from the urinary tract and renal tubular epithelial cells. Therefore, urine covers related diseases of the blood, kidneys, bladder, and urinary tract. One advantage of using urine as a clinical sample is that it can be collected in large quantities in a non-invasive manner. It is also possible to collect urine samples repeatedly, which aids in time-resolved studies. Furthermore, urine has a 1000-fold lower proteome complexity ( $0.08 \text{ mg}\cdot\text{mL}^{-1}$ ) than serum or plasma [27]. Finally, proteins or peptides in urine are stable because the urine proteome remains in the bladder for a considerable time before excretion, after the proteolytic process mediated by endogenous proteases. These characteristics make urine a convenient source for discovering biomarkers. To date, the major targets in urine proteomics have involved the screening of biomarkers of renal and urogenital dysfunctions [28], but it has also been applied to other infectious diseases described later [29–33]. Saliva is a fluid secreted by the salivary glands and gingival crevice [34]. It contains more than 1000 proteins that originate from glands and plasma [16]. Cystatins,  $\alpha$ -amylase, mucin, albumin, globulins, and serotransferrin are the major proteins [35]. As the protein complexity of saliva is relatively low, it is preferred for biomarker screening by proteomic analysis. Like urine, saliva has the advantage of easy and non-invasive sample collection. Periodontal diseases, such as periodontitis, oral cancer, autoimmune disease, and diabetes mellitus, are popular targets of saliva proteomics [34–36].

The standard operating procedures (SOPs) of authorized institutions should be considered to obtain reproducible results (Table 1). This will be helpful to minimize pre-analytical variables while ensuring quality and uniformity among samples. There are some excellent reviews summarizing detailed information on SOPs [14,19–22]. Although each body fluid has unique characteristics, several steps for obtaining high-quality results are similar. The commonly considered factors related to protein sample preparation are summarized in Table 2.

**Table 2.** Considered factors for protein sample preparation.

Factors	Description	Ref.
Immunodepletion	Immunodepletion is generally applied to remove high-abundance proteins and enrich low-abundance proteins.	[37–40]
Solubility of target proteins	MS-grade detergent can be applied to target low-abundance and hydrophobic proteins, such as membrane proteins.	[41]
Efficiency of protein preparation	The applicability of automation of protein isolation methods or extraction efficiency is critical to large-scale projects.	[42–44]
Peptide prefractionation or enrichment	Enrichment methods based on affinity binding require large starting protein amounts.	[18,45,46]

For example, sample fractionation or depletion is commonly required before tandem mass spectroscopy (MS/MS) analysis regardless of the type of body fluids because of the proteome complexity or the presence of high abundance proteins in body fluids [12,27]. In the case of blood, the dynamic ranges differ by a factor of  $10^{10}$  between serum albumin (35–50 mg/mL in normal conditions), the most abundant protein, and cytokines (low pg/mL range). Unfortunately, this needs to be considered because immunodepletion can also co-deplete low-abundance proteins. Several studies have shown limited success in depleting high-abundance proteins and enriching low-abundance proteins [37–40]. Alternative methods to simplify sample complexity, such as fractionation, have been applied optionally, which increases the number of LC-MS/MS runs. The advantages and disadvantages of each fractionation method have been described well previously [18,45,47].

The solubility of target proteins needs to be considered because avoiding sample loss owing to precipitation or aggregation and efficient digestion into peptides by proteases are prerequisites for MS-based bottom-up analysis [48]. Several MS-compatible reagents are commercially available to enhance the solubility of proteins or enzymatic digestions. Waas and colleagues evaluated the efficiency of eight commercially available reagents based

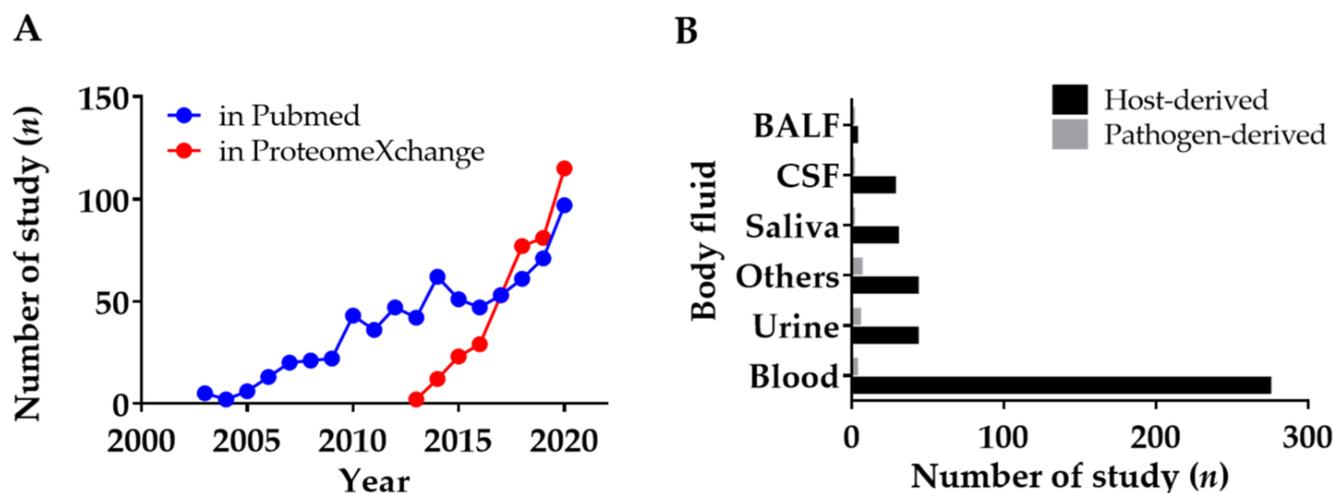
on the number of peptides and proteins identified, total protein sequence coverage, and digestion specificity in various conditions [41].

The digestion methods are the major bottle necks for large-scale projects. They are also the source of experimental variability and need to be performed as consistently as possible. To develop an optimal process, an extensive set of protocols have been developed which are divided into two categories: in-solution process and cleanup methods [49]. In-solution processing methods include in-solution digestion [50] and integrated StageTip workflow [51]. Those have simple protocols and can be adapted to high-throughput processes, but the presence of reagents can hinder downstream processes. Cleanup methods include in-gel methods, filter-aided sample preparation [52], suspension trapping [53], single-pot, solid-phase-enhance sample-preparation (SP3) [49], precipitation, and other affinity-based methods. Those methods provide high-quality results but require time consuming and laborious processes. Recently, Müller and colleagues proposed fully automated high-throughput and streamlined workflows for clinical samples using SP3 methods, which can deal with 96 samples in 3.5 h [44]. A detailed comparison of these methods is beyond the scope of this paper. However, several studies have described the differences among such methods [18,42,43,46,49,53].

Over 500,000 peptides can be derived from body fluid proteins per sample. The hundreds of co-eluting peptides were ionized and analyzed together. The high dynamic range and difference in ionization efficiency of the co-eluting peptides can affect MS/MS analysis [54]. Peptide prefractionation or enrichment is one way to solve this problem. Those procedures are usually achieved by chromatographic and electrophoretic fractionation to reduce the number of co-elutes. Strong-cation exchange coupled with reverse phase (RP), high-pH RP coupled with RP, or hydrophilic interaction and RP chromatography are commonly applied for peptide prefractionation [18,46]. An interesting reference paper by Wasinger and colleagues provides more detail of the separation of peptides using various one- or multidimensional methods for LC-MS analysis [45].

## 2.2. Increase in Applications of Body Fluids for Proteomics of Infectious Diseases

With the spread of COVID-19, scientists and clinicians are now paying more attention to the rapid and effective diagnosis of infectious diseases than ever before. Body fluid proteomics has been considered an emerging technology to identify novel biomarkers. For this review of body fluid proteomics, more than 630 research papers related to body fluid proteomics were identified in PubMed using the following words: ((Biomarker) AND (Proteomics)) AND (Infectious disease OR Emerging disease) AND ((Plasm\*) OR (Seru\*) OR (bloo\*) OR (Urin\*) OR (Cerebro\*) OR (Bronchoalveolar\*) OR (Body fluid) OR (Liquid biopsy)) NOT (Cancer). Additionally, 422 deposited datasets related to infectious diseases were also found in the data repository proteomeXchange, which is a data repository for proteomics studies [55,56]. Continuous increases in research papers and deposited datasets related to body fluid proteomics for infectious diseases can thus be seen (Figure 2A). The deposited datasets were manually curated according to the type of body fluid and the type of targets in Figure 2B. Unexpectedly, only a few studies were discovered that targeted pathogen-derived proteins, which are major interests of this review. Most proteomic studies have focused on infected human or host proteomics.



**Figure 2.** Current statistics of studies on body fluid proteomics for infectious diseases. (A) Gradual increases in research papers and deposited datasets on body fluid proteomics for infectious diseases. These results were based on PubMed and ProteomeXchange searches conducted in 2020 for the past 20 years (2002–2020). (B) Utilization of various body fluids for diagnosing infectious diseases divided based on the targets. These results were based on ProteomeXchange searches. The statistics showed that body fluid proteomics is an emerging proteomics field for pathogen-originated biomarker discovery.

### 3. Application of LC-MS Proteomic Analysis for Identification of Pathogens Using Body Fluids Associated with Infectious Diseases

MS analysis can be divided into three methods according to the acquisition method: DDA, DIA, and targeted-mass spectrometry (parallel reaction monitoring [PRM] and multiple reaction monitoring [MRM]). The characteristics of the acquisition methods are summarized in Table 3. Figure 1B also shows the fundamental concepts that each acquisition method. In this section, the characteristics of each method are briefly described, focusing on body fluid proteomics studies of infectious diseases that apply these MS methods for biomarker discovery, as summarized in Table 4.

**Table 3.** Characteristics of LC-MS acquisition method.

Acquisition Methods	DDA <sup>a</sup>	MRM/PRM <sup>b</sup>	DIA <sup>c</sup>	Ref.
Requirement for high-quality instruments	High	Moderate/High	Highest	[57]
Accuracy of protein quantification	Low	Highest	High	[58,59]
Reproducibility between replicates	Low	Highest	High	[59,60]
Depth of protein identification	Highest	Low	High	[58,59]
Ease of data analysis	Easy	Moderate	Hard	[61–64]

DDA <sup>a</sup>, data-dependent acquisition; MRM <sup>b</sup>, multiple reaction monitoring; PRM, parallel reaction monitoring; DIA <sup>c</sup>, data-independent acquisition.

**Table 4.** Summary of body fluid proteomics for targeting pathogen-derived proteins.

Body Fluid	Study Groups	Sample Size	Target Pathogen	Major Findings	Method	Instrument	Ref.
Urine	Active pulmonary tuberculosis (TB)	9, 21	<i>Mycobacterium tuberculosis</i>	Four mycobacterial proteins were identified from the urine of nine patients. One of the candidate proteins was reconfirmed in urine from 21 clinical samples.	DDA	LCQ	[29,30]
Urine	Active TB vs. latent TB vs. non-TB	21 vs. 24 vs. 18	<i>Mycobacterium tuberculosis</i>	Ten mycobacterial proteins of active TB and six mycobacterial proteins of latent TB were identified.	DDA	LTQ-Orbitrap Velos Pro	[31]
CSF, Urine, Serum, and Saliva	Sleeping sickness early-stage vs. late-stage vs. uninfected	3 vs. 4 vs. 3	<i>Trypanosoma brucei gambiense</i>	Parasite proteins were identified but not further analyzed because of a lack of validity.	DDA	Q Exactive	[65]
Urine	Syphilis patient vs. Healthy	54 vs. 6	<i>Treponema pallidum</i>	The 26 unique peptides derived from 4 unique <i>T. pallidum</i> proteins were identified. These proteins have low sequence similarity to the human protein.	DDA, DIA	Synapt MS	[32]
Blood	Malaria patient	7	<i>Plasmodium vivax</i>	Five parasite-derived proteins of <i>P. vivax</i> were identified in 80% of patients.	DDA	6550 iFunnel Q-TOF	[66]
Urine	Urinary tract infection (UTI) patient	27	15 bacterial species <sup>a</sup>	Eighty-two peptides were selected using machine learning classification and used for finding predominant pathogens from UTI patients.	DIA, PRM	Orbitrap Fusion, Q Exactive HF-X	[33]
Serum	pulmonary TB vs. extrapulmonary TB vs. latent TB vs. non-TB	31 vs. 10 vs. 9 vs. 9, 40	<i>Mycobacterium tuberculosis</i>	Twenty mycobacterial proteins were identified in the serum exosome of TB patients. The MRM assay can detect targets in the range of attomolar to femtomolar combined with isotope labeling.	MRM	Xevo TQ-S, LTQ-Orbitrap Velos	[67–69]
Nasopharyngeal and nasal swab	Respiratory tract infections patients	218	4 respiratory tract infection (RTI)-related bacterial species <sup>b</sup>	Top 16–18 peptide biomarker candidates were selected for each of the four pathogens and verified using clinical samples.	PRM	Q Exactive, Q Exactive HF	[70,71]
BALF	Pneumonia patients	1	5 RTI-related bacterial species <sup>c</sup>	Five unique peptides for each pathogen were selected according to abundance and applied for direct detection of pathogens.	MRM	Q-Exactive, Xevo TQ-S	[72]
endotracheal aspirate	VAP patients	37	6 RTI-related bacterial species <sup>d</sup>	Ninety-seven species-specific peptides of the six pathogens, selected based on the proteotypicity and high ionization yield, were monitored and verified in clinical samples. The targeted proteomics assay showed 76% sensitivity and 100% specificity.	MRM	TripleTOF®5600 MS	[73]
nasopharyngeal swab	COVID-19 patient	9	SARS-CoV-2	To develop an assay, nasopharyngeal swabs with different quantities of viral material were used. The two peptides of N protein were selected. They can be obtained within 3 min of elution.	DDA	Q Exactive HF	[74,75]
nasopharyngeal swab	COVID-19 patient	103	SARS-CoV-2	The two peptides of the S protein were selected and monitored. The targeted assay showed 90.5% sensitivity and 100% specificity in a 2-min gradient run.	MRM	TripleTOF 6600	[76]

Table 4. Cont.

Body Fluid	Study Groups	Sample Size	Target Pathogen	Major Findings	Method	Instrument	Ref.
nasopharyngeal swab	COVID-19 patient	985	SARS-CoV-2	Fully automated sample preparation (SP3) and sample-cleanup methods (turbulent flow) were applied. The two peptides of the N protein were validated in a qualitative (Tier 3) and quantitative (Tier 1) manner. The targeted assay showed 84% sensitivity and 97% specificity in a 2.5-min gradient run.	PRM	Q Exactive HF-X	[77]
nasopharyngeal swab	COVID-19 patient vs. Healthy	88 vs. 88	SARS-CoV-2	Automated immunoaffinity-based sampling was applied. The two peptides of the N protein were selected for the targeted assay. The targeted assay was qualified using the ensemble method and showed 98% sensitivity and 100% specificity in a 5-min gradient run.	PRM	Orbitrap Exploris 480	[78]

<sup>a</sup> *Citrobacter freundii*, *Enterobacter cloacae*, *Escherichia coli*, *Klebsiella aerogenes*, *Klebsiella oxytoca*, *Klebsiella pneumoniae*, *Pseudomonas aeruginosa*, *Proteus mirabilis*, *Enterococcus faecalis*, *Streptococcus agalactiae*, *Staphylococcus aureus*, *Staphylococcus epidermidis*, *Staphylococcus haemolyticus*, *Streptococcus mitis*, and *Staphylococcus saprophyticus*; <sup>b</sup> *S. aureus*; *Moraxella catarrhalis*; *Haemophilus influenzae* and *Streptococcus pneumoniae*; <sup>c</sup> *Acinetobacter baumannii*, *M. catarrhalis*, *P. aeruginosa*, *Stenotrophomonas maltophilia*, and *K. pneumoniae*; <sup>d</sup> *A. baumannii*, *E. coli*, *H. influenzae*, *P. aeruginosa*, *S. aureus*, and *S. pneumoniae*.

### 3.1. Application of DDA for Proteomics of Infectious Diseases

LC-MS-based proteomics has evolved into two analytical methods: (1) discovery proteomics and (2) targeted proteomics [79]. The DDA, so-called “shotgun proteomics”, is a suitable method for discovery studies because it allows the comprehensive identification of bacterial proteins. In traditional DDA, protein samples are tryptic digested, following which the peptide mixtures are fractionated and analyzed by LC-MS/MS. The most abundant precursor ions in a given spectrum are then selected and fragmented into MS/MS for further analysis (Figure 1B) [80]. Various protein identification programs have been developed [64,81]. The most common approach for protein identification is the sequence database matching algorithm, in which real spectra obtained from MS/MS analysis are comparatively analyzed with in silico spectra derived from peptide sequences from a reference database. Therefore, using the correct high-quality searching algorithms and reference databases is essential for determining the quality of search results when using DDA. The high accessibility and wide coverage of DDA have made it the most widely used method (Table 3). However, stochastic sampling is the main limitation of DDA; it complicates the identification of low-abundance proteins in complex samples and, in some cases, low-abundance proteins are frequently ignored [82]. Owing to this problem, data are plagued with numerous missing values, therefore requiring imputation and resulting in the loss of statistical power when the sample size is increased. The depletion of abundant proteins or the fractionation of protein mixtures has commonly been applied to overcome this technical limitation [83]. Label-based protein quantification methods, such as tandem mass tags and isobaric tags for relative and absolute quantitation, are also routinely applied to the comparative quantitative analysis of the infected host proteome [84,85]. Optimizing MS/MS measurement conditions in LC/MS is also considered to be an important factor in expanding the usefulness of DDA [86,87].

There have been several important studies regarding the discovery of proteins of pathogen origin from clinical samples of infectious diseases. Kashino and colleagues applied the DDA approach to urine proteomics in patients with pulmonary tuberculosis (TB) [29]. The urine samples were prepared by filtration through a 5 kDa molecular weight cut-off (MWCO) filter. They found four proteins (MT\_1721, MT\_1694, MT\_3444, and

MT\_2462) of *Mycobacterium tuberculosis* (*Mtb*) from nine patients with culture-confirmed pulmonary TB. Further validation of these proteins was performed by western blotting using anti-sera from patients with TB. Pollock and colleagues selected one candidate protein (MT\_1721 or Rv1681) for further study [30]. This protein was confirmed by LC-MS analysis, and the full length of the target protein was validated using immunoaffinity precipitation MS analysis [30]. However, because of the low sensitivity of the DDA approach used in this study, the detection rate of the target protein (MT\_1721) in the group of patients was not significantly high (less than 20%). The antibody of the target protein (MT\_1721) was subjected to an enzyme-linked immunosorbent assay (ELISA) using approximately 100 clinical samples. ELISAs for the target proteins showed a detection rate of <50%. However, the authors confirmed the complete absence of urine reactivity in the negative controls. Young and colleagues also performed urine proteomics to discover TB-specific biomarkers using clinical samples obtained from patients with TB ( $n = 63$ ) [31]. TB patients were categorized as having definite TB ( $n = 21$ ), presumed latent TB ( $n = 24$ ), or presumed non-TB ( $n = 18$ ). The clinical samples were pretreated by filtration (50 kDa MWCO filter) and concentration (3 kDa MWCO filter) to deplete highly abundant proteins before proteomic analysis. Using the DDA approach, the authors discovered 16 proteins originating from *Mtb*. Additionally, 27 human proteins were selectively identified in patients with active pulmonary TB.

However, although the previously described body fluid proteomics studies succeeded in identifying bacterial-derived markers, in many cases researchers failed to identify bacterial proteins because of intrinsic limitations, low quantity target proteins relative to the host proteins, and/or the absence of target proteins in existing databases, as mentioned above [65]. Spectral library searching is an alternative method for overcoming sensitivity-related limitations [81]. This is described in more detail in the next section. In brief, this technique is typically more sensitive and faster than the sequence database searching approach because it directly matches the spectra of peptide ions to spectra contained in libraries [88,89]. Hentschker and colleagues reported improved and faster results based on the proteome and phosphoproteome of pneumococci [90]. They applied a spectral library instead of a sequence database to identify more unidentified bacterial proteins. The spectral library was derived from MS/MS analysis of the culture cells; it was validated using synthetic peptides. They identified 76% of the theoretical proteome and 128 phosphorylated proteins in *Streptococcus pneumoniae*. This method is expected to be useful for body fluid proteomics.

### 3.2. Application of DIA for Proteomics of Infectious Diseases

Following its introduction in 2004, DIA has become a new strategy for systemically analyzing complex protein mixtures [91]. Unlike DDA, in DIA all ions present in a certain range of the  $m/z$  window are co-fragmented and collectively analyzed (Figure 1B). The DIA approach makes it possible to expand the profiles of proteomes and accurately quantify targeted proteins. This method can result in better experimental reproducibility than DDA methods [60,92–94]. DIA has the merits of both DDA and targeted approaches (selected reaction monitoring [SRM]/MRM and PRM). Therefore, it has become a popular technology in proteomics research [95]. However, it is still unable to overcome the depth of proteome coverage in DDA and the accuracy of MRM or PRM in measuring very low-abundance proteins (Table 3). High-resolution MS/MS acquisition at fast scan speeds is required for DIA-MS experiments. The most widely used hybrid instruments, QExactive and QE plus, are believed to have sufficient performance for DIA analysis. Although DIA is an extremely powerful method, it is more complex than DDA because of the difficulties of MS/MS spectral data analysis. Previously used peptide identification algorithms are not appropriate for DIA because of the complexity of the MS/MS spectrum of DIA [58,81]. In order to deconvolute complex spectra, spectral libraries are essential as reference databases. In general, spectral libraries contain intensity and peak information of non-canonical fragment ions generated by multiple DDA analyses of target samples [59]. Unfortunately,

standardized pipelines have not yet been established [58,64]. The contents necessary for the practical application of DIA have been described in more detail in recent review papers [94,96].

Roux-Dalvai and colleagues conducted urine proteomics using DIA analysis and machine learning to identify pathogens in the urinary tract. In the first step, spectral libraries containing 31,096 peptides from 15 pathogen colonies were obtained. Then, the authors prepared 12 artificial urine sample replicates spiked with 15 bacterial species to verify the spectral library. As a result, 4319 peptides were obtained as detected spectra. To select informative features among them, machine learning was used to identify peptide signatures; 82 peptides were selected. Further validation of the selected peptides was conducted using PRM. They successfully predicted the predominant bacteria in clinical samples ( $n = 27$ ) [33]. DIA has been applied for the proteome analysis of infectious diseases by targeting host proteins; it has also been applied for the rapid diagnosis of identified pathogens [97].

### 3.3. Application of Targeted-MS for Proteomics of Infectious Diseases

DDA has been routinely used to discover biomarkers from clinical samples, with further validation being achieved through rigorous statistical methods. This validation process requires accurate, reproducible, and highly robust methods for quantifying candidate biomarkers. However, the abovementioned major limitations of DDA, related to irreproducibility and imprecision, result from stochastic problems. Targeted proteomics, meanwhile, have been devised for the precise quantitative analysis of specific proteins or protein complexes. Representative targeted proteomics include SRM, MRM, and PRM [98,99]. SRM/MRM technology eliminates most non-targeted detection methods, which can reduce the noise signal and improve the detection sensitivity. In general, a triple quadrupole instrument is used for these technologies. Monitoring specific transition windows (a small range of  $m/z$  values of precursor/fragment ion pairs; Figure 1B) results in increased selectivity and sensitivity compared to those with DDA and DIA approaches. It is known that targeted methods are at least 5–10 times more sensitive than DDA when analyzing whole-cell lysates [92,100] (Table 3). However, the bottleneck in the development of SRM/MRM-based assays is the complicated procedure of the optimization process [101–104]. For example, it is important to choose the prototypic peptides, which are the unique peptides that empirically have a high chance of being observed.

PRM technology has been optimized based on quadrupole-orbitrap instruments to deliver an improved version of targeted proteomics. Unlike SRM/MRM, PRM involves the acquisition of full MS/MS scans of product ions in orbitrap, rather than selected fragment ions from predefined precursor ions. Therefore, this technology is more convenient because it does not require the selection and optimization of fragment ions. It can also be used for qualitative purposes, as in DDA approaches, to avoid false positives. In summary, this technique provides simplified and robust workflows but requires time-consuming optimization steps. Therefore, it is not suitable for discovery-based applications but is very useful for validation applications targeting low-abundance proteins present in body fluids [105]. Targeted-MS based diagnosis has inherent strength compared to immunoassays in that it can perform the analysis in a multiplexed manner with high selectivity and sensitivity, without an antibody, at a low cost if the lab has appropriate instruments and has developed the assay [7,106].

Several studies have successfully employed targeted proteomics to quantify biomarkers exposed in body fluids for infectious diseases. Kruh-Garcia and colleagues first developed an MRM assay for the antigen 85 complex (Ag85) mycobacterial proteins that are potential diagnostic biomarkers for TB. They compared the amount of the Ag85 complex (represented by Ag85A, Ag85B, and Ag85C proteins), in the secretome of various clades of *Mtb*, revealing precise discrimination among those highly homologous proteins [67]. In a further study, they expanded their proteomic results in the secretome to include TB patient serum [68]. They identified 250 targeted peptides using DDA proteomics of

*Mtb*-infected macrophages and a mouse model. After a thorough optimization process aided by in silico analysis, they selected 76 peptides as target peptides, representing 33 mycobacterial proteins (including Ag85). Then, they performed an MRM assay, using serum exosomes from TB patients as clinical samples. As a result, for the first time, they suggested 20 mycobacterial proteins present in the serum exosomes of TB patients as potential biomarkers ( $n = 41$ ). The same research team developed refined MRM assays using isotope-labeled peptide standards [69]; these assays can detect mycobacterial proteins in serum exosomes in the attomolar to femtomolar range.

Karlsson and colleagues successfully selected species-unique peptides of the Mitis group of the genus *Streptococcus*, using proteogenomic analysis. They characterized and identified more than 200 unique peptides from cell lysates of cultured cells using DDA proteomics [70]. They then expanded their platform to discover peptide biomarkers of representative respiratory tract pathogens, including *S. pneumoniae*, *Haemophilus influenzae*, *Moraxella catarrhalis*, and *Staphylococcus aureus*. For the discovery phase, representative genetic variations were preselected as MS-inclusion lists and validated in bacterial culture proteomics. Finally, the targeted peptides of each of the four pathogens were confirmed in 218 clinical samples [71].

Wang and colleagues used a similar approach to identify five gram-negative pathogens in the BALF, including *Acinetobacter baumannii*, *M. catarrhalis*, *Pseudomonas aeruginosa*, *Stenotrophomonas maltophilia*, and *Klebsiella pneumoniae* [72]. Bardet and colleagues, meanwhile, developed an SRM-based method to rapidly and reliably identify pathogens using endotracheal aspirate samples of ventilator-associated pneumonia (VAP) [73]. Based on the high ionization yields of the unique peptides confirmed in DDA experiments, 97 species-specific peptides from the six most frequent bacterial species (*A. baumannii*, *Escherichia coli*, *H. influenzae*, *Pseudomonas aeruginosa*, *S. aureus*, and *S. pneumoniae*) responsible for VAP were selected and monitored using the developed SRM assay.

### 3.4. Application of LC-MS/MS for COVID-19 Diagnosis

The current COVID-19 pandemic, which is caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), has justified the need for the development of diagnostic technology for infectious diseases. Molecular diagnostics such as polymerase chain reaction (PCR) have to date been used as the gold standard for the detection of SARS-CoV-2. However, novel alternative approaches have been introduced. Proteomics researchers have introduced novel LC-MS/MS-based diagnostic approaches for COVID-19 (Table 4).

Gouveia and colleagues first reported 101 tryptic peptides derived from six viral proteins identified from SARS-CoV-2-infected Vero E6 cells, using DDA analysis [74]. Through further curation, 14 peptides from nucleocapsid phosphoprotein (N protein), spike protein (S protein), and membrane glycoprotein (M protein) of the virus were recommended for further targeted MS. In a subsequent study, they proposed a time-efficient diagnostic method for COVID-19 clinical samples using LC-MS/MS as alternative methodologies to PCR or immunodiagnostic assays. They applied artificial nasopharyngeal swabs to evaluate 14 peptides. Among these 14 peptides, two peptides of the N protein were selected as attractive candidates [75]. Interestingly, the same target peptides were confirmed by independent groups using the PRM method [107,108]. However, neither approach could overcome the problems of the low detection rate (approximately 20% of the PCR assay) and low throughput analysis (20 min per sample). Thus, further investigations should aim to improve practical usage. Singh and colleagues also reported MRM assays using two other peptides derived from the S protein and replicase polyprotein, achieving significant results of 100% specificity and 90.5% sensitivity in a 2 min gradient run ( $n = 103$ ) [76]. However, MRM measurements are limited by their low resolution, which makes it impossible to verify the peptide spectrum itself.

Cazares and colleagues reported a PRM assay for the detection of viral proteins in virus-spiked mucus samples and found that the limit of detection (LOD) and limit of quantitation (LOQ) were approximately 200 and 390 attomoles, respectively [109]. These

values indicated that the assay could detect approximately  $2 \times 10^5$  viral particles/mL in a sample, showing comparable performance to the RT-PCR method.

Fully automated sample preparation and sample-cleanup methods with high-resolution MS seem to overcome these problems. Cardozo and colleagues developed a fully automated magnetic-based sample preparation method for nasopharyngeal and oropharyngeal swabs that could be completed within 4 h using a robotic liquid handler. Turbulent flow chromatography coupled with tandem mass spectrometry (TFC-MS/MS) can provide an efficient online sample cleanup method. This workflow can analyze four samples in a row within 10 min (in other words, more than 500 samples per day). The authors evaluated the target peptides of the SARS-CoV2 N protein qualitatively and quantitatively using PRM methods. The LOD and LOQ were reported to be 2–3 and 4–6 ng/mL, respectively. Compared to an RT-PCR-validated cohort, this workflow could detect up to 84% of the positive cases with a specificity of up to 97% ( $n = 985$ ) [77]. Renuse and colleagues introduced automated immunoaffinity-based sampling combined with targeted high field asymmetric waveform ion mobility spectrometry (FAIMS) [78]. Acquired PRM data were used to model an “ensemble” machine learning-based classification method. This method obtained high-quality results, delivering 98% (86/88) sensitivity and 100% (88/88) specificity [78].

Rajczewski and colleagues thoroughly evaluated 636 viral peptides identified in datasets using Galaxy-based workflows [110]. Galaxy is a web-based platform that provides reproducible computational research and numerous bioinformatics tools. Using *in vitro* and clinical source datasets deposited in the public repository proteomeXchange, they selected four peptides derived from N and M proteins. These peptides were consistently detected across all datasets used in the study and were proposed as potential diagnostic biomarkers.

Additional studies from nasopharyngeal swabs, gargle solutions, or other human samples have also been published [109,111–114]. However, the results are limited, except for those of nasopharyngeal swabs, compared with the results of a PCR-based study [115,116]. During the initial phase, Ihling and colleagues reported PRM-based identification of N protein from patient gargle solutions [114]. Recently, Kipping and colleagues proposed an improved sample preparation protocol and developed MRM methods using a synthetic peptide library to target the N protein from gargle solutions and saliva [117]. Based on these results, LC-MS-based diagnostics seem to be in the beginning stage, except for the use of nasopharyngeal swabs. The SARS-CoV2 peptides that have been introduced as potential biomarkers in recent studies have been summarized in two previous review papers [118,119].

#### 4. Concluding Remarks and Future Outlook

This review summarizes LC-MS-based proteomics for discovering biomarkers of infectious diseases, using various body fluids. Body fluid proteomics is an attractive method for monitoring patient status. The direct detection of pathogen-derived proteins or peptides from body fluids could also prove to be an optimal tool for identifying infectious diseases. However, studies into infectious diseases focusing on body fluid proteomics have not been actively performed due to technical difficulties. The first part of this review introduces the characteristics of the representative body fluids. Recent improvements in sample preparation methods have increased the coverage of proteome discovery. The second part describes the characteristics of representative MS acquisition methods such as DDA, DIA, MRM, and PRM. The application of body fluid proteomics to infectious diseases is also introduced here. Given the importance of the COVID-19 pandemic, recent results of COVID-19 studies using body fluids are also summarized here. Researchers have applied cutting-edge sample preparation methods and proteomic technologies to discover biomarkers and have reported improved results. For example, compared to the MRM assay developed by Kruh-Garcia in 2014 [68], which can process one sample per day, the automatic workflow developed by Cardozo in 2020 can handle 500 samples per day without laborious work [77]. In these studies, micro-flow LC with a short separation time

was applied. This is contrary to the mainstream method of using nano-flow LC with a long separation time to obtain very high sensitivity. This method was introduced to increase the sample processing throughput using micro-flow LC [120–123]. As a result, a moderate loss of sensitivity, which was the key reason for using nano-flow LC, can result in improved robustness, throughput, and reproducibility. Recently, the same research groups proved the robustness of micro-flow LC and the potential for high-throughput clinical applications based on more than 38,000 proteomic samples collected over the past 2 years [124].

Despite the remarkable results of LC-MS-based diagnosis, further innovations in instruments and informatics are required for practical applications. The following section, therefore, focuses on informatics, which should be improved for future proteomics applications.

#### 4.1. Data Deposition and Sharing Using Public Repositories

Data management has become important because of the rapid accumulation of MS data worldwide, produced by high-throughput MS (Figure 1C). However, enormous amounts of raw MS data can be useless if they are not quality controlled and well-organized using predefined terms, clinical metadata, and parameters used for analysis. Data repositories such as ProteomeXchange, Panorama, PeptideAtlas [125–127], and PRoteomics IDentifications database (PRIDE) [128–130] are playing important roles in proteomics. The main characteristics or functionalities of each repository have been summarized in a recent paper [131]. Recently, most leading journals have mandated the deposition of raw data and analysis results in these public repositories [130]. Deposited data should be processed in a standard format that can be reused for further analysis. Accumulated datasets should be easily accessible and should be capable of being reanalyzed; this could be achieved by using an improved pipeline or other processes in a high-performance cloud computing environment [110]. These databases should also support other independent research results or benchmarking for new algorithms [110,132,133]. The integration of proteomic datasets with other omics datasets would expand the scope of our understanding of infectious diseases.

#### 4.2. Expanding Community-Level Spectral Libraries

Similar to that with next-generation sequencing, it remains challenging to identify differentially abundant proteins, especially in the cases of proteins that occur at low abundance levels. Moreover, the larger the number of samples, the greater the number of missing values in the proteomics results, which will inhibit downstream analysis. To deal with this problem, missing values are currently replaced with reasonable values using various imputation methods, and statistical methods of transcriptomics are adopted. However, this can result in over-confident predictions [132,134–136]. Recently, intensity-dependent probabilistic modeling without imputation has been proposed to overcome this limitation [137].

However, the best answer for low-abundance proteins would be to increase proteome coverage to a reliable level. For this purpose, from an informatics perspective, spectral library approaches or hybrid searches could be applied to increase the proteomics depth; these approaches have been successfully used in DDA experiments [33,89,138–141]. Recent results have shown that search strategies can be critical for reproducibility, regardless of the acquisition method (DDA or DIA). This shows the importance of the spectral library approach [88,142]. However, the quality of spectral libraries from individual experiments is questionable, and the lack of publicly available libraries makes it difficult to apply spectral library search strategies to common proteomic data analysis [81,88,89,143–146]. For this reason, several synthetic peptide libraries, library-free approaches, or spectral prediction approaches have been developed as alternatives [147–152]. Although further improvements on deep learning-based methods have been achieved, these *in silico* approaches cannot completely replace experimentally derived libraries [150,153]. In addition, one of the major hindrances of this approach is that the series of library generation workflows is rather complicated, making it difficult for researchers to use. MaxDIA has recently enabled

library-based and library-free DIA proteomics in the MaxQuant environment, making these approaches more intuitive for researchers to use [62]. Therefore, improving and accumulating public spectral libraries will aid in the development of the next generation of proteomics.

**Author Contributions:** Conceptualization, S.I.K.; Manuscript Preparation, H.L. and S.I.K. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by grants from the Korea Basic Science Institute research program [grant number C180310], [grant number C180600], and Korea Health Technology R&D Project through the Korea Health Industry Development Institute, funded by the Ministry of Health & Welfare, Korea [Grant No. HI20C0033]. The funding sources had no involvement in the research project or the preparation and submission of this article for publication.

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

## References

- Chen, B.; Zhang, D.; Wang, X.; Ma, W.; Deng, S.; Zhang, P.; Zhu, H.; Xu, N.; Liang, S. Proteomics progresses in microbial physiology and clinical antimicrobial therapy. *Eur. J. Clin. Microbiol. Infect. Dis.* **2017**, *36*, 403–413. [[CrossRef](#)] [[PubMed](#)]
- Bloom, D.E.; Cadarette, D. Infectious disease threats in the twenty-first century: Strengthening the global response. *Front. Immunol.* **2019**, *10*, 1–12. [[CrossRef](#)]
- Afzal, A. Molecular diagnostic technologies for COVID-19: Limitations and challenges. *J. Adv. Res.* **2020**, *26*, 149–159. [[CrossRef](#)]
- Excler, J.L.; Saville, M.; Berkley, S.; Kim, J.H. Vaccine development for emerging infectious diseases. *Nat. Med.* **2021**, *27*, 591–600. [[CrossRef](#)]
- Badiee, P. Evaluation of human body fluids for the diagnosis of fungal infections. *BioMed Res. Int.* **2013**, *2013*, 8325. [[CrossRef](#)]
- Camerini, S.; Mauri, P. The role of protein and peptide separation before mass spectrometry analysis in clinical proteomics. *J. Chromatogr. A* **2015**, *1381*, 1–12. [[CrossRef](#)] [[PubMed](#)]
- Güzel, C.; Govorukhina, N.I.; Stingl, C.; Dekker, L.J.M.; Boichenko, A.; van der Zee, A.G.J.; Bischoff, R.P.H.; Luider, T.M. Comparison of Targeted Mass Spectrometry Techniques with an Immunoassay: A Case Study for HSP90 $\alpha$ . *Proteom. Clin. Appl.* **2018**, *12*, 1–12. [[CrossRef](#)] [[PubMed](#)]
- Han, D.; Li, R.; Shi, J.; Tan, P.; Zhang, R.; Li, J. Liquid biopsy for infectious diseases: A focus on microbial cell-free DNA sequencing. *Theranostics* **2020**, *10*, 5501–5513. [[CrossRef](#)]
- Ko, E.R.; Yang, W.E.; McClain, M.T.; Woods, C.W.; Ginsburg, G.S.; Tsalik, E.L. What was old is new again: Using the host response to diagnose infectious disease. *Expert Rev. Mol. Diagn.* **2015**, *15*, 1143–1158. [[CrossRef](#)]
- Qureshi, A.; Niazi, J.H. Biosensors for detecting viral and bacterial infections using host biomarkers: A review. *Analyst* **2020**, *145*, 7825–7848. [[CrossRef](#)]
- Wuen Ma, J.Y.; Sze, Y.H.O.N.; Bian, J.F.; Lam, T.C. Critical role of mass spectrometry proteomics in tear biomarker discovery for multifactorial ocular diseases (Review). *Int. J. Mol. Med.* **2021**, *47*, 4916. [[CrossRef](#)]
- Hu, S.; Loo, J.A.; Wong, D.T. Human body fluid proteome analysis. *Proteomics* **2006**, *6*, 6326–6353. [[CrossRef](#)] [[PubMed](#)]
- Sataloff, R.T.; Johns, M.M.; Kost, K.M. *Integration of Omics Approaches and Systems Biology for Clinical Applications*; Vlahou, A., Mischak, H., Zoidakis, J., Magni, F., Eds.; John Wiley & Sons: Hoboken, NJ, USA, 2018, ISBN 9781119183952.
- Lygirou, V.; Makridakis, M.; Vlahou, A. Biological sample collection for clinical proteomics: Existing SOPs. In *Methods in Molecular Biology*; Humana Press: New York, NY, USA, 2015; Volume 1243, pp. 3–27, ISBN 978-1-4939-1872-0.
- Kroksveen, A.C.; Opsahl, J.A.; Aye, T.T.; Ulvik, R.J.; Berven, F.S. Proteomics of human cerebrospinal fluid: Discovery and verification of biomarker candidates in neurodegenerative diseases using quantitative proteomics. *J. Proteom.* **2011**, *74*, 371–388. [[CrossRef](#)]
- Schulz, B.L.; Cooper-White, J.; Punyadeera, C.K. Saliva proteome research: Current status and future outlook. *Crit. Rev. Biotechnol.* **2013**, *33*, 246–259. [[CrossRef](#)] [[PubMed](#)]
- Sabounchi-Schütt, F.; Åström, J.; Hellman, U.; Eklund, A.; Grunewald, J. Changes in bronchoalveolar lavage fluid proteins in sarcoidosis: A proteomics approach. *Eur. Respir. J.* **2003**, *21*, 414–420. [[CrossRef](#)]
- Filip, S.; Zoidakis, J. Proteomics of Body Fluids. In *Integration of Omics Approaches and Systems Biology for Clinical Applications*; John Wiley & Sons: Hoboken, NJ, USA, 2018; pp. 93–112.
- Rai, A.J.; Gelfand, C.A.; Haywood, B.C.; Warunek, D.J.; Yi, J.; Schuchard, M.D.; Mehig, R.J.; Cockrill, S.L.; Scott, G.B.I.; Tammen, H.; et al. HUPO Plasma Proteome Project specimen collection and handling: Towards the standardization of parameters for plasma proteome samples. *Proteomics* **2005**, *5*, 3262–3277. [[CrossRef](#)]
- Kim, M.-R.; Kim, C.-W. Human blood plasma preparation for two-dimensional gel electrophoresis. *J. Chromatogr. B Anal. Technol. Biomed. Life Sci.* **2007**, *849*, 203–210. [[CrossRef](#)]
- Di Domenico, M.; Scumaci, D.; Grasso, S.; Gaspari, M.; Curcio, A.; Oliva, A.; Ausania, F.; Di Nunzio, C.; Ricciardi, C.; Santini, A.C.; et al. Biomarker discovery by plasma proteomics in familial Brugada Syndrome. *Front. Biosci. Landmark Ed.* **2013**, *18*, 564–571. [[CrossRef](#)]

22. Teunissen, C.E.; Petzold, A.; Bennett, J.L.; Berven, F.S.; Brundin, L.; Comabella, M.; Franciotta, D.; Frederiksen, J.L.; Fleming, J.O.; Furlan, R.; et al. A consensus protocol for the standardization of cerebrospinal fluid collection and biobanking. *Neurology* **2009**, *73*, 1914–1922. [[CrossRef](#)]
23. Ray, S.; Reddy, P.J.; Jain, R.; Gollapalli, K.; Moiyadi, A.; Srivastava, S. Proteomic technologies for the identification of disease biomarkers in serum: Advances and challenges ahead. *Proteomics* **2011**, *11*, 2139–2161. [[CrossRef](#)]
24. Zhu, P.; Bowden, P.; Zhang, D.; Marshall, J.G. Mass spectrometry of peptides and proteins from human blood. *Mass Spectrom. Rev.* **2011**, *30*, 685–732. [[CrossRef](#)] [[PubMed](#)]
25. Sim, S.Y.; Choi, Y.R.; Lee, J.H.; Lim, J.M.; Lee, S.; Kim, K.P.; Kim, J.Y.; Lee, S.H.; Kim, M. In-Depth Proteomic Analysis of Human Bronchoalveolar Lavage Fluid toward the Biomarker Discovery for Lung Cancers. *Proteom. Clin. Appl.* **2019**, *13*, 1900028. [[CrossRef](#)] [[PubMed](#)]
26. Lausted, C.; Lee, I.; Zhou, Y.; Qin, S.; Sung, J.; Price, N.D.; Hood, L.; Wang, K. Systems approach to neurodegenerative disease biomarker discovery. *Annu. Rev. Pharmacol. Toxicol.* **2014**, *54*, 457–481. [[CrossRef](#)] [[PubMed](#)]
27. Filip, S.; Zoidakis, J.; Vlahou, A.; Mischak, H. Advances in urinary proteome analysis and applications in systems biology. *Bioanalysis* **2014**, *6*, 2549–2569. [[CrossRef](#)] [[PubMed](#)]
28. Kim, M.J.; Frankel, A.H.; Tam, F.W.K. Urine proteomics and biomarkers in renal disease. *Nephron Exp. Nephrol.* **2011**, *119*, e1–e7. [[CrossRef](#)]
29. Kashino, S.S.; Pollock, N.; Napolitano, D.R.; Rodrigues, V.; Campos-Neto, A. Identification and characterization of Mycobacterium tuberculosis antigens in urine of patients with active pulmonary tuberculosis: An innovative and alternative approach of antigen discovery of useful microbial molecules. *Clin. Exp. Immunol.* **2008**, *153*, 56–62. [[CrossRef](#)]
30. Pollock, N.R.; Macovei, L.; Kanunfre, K.; Dhiman, R.; Restrepo, B.I.; Zarate, I.; Pino, P.A.; Mora-Guzman, F.; Fujiwara, R.T.; Michel, G.; et al. Validation of Mycobacterium tuberculosis Rv1681 protein as a diagnostic marker of active pulmonary tuberculosis. *J. Clin. Microbiol.* **2013**, *51*, 1367–1373. [[CrossRef](#)]
31. Young, B.L.; Mlamla, Z.; Gqamana, P.P.; Smit, S.; Roberts, T.; Peter, J.; Theron, G.; Govender, U.; Dheda, K.; Blackburn, J. The identification of tuberculosis biomarkers in human urine samples. *Eur. Respir. J.* **2014**, *43*, 1719–1729. [[CrossRef](#)]
32. Osbak, K.K.; Van Raemdonck, G.A.; Dom, M.; Cameron, C.E.; Meehan, C.J.; Deforce, D.; Van Ostade, X.; Kenyon, C.R.; Dhaenens, M. Candidate Treponema pallidum biomarkers uncovered in urine from individuals with syphilis using mass spectrometry. *Future Microbiol.* **2018**, *13*, 1497–1510. [[CrossRef](#)]
33. Roux-Dalvai, F.; Gotti, C.; Leclercq, M.; Hélie, M.C.; Boissinot, M.; Arrey, T.N.; Daully, C.; Fournier, F.; Kelly, I.; Marcoux, J.; et al. Fast and accurate bacterial species identification in urine specimens using LC-MS/MS mass spectrometry and machine learning. *Mol. Cell. Proteom.* **2019**, *18*, 2492–2505. [[CrossRef](#)]
34. Spielmann, N.; Wong, D. Saliva: Diagnostics and therapeutic perspectives. *Oral Dis.* **2011**, *17*, 345–354. [[CrossRef](#)]
35. Liu, J.; Duan, Y. Saliva: A potential media for disease diagnostics and monitoring. *Oral Oncol.* **2012**, *48*, 569–577. [[CrossRef](#)] [[PubMed](#)]
36. Shin, M.S.; Kim, Y.G.; Shin, Y.J.; Ko, B.J.; Kim, S.; Kim, H.D. Deep sequencing salivary proteins for periodontitis using proteomics. *Clin. Oral Investig.* **2019**, *23*, 3571–3580. [[CrossRef](#)]
37. Tu, C.; Rudnick, P.A.; Martinez, M.Y.; Cheek, K.L.; Stein, S.E.; Slebos, R.J.C.; Liebler, D.C. Depletion of Abundant Plasma Proteins and Limitations of Plasma Proteomics. *J. Proteom. Res.* **2010**, *9*, 4982–4991. [[CrossRef](#)] [[PubMed](#)]
38. Afkarian, M.; Bhasin, M.; Dillon, S.T.; Guerrero, M.C.; Nelson, R.G.; Knowler, W.C.; Thadhani, R.; Libermann, T.A. Optimizing a proteomic platform for urine biomarker discovery. *Mol. Cell. Proteom.* **2010**, *9*, 2195–2204. [[CrossRef](#)]
39. Kullolli, M.; Warren, J.; Arampatzidou, M.; Pitteri, S.J. Performance evaluation of affinity ligands for depletion of abundant plasma proteins. *J. Chromatogr. B Anal. Technol. Biomed. Life Sci.* **2013**, *939*, 10–16. [[CrossRef](#)] [[PubMed](#)]
40. Filip, S.; Vougas, K.; Zoidakis, J.; Latosinska, A.; Mullen, W.; Spasovski, G.; Mischak, H.; Vlahou, A.; Jankowski, J. Comparison of depletion strategies for the enrichment of low-Abundance proteins in urine. *PLoS ONE* **2015**, *10*, e0133773. [[CrossRef](#)]
41. Waas, M.; Bhattacharya, S.; Chuppa, S.; Wu, X.; Jensen, D.R.; Omasits, U.; Wollscheid, B.; Volkman, B.F.; Noon, K.R.; Gundry, R.L. Combine and conquer: Surfactants, solvents, and chaotropes for robust mass spectrometry based analyses of membrane proteins. *Anal. Chem.* **2014**, *86*, 1551–1559. [[CrossRef](#)]
42. Ding, H.; Fazelinia, H.; Spruce, L.A.; Weiss, D.A.; Zderic, S.A.; Seeholzer, S.H. Urine Proteomics: Evaluation of Different Sample Preparation Workflows for Quantitative, Reproducible, and Improved Depth of Analysis. *J. Proteom. Res.* **2020**, *19*, 1857–1862. [[CrossRef](#)]
43. Sielaff, M.; Kuharev, J.; Bohn, T.; Hahlbrock, J.; Bopp, T.; Tenzer, S.; Distler, U. Evaluation of FASP, SP3, and iST Protocols for Proteomic Sample Preparation in the Low Microgram Range. *J. Proteom. Res.* **2017**, *16*, 4060–4072. [[CrossRef](#)]
44. Müller, T.; Kalxdorf, M.; Longuespée, R.; Kazdal, D.N.; Stenzinger, A.; Krijgsveld, J. Automated sample preparation with SP 3 for low-input clinical proteomics. *Mol. Syst. Biol.* **2020**, *16*, 1–19. [[CrossRef](#)] [[PubMed](#)]
45. Ly, L.; Wasinger, V.C. Protein and peptide fractionation, enrichment and depletion: Tools for the complex proteome. *Proteomics* **2011**, *11*, 513–534. [[CrossRef](#)] [[PubMed](#)]
46. Nakayasu, E.S.; Gritsenko, M.; Piehowski, P.D.; Gao, Y.; Orton, D.J.; Schepmoes, A.A.; Fillmore, T.L.; Frohnert, B.I.; Rewers, M.; Krischer, J.P.; et al. Tutorial: Best practices and considerations for mass-spectrometry-based protein biomarker discovery and validation. *Nat. Protoc.* **2021**, *16*, 3737–3760. [[CrossRef](#)] [[PubMed](#)]

47. El Rassi, Z.; Puangpila, C. Liquid-phase based separation systems for depletion, prefractionation, and enrichment of proteins in biological fluids and matrices for in-depth proteomics analysis-An update covering the period 2014–2016. *Electrophoresis* **2017**, *38*, 150–161. [[CrossRef](#)] [[PubMed](#)]
48. Macklin, A.; Khan, S.; Kislinger, T. Recent advances in mass spectrometry based clinical proteomics: Applications to cancer research. *Clin. Proteom.* **2020**, *17*, 1–25. [[CrossRef](#)]
49. Hughes, C.S.; Moggridge, S.; Müller, T.; Sorensen, P.H.; Morin, G.B.; Krijgsveld, J. Single-pot, solid-phase-enhanced sample preparation for proteomics experiments. *Nat. Protoc.* **2019**, *14*, 68–85. [[CrossRef](#)] [[PubMed](#)]
50. Andersen, R.F.; Palmfeldt, J.; Jespersen, B.; Gregersen, N.; Rittig, S. Plasma and urine proteomic profiles in childhood idiopathic nephrotic syndrome. *Proteom. Clin. Appl.* **2012**, *6*, 382–393. [[CrossRef](#)] [[PubMed](#)]
51. Kulak, N.A.; Pichler, G.; Paron, I.; Nagaraj, N.; Mann, M. Minimal, encapsulated proteomic-sample processing applied to copy-number estimation in eukaryotic cells. *Nat. Methods* **2014**, *11*, 319–324. [[CrossRef](#)]
52. Wiśniewski, J.R.; Zougman, A.; Nagaraj, N.; Mann, M. Universal sample preparation method for proteome analysis. *Nat. Methods* **2009**, *6*, 359–362. [[CrossRef](#)]
53. HaileMariam, M.; Eguez, R.V.; Singh, H.; Bekele, S.; Ameni, G.; Pieper, R.; Yu, Y. S-Trap, an Ultrafast Sample-Preparation Approach for Shotgun Proteomics. *J. Proteom. Res.* **2018**, *17*, 2917–2924. [[CrossRef](#)]
54. Bekker-Jensen, D.B.; Kelstrup, C.D.; Batth, T.S.; Larsen, S.C.; Haldrup, C.; Bramsen, J.B.; Sørensen, K.D.; Høyer, S.; Ørntoft, T.F.; Andersen, C.L.; et al. An Optimized Shotgun Strategy for the Rapid Generation of Comprehensive Human Proteomes. *Cell Syst.* **2017**, *4*, 587–599.e4. [[CrossRef](#)] [[PubMed](#)]
55. Vizcaíno, J.A.; Deutsch, E.W.; Wang, R.; Csordas, A.; Reisinger, F.; Ríos, D.; Dianes, J.A.; Sun, Z.; Farrah, T.; Bandeira, N.; et al. ProteomeXchange provides globally coordinated proteomics data submission and dissemination. *Nat. Biotechnol.* **2014**, *32*, 223–226. [[CrossRef](#)] [[PubMed](#)]
56. Deutsch, E.W.; Csordas, A.; Sun, Z.; Jarnuczak, A.; Perez-Riverol, Y.; Ternent, T.; Campbell, D.S.; Bernal-Llinares, M.; Okuda, S.; Kawano, S.; et al. The ProteomeXchange consortium in 2017: Supporting the cultural change in proteomics public data deposition. *Nucleic Acids Res.* **2017**, *45*, D1100–D1106. [[CrossRef](#)] [[PubMed](#)]
57. Liebler, D.C.; Zimmerman, L.J. Targeted quantitation of proteins by mass spectrometry. *Biochemistry* **2013**, *52*, 3797–3806. [[CrossRef](#)]
58. Wolf-Yadlin, A.; Hu, A.; Noble, W.S. Technical advances in proteomics: New developments in data-independent acquisition. *F1000Research* **2016**, *5*, 1–12. [[CrossRef](#)]
59. Krasny, L.; Huang, P.H. Data-independent acquisition mass spectrometry (DIA-MS) for proteomic applications in oncology. *Mol. Omi* **2021**, *17*, 29–42. [[CrossRef](#)]
60. Bruderer, R.; Bernhardt, O.M.; Gandhi, T.; Miladinović, S.M.; Cheng, L.-Y.; Messner, S.; Ehrenberger, T.; Zanotelli, V.; Butscheid, Y.; Escher, C.; et al. Extending the Limits of Quantitative Proteome Profiling with Data-Independent Acquisition and Application to Acetaminophen-Treated Three-Dimensional Liver Microtissues. *Mol. Cell. Proteom.* **2015**, *14*, 1400–1410. [[CrossRef](#)] [[PubMed](#)]
61. McIlwain, S.; Tamura, K.; Kertesz-Farkas, A.; Grant, C.E.; Diamant, B.; Frewen, B.; Howbert, J.J.; Hoopmann, M.R.; Käll, L.; Eng, J.K.; et al. Crux: Rapid open source protein tandem mass spectrometry analysis. *J. Proteom. Res.* **2014**, *13*, 4488–4491. [[CrossRef](#)]
62. Sinitcyn, P.; Hamzeiy, H.; Salinas Soto, F.; Itzhak, D.; McCarthy, F.; Wichmann, C.; Steger, M.; Ohmayer, U.; Distler, U.; Kasper-Schoenefeld, S.; et al. MaxDIA enables library-based and library-free data-independent acquisition proteomics. *Nat. Biotechnol.* **2021**, *21*, 968. [[CrossRef](#)]
63. 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](#)]
64. Chen, C.; Hou, J.; Tanner, J.J.; Cheng, J. Bioinformatics methods for mass spectrometry-based proteomics data analysis. *Int. J. Mol. Sci.* **2020**, *21*, 2873. [[CrossRef](#)] [[PubMed](#)]
65. Bonnet, J.; Garcia, C.; Leger, T.; Couquet, M.P.; Vignoles, P.; Vatunga, G.; Ndung'u, J.; Boudot, C.; Bisser, S.; Courtioux, B. Proteome characterization in various biological fluids of Trypanosoma brucei gambiense-infected subjects. *J. Proteom.* **2019**, *196*, 150–161. [[CrossRef](#)]
66. Venkatesh, A.; Lahiri, A.; Reddy, P.J.; Shastri, J.; Bankar, S.; Patankar, S.; Srivastava, S. Identification of Highly Expressed Plasmodium Vivax Proteins from Clinical Isolates Using Proteomics. *Proteom. Clin. Appl.* **2018**, *12*, 1–7. [[CrossRef](#)]
67. Kruh-Garcia, N.A.; Murray, M.; Prucha, J.G.; Dobos, K.M. Antigen 85 variation across lineages of Mycobacterium tuberculosis-Implications for vaccine and biomarker success. *J. Proteom.* **2014**, *97*, 141–150. [[CrossRef](#)] [[PubMed](#)]
68. Kruh-Garcia, N.A.; Wolfe, L.M.; Chaisson, L.H.; Worodria, W.O.; Nahid, P.; Schorey, J.S.; Davis, J.L.; Dobos, K.M. Detection of Mycobacterium tuberculosis peptides in the exosomes of patients with active and latent M. tuberculosis infection using MRM-MS. *PLoS ONE* **2014**, *9*, e0103811. [[CrossRef](#)]
69. Mehaffy, C.; Dobos, K.M.; Nahid, P.; Kruh-Garcia, N.A. Second generation multiple reaction monitoring assays for enhanced detection of ultra-low abundance Mycobacterium tuberculosis peptides in human serum. *Clin. Proteom.* **2017**, *14*, 1–10. [[CrossRef](#)]
70. Karlsson, R.; Gonzales-Siles, L.; Gomila, M.; Busquets, A.; Salvà-Serra, F.; Jaén-Luchoro, D.; Jakobsson, H.E.; Karlsson, A.; Boulund, F.; Kristiansson, E.; et al. Proteotyping bacteria: Characterization, differentiation and identification of pneumococcus and other species within the Mitis Group of the genus Streptococcus by tandem mass spectrometry proteomics. *PLoS ONE* **2018**, *13*, e0208804. [[CrossRef](#)] [[PubMed](#)]

71. Karlsson, R.; Thorsell, A.; Gomila, M.; Salvà-Serra, F.; Jakobsson, H.E.; Gonzales-Siles, L.; Jaén-Luchoro, D.; Skovbjerg, S.; Fuchs, J.; Karlsson, A.; et al. Discovery of species-unique peptide biomarkers of bacterial pathogens by tandem mass spectrometry-based proteotyping. *Mol. Cell. Proteom.* **2020**, *19*, 518–528. [[CrossRef](#)] [[PubMed](#)]
72. Wang, H.; Drake, S.K.; Yong, C.; Gucek, M.; Lyes, M.A.; Rosenberg, A.Z.Z.; Soderblom, E.; Moseley, M.A.; Dekker, J.P.; Suffredini, A.F. A genoproteomic approach to detect peptide markers of bacterial respiratory pathogens. *Clin. Chem.* **2017**, *63*, 1398–1408. [[CrossRef](#)]
73. Bardet, C.; Barraud, O.; Clavel, M.; Fortin, T.; Charrier, J.P.; Rodrigue, M.; François, B.; Yugueros-Marcos, J.; Lemoine, J.; Ploy, M.C.; et al. Early and specific targeted mass spectrometry-based identification of bacteria in endotracheal aspirates of patients suspected with ventilator-associated pneumonia. *Eur. J. Clin. Microbiol. Infect. Dis.* **2021**, *40*, 1291–1301. [[CrossRef](#)]
74. Gouveia, D.; Grenga, L.; Gaillard, J.; Gallais, F.; Bellanger, L.; Pible, O.; Armengaud, J. Shortlisting SARS-CoV-2 Peptides for Targeted Studies from Experimental Data-Dependent Acquisition Tandem Mass Spectrometry Data. *Proteomics* **2020**, *20*, 2000107. [[CrossRef](#)] [[PubMed](#)]
75. Gouveia, D.; Miotello, G.; Gallais, F.; Gaillard, J.C.; Debroas, S.; Bellanger, L.; Lavigne, J.P.; Sotto, A.; Grenga, L.; Pible, O.; et al. Proteotyping SARS-CoV-2 Virus from Nasopharyngeal Swabs: A Proof-of-Concept Focused on a 3 Min Mass Spectrometry Window. *J. Proteom. Res.* **2020**, *19*, 4407–4416. [[CrossRef](#)]
76. Singh, P.; Chakraborty, R.; Marwal, R.; Radhakrishnan, V.S.; Bhaskar, A.K.; Vashisht, H.; Dhar, M.S.; Pradhan, S.; Ranjan, G.; Imran, M.; et al. A rapid and sensitive method to detect SARS-CoV-2 virus using targeted-mass spectrometry. *J. Proteins Proteom.* **2020**, *11*, 159–165. [[CrossRef](#)] [[PubMed](#)]
77. Cardozo, K.H.M.; Lebkuchen, A.; Okai, G.G.; Schuch, R.A.; Viana, L.G.; Olive, A.N.; Lazari, C.D.S.; Fraga, A.M.; Granato, C.F.H.; Pintão, M.C.T.; et al. Establishing a mass spectrometry-based system for rapid detection of SARS-CoV-2 in large clinical sample cohorts. *Nat. Commun.* **2020**, *11*, 6201. [[CrossRef](#)]
78. Renuse, S.; Vanderboom, P.M.; Maus, A.D.; Kemp, J.V.; Gurtner, K.M.; Madugundu, A.K.; Chavan, S.; Peterson, J.A.; Madden, B.J.; Mangalparthi, K.K.; et al. A mass spectrometry-based targeted assay for detection of SARS-CoV-2 antigen from clinical specimens. *EBioMedicine* **2021**, *69*, 103465. [[CrossRef](#)] [[PubMed](#)]
79. Yates, J.R. 3rd Mass spectrometry and the age of the proteome. *J. Mass Spectrom.* **1998**, *33*, 1–19. [[CrossRef](#)]
80. Stahl, D.C.; Swiderek, K.M.; Davis, M.T.; Lee, T.D. Data-controlled automation of liquid chromatography/tandem mass spectrometry analysis of peptide mixtures. *J. Am. Soc. Mass Spectrom.* **1996**, *7*, 532–540. [[CrossRef](#)]
81. Tariq, M.U.; Haseeb, M.; Aledhari, M.; Razzak, R.; Parizi, R.M.; Saeed, F. Methods for Proteogenomics Data Analysis, Challenges, and Scalability Bottlenecks: A Survey. *IEEE Access* **2021**, *9*, 5497–5516. [[CrossRef](#)]
82. Michalski, A.; Cox, J.; Mann, M. More than 100,000 detectable peptide species elute in single shotgun proteomics runs but the majority is inaccessible to data-dependent LC-MS/MS. *J. Proteom. Res.* **2011**, *10*, 1785–1793. [[CrossRef](#)]
83. Distler, U.; Kuharev, J.; Navarro, P.; Levin, Y.; Schild, H.; Tenzer, S. Drift time-specific collision energies enable deep-coverage data-independent acquisition proteomics. *Nat. Methods* **2014**, *11*, 167–170. [[CrossRef](#)]
84. Thompson, A.; Schäfer, J.; Kuhn, K.; Kienle, S.; Schwarz, J.; Schmidt, G.; Neumann, T.; Hamon, C. Tandem Mass Tags: A Novel Quantification Strategy for Comparative Analysis of Complex Protein Mixtures by MS/MS. *Anal. Chem.* **2003**, *75*, 1895–1904. [[CrossRef](#)] [[PubMed](#)]
85. 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](#)]
86. Hsieh, E.J.; Bereman, M.S.; Durand, S.; Valaskovic, G.A.; MacCoss, M.J. Effects of column and gradient lengths on peak capacity and peptide identification in nanoflow LC-MS/MS of complex proteomic samples. *J. Am. Soc. Mass Spectrom.* **2013**, *24*, 148–153. [[CrossRef](#)] [[PubMed](#)]
87. Moruz, L.; Pichler, P.; Stranzl, T.; Mechtler, K.; Käll, L. Optimized nonlinear gradients for reversed-phase liquid chromatography in shotgun proteomics. *Anal. Chem.* **2013**, *85*, 7777–7785. [[CrossRef](#)]
88. Deutsch, E.W.; Perez-Riverol, Y.; Chalkley, R.J.; Wilhelm, M.; Tate, S.; Sachsenberg, T.; Walzer, M.; Käll, L.; Delanghe, B.; Böcker, S.; et al. Expanding the Use of Spectral Libraries in Proteomics. *J. Proteom. Res.* **2018**, *17*, 4051–4060. [[CrossRef](#)]
89. Griss, J. Spectral library searching in proteomics. *Proteomics* **2016**, *16*, 729–740. [[CrossRef](#)]
90. Hentschker, C.; Maaß, S.; Junker, S.; Hecker, M.; Hammerschmidt, S.; Otto, A.; Becher, D. Comprehensive Spectral Library from the Pathogenic Bacterium *Streptococcus pneumoniae* with Focus on Phosphoproteins. *J. Proteom. Res.* **2020**, *19*, 1435–1446. [[CrossRef](#)]
91. Venable, J.D.; Dong, M.-Q.; Wohlschlegel, J.; Dillin, A.; Yates, J.R. Automated approach for quantitative analysis of complex peptide mixtures from tandem mass spectra. *Nat. Methods* **2004**, *1*, 39–45. [[CrossRef](#)] [[PubMed](#)]
92. Gillet, L.C.; Navarro, P.; Tate, S.; Röst, H.; Selevsek, N.; Reiter, L.; Bonner, R.; Aebersold, R. Targeted Data Extraction of the MS/MS Spectra Generated by Data-independent Acquisition: A New Concept for Consistent and Accurate Proteome Analysis. *Mol. Cell. Proteom.* **2012**, *11*, O111.016717. [[CrossRef](#)] [[PubMed](#)]
93. Barkovits, K.; Pacharra, S.; Pfeiffer, K.; Steinbach, S.; Eisenacher, M.; Marcus, K.; Uszkoreit, J. Reproducibility, Specificity and Accuracy of Relative Quantification Using Spectral Library-based Data-independent Acquisition. *Mol. Cell. Proteom.* **2020**, *19*, 181–197. [[CrossRef](#)] [[PubMed](#)]
94. Ludwig, C.; Gillet, L.; Rosenberger, G.; Amon, S.; Collins, B.C.; Aebersold, R. Data-independent acquisition-based SWATH-MS for quantitative proteomics: A tutorial. *Mol. Syst. Biol.* **2018**, *14*, 1–23. [[CrossRef](#)]

95. Doerr, A. DIA mass spectrometry. *Nat. Methods* **2015**, *12*, 35. [[CrossRef](#)]
96. Reubsæet, L.; Sweredoski, M.J.; Moradian, A. Data-Independent Acquisition for the Orbitrap Q Exactive HF: A Tutorial. *J. Proteom. Res.* **2019**, *18*, 803–813. [[CrossRef](#)] [[PubMed](#)]
97. Messner, C.B.; Demichev, V.; Wendisch, D.; Michalick, L.; White, M.; Freiwald, A.; Textoris-Taube, K.; Vernardis, S.I.; Egger, A.S.; Kreidl, M.; et al. Ultra-High-Throughput Clinical Proteomics Reveals Classifiers of COVID-19 Infection. *Cell Syst.* **2020**, *11*, 11–24.e4. [[CrossRef](#)] [[PubMed](#)]
98. Picotti, P.; Bodenmiller, B.; Mueller, L.N.; Domon, B.; Aebersold, R. Full Dynamic Range Proteome Analysis of *S. cerevisiae* by Targeted Proteomics. *Cell* **2009**, *138*, 795–806. [[CrossRef](#)]
99. Peterson, A.C.; Russell, J.D.; Bailey, D.J.; Westphall, M.S.; Coon, J.J. Parallel reaction monitoring for high resolution and high mass accuracy quantitative, targeted proteomics. *Mol. Cell. Proteom.* **2012**, *11*, 1475–1488. [[CrossRef](#)]
100. Bauer, M.; Ahrné, E.; Baron, A.P.; Glatter, T.; Fava, L.L.; Santamaria, A.; Nigg, E.A.; Schmidt, A. Evaluation of data-dependent and -independent mass spectrometric workflows for sensitive quantification of proteins and phosphorylation sites. *J. Proteom. Res.* **2014**, *13*, 5973–5988. [[CrossRef](#)]
101. Schnatbaum, K.; Zerweck, J.; Nehmer, J.; Wenschuh, H.; Schutkowski, M.; Reimer, U. SpikeTides<sup>TM</sup>—Proteotypic peptides for large-scale MS-based proteomics. *Nat. Methods* **2011**, *8*, 1–23. [[CrossRef](#)]
102. Lam, H.; Deutsch, E.W.; Eddes, J.S.; Eng, J.K.; King, N.; Stein, S.E.; Aebersold, R. Development and validation of a spectral library searching method for peptide identification from MS/MS. *Proteomics* **2007**, *7*, 655–667. [[CrossRef](#)]
103. Kim, Y.; Jeon, J.; Mejia, S.; Yao, C.Q.; Ignatchenko, V.; Nyalwidhe, J.O.; Gramolini, A.O.; Lance, R.S.; Troyer, D.A.; Drake, R.R.; et al. Targeted proteomics identifies liquid-biopsy signatures for extracapsular prostate cancer. *Nat. Commun.* **2016**, *7*, 1–10. [[CrossRef](#)]
104. Hoofnagle, A.N.; Whiteaker, J.R.; Carr, S.A.; Kuhn, E.; Liu, T.; Massoni, S.A.; Thomas, S.N.; Reid Townsend, R.; Zimmerman, L.J.; Boja, E.; et al. Recommendations for the generation, quantification, storage, and handling of peptides used for mass spectrometry-based assays. *Clin. Chem.* **2016**, *62*, 48–69. [[CrossRef](#)] [[PubMed](#)]
105. Kockmann, T.; Trachsel, C.; Panse, C.; Wahlander, A.; Selevsek, N.; Grossmann, J.; Wolski, W.E.; Schlapbach, R. Targeted proteomics coming of age—SRM, PRM and DIA performance evaluated from a core facility perspective. *Proteomics* **2016**, *16*, 2183–2192. [[CrossRef](#)] [[PubMed](#)]
106. Cross, T.G.; Hornshaw, M.P. Can LC and LC-MS ever replace immunoassays? *J. Appl. Bioanal.* **2016**, *2*, 108–116. [[CrossRef](#)]
107. Zecha, J.; Lee, C.Y.; Bayer, F.P.; Meng, C.; Grass, V.; Zerweck, J.; Schnatbaum, K.; Michler, T.; Pichlmair, A.; Ludwig, C.; et al. Data, Reagents, Assays and Merits of Proteomics for SARS-CoV-2 Research and Testing. *Mol. Cell. Proteom.* **2020**, *19*, 1503–1522. [[CrossRef](#)] [[PubMed](#)]
108. Bezstarosti, K.; Lamers, M.M.; Doff, W.A.S.; Wever, P.C.; Thai, K.T.D.; van Kampen, J.J.A.; Haagmans, B.L.; Demmers, J.A.A. Targeted proteomics as a tool to detect SARS-CoV-2 proteins in clinical specimens. *PLoS ONE* **2021**, *16*, e0259165. [[CrossRef](#)]
109. Cazares, L.H.; Chaerkady, R.; Samuel Weng, S.H.; Boo, C.C.; Cimbri, R.; Hsu, H.E.; Rajan, S.; Dall’acqua, W.; Clarke, L.; Ren, K.; et al. Development of a Parallel Reaction Monitoring Mass Spectrometry Assay for the Detection of SARS-CoV-2 Spike Glycoprotein and Nucleoprotein. *Anal. Chem.* **2020**, *92*, 13813–13821. [[CrossRef](#)] [[PubMed](#)]
110. Rajczewski, A.T.; Mehta, S.; Nguyen, D.D.A.; Grüning, B.; Johnson, J.E.; McGowan, T.; Griffin, T.J.; Jagtap, P.D. A rigorous evaluation of optimal peptide targets for MS-based clinical diagnostics of Coronavirus Disease 2019 (COVID-19). *Clin. Proteom.* **2021**, *18*, 1–16. [[CrossRef](#)] [[PubMed](#)]
111. Nikolaev, E.N.; Indeykina, M.I.; Brzhozovskiy, A.G.; Bugrova, A.E.; Kononikhin, A.S.; Starodubtseva, N.L.; Petrotchenko, E.V.; Kovalev, G.I.; Borchers, C.H.; Sukhikh, G.T. Mass-Spectrometric Detection of SARS-CoV-2 Virus in Scrapings of the Epithelium of the Nasopharynx of Infected Patients via Nucleocapsid N Protein. *J. Proteom. Res.* **2020**, *19*, 4393–4397. [[CrossRef](#)] [[PubMed](#)]
112. Rivera, B.; Leyva, A.; Portela, M.M.; Moratorio, G.; Moreno, P.; Durán, R.; Lima, A. Quantitative proteomic dataset from oro- and naso-pharyngeal swabs used for COVID-19 diagnosis: Detection of viral proteins and host’s biological processes altered by the infection. *Data BR* **2020**, *32*, 106121. [[CrossRef](#)]
113. Nachtigall, F.M.; Pereira, A.; Trofymchuk, O.S.; Santos, L.S. Detection of SARS-CoV-2 in nasal swabs using MALDI-MS. *Nat. Biotechnol.* **2020**, *38*, 1168–1173. [[CrossRef](#)]
114. Ihling, C.; Tänzler, D.; Hagemann, S.; Kehlen, A.; Hüttelmaier, S.; Arlt, C.; Sinz, A. Mass Spectrometric Identification of SARS-CoV-2 Proteins from Gargle Solution Samples of COVID-19 Patients. *J. Proteom. Res.* **2020**, *19*, 4389–4392. [[CrossRef](#)] [[PubMed](#)]
115. Kapoor, P.; Chowdhry, A.; Kharbanda, O.P.; Bablani Popli, D.; Gautam, K.; Saini, V. Exploring salivary diagnostics in COVID-19: A scoping review and research suggestions. *BDJ Open* **2021**, *7*, 8. [[CrossRef](#)]
116. Fernandes, L.L.; Pacheco, V.B.; Borges, L.; Athwal, H.K.; de Paula Eduardo, F.; Bezinelli, L.; Correa, L.; Jimenez, M.; Dame-Teixeira, N.; Lombaert, I.M.A.; et al. Saliva in the Diagnosis of COVID-19: A Review and New Research Directions. *J. Dent. Res.* **2020**, *99*, 1435–1443. [[CrossRef](#)] [[PubMed](#)]
117. Kipping, M.; Tänzler, D.; Sinz, A. A rapid and reliable liquid chromatography/mass spectrometry method for SARS-CoV-2 analysis from gargle solutions and saliva. *Anal. Bioanal. Chem.* **2021**, *413*, 6503–6511. [[CrossRef](#)] [[PubMed](#)]
118. Praissman, J.L.; Wells, L. Proteomics-Based Insights Into the SARS-CoV-2-Mediated COVID-19 Pandemic: A Review of the First Year of Research. *Mol. Cell. Proteom.* **2021**, *20*, 100103. [[CrossRef](#)]
119. Rais, Y.; Fu, Z.; Drabovich, A.P. Mass spectrometry-based proteomics in basic and translational research of SARS-CoV-2 coronavirus and its emerging mutants. *Clin. Proteom.* **2021**, *18*, 1–16. [[CrossRef](#)]

120. Messner, C.B.; Demichev, V.; Bloomfield, N.; Yu, J.S.L.; White, M.; Kreidl, M.; Egger, A.S.; Freiwald, A.; Ivosev, G.; Wasim, F.; et al. Ultra-fast proteomics with Scanning SWATH. *Nat. Biotechnol.* **2021**, *39*, 846–854. [[CrossRef](#)]
121. Bian, Y.; Zheng, R.; Bayer, F.P.; Wong, C.; Chang, Y.-C.; Meng, C.; Zolg, D.P.; Reinecke, M.; Zecha, J.; Wiechmann, S.; et al. Robust, reproducible and quantitative analysis of thousands of proteomes by micro-flow LC–MS/MS. *Nat. Commun.* **2020**, *11*, 157. [[CrossRef](#)]
122. Chambers, A.G.; Percy, A.J.; Yang, J.; Borchers, C.H. Multiple reaction monitoring enables precise quantification of 97 proteins in dried blood spots. *Mol. Cell. Proteom.* **2015**, *14*, 3094–3104. [[CrossRef](#)]
123. Lenčo, J.; Vajrychová, M.; Pimková, K.; Prokšová, M.; Benková, M.; Klimentová, J.; Tambor, V.; Soukup, O. Conventional-Flow Liquid Chromatography-Mass Spectrometry for Exploratory Bottom-Up Proteomic Analyses. *Anal. Chem.* **2018**, *90*, 5381–5389. [[CrossRef](#)]
124. Kuster, B.; Bian, Y.; Bayer, F.P.; Chang, Y.C.; Meng, C.; Hoefler, S.; Deng, N.; Zheng, R.; Boychenko, O. Robust microflow LC-MS/MS for proteome analysis: 38 000 runs and counting. *Anal. Chem.* **2021**, *93*, 3686–3690. [[CrossRef](#)]
125. Desiere, F.; Deutsch, E.W.; Nesvizhskii, A.I.; Mallick, P.; King, N.L.; Eng, J.K.; Aderem, A.; Boyle, R.; Brunner, E.; Donohoe, S.; et al. Integration with the human genome of peptide sequences obtained by high-throughput mass spectrometry. *Genome Biol.* **2005**, *6*, 1–12. [[CrossRef](#)]
126. Desiere, F.; Deutsch, E.W.; King, N.L.; Nesvizhskii, A.I.; Mallick, P.; Eng, J.; Chen, S.; Eddes, J.; Loevenich, S.N.; Aebersold, R. The PeptideAtlas project. *Nucleic Acids Res.* **2006**, *34*, 655–658. [[CrossRef](#)] [[PubMed](#)]
127. Deutsch, E.W.; Sun, Z.; Campbell, D.; Kusebauch, U.; Chu, C.S.; Mendoza, L.; Shteynberg, D.; Omenn, G.S.; Moritz, R.L. State of the human proteome in 2014/2015 As viewed through peptideatlas: Enhancing accuracy and coverage through the atlas prophet. *J. Proteom. Res.* **2015**, *14*, 3461–3473. [[CrossRef](#)]
128. Martens, L.; Hermjakob, H.; Jones, P.; Adamsk, M.; Taylor, C.; States, D.; Gevaert, K.; Vandekerckhove, J.; Apweiler, R. PRIDE: The proteomics identifications database. *Proteomics* **2005**, *5*, 3537–3545. [[CrossRef](#)]
129. Vizcaino, J.A.; Csordas, A.; Del-Toro, N.; Dianes, J.A.; Griss, J.; Lavidas, I.; Mayer, G.; Perez-Riverol, Y.; Reisinger, F.; Ternent, T.; et al. 2016 update of the PRIDE database and its related tools. *Nucleic Acids Res.* **2016**, *44*, D447–D456. [[CrossRef](#)]
130. Perez-Riverol, Y.; Csordas, A.; Bai, J.; Bernal-Llinares, M.; Hewapathirana, S.; Kundu, D.J.; Inuganti, A.; Griss, J.; Mayer, G.; Eisenacher, M.; et al. The PRIDE database and related tools and resources in 2019: Improving support for quantification data. *Nucleic Acids Res.* **2019**, *47*, D442–D450. [[CrossRef](#)]
131. Deutsch, E.W.; Bandeira, N.; Sharma, V.; Perez-Riverol, Y.; Carver, J.J.; Kundu, D.J.; García-Seisdedos, D.; Jarnuczak, A.F.; Hewapathirana, S.; Pullman, B.S.; et al. The ProteomeXchange consortium in 2020: Enabling “big data” approaches in proteomics. *Nucleic Acids Res.* **2020**, *48*, D1145–D1152. [[CrossRef](#)]
132. Smyth, G.K. limma: Linear models for microarray data. In *Bioinformatics and Computational Biology Solutions Using R and Bioconductor*; Springer: New York, NY, USA, 2005; pp. 397–420, ISBN 978-0-387-29362-2.
133. Navarro, P.; Kuharev, J.; Gillet, L.C.; Bernhardt, O.M.; MacLean, B.; Röst, H.L.; Tate, S.A.; Tsou, C.C.; Reiter, L.; Distler, U.; et al. A multicenter study benchmarks software tools for label-free proteome quantification. *Nat. Biotechnol.* **2016**, *34*, 1130–1136. [[CrossRef](#)]
134. Love, M.I.; Huber, W.; Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **2014**, *15*, 1–21. [[CrossRef](#)]
135. Ritchie, M.E.; Phipson, B.; Wu, D.; Hu, Y.; Law, C.W.; Shi, W.; Smyth, G.K. Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **2015**, *43*, e47. [[CrossRef](#)] [[PubMed](#)]
136. Robinson, M.D.; McCarthy, D.J.; Smyth, G.K. edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **2009**, *26*, 139–140. [[CrossRef](#)] [[PubMed](#)]
137. Ahlmann-Eltze, C.; Anders, S. proDA: Probabilistic Dropout Analysis for Identifying Differentially Abundant Proteins in Label-Free Mass Spectrometry. *bioRxiv* **2020**, *14*, 661496. [[CrossRef](#)]
138. Zhang, Z.; Burke, M.; Mirokhin, Y.A.; Tchekhovskoi, D.V.; Markey, S.P.; Yu, W.; Chaerkady, R.; Hess, S.; Stein, S.E. Reverse and Random Decoy Methods for False Discovery Rate Estimation in High Mass Accuracy Peptide Spectral Library Searches. *J. Proteom. Res.* **2018**, *17*, 846–857. [[CrossRef](#)]
139. Burke, M.C.; Mirokhin, Y.A.; Tchekhovskoi, D.V.; Markey, S.P.; Heidbrink Thompson, J.; Larkin, C.; Stein, S.E. The Hybrid Search: A Mass Spectral Library Search Method for Discovery of Modifications in Proteomics. *J. Proteom. Res.* **2017**, *16*, 1924–1935. [[CrossRef](#)]
140. Cho, J.-Y.; Lee, H.-J.; Jeong, S.-K.; Paik, Y.-K. Epsilon-Q: An Automated Analyzer Interface for Mass Spectral Library Search and Label-Free Protein Quantification. *J. Proteom. Res.* **2017**, *16*, 4435–4445. [[CrossRef](#)]
141. Guan, S.; Taylor, P.P.; Han, Z.; Moran, M.F.; Ma, B. Data Dependent-Independent Acquisition (DDIA) Proteomics. *J. Proteom. Res.* **2020**, *19*, 3230–3237. [[CrossRef](#)]
142. Fernández-Costa, C.; Martínez-Bartolomé, S.; McClatchy, D.B.; Saviola, A.J.; Yu, N.K.; Yates, J.R. Impact of the Identification Strategy on the Reproducibility of the DDA and DIA Results. *J. Proteom. Res.* **2020**, *19*, 3153–3161. [[CrossRef](#)]
143. Shiferaw, G.A.; Vandermarliere, E.; Hulstaert, N.; Gabriels, R.; Martens, L.; Volders, P.J. COSS: A Fast and User-Friendly Tool for Spectral Library Searching. *J. Proteom. Res.* **2020**, *19*, 2786–2793. [[CrossRef](#)]
144. Liu, J.; Bell, A.W.; Bergeron, J.J.M.; Yanofsky, C.M.; Carrillo, B.; Beaudrie, C.E.H.; Kearney, R.E. Methods for peptide identification by spectral comparison. *Proteome Sci.* **2007**, *5*, 1–12. [[CrossRef](#)]

145. Lee, S.; Kwon, M.S.; Lee, H.J.; Paik, Y.K.; Tang, H.; Lee, J.K.; Park, T. Enhanced peptide quantification using spectral count clustering and cluster abundance. *BMC Bioinform.* **2011**, *12*, 423. [[CrossRef](#)] [[PubMed](#)]
146. Wang, M.; Wang, J.; Carver, J.; Pullman, B.S.; Cha, S.W.; Bandeira, N. Assembling the Community-Scale Discoverable Human Proteome. *Cell Syst.* **2018**, *7*, 412–421.e5. [[CrossRef](#)]
147. Zauber, H.; Kirchner, M.; Selbach, M. Picky: A simple online PRM and SRM method designer for targeted proteomics. *Nat. Methods* **2018**, *15*, 156–157. [[CrossRef](#)] [[PubMed](#)]
148. 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](#)] [[PubMed](#)]
149. Tiwary, S.; Levy, R.; Gutenbrunner, P.; Salinas Soto, F.; Palaniappan, K.K.; Deming, L.; Berndl, M.; Brant, A.; Cimermancic, P.; Cox, J. High-quality MS/MS spectrum prediction for data-dependent and data-independent acquisition data analysis. *Nat. Methods* **2019**, *16*, 519–525. [[CrossRef](#)]
150. Yang, Y.; Liu, X.; Shen, C.; Lin, Y.; Yang, P.; Qiao, L. In silico spectral libraries by deep learning facilitate data-independent acquisition proteomics. *Nat. Commun.* **2020**, *11*, 1–11. [[CrossRef](#)]
151. Searle, B.C.; Swearingen, K.E.; Barnes, C.A.; Schmidt, T.; Gessulat, S.; Küster, B.; Wilhelm, M. Generating high quality libraries for DIA MS with empirically corrected peptide predictions. *Nat. Commun.* **2020**, *11*, 1–10. [[CrossRef](#)]
152. Lou, R.; Tang, P.; Ding, K.; Li, S.; Tian, C.; Li, Y.; Zhao, S.; Zhang, Y.; Shui, W. Hybrid Spectral Library Combining DIA-MS Data and a Targeted Virtual Library Substantially Deepens the Proteome Coverage. *iScience* **2020**, *23*, 100903. [[CrossRef](#)]
153. Gessulat, S.; Schmidt, T.; Zolg, D.P.; Samaras, P.; Schnatbaum, K.; Zerweck, J.; Knaute, T.; Rechenberger, J.; Delanghe, B.; Huhmer, A.; et al. ProSIT: Proteome-wide prediction of peptide tandem mass spectra by deep learning. *Nat. Methods* **2019**, *16*, 509–518. [[CrossRef](#)]