Current Pharmaceutical Design, 2018, *24*, 3842-3858

REVIEW ARTICLE

Computational Advances in the Label-free Quantification of Cancer Proteomics Data

Jing Tang[a,b], Yang Zhang[a,b], Jianbo Fu[b], Yunxia Wang[b], Yi Li[b], Qingxia Yang[b], Lixia Yao[c], Weiwei Xue*[,a] and Feng Zhu*[,a,b]

[a]*School of Pharmaceutical Sciences, Chongqing University, Chongqing 401331, China;* [b]*College of Pharmaceutical Sciences, Zhejiang University, Hangzhou 401331, China;* [c]*Department of Health Sciences Research, Mayo Clinic, Rochester MN 55905, United States*

Abstract: ***Background*:** Due to its ability to provide quantitative and dynamic information on tumor genesis and development by directly profiling protein expression, the proteomics has become intensely popular for characterizing the functional proteins driving the transformation of malignancy, tracing the large-scale protein alterations induced by anticancer drug, and discovering the innovative targets and first-in-class drugs for oncologic disorders.

***Objective*:** To quantify cancer proteomics data, the label-free quantification (LFQ) is frequently employed. However, low precision, poor reproducibility and inaccuracy of the LFQ of proteomics data have been recognized as the key "technical challenge" in the discovery of anticancer targets and drugs. In this paper, the recent advances and development in the computational perspective of LFQ in cancer proteomics were therefore systematically reviewed and analyzed.

***Methods*:** PubMed and Web of Science database were searched for label-free quantification approaches, cancer proteomics and computational advances.

***Results*:** First, a variety of popular acquisition techniques and state-of-the-art quantification tools are systematically discussed and critically assessed. Then, many processing approaches including transformation, normalization, filtering and imputation are subsequently discussed, and their impacts on improving LFQ performance of cancer proteomics are evaluated. Finally, the future direction for enhancing the computation-based quantification technique for cancer proteomics are also proposed.

***Conclusion*:** There is a dramatic increase in LFQ approaches in recent year, which significantly enhance the diversity of the possible quantification strategies for studying cancer proteomics.

ARTICLE HISTORY

Received: October 10, 2018
Accepted: November 2, 2018

DOI:
10.2174/1381612824666181102125638

Keywords: Cancer proteomics, label-free quantification, target discovery, anticancer drug, computation, mass spectrometry.

1. INTRODUCTION

The primary therapeutic targets for most anticancer drugs (both approved and in clinical trial) are proteins [1-4]. There are millions of distinct proteins in human cells, which requires the qualitative and quantitative analyses of proteome to discover target for anticancer drugs [5]. Quantitative proteomics is thus developed to detect protein concentrations in a variety of experimental samples by integrating cutting-edge analytical technique with computational algorithms [6]. This technique has contributed to the understanding of tumor genesis and development [7-9]. In preclinical anticancer drug discovery, proteomics has unique advantages in understanding the interaction mechanism between drug and target and illustrating the molecular process underlying studied phenotypes [10-12]. Tumorigenesis and metastasis have been found to be closely associated with the dynamics of large protein network [13-15], which make quantitative proteomics greatly attractive to anticancer drug discovery [16]. So far, the proteomics has evolved into a powerful tool and been increasingly adopted by cancer-related research [17-19]. In particular, it has been adopted to characterize the functional proteins driving malignancy transformation [20], trace large-scale protein alteration induced by anticancer drug [21], and discover the innovative targets and first-in-class drugs for oncologic disorder [22, 23].

*Address correspondence to these authors at the College of Pharmaceutical Sciences, Zhejiang University, Hangzhou, China; Fax: +86-23-6567-8450; E-mails: zhufeng@zju.edu.cn (F. ZHU); xueww@cqu.edu.cn (W. XUE)

The large-scale protein concentrations and expressions required in the cancer proteomics studies drive the fast development of quantitative proteomics, and the qualitative technique is often found to be limited in illustrating the full landscape of complex biological processes [6]. Protein expression intensities can then facilitate the identification of potential biomarkers by analyzing differential expression proteins between patients and control subjects [24, 25]. These biomarkers are very useful for choosing the appropriate anticancer therapeutic targets [25-27]. Till now, various established, clinical trial or investigative targets of anticancer drugs have been discovered (directly or indirectly) by quantifying the proteome at the level of both cancer cells and tumor tissues [13, 27]. These remarkable advances have significantly and effectively accelerated the discovery process of anticancer drugs [5, 28].

Diverse quantitative techniques have been employed to quantify the proteins which thereby facilitated the discovery of proteomics biomarker of drug targets for therapeutic developments [23, 29], which included label-free approaches [30-32] and labeling approaches (*e.g.* isobaric [33-35] or isotopic labeling [36-38]). Compared with the proteome quantitation based on labeling approaches, the label-free proteome quantitation (LFQ) approach demonstrates the advantages of allowing a simultaneous detection of proteome without the time and money-consuming procedure for preparing experimental samples by introducing stable isotopes [39]. Moreover, LFQ is capable of processing the large cohort of samples [7] and treating the wide range of sample sources [40, 41]. These distinguishing features make it the most frequently employed proteome quantification in cancer proteomics [30, 42-44]. For exam-

ple, mass spectrometry-based on liquid chromatography (LC-MS) was adopted for conducting LFQ to study the stem cells of colon cancer and discover a key regulator of drug resistance [45-48]. Moreover, LFQ has been applied to identify the activator of human mutant ERα as potential therapeutic targets for the treatment of breast cancer [49].

Although the extensively expanded application of LFQ to the various aspects of current anticancer researches, many great challenges still existed in this research sphere [13, 50-52]. In particular, low precision [53] (substantial change of the detected concentrations among replicates), poor reproducibility [39, 54] (low robustness among identified markers) and inaccuracy [55-58] (extensive deviation from presumed protein abundance) of the LFQ have been recognized as key "technical challenge" in the discovery of targets and drugs for treating cancer. All the issues may be attributed to several factors, which included (*a*) extremely large dynamic range of protein abundances [59], (*b*) large-scale drift of protein peaks of mass spectrometry (MS) platforms [6], (*c*) variations among instrumental runs [6, 58] and (*d*) divergences of different experimental preparations [60]. To address these above issues, mass spectrometry (MS) and several computational approaches (such as quantification tool, transformation, normalization and missing value imputation strategies) were developed and extensively employed for LFQ analyses [7], and were especially applied to identify diagnostic, prognostic, and therapeutic biomarkers for anticancer drug discovery [3, 7, 61].

In this article, the most recent computational progressions in the application of LFQ to cancer proteomics studies were systematically described and critically assessed from multiple perspectives. First, a variety of popular acquisition techniques and state-of-the-art quantification tools are comprehensively discussed and evaluated. Then, a variety of processing approaches including transformation, normalization, missing values filtering and imputation are subsequently discussed, and their impacts on improving performances of LFQ on current cancer proteomics are evaluated. Finally, future directions for enhancing computation-based LFQ technique for cancer proteomics are also proposed.

2. MASS SPECTROMETRY APPLIED TO CANCER PROTEOMICS

Anticancer drug discovery is substantially accelerated by MS-based techniques, and two complementary approaches for such analyses of proteins (bottom-up and top-down) are of great importance [62]. Bottom-up approach has been adopted to discover protein biomarkers for cancer diagnosis and treatment, which is found to yield a larger number of protein markers than the top-down one [63]. It has been successfully used to identify protein candidates differentiating breast cancer stem cells from normal ones [64], and discover the protein fingerprint indicating cancer subtypes [65] and homeostasis [66]. However, there is a significant loss of protein intensities information by bottom-up approach [67]. To cope with this problem, the top-down approach is proposed [68]. The top-down approach aims at identifying proteins together with post-translational modifications (PTMs) [69]. PTMs are found to be very important in tumorigenesis and cancer development, and the top-down approach has been widely utilized to identify markers and analyze the efficacies of anticancer drugs [70-72].

To acquire the raw protein quantification data for cancer research, two modes of the acquisition have been developed, which include the data-dependent (DDA) and the data-independent (DIA) acquisitions [7]. The DDA detects each ion-precursor by intensity, and the DIA implements a complete record of samples [7]. For the proteins quantification based on DDA, the peak intensity and spectral counting were two mainly relative quantification methods [73]. The approach of peak intensity quantification relies on to extract the intensity from MS1 full scan [74]. The label-free approach based on the MS2 quantification depends heavily on the total amount of protein identified [75]. MS1 is the first stage of MS and MS2 is the second stage of MS. It is shown that spectral counting quantifications are extensively efficient for relatively quantify the cancer proteomics data since they can be used to process the dataset specifically collected for discovery [75]. Moreover, as one of the new DIA-based methods, the all theoretical MS acquired by sequential window acquisition (SWATH-MS) is constructed for overcoming DDA problems. To give a comprehensive review, their advantage and disadvantage in cancer proteomics study are discussed as the following, and three modes of acquisition applied in cancer proteomics together with their representative quantification tools were illustrated in Fig. **1**.

The amount of sampled proteins is restricted by the processes of MS/MS sampling [76], but the MS is not capable of acquiring MS spectra of high quality for the large-scale proteins in specific samples [77]. DDA is reported to induce great compromise in MS sensitivity [78]. Recently, several DIA mass spectrometric methods,

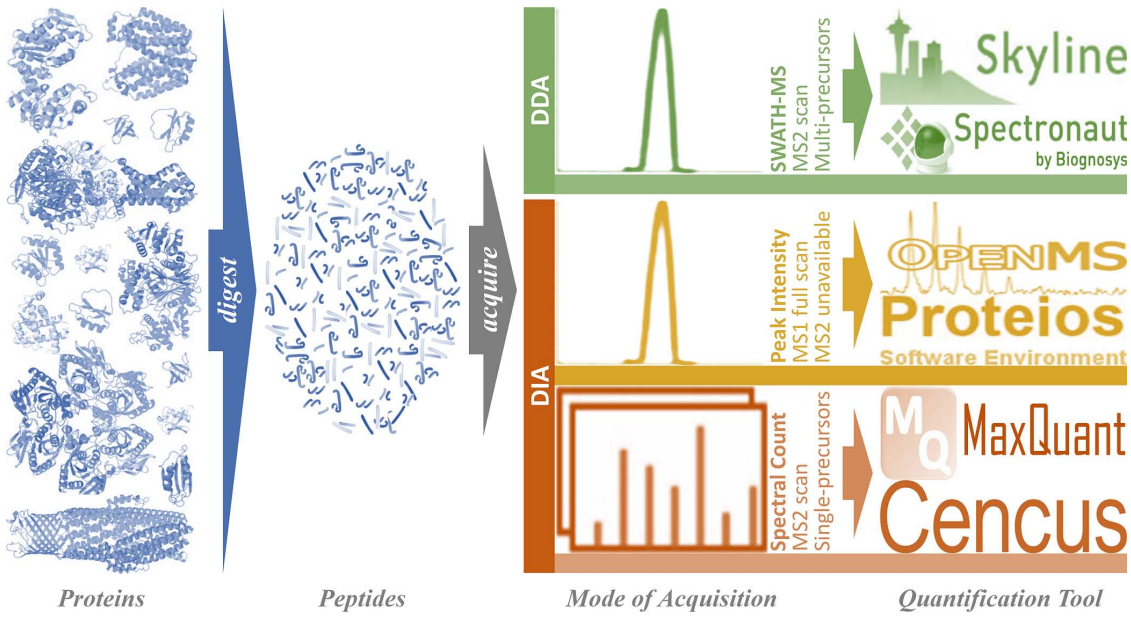

Fig. (1). Three modes of acquisition used in cancer proteomics together with their representative quantification tools.

like SWATH-MS, HDMSE (high definition MSE), AIF (all-ion fragmentation), were established and circumvent the problems of DDA methods [7]. However, the application of DDA for LFQ is reported to be highly dependent on several key factors such as low robustness induced by insufficient sampling and false identification [79].

As one of the most popular DDA-based approaches, peak intensity provides extended dynamic range and better accuracy than spectrum counting [80], and, for higher resolution machines, the protein concentration assessment using peak intensity is reported to be more accurate [81]. However, the precision of approaches based on peak intensity is undermined for low-resolution MS due to significant amounts of thermal noise [82, 83]. Moreover, the intensity-based approach may be limited by its time-consuming analytical process, since the corresponding quantification tools are not developed [84]. Another popular DDA-based method is spectral count, which is the favorable label-free approach for the MS of moderate resolution [85, 86]. It demonstrates the best robustness of biomarker discovery [87] and optimizes the total protein identification [83]. However, the spectral count may result in insufficient sampling when protein concentrations are changed greatly [88, 89].

As new emerging technique [40, 90], SWATH-MS has become increasingly popular by offering enhanced quantification and improved detection of protein intensities compared with conventional methods adopted for analyzing cancer proteomics data [32, 90, 91]. It has emerged as a powerful and effective approach for discovering therapeutic targets [14, 90] and drugs [32, 92, 93] for treating cancer. Moreover, this technique has been used to construct an assay library for profiling cancer proteomics data [94], and recognize alterations induced by anticancer drugs [95, 96]. However, the quantification of cancer proteomics data by SWATH-MS has been found to suffer from inaccuracy [55, 97, 98] and limitation in the dynamic range [90], which should be carefully considered during the LFQ of proteomics data [99].

3. QUANTIFICATION TOOL APPLIED TO CANCER PROTEOMICS

Over the past decade, a number of quantification tools have been developed to analyze cancer proteomics data, which contain both freely accessible software tools and the commercial ones. These tools are the software packages equipped with different sets of statistic algorithm for processing cancer proteomics data acquired by a variety of modes. So far, 18 quantification tools popular in pre-processing proteomics raw data acquired by 3 modes of acquisition have been developed, which are described as follow. List of these tools (SWATH-MS, Peak Intensity and Spectral Counting) are illustrated in Table **1**.

3.1. Tools Pre-processing the Cancer Proteomics Data Acquired by Multiple Modes of Acquisition

Three quantification software tools capable of pre-processing the data acquired by multiple modes of acquisition are available, which include MaxQuant, MFPaQ and Scaffold. All three tools are able to process the data acquired by both peak intensity and spectral counting.

MaxQuant shows advantages of integrating popular algorithms for quantify proteins from high resolution MS-based instrument and enabling match of protein across different samples [100]. Nowadays, MaxQuant is one of the most frequently adopted software for analyzing cancer proteomics data [101, 102]. It is widely used to analyze tandem spectra generated by the collision-induced (CID), high-energy collisional (HECD) and electron-transfer (ETD) dissociation [103] in the cancer proteomics. MaxQuant is used for analyzing the cancer proteomics derived from relative quantification techniques, including label-free quantification [102], labeling readouts from the level of MS1 and MS2 [104]. It was used to identify differentially expressed proteins across NSCLC cells and study dysregulated cellular processes in prostate cancer [105].

MFPaQ is popular for quantifying the cancer proteomics data and is implemented under the condition of Mascot server and Perl program environment [106]. It can extract peak intensity from MS proteomics data based on Extract Daemon Module (EDM), which is a key feature distinguished from other label-free quantification tools [107]. MFPaQ is a tool capable of assisting the identification outputs of Mascot and providing various functions on assessing protein intensities [108]. It quantifies protein concentrations from the raw data files acquired using LC-MS/MS [108] and has been applied to large-scale study on inflammatory endothelial cell [107]. So far, it has been widely used to quantify membrane proteins from primary human endothelial cells [108], and identify novel drug targets for metastatic breast cancer [109].

Scaffold is a commercial bioinformatic tool providing high accuracy on protein identification via applying various statistical methods [110]. It supports various search engines and provides multiple approaches for validating the accuracy of peptides/proteins identification from primary databases [111]. Scaffold has been applied to reveal NS4B-cyclophilin A interaction as a new drug target for the treatment of yellow fever virus infection by inhibiting their replications [112]. Moreover, it has also been used to analyze the follicle fluid proteome to identify the related pathways that are beneficial to the embryo quality [113]. Furthermore, it has been adopted to identify the effects of cadmium exposure on the gill proteome of Cottusgobio [114].

3.2. Tools Pre-processing the Cancer Proteomics Data Acquired Based on SWATH-MS

As a freely accessible quantification tool for processing MS-based raw cancer proteomics data acquired by data-independent acquisition (DIA) [77], the DIA-UMPIRE is extensively functional for the untargeted protein quantification using the SWATH-MS based proteomics dataset obtained via Orbitrap family of MS, and is capable of extracting quantitative data according to proteins discovered in just one sample set [115]. Thus, this tool is capable of getting robust protein quantification across various sets of samples [77]. Compared with the traditional tools of data-dependent acquisition (DDA) [116], this software has been widely applied to discover the similar amount of proteins with greatly improved discovery robustness among various samples. Moreover, it has been frequently applied to process untargeted data for identifying host cell proteins [117] and to export the peptide identification results of pseudo-MS2 spectra [118].

OpenSWATH is high-throughput, open-accessible and automated software tool ensuring a comprehensive analysis of cancer proteomics based on the acquisition mode of SWATH-MS [119]. Particularly, its language of programming is C++, and it is designed as able to work across different platforms, which supports the analysis of dataset from a variety of software developers and is integrated and distributed together with OpenMS [7]. It has been frequently adopted to process bacterial proteomics dataset [119] and estimate *q*-values of protein level [120]. Its generic utility for all types of modification and its scalability enable confident quantification of post-translational modifications in DIA-based large-scale studies [120].

Among these commercial quantification tools aiming at processing raw MS data based on DIA technique, the PeakView demonstrates unique advantages of integrating most of the *in-silico* processing algorithms and offers certain functions of statistical analyses [73, 121]. In particular, this quantification tool is capable of selecting these appreciate transitions or protein ions for quantifying the complex proteome by filtering the basic ion library based on corresponding parameter settings [90], Currently, PeakView has emerged as a powerful quantification tool for processing cancer

Table 1. Eighteen quantification tools popular in pre-processing proteomic raw data acquired by 3 modes of acquisition.

Quantification Tool	Tool Type (*Language*)	Operating System	Type of Input (*File Format*)	Developer	References
(1) SWATH-MS					
DIA-Umpire	Open Source (*editable Java*)	Windows; Linux; OSX	MS2 (*mzXML; wiff*)	University of Michigan	*Nat Methods*. 12:258-64, 2015
OpenSWATH	Open Source (*editable C++*)	Windows; Linux	MS/MS (*mzML; traML*)	ETH Zurich	*Nat Biotechnol*. 32:219-23, 2014
PeakView	Commercial (*uneditable*)	Windows	LC-MS/MS (*wiff*)	SCIEX	*Sci Data.* 1:140031, 2014
Skyline	Open Source (*editable C#*)	Windows	LC/MS (*mzXML*; *pepXML*)	University of Washington	*Bioinformatics.* 30:2521-3, 2014
Spectronaut	Commercial (*uneditable*)	Windows	HTRMS (*raw*)	Biognosys	*Mol Cell Proteomics.* 14:1400-10, 2015
(2) Peak Intensity					
MaxQuant	Open Source (*editable C#*)	Windows; Linux	MS1/MS2 (*raw*)	Max-Planck Institute	*Nat Protoc.* 11:2301-19, 2016
MFPaQ	Open Source (*editable Perl*)	Windows; Linux	LC-MS/MS (*dat*)	IPBS Toulouse	*Mol Cell Proteomics.* 6:1621-37, 2007
OpenMS	Open Source (*editable C++*)	Windows; Linux; OSX	MS1/MS2 (*dat*; *mzXML*)	University of Tübingen	*Nat Methods.* 13:741-8, 2016
PEAKS	Commercial (*uneditable*)	Windows	LC-MS/MS (*raw*; *wiff*)	Bioinformatics Solutions	*Mol Cell Proteomics*. 11:111.10587, 2012
Progenesis	Commercial (*uneditable*)	Windows	LC-MS (*mzXML*; *mzML*)	University of Liverpool	*OMICS*. 16:489-95, 2012
Proteios SE	Open Source (*editable Java*)	Windows; Linux; OSX	MS1/MS2 (*mzML*)	Wellcome Trust Genome Campus	*Nucleic Acids Res.* 45:1100-6, 2017
Proteome Discoverer	Commercial (*uneditable*)	Windows; Linux	MS1/MS2 (*raw*)	Thermo Fisher	*J Proteome Res.* 10:3840-3, 2011
Scaffold	Commercial (*uneditable*)	Windows	Thermo SCIEX (*raw*; *wiff*)	Proteome Software	*Proteomics.* 10:1265-9, 2010
(3) Spectral Count					
Abacus	Open Source (*editable Java*)	Windows; Linux; OSX	MS (*fasta*)	University of Michigan	*Proteomics.* 11:1340-5, 2011
Census	Open Source (*editable Java*)	Windows; Linux; OSX	MS1/MS2 (*pepXML*; *mzXML*)	Scripps	*Bioinformatics.* 30:2208-9, 2014
DTASelect	Open Source (*editable Perl*)	Windows	LC/MS/MS (*fasta*)	Scripps	J Proteome Res. 1:21-6, 2002
IRMa-hEIDI	Open Source (*editable Java*)	Windows	LC-MS/MS (*dat*)	Fondation Rhône-Alpes Futur	*Bioinformatics.* 25:1980-1, 2009

(Table 1) Contd....

Quantification Tool	Tool Type (*Language*)	Operating System	Type of Input (*File Format*)	Developer	References
MaxQuant	Open Source (*editable C#*)	Windows; Linux	MS1/MS2 (*raw*)	Max-Planck Institute	*Nat Protoc.* 11:2301-19, 2016
MFPaQ	Open Source (*editable Perl*)	Windows; Linux	LC-MS/MS (*dat*)	IPBS Toulouse	*Mol Cell Proteomics.* 6:1621-37, 2007
Scaffold	Commercial (*uneditable*)	Windows	Thermo SCIEX (*raw*; *wiff*)	Proteome Software	*Proteomics.* 10:1265-9, 2010

proteomics data, especially in the fields of diagnostic, prognostic, and therapeutic biomarkers identification [90]. For example, it was applied to fulfill enrichment analysis of N-linked glycoproteins [122], evaluate the sample volume needed for SWATH-MS analysis [123] and identify methods used for extracting green algae [124].

The freely accessible quantification tool Skyline can not only be useful for processing the datasets acquired by three reaction monitoring techniques (selected (SRM), multiple (MRM) and parallel (PRM)), but also is capable of analyzing SWATH-MS data and targeted DDA data based on MS1 quantification information [124]. This tool can facilitate targeted cancer proteomics study [125-128]. So far, it has been applied to the protein quantifications of targeted cancer proteomics [129], including the proteomics profiling of different cancer cell lines [18], discovery of certain proteins associated with pancreatic cancer [130] and prediction of drug responses to anticancer therapeutic targets [131].

Another widely applied quantification tool for targeted analysis of DIA measurement is the Spectronaut, which is designed for targeted analysis of DIA measurement based on SWATH-MS independent of mass spectrometer [116, 132]. It is very powerful in peak picking and automatic interference correction utilizing specific spectral library, which was mainly produced in the data acquisition across different MS analysis platforms, and specifically applied to support the workflow without a spectral library and targeted analysis of OMICs data by hyper reaction monitoring [7, 73]. It is widely applied to DIA-based quantitative protein profiling [116], proteomics quantifications enhanced by sequential window acquisitions [73] and retention time prediction in targeted DIA analysis indexed by high-precision [132].

3.3. Tools Pre-processing the Cancer Proteomics Data Acquired Based on Peak Intensity

As freely accessible quantification software for processing MS-based raw cancer proteomics dataset, OpenMS has robust and high-throughput characteristics and is thus suitable for analyzing cancer proteomics dataset with improved reproducibility [104]. It supports processing procedures by submitting various standardized MS raw dataset formats and provide a well access interface [133]. It is widely applied to the quantitative and variant enabled mapping of protein to genome [134], analyses of cerebrospinal fluids proteome in Alzheimer's disease [135], identification of key proteins involved in the microbial-host interaction based on label-free LC-MS data [136] and screening of altered plasma proteins expression in colorectal cancer [137].

PEAKS is a software platform with a complete solution for the discovery proteomics, which conducts the identification of proteins using protein de novo sequencing searching engine approaches [138]. It can efficiently estimate the optimal protein sequence due to their fragment ions can well reflect the peaks in tandem MS spectrum based on a dynamic programming [139]. It has emerged as a powerful software for identification and quantification of protein from cancer proteomics dataset [138]. It matured into a comprehensive proteomics platform supporting the analysis of label-free and labeling based proteomics dataset. Compared with other quantification software, PEAKS stands out by generating the high accuracy and sensitivity in protein quantification [140].

As a commercial tool for processing MS-based raw cancer proteomics data, the Progenesis has emerged as the new generation of bioinformatics vehicle targeting small molecule analysis for both metabolomics and proteomics, which quantifies protein concentrations by MS1 ion intensity [141]. It supports parameter settings to align peak ion signals across different runs [142]. It provided the function of protein label-free quantification and ion detection based on a high sensitivity algorithm, which can be suitable for data with noise [143]. Nowadays, Progenesis has been widely applied for cancer proteomics study, including the identification of potential serum biomarkers for improving the diagnostic accuracy of ovarian cancer [144] and discovery of potential biomarkers associated with NSCLC which are possibly regarded as drug targets for drug-induced cell apoptosis [145].

Proteios SE is free and open source quantification tool, which can process two types of cancer proteomics data [146]. During the whole process of cancer proteomics quantification using this tool, it allows not only the identification of proteins using search engine approaches but also provides the continuous annotations as well as quantitation data [147]. This tool has become the standard analysis platform for analyzing cancer proteomics data due to the characteristics of shared data and tracking samples. More importantly, it provides links which automatically access various proteomics processing procedures [148], and enlarges coverage of proteins via supporting identification based on a variety of common search engines, and automatically generates the proteins identification reports containing the information required for publication of proteomics results [149]. These advantages make it widely adopted by various aspects of cancer proteomics, including identification of potential portraits or differential expression proteins for breast cancer [150].

By providing the workflow-driven analysis of the cancer proteomics dataset, Thermo Proteome Discoverer automatically completes multiple processing procedure [151], such as the tandem MS spectrum extraction, protein identification and quantification [152]. It has a convenient graphical user interface [153]. The users can directly submit the MS raw (Thermo) data from instrument, and this tool allows the identification and quantification of proteins via multiple search engines [154]. It is suitably applied to diverse quantification techniques (iTRAQ, TMT and SILAC) [155, 156]. Proteome Discoverer has been applied for studying the effect of ERBB2 gene expression of on gastric cancer [157].

3.4. Tools Pre-processing the Cancer Proteomics Data Acquired Based on Spectral Count

Abacus is an open source tool for processing proteomics data [96]. Compared the protein quantification based on the MS1 peak intensity, it extracts and processes spectral count from MS/MS spectrum for label-free proteome quantification [158]. The abacus mainly focused on providing a streamlining, automatic analysis and

user-friendly workflow for protein quantification by spectral count [159]. The convenient and efficient quantification workflow generated quantification report or result that is well suitable for the downstream bioinformatic analysis [158]. However, this method also has the shortcoming such as missing information because of analysis abounding spectra numbers based on relatively small sets of differential spectrums, and it is widely applied in cancer proteomics studies [158] to identify biomarkers or therapeutic targets for improving survival hormone-refractory prostate cancer [160].

As commercial quantification tool for protein quantification based on the spectral count, Census not only can process the shotgun cancer proteomics data with label-free but also is available for various stable isotope labeling experiments [161]. Wide coverage of quantification strategies and multiple statistical algorithms for improving quantification quality makes it differentiated most from other spectral count quantification tools [162]. Census can be used for identifying altered expression proteins associated with drug treatment in Plasmodium falciparum [162] and investigating protein turnover using metabolic labeling strategy [163]. DTASelect is developed using Java language and can be applied to analyze and validate identification of the proteins which generated by tandem MS database search engine (SEQUEST) [164]. SEQUEST is one of the most widely applied proteins search engines [165]. The procedures of DTASelect included filtering, establishing, visualization of a huge number of tandem mass spectra from a simple bio-sample [166]. This method focuses on the proteins of interest by eliminating the unlike identification and thus improving protein quantification based the accuracy peptide data [166]. It makes more complex experiments feasible by streamlining data analysis [167], and it can be applied to a variety of cancer proteomics studies with a lower false positive [168] and identifying the large-scale palmitoylated proteins [169].

IRMa toolbox is similar toDTASelect and can also analyze and validate the accuracy of protein identification, but protein identification is generated from different Mascot search engines [170]. The IRMa can automatically filter inaccuracy identifications from the primary Mascot identification searching and ensure the accuracy of peptide identification with significantly low false discovery rate [32]. At the same time, it also provides the supporting of manual confirmation or elimination of peptide-spectrum matches (PSMs) [171]. Its main originality is to filter the matches rather than identified proteins and its features are easy navigation within identification result and batch mode to automatically validate multiple results [171]. Filtered results based on IRMa needed to be processed using the in-house tool (hEIDI), which can make compilation, grouping and comparison of protein intensities across different samples [172]. IRMa-hEIDI has been widely used for investigating the relationship between triads and microtubules [173]. Moreover, the ProteinProphet sets up a statistic package to compute the percentage of chances which proteins are available in the studied target [174]. ProteinProphet can be applied to filter the large-scale cancer proteomics data with significantly reduced false-discovery rate [174]. It has been applied to differentiate the correct identification from the false one [175] and also used to calculate the possibility of a protein successfully identified [174].

3.5. Application of Quantification Tools in Cancer Immunotherapy

The immunotherapy is a very hot topic recently, especially for cancer treatment. There are many applications of proteomics on this topic. Especially, many quantification tools have been applied to this particular research direction. In particular, a variety of quantification tools were frequently used to enhance effective cancer immunotherapy [176]. MaxQuant has been widely applied to investigate the results of the protein or metabolite level of the studied inhibition of protein PC1/3 in macrophage, and identified the suppression of this studied protein demonstrates significant potentials in applying to the discovery of novel immunotherapy for cancer patients [177]. Progenesis has been applied for investigating the allergen composition in certain crops for oral immunotherapy [178].

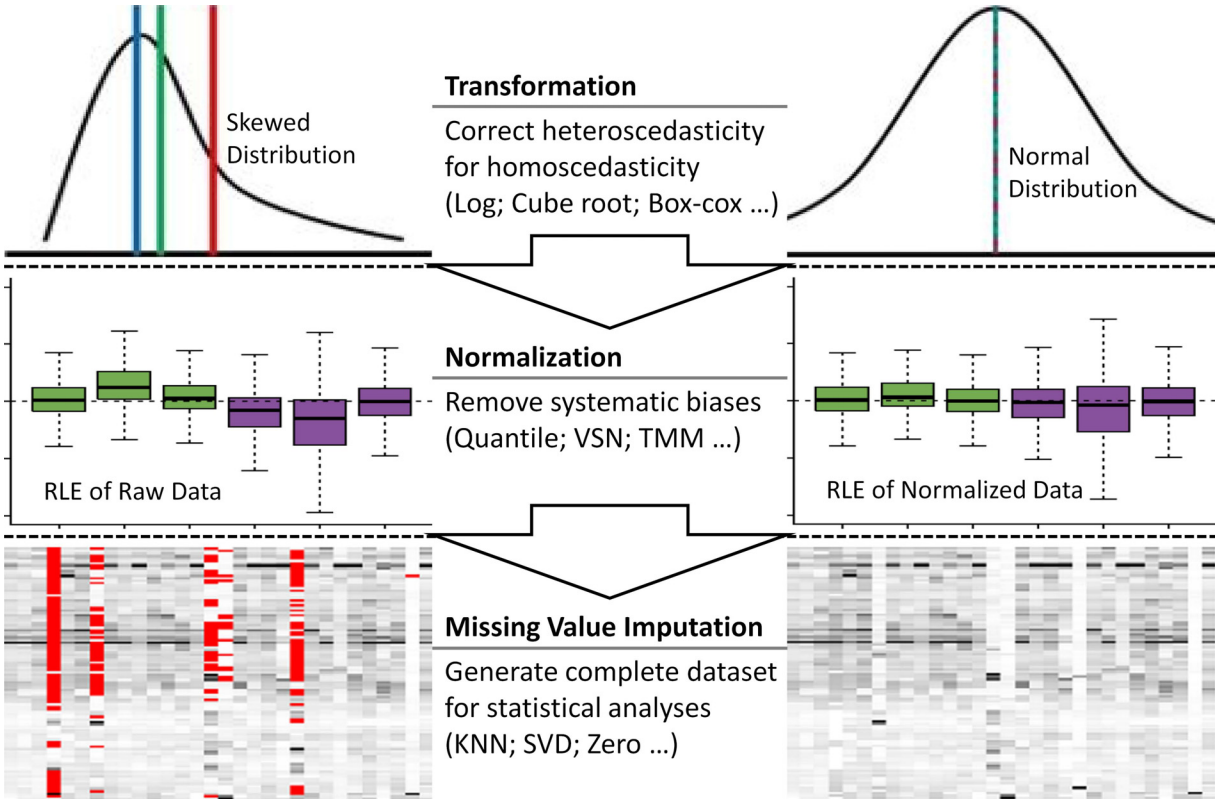

Fig. (2). Data processing methods sequentially applied in cancer proteomics.

Thermo Proteome Discoverer has been used for identifying circulating protein and antibody biomarker for personalized cancer immunotherapy [179].

4. PROCESSING METHODS IN CANCER PROTEOMICS

There are three types of processing methods currently developed for cancer proteomics (transformation, normalization, and missing value imputation). These methods are sequentially applied, which were illustrated in Fig. **2**. Detailed description of those applied methods and their application in current cancer proteomic studies were further provided in the following sections.

4.1. Data Transformation Applied for Processing Cancer Proteomics Data

Before normalization, cancer proteomics data often need to be transformed [180]. Protein abundances in data matrix are found to be distributed in the right-skewed manner [180]. Thus, a proper usage of the transformation method is essential for resulting in a distribution of improved symmetry. In other words, transformation methods can make the distribution of protein intensities more normal and symmetric [181]. Currently, 4 transformation methods (Box-cox, Cube Root, Log and Power) frequently applied to process the label-free proteomics data. Explanations on each method are provided in Table **2**.

As a method capable of stabilizing variances of protein intensities, arcsine transformation is proposed by Snedecor*et al*, which is well suitable for processing proportion or percentages dataset which is not in normal but a skewed distribution [182]. The application of arcsine transform requires the total amount of trials to be the same for different data-point, but the effectiveness of this method for processing proportional data is highly susceptible to the size of samples [183, 184]. Moreover, the arcsine transformation will lead to extrapolation of the calculated values which cannot be sensitively detected in the anticancer research [185]. The method has been widely applied to process the datasets of both binomial and non-binomial data and been used to enhance the understanding of LAR's way regulating cell adhesion in proteomics data [186] and discover biomarkers closely associated with the process of the cryopreservation of fish sperm [187]. Moreover, a method with the ability of parametric power transformation aiming at getting rid of multiple anomalies [188], the box-cox transformation has received extensive researches and utilities to various cancer proteomics studies [188]. Moreover, the novel biomarkers and emerging therapeutic targets towards several important hepatic diseases can be identified by box-cox through the examples of relevant omics datasets [188, 189].

Mean and variance of the distribution using N_{th} power transformation by substituting N=1/3 have been applied to treat the cancer proteomics data using the cube root transformation (CUB), which is primarily developed based on probability density function [190]. CUB has been used to improve the peak detection of proteomics and quantifications of mass spectrometry-based cancer proteomics datasets that are mainly obtained from surface-enhanced laser desorption [191]. Additionally, a symmetric distribution prior to statistical analysis is generally acquired from the log transformation (LOG), which is suitable for the data that the residuals become bigger for values of the dependent variables [192]. Such tendency happens usually in the residuals due to errors or changes in the value of the result variable is usually a percentage of value instead of an absolute value [192]. The log transformation has been used to the cancer proteomics analyses of colorectal cancer patients and to quantify thousands of proteins among patients mainly with normal mucosa, primary carcinoma, and nodal metastases [192, 193].

The normal linear model can be transformed through power transformation (POW) [194]. It usually possesses a series of functions that can be applied to carry out a monotonic transformation [194] and is a powerful data transformation technique with the capacity of stabilizing variance [195]. This transformation converts the original data distribution into a normal one and further enhances and increases the association between alternating quantities and some alternative procedures of datasets stabilization [194]. This approach has been applied to quantitatively demonstrate how the observational data alters the findings derived from synthesized evidence from RCT [196]. It has also been used to relatively estimate the protein intensities acquired by bottom-up MS information incorporating data [197].

Table 2. Five transformation methods currently available for LFQ-based cancer proteomics.

Methods	Abbr.	Packages (*Function*)	Brief Descriptions	References
Arcsine	ARC	metafor (*transf.arcsine*)	The ARC can make variances more constant. For proportions or percentages data, ARC is often used. The numbers to be arcsine transformed must be in the range 0 to 1.	*J Cell Sci.* 29:2962-71, 2016
Box-cox	BOX	AID (*boxcoxfr*)	Box-Cox is used as a metric to quantify how normal or log-normal certain data. The Box-Cox fulfils the basic assumptions of linearity, normality and homoscedasticity simultaneously The LOG can be applied for heavilyleft-skewed data distributions	*The Statistician*. 41:169-178, 1992
Cube Root	CUB	pamr (*pamr.cube.roo*)	The CUB transformation is strong, which applied for right-skewed data and improves distribution of the data somewhat. For simple count data, CUB transform is often used.	*Chemistry*. 22:2501-6, 2016
Log	LOG	metabolomic (*LogTransform*)	The LOG is a special case of Box-Cox. The LOG is a relatively strong transformation. Difficulties with values with large relative standard deviation and zeros. The LOG was applied for right-skewed distribution.	*Anal Chem*. 84:10768-76, 2012
Power	POW	car (*bcPower*)	Powertransformation technique can be widely applied for obtaining stable variances, which aimed at generating more normal distribution.Choice for square root is arbitrary.No problems with small values	*Atmos Environ*. 71:54-63, 1994

4.2. Data Normalization Applied for Processing Cancer Proteomics Data

Systematic biases are reported to be prevalent in cancer proteomics data due to the semi-stochastic property of DDA-MS method [198]. The normalization techniques can remove any excess technical variability and has gradually become popular in cancer proteomics[198]. Normalization has been widely considered as an integral part of LFQ for improving accuracies for relative protein quantification [198]. So far, 16 normalization methods have been developed and popular in analyzing cancer proteomics data. Detail explanations on each normalization method are provided in Table **3**.

As the simplest approach regulating the proteomics variance, the auto scaling (ATO, unit variance scaling) can scale protein intensities according to standard deviation of cancer proteomics dataset [199]. Such approach scales the protein intensities into unit variances, and all intensities are equally important and comparably scaled [200]. The data is analyzed on the basis of correlations and standard deviations of all intensities, but it is necessary to pay attention to the amplification of the analytical variations because of dilution effects [199]. This method has also been adopted to identify proteomics biomarkers for psoriasis and psoriasis arthritis [201] and normalize LC-MS proteomics data based on scan-level information [202].

Based on the combination of MA-plots and logged Bland-Altman plots obtained through the assumption of non-linear bias existences [199], the cyclic loess (CYC, cyclic locally weighted regression,) is obtained to estimate regression surface using multivariate smoothing procedures [203]. But the time-consuming process of cyclic loess should be carefully considered, and the consumption of time rises exponentially with the increase in the total number of samples [204]. CYC has been applied to proteomics profiling in the context of common experimental designs for anticancer research [205].

The bias of unknown complexity from cancer proteomics data based on LC/MS can be removed by EigenMS (EIG), and the sensitivity of differential analysis is improved [206]. EigenMS normalization aims at preserving original difference while removing the bias from the data [207], and works via three steps [208]: (*a*) it retains true difference of proteomics data through evaluating an ANOVA model effectiveness; (*b*) the bias trends can be determined by singular value decomposition of residuals matrix; (*c*) a permutation test is used to estimate the number of bias trends as well as eliminating the bias trends. EIG has been applied in the profiling of MS-based quantitative label-free proteomics and LC-based proteomics [209, 210].

Each spectrum can be mapped to the baseline by linear baseline (LIN, linear baseline Scaling,) based on the hypothesis of a constant linear relation between a given spectrum's features and baseline [199]. Baseline refers to the median value of protein intensities across the whole spectrum, and the factor of scaling is then calculated by assessing the percentage of mean protein concentration in the spectrum mean intensities [199]. Nevertheless, it may be oversimplified to assume a linear-type of correlation among samples [199].

Two-color expression data are normalized by locally weighted scatterplot smoothing with compensation for non-linear dye-bias. In such method, the lowess fitted value can adjust the log-ratio for each sample [211], and the normalization hypothesizes that the appearance of dye bias relies on spot intensity [211]. This normalization can be applied to complete or incomplete datasets and may be applied to a two-color array expression dataset [211]. This method has been used in MS-based cancer proteomics [209].

Data can also be normalized by the mean normalization (MEA) using mean value of all signals to eliminate background effects [212]. The intensity of each protein in a given sample is adopted by the mean intensity of all variables in the sample [192]. To make the samples comparable, the means of intensities for each experimental run are forced to be equal to one another using this method [213]. Each sample is scaled such that the mean of all abundances in one sample equals one [192]. This method has been used in the profiling of urine peptidome [214].

Based on the assumptions that the samples of a dataset are separated by a constant, median normalization (MED) is proposed to scale samples so that they have the same median [215]. For instance, the median of protein intensities in the sample equals one [216]. The median normalization, the commonly used method without the need for internal standards, is more practical than sum normalization especially in these conditions where several saturated abundances may be related to the factors of interest [216]. It has previously been used in MS-based label-free proteomics analysis for removing those biases closely related to MS-based instruments [217].

As a robust measure of the data spread, Median Absolute Deviation (MAD) can be applied to evaluate the standard deviation of sample when scaled by the factor of 1.483, and it is a simple way to quantify variation [218]. Moreover, the quality control processes of proteomics data based on the peptide-centric LC-MS can be improved by such approach, and this method has been used to improve QC procedure of protein-centric LC/MS proteomics [218]. Moreover, the standard deviation of the sample can be utilized by Pareto scaling (PAR) as a scaling factor [199]. PAR is capable of reducing the weight of large fold changes in the protein intensities, which is more significant than auto-scaling [199]. However, as dominant weight, the extremely large fold changes may not change [199]. Therefore, the disadvantage of Pareto scaling is the sensitivity to the large fold changes [219]. The data based on the information of scan-level can be applied to normalize LC-MS proteomics data in the Gaussian process regression model [202].

Based on the systematic estimation of the most likely dilutions, the proteomics spectra can be transformed by probabilistic quotient normalization (PQN) [220]. In contrast to the normalization based on the integral and the vector length, PQN algorithm has been pointed out to have remarkable robustness and accuracy [220]. There are three steps in the procedure of PQN: (*a*) each spectrum should be integrally normalized, then a reference spectrum (median spectrum) will be selected; (*b*) calculate the quotients between experimental spectra and the control ones, then estimate median values of the quotients for each variable; (*c*) the median quotient can be used to divide the whole variables of the test spectrum. PQN has been applied in MALDI-TOF mass spectrometry knowledge discovery [221].

Equal distribution of the protein intensities crossing whole samples can be obtained by the quantile (QUA, quantile normalization), and the quantile-quantile plots embedded in this method can be used to visualize the similarity of such distributions [199]. QUA is motivated by the idea that the distribution of two data vectors is equal if the quantile-quantile plot forms a straight diagonal line [216]. While a common and non-data driven distribution is generated by quantile normalization, an agreed standard could not be reached [216]. Systematic biases related to mass spectrometry and label-free proteomics can be removed by this method [217]. In addition, as a transference approach, robust linear regression (RLR) is used for rescaling one reference interval to another scale. RLR is more robust against the outliers in the data than linear regression using least squares estimation [207]. This method has been used to reduce plate effects from data of suspension bead array [222].

Compared with other normalization methods widely applied to cancer proteomics data, Total Ionic Current can normalize proteomics data based on estimating the sum of the whole peak intensities of proteins identical to a particular sample [223]. TIC assumed that all peptides/proteins are of the same importance in a specific

Table 3. Sixteennormalization methods currently availablefor LFQ-based cancer proteomics.

Methods	Abbr.	Packages	Brief Descriptions	References
Auto Scaling	ATO	Metabolomics	This method scales all peptides/proteins to unit variance, and all proteins are the same important and comparably scaled. The disadvantage is that the method may be unsuitable when the assumption does not hold.	*Metabolomics*. 11:684–695, 2015
Cyclic Loess	CYC	Limma Affy	Cyclic Loess has thedisadvantage of time-consuming especially for the large number of bio-samples or high-dimensional peptides/proteins features.	*Metabolomics*. 10:897-908, 2014
EigenMS	EIG	DanteR	EigenMS aims at preserving original differences and removing bias from data. It can preserve true differences by constructing statistic model compared with other normalization methods.	*Brief Bioinform*. 19:1-11, 2018
Linear Baseline	LIN	Affy	LIN assumed that the peptides/proteins features are linear correlated in the specific bio-samples. Thus, the EIG has may be unsuitable when the peptides/proteins features are not linear correlated.	*Metabolomics*. 8:146-160, 2012
Locally Weighted Scatterplot Smoothing	LOW	LPE	Lowess assumed that the variation was relied on peptides/proteins intensity. It is a non-linear normalization method, and the log-ratios (intensity) was corrected via the fitted values.	*Nucleic Acids Res*. 30:e15, 2002
Mean	MEA	mixOmics; Normalyzer	Mean normalizes data by mean value of all signals to eliminate background effect. To make the samples comparable, the means of the intensities for each experimental run are forced to be equal to one another using this method.	*Plant Cell Rep*. 25:71-9, 2006
Median	MED	Normalyzer mixOmics	Median assumes that the samples of a dataset are separated by a constant. It scales the samples so that they have the same median, which is practical especially when several saturated abundances may be associated with some factors of interest.	*Bioinformatics*. 19:185-93, 2003
Median Absolute Deviation	MAD	stats	MAD is a robust normalization method based on the estimation on sample standard deviation. This method has advantage of processing asymmetric proteomics data.	*Bioinformatics*. 27:2866-72, 2011
Pareto Scaling	PAR	BioMark	PAR can decrease the important of large fold change in the large peptides/proteins. Thus, it may be too sensitive to large fold change proteins.	*BMC Genomics*. 7:142, 2006
PQN	PQN	KODAMA MALDIquant mQTL	PQN had advance advantages of high robust and accuracy, which can normalize proteomic data via choosing a specific reference sample as the median one.	*Anal Chem*. 78:4281-90, 2006
Quantile	QUA	Normalyzer	Quantile can make the distributions of peptides/proteins intensities be similarity across different MS runs. Its disadvantages generated large protein intensity values after normalization.	*J Proteome Res*. 5:277-86, 2006
Robust Linear Regression	RLR	Normalyzer	RLR is used for transference when you want to rescale one reference interval to another scale. The robust linear regression is more robust against outliers in the data than linear regression using least squares estimation.	*J Proteome Res*. 15:3473-3480, 2016
Total Ion Current	TIC	Normalyzer	TIC assumed that all peptides/proteins are the same important in a specific bio-sample and generated lower peak intensities. Its disadvantages not suitable this situation when the assumption does not hold.	*Anal Chem*. 88:11568-74, 2016
Trimmed Mean of M Values	TMM	edgeR	TMM normalized proteomic data based on estimating relative protein peak intensity and often was incorporated bioinformatic analysis for identifying differential expression protein.	*BMC Genomics*. 17:28, 2016
VSN	VSN	vsn	VSN, a non-linear method, aims at maintaining variance constant across whole ranges. It performs linear transformation behavior to make variance unchanged and can reduce sample-to-sample variation and adjust variance of different proteins.	*Bioinformatics*. 18:S96-104, 2002
Z-score	ZSC	mosaic	ZSC normalizes data based on the mean and standard deviation and has the advantage of allowing comparison of proteomic data independent of raw protein abundances.	*Mol Cell Proteomics*. 8:2285-95, 2009

bio-sample and generated lower peak intensities after normalization [223]. It has been applied to MALDI- and SELDI-TOF mass spectra proteomics profiling [224]. Moreover, as a popular normalization method, trimmed mean of M values (TMM) is easy and efficient to process the RNA-sequence data [225]. It can be used to estimate scaling factors among data and can be embedded in statistic method [225], which is susceptible to the removal of genes of low-expression from the dataset in RNA-sequence data [225].

Variance can be a constant over the whole data range by variance stabilization normalization (VSN), and it is well suitable for processing large feature values to remove the heteroscedasticity using the inverse hyperbolic sine [199, 226]. For small intensities, VSN performs linear transformations behavior to make the variances unchanged [199], which was originally developed as normalization for the relative LFQ of endogenous peptide [199, 227]. Moreover, data can be normalized by Z-score normalization (ZSC) based on the mean and standard deviation [228]. ZSC offers an approach of data standardization as well as comparing the microarray data which is independence of the intensities of original hybridization [228]. Normalized data by such a method can be applied to directly calculate the remarkable changes between two distinct groups [229]. In addition, this method has been used in proteomics experiments based on LC-MS to assess the outcomes of data normalization, which can decrease the possibility of the bias introduction and determine the suitable approach of normalization [230].

4.3. Missing Value Filtering & Imputation Applied for Processing Cancer Proteomics Data

Cancer proteomics data are sparsely distributed [231], namely a typical proteomics data matrix contain many missing values in many cases [73]. Missing values can occur due to serval causes. For example, the concentration of proteins is lower compared to the detection limit of the instrument [232], the identification of the incorrect peptide [209], various biological factors or technical/analytical mistakes, or the missing peptide or proteins abundances may not appear in the samples [209]. Thus, data filtering and missing value imputation strategies often are available for addressing these issues [232]. Currently, there are 6 imputation approaches that are often used to treat the missing values, including Bayesian Principal Component Imputation, Censored Imputation, K-nearest Neighbor Imputation, Local Least Squares Imputation, Singular Value Decomposition and Zero Imputation. A detail explanation of each imputation method is provided in Table **4**.

The condition of missing protein values due to the smaller concentrations in the samples can be stimulated by Background Imputation (BAK) [233]. Missing values can be displaced with the lowest values of the dataset, and the lowest values can be used to impute the missing values [233]. Moreover, this method has been used in some cancer proteomics analysis software for label-free cancer proteomics quantification and imputation [73]. Moreover, as one of the most popular filtering method, Basic Filtering (filtered) has been integrated into proteomics analysis [73]. "Not missing at random" mainly refers the proteins with not merely a missing value per particular sample group which contains 3 technical replicates in every dataset, which are filtered out to analyze the differential expression between the datasets [73]. While, as for the "missing completely at random", there are no values imputed [73].

As an imputation method, Bayesian Principal Component Imputation (BPCA) out-performs the KNN and SVD approaches. Compared to KNN and SVD, BPCA has the advantages of auto setting parameters of estimation, which makes BPCA easy to operate and perform well [233]. This method also produces improved estimation performance when the number of samples is huge [233]. In addition, this method has been used to process missing values of multivariable statistical analysis of proteomics data [233, 234]. Moreover, as "complete missing at random", there is no values imputed for Censored Imputation (CEN) only when a single missing value for the given protein in sample group appears [73]. As for this situation, namely the given protein consisting of not merely one missing value in a sample group, this CEN strategy can address and impute the missing value due to lower concentration peptides or proteins based on lowest values in a specific proteomics data [73]. This method has been used to improve detection of differentially abundant proteins [235].

Table 4. Six missing imputation methods currently available for LFQ-based cancer proteomics.

Methods	Abbr.	Packages (*Function*)	Brief Descriptions	References
Bayesian Principal Component	BPCA	pcaMethods (bpca)	The missing values are estimated based on a variant Bayes algorithm. The imputation strategy was well suitable for the large number of studied samples	*Malays J Med Sci.* 21:20-7, 2014
Censored Imputation	CEN	imputeLCMD (impute.MAR.)	The lowest intensity value in the data set was imputed for the missing values when were considered non- missing completely at random	*Brief Bioinform.* doi: 10.1093, 2017
K-nearest Neighbor	KNN	imputation (knnImputation) VIM Packages (kNN)	The missing values are estimated with a weighted average over k proteins. The k most similar proteins were found by k-nearest neighbors algorithm.	*BMC Bioinformatics.* 17:247, 2016
Local Least Squares	LLS	pcaMethods (llsImpute)	The missing values are estimated with least squares regression as a linear combination of the values of these k proteins.	*Bioinformatics.* 21:187-98, 2005
Singular Value Decomposition	SVD	pcaMethods (svdImpute)	The missing values are estimated based on a linear consideration. This most significantly expressed eigenproteins were applied for linear regression	*Proc Natl Acad Sci.* 97:10101-6, 2000
Zero	ZER	imputeLCMD (impute.ZERO)	The missing values are estimated as zero not consideration using above algorithms	*Nucleic Acids Res.* 34:1608-19, 2006

K proteins analogous to proteins with missing values can be identified by K-nearest Neighbor Imputation (KNN). Euclidean distance measure can be used to estimate the similarity between the proteins, and the values from weighted average of the neighboring proteins can be used to impute the missing values [233]. The methods based on KNN have the capacity to select the most similar proteins with expression profiles to the desired proteins to impute missing values, and as for the relatively small size samples, KNN presents some advantages compared to BPCA and LLS [233]. This method has been used in integrative analysis of omics data [236]. Moreover, Local similar structures together with the optimization treatment by least squares in the given data can be exploited by Local Least Squares Imputation (LLS) [233]. LLS can impute the missing values based on three mainly procedures: (1) choosing N most similar proteins by k-nearest neighbors, (2) making a linear regression based on these N proteins and (3) estimating the missing values via the least squares algorithm [237]. This method has been utilized in the treatment of missing values for data with the form of matrix, such as NGS data [237, 238].

Singular Value Decomposition (SVD) is an imputation method based on a linear relationship across different peptides or proteins of a specific sample [239]. Compared with KNN using the local pairwise information from proteins expression, SVD forecasts the missing values mainly through the global information acquired from the whole matrix [239]. SVD contributes accuracies in quantitative comparisons of protein intensity levels [232, 240]. Moreover, displacing missing values with zeros (zero imputation) was regarded as the simplest among above-described methods. The ZER is not dependent on any information about the data [240]. In reality, the incorrect or inappropriate relation among the proteins can be generated because of human factors when imputing, which negatively impact integrity and usefulness of the data [240]. This method has been used in the analysis of experiments using isobaric tagging based on quantitative proteomics [241].

CONCLUSION

High-throughput mass spectrometry technology has been developed to mature the analytical platform for qualitative and quantitative analyses of proteins. The large-scale protein differential expressions analysis not only can facilitate to identify the potential biomarkers for the cancer diagnosis and treatment, but also provide the new insights into molecular mechanisms underlying disease process and development. Moreover, these potential markers may be possible to be chosen the most suitable anticancer therapeutic targets for improving the prognosis and survival time of cancer patients. Recent advances in computation methods of LFQ significantly enhance the diversity of possible quantification strategies for studying cancer proteomics, and many processing approaches including transformation, normalization, filtering and imputation and their impacts on improving LFQ performance of cancer proteomics are discussed and evaluated

PROSPECTS

It is expected that incremental improvements of data acquisition techniques (DDA and DIA) and emerging of advanced computation methods (quantification tool and data processing) can significantly improve the proteomics analysis. Those tremendous advances discussed above could make MS-based proteomics more widely applied to identify diagnostic, prognostic, and therapeutic biomarkers for anticancer drug discovery.

LIST OF ABBREVIATIONS

AIF	=	All-ion Fragmentation
CYC	=	Cyclic Locally Weighted Regression
DIA / DDA	=	Data-independent / dependent Acquisition
KNN	=	K-nearest Neighbor
LFQ	=	Label-free Proteome Quantification
Log	=	Logarithmic
MS	=	mass spectrometry
MS1	=	the first stage of mass spectrometry
MS2	=	the second stage of mass spectrometry
RLE	=	Relative log expression
SVD	=	Singular Value Decomposition
TMM	=	Trimmed Mean of M Values
VSN	=	Variance Stabilization Normalization

CONSENT FOR PUBLICATION

Not applicable.

CONFLICT OF INTEREST

The author declares no conflict of interest, financial or otherwise.

ACKNOWLEDGEMENTS

Funded by the National Natural Science Foundation of China (81872798); Innovation Project on Industrial Generic Key Technologies of Chongqing (cstc2015zdcy-ztzx120003); and Fundamental Research Funds for Central Universities (10611CDJXZ238826, CDJZR14468801, CDJKXB14011, 2015CDJXY).

REFERENCES

[1] Santos R, Ursu O, Gaulton A, *et al.* A comprehensive map of molecular drug targets. Nat Rev Drug Discov 2017; 16(1): 19-34.

[2] Khan A, Ali A, Junaid M, *et al.* Identification of novel drug targets for diamond-blackfan anemia based on RPS19 gene mutation using protein-protein interaction network. BMC Syst Biol 2018; 12(Suppl. 4): 39.

[3] Li YH, Yu CY, Li XX, *et al.* Therapeutic target database update 2018: enriched resource for facilitating bench-to-clinic research of targeted therapeutics. Nucleic Acids Res 2018; 46(D1): D1121-7.

[4] Zeng X, Lin W, Guo M, Zou Q. A comprehensive overview and evaluation of circular RNA detection tools. PLOS Comput Biol 2017; 13(6): e1005420.

[5] Liu Y, Guo M. Chemical proteomic strategies for the discovery and development of anticancer drugs. Proteomics 2014; 14(4-5): 399-411.

[6] Distler U, Kuharev J, Navarro P, Tenzer S. Label-free quantification in ion mobility-enhanced data-independent acquisition proteomics. Nat Protoc 2016; 11(4): 795-812.

[7] Navarro P, Kuharev J, Gillet LC, *et al.* A multicenter study benchmarks software tools for label-free proteome quantification. Nat Biotechnol 2016; 34(11): 1130-6.

[8] Liu W, Yang X, Wang N, *et al.* Multiple immunosuppressive effects of CpG-c41 on intracellular TLR-mediated inflammation. Mediators Inflamm 2017; 2017: 6541729.

[9] Guo Y, Wang H, Li Y, *et al.* Genome of Helicobacter pylori strain XZ274, an isolate from a tibetan patient with gastric cancer in China. J Bacteriol 2012; 194(15): 4146-7.

[10] Schirle M, Bantscheff M, Kuster B. Mass spectrometry-based proteomics in preclinical drug discovery. Chem Biol 2012; 19(1): 72-84.

[11] Arora D, Chaudhary R, Singh A. System biology approach to identify potential receptor for targeting cancer and biomolecular interaction studies of indole[2, 1-a] isoquinoline derivative as anticancerous drug candidate against it. Interdiscip Sci 2017; doi: 10.1007/s12539-017-0249.

[12] Zhao M, Wei DQ. Rare Diseases: Drug Discovery and Informatics Resource. Interdiscip Sci 2018; 10(1): 195-204.

[13] Guo S, Zou J, Wang G. Advances in the proteomic discovery of novel therapeutic targets in cancer. Drug Des Devel Ther 2013; 7: 1259-71.

[14] Li YH, Wang PP, Li XX, *et al.* The human kinome targeted by FDA approved multi-target drugs and combination products: A

comparative study from the drug-target interaction network perspective. PLoS One 2016; 11(11): e0165737.

[15] Xu J, Wang P, Yang H, *et al.* Comparison of FDA approved kinase targets to clinical trial ones: Insights from their system profiles and drug-target interaction networks. BioMed Res Int 2016; 2016: 2509385.

[16] Kondo T. Current status of proteomics in Ewing's sarcoma. Proteomics Clin Appl 2018; 10: e1700130.

[17] Raghavendra NM, Pingili D, Kadasi S, Mettu A, Prasad SVUM. Dual or multi-targeting inhibitors: The next generation anticancer agents. Eur J Med Chem 2018; 143: 1277-300.

[18] Tao L, Zhu F, Xu F, Chen Z, Jiang YY, Chen YZ. Co-targeting cancer drug escape pathways confers clinical advantage for multi-target anticancer drugs. Pharmacol Res 2015; 102: 123-31.

[19] Zhu F, Zheng CJ, Han LY, *et al.* Trends in the exploration of anti-cancer targets and strategies in enhancing the efficacy of drug targeting. Curr Mol Pharmacol 2008; 1(3): 213-32.

[20] Ng PK, Li J, Jeong KJ, *et al.* Systematic functional annotation of somatic mutations in cancer. Cancer Cell 2018; 33(3): 450-462.e10.

[21] Lang JD, Hendricks WPD, Orlando KA, *et al.* Ponatinib shows potent antitumor activity in small cell carcinoma of the ovary hypercalcemic type (SCCOHT) through multikinase inhibition. Clin Cancer Res 2018; 24(8): 1932-43.

[22] Perna F, Berman SH, Soni RK, *et al.* Integrating Proteomics and Transcriptomics for Systematic Combinatorial Chimeric Antigen Receptor Therapy of AML. Cancer Cell 2017; 32(4): 506-519.e5.

[23] Zhu F, Li XX, Yang SY, Chen YZ. clinical success of drug targets prospectively predicted by in silico study. Trends Pharmacol Sci 2018; 39(3): 229-31.

[24] Tang W, Wan S, Yang Z, Teschendorff AE, Zou Q. Tumor origin detection with tissue-specific miRNA and DNA methylation markers. Bioinformatics 2018; 34(3): 398-406.

[25] Li B, Tang J, Yang Q, *et al.* Performance evaluation and online realization of data-driven normalization methods used in LC/MS based untargeted metabolomics analysis. Sci Rep 2016; 6: 38881.

[26] Chen L, Su W, Chen H, *et al.* Proteomics for biomarker identification and clinical application in kidney disease. Adv Clin Chem 2018; 85: 91-113.

[27] Zhu F, Shi Z, Qin C, *et al.* Therapeutic target database update 2012: A resource for facilitating target-oriented drug discovery. Nucleic Acids Res 2012; 40(Database issue): D1128-36.

[28] Singh S, Singh DB, Singh A, *et al.* An approach for identification of novel drug targets in streptococcus pyogenes SF370 through pathway analysis. Interdiscip Sci 2016; 8(4): 388-94.

[29] Ong SE, Mann M. Mass spectrometry-based proteomics turns quantitative. Nat Chem Biol 2005; 1(5): 252-62.

[30] Lichtman JS, Sonnenburg JL, Elias JE. Monitoring host responses to the gut microbiota. ISME J 2015; 9(9): 1908-15.

[31] Selvaraj G, Kaliamurthi S, Cakmak ZE, Cakmak T. In silico validation of microalgal metabolites against Diabetes mellitus. Diabetes Mellitus 2017; 20: 301-7.

[32] Li B, Tang J, Yang Q, *et al.* NOREVA: normalization and evaluation of MS-based metabolomics data. Nucleic Acids Res 2017; 45(W1): W162-70.

[33] Moulder R, Bhosale SD, Goodlett DR, Lahesmaa R. Analysis of the plasma proteome using iTRAQ and TMT-based Isobaric labeling. Mass Spectrom Rev 2018; 37(5): 583-606.

[34] Jia C, Zuo Y, Zou Q. O-GlcNAcPRED-II: An integrated classification algorithm for identifying O-GlcNAcylation sites based on fuzzy undersampling and a K-means PCA oversampling technique. Bioinformatics 2018; 34(12): 2029-36.

[35] Shen W, Li Y. A novel algorithm for detecting multiple covariance and clustering of biological sequences. Sci Rep 2016; 6: 30425.

[36] Zhang X, Ning Z, Mayne J, *et al.* In Vitro Metabolic Labeling of Intestinal Microbiota for Quantitative Metaproteomics. Anal Chem 2016; 88(12): 6120-5.

[37] Shen W, Le S, Li Y, Hu F. SeqKit: A Cross-Platform and Ultrafast Toolkit for FASTA/Q File Manipulation. PLoS One 2016; 11(10): e0163962.

[38] Selvaraj G, Kaliamurthi S, Cakmak ZE, Cakmak T. Computational screening of dipeptidyl peptidase IV inhibitors from micoroalgal metabolites by pharmacophore modeling and molecular docking. Phycological Res 2016; 64: 291-9.

[39] Cretu D, Prassas I, Saraon P, *et al.* Identification of psoriatic arthritis mediators in synovial fluid by quantitative mass spectrometry. Clin Proteomics 2014; 11(1): 27.

[40] Fu J, Tang J, Wang Y, *et al.* Discovery of the Consistently Well-Performed Analysis Chain for SWATH-MS Based Pharmacoproteomic Quantification. Front Pharmacol 2018; 9: 681.

[41] Kaliamurthi S, Selvaraj G, Thirugnasambandan R. Inhibitory Effect of Excoecaria Agallocha L. Extracts on Elastase and Collagenase and Identification of Metabolites Using HPLC-UV-MS Techniques. Pharm Chem J 2018; 51: 960-4.

[42] von Bergen M, Jehmlich N, Taubert M, *et al.* Insights from quantitative metaproteomics and protein-stable isotope probing into microbial ecology. ISME J 2013; 7(10): 1877-85.

[43] VerBerkmoes NC, Denef VJ, Hettich RL, Banfield JF. Systems biology: Functional analysis of natural microbial consortia using community proteomics. Nat Rev Microbiol 2009; 7(3): 196-205.

[44] Pan JB, Hu SC, Wang H, Zou Q, Ji ZL. PaGeFinder: quantitative identification of spatiotemporal pattern genes. Bioinformatics 2012; 28(11): 1544-5.

[45] Van Houdt WJ, Emmink BL, Pham TV, *et al.* Comparative proteomics of colon cancer stem cells and differentiated tumor cells identifies BIRC6 as a potential therapeutic target. Mol Cell Proteomics 2011; 10 M111.011353.

[46] Zeng X, Liu L, Lü L, Zou Q. Prediction of potential disease-associated microRNAs using structural perturbation method. Bioinformatics 2018; 34(14): 2425-32.

[47] Zou Q, Hu Q, Guo M, Wang G. HAlign: Fast multiple similar DNA/RNA sequence alignment based on the centre star strategy. Bioinformatics 2015; 31(15): 2475-81.

[48] Li Y, Chen M, Cao H, Zhu Y, Zheng J, Zhou H. Extraordinary GU-rich single-strand RNA identified from SARS coronavirus contributes an excessive innate immune response. Microbes Infect 2013; 15(2): 88-95.

[49] Gates LA, Gu G, Chen Y, *et al.* Proteomic profiling identifies key coactivators utilized by mutant ERα proteins as potential new therapeutic targets. Oncogene 2018; 37(33): 4581-98.

[50] Zeng X, Zhang X, Zou Q. Integrative approaches for predicting microRNA function and prioritizing disease-related microRNA using biological interaction networks. Brief Bioinform 2016; 17(2): 193-203.

[51] Cai Y, Wang N, Wu X, Zheng K, Li Y. Compensatory variances of drug-induced hepatitis B virus YMDD mutations. Springerplus 2016; 5(1): 1340.

[52] Shen W, Chen M, Wei G, Li Y. MicroRNA prediction using a fixed-order Markov model based on the secondary structure pattern. PLoS One 2012; 7(10): e48236.

[53] Li Z, Adams RM, Chourey K, Hurst GB, Hettich RL, Pan C. Systematic comparison of label-free, metabolic labeling, and isobaric chemical labeling for quantitative proteomics on LTQ Orbitrap Velos. J Proteome Res 2012; 11(3): 1582-90.

[54] Ahmad S, Navid A, Akhtar AS, Azam SS, Wadood A, Pérez-Sánchez H. Subtractive Genomics, Molecular Docking and Molecular Dynamics Simulation Revealed LpxC as a Potential Drug Target Against Multi-Drug Resistant Klebsiella pneumoniae. Interdiscip Sci 2018; doi: 10.1007/s12539-018-0299-y.

[55] Huang Q, Yang L, Luo J, *et al.* SWATH enables precise label-free quantification on proteome scale. Proteomics 2015; 15(7): 1215-23.

[56] Zhang XQ, Yuan JN, Selvaraj G, Ji GF, Chen XR, Wei DQ. Towards the low-sensitive and high-energetic co-crystal explosive CL-20/TNT: from intermolecular interactions to structures and properties. Phys Chem Chem Phys 2018; 20(25): 17253-61.

[57] Selvaraj G, Kaliamurthi S, Thirugnasambandan R. Effect of Glycosin alkaloid from Rhizophora apiculata in non-insulin dependent diabetic rats and its mechanism of action: In vivo and in silico studies. Phytomedicine 2016; 23(6): 632-40.

[58] Kaliamurthi S, Selvaraj G, Cakmak ZE, Cakmak T. Production and characterization of spherical thermostable silver nanoparticles from Spirulina platensis (Cyanophyceae). Phycologia 2016; 55: 568-76.

[59] Geyer PE, Kulak NA, Pichler G, Holdt LM, Teupser D, Mann M. Plasma Proteome Profiling to Assess Human Health and Disease. Cell Syst 2016; 2(3): 185-95.

[60] Piehowski PD, Petyuk VA, Orton DJ, *et al.* Sources of technical variability in quantitative LC-MS proteomics: human brain tissue sample analysis. J Proteome Res 2013; 12(5): 2128-37.

[61] Barschke P, Oeckl P, Steinacker P, Ludolph A, Otto M. Proteomic studies in the discovery of cerebrospinal fluid biomarkers for amyotrophic lateral sclerosis. Expert Rev Proteomics 2017; 14(9): 769-77.

[62] Ahlf DR, Thomas PM, Kelleher NL. Developing top down proteomics to maximize proteome and sequence coverage from cells and tissues. Curr Opin Chem Biol 2013; 17(5): 787-94.

[63] Ntai I, LeDuc RD, Fellers RT, *et al.* Integrated Bottom-Up and Top-Down Proteomics of Patient-Derived Breast Tumor Xenografts. Mol Cell Proteomics 2016; 15(1): 45-56.

[64] Nie S, McDermott SP, Deol Y, Tan Z, Wicha MS, Lubman DM. A quantitative proteomics analysis of MCF7 breast cancer stem and progenitor cell populations. Proteomics 2015; 15(22): 3772-83.

[65] Suman S, Basak T, Gupta P, *et al.* Quantitative proteomics revealed novel proteins associated with molecular subtypes of breast cancer. J Proteomics 2016; 148: 183-93.

[66] Pozniak Y, Balint-Lahat N, Rudolph JD, *et al.* System-wide Clinical Proteomics of Breast Cancer Reveals Global Remodeling of Tissue Homeostasis. Cell Syst 2016; 2(3): 172-84.

[67] Nesvizhskii AI, Aebersold R. Interpretation of shotgun proteomic data: the protein inference problem. Mol Cell Proteomics 2005; 4(10): 1419-40.

[68] Fleisher B, Brown AN, Ait-Oudhia S. Application of pharmacometrics and quantitative systems pharmacology to cancer therapy: The example of luminal a breast cancer. Pharmacol Res 2017; 124: 20-33.

[69] Tran JC, Zamdborg L, Ahlf DR, *et al.* Mapping intact protein isoforms in discovery mode using top-down proteomics. Nature 2011; 480(7376): 254-8.

[70] Cheon DH, Nam EJ, Park KH, *et al.* Comprehensive Analysis of Low-Molecular-Weight Human Plasma Proteome Using Top-Down Mass Spectrometry. J Proteome Res 2016; 15(1): 229-44.

[71] Chen Y, Hoover ME, Dang X, *et al.* Quantitative Mass Spectrometry Reveals that Intact Histone H1 Phosphorylations are Variant Specific and Exhibit Single Molecule Hierarchical Dependence. Mol Cell Proteomics 2016; 15(3): 818-33.

[72] Yang F, Zheng G, Fu T, *et al.* Prediction of the binding mode and resistance profile for a dual-target pyrrolyl diketo acid scaffold against HIV-1 integrase and reverse-transcriptase-associated ribonuclease H. Phys Chem Chem Phys 2018; 20(37): 23873-84.

[73] Välikangas T, Suomi T, Elo LL. A comprehensive evaluation of popular proteomics software workflows for label-free proteome quantification and imputation. Brief Bioinform 2017; 19(6): 1344-55.

[74] Nikolov M, Schmidt C, Urlaub H. Quantitative mass spectrometry-based proteomics: An overview. Methods Mol Biol 2012; 893: 85-100.

[75] Arike L, Peil L. Spectral counting label-free proteomics. Methods Mol Biol 2014; 1156: 213-22.

[76] Egertson JD, Kuehn A, Merrihew GE, *et al.* Multiplexed MS/MS for improved data-independent acquisition. Nat Methods 2013; 10(8): 744-6.

[77] Tsou CC, Avtonomov D, Larsen B, *et al.* DIA-Umpire: comprehensive computational framework for data-independent acquisition proteomics. Nat Methods 2015; 12(3): 258-264, 7, 264.

[78] Bilbao A, Varesio E, Luban J, *et al.* Processing strategies and software solutions for data-independent acquisition in mass spectrometry. Proteomics 2015; 15(5-6): 964-80.

[79] Rardin MJ, Schilling B, Cheng LY, *et al.* MS1 Peptide Ion Intensity Chromatograms in MS2 (SWATH) Data Independent Acquisitions. Improving Post Acquisition Analysis of Proteomic Experiments. Mol Cell Proteomics 2015; 14(9): 2405-19.

[80] Asara JM, Christofk HR, Freimark LM, Cantley LC. A label-free quantification method by MS/MS TIC compared to SILAC and spectral counting in a proteomics screen. Proteomics 2008; 8(5): 994-9.

[81] America AH, Cordewener JH. Comparative LC-MS: A landscape of peaks and valleys. Proteomics 2008; 8(4): 731-49.

[82] Fu T, Zheng G, Tu G, *et al.* Exploring the Binding Mechanism of Metabotropic Glutamate Receptor 5 Negative Allosteric Modulators in Clinical Trials by Molecular Dynamics Simulations. ACS Chem Neurosci 2018; 9(6): 1492-502.

[83] Bantscheff M, Schirle M, Sweetman G, Rick J, Kuster B. Quantitative mass spectrometry in proteomics: A critical review. Anal Bioanal Chem 2007; 389(4): 1017-31.

[84] Geis-Asteggiante L, Ostrand-Rosenberg S, Fenselau C, Edwards NJ. Evaluation of Spectral Counting for Relative Quantitation of Proteoforms in Top-Down Proteomics. Anal Chem 2016; 88(22): 10900-7.

[85] Xue W, Yang F, Wang P, *et al.* What Contributes to Serotonin-Norepinephrine Reuptake Inhibitors' Dual-Targeting Mechanism? The Key Role of Transmembrane Domain 6 in Human Serotonin and Norepinephrine Transporters Revealed by Molecular Dynamics Simulation. ACS Chem Neurosci 2018; 9(5): 1128-40.

[86] Mueller LN, Brusniak MY, Mani DR, Aebersold R. An assessment of software solutions for the analysis of mass spectrometry based quantitative proteomics data. J Proteome Res 2008; 7(1): 51-61.

[87] Bantscheff M, Lemeer S, Savitski MM, Kuster B. Quantitative mass spectrometry in proteomics: critical review update from 2007 to the present. Anal Bioanal Chem 2012; 404(4): 939-65.

[88] Grossmann J, Roschitzki B, Panse C, *et al.* Implementation and evaluation of relative and absolute quantification in shotgun proteomics with label-free methods. J Proteomics 2010; 73(9): 1740-6.

[89] Li S, Cao Q, Xiao W, *et al.* Optimization of Acquisition and Data-Processing Parameters for Improved Proteomic Quantification by Sequential Window Acquisition of All Theoretical Fragment Ion Mass Spectrometry. J Proteome Res 2017; 16(2): 738-47.

[90] Anjo SI, Santa C, Manadas B. SWATH-MS as a tool for biomarker discovery: From basic research to clinical applications. Proteomics 2017; 17(3-4)

[91] Zhu F, Han LY, Chen X, *et al.* Homology-free prediction of functional class of proteins and peptides by support vector machines. Curr Protein Pept Sci 2008; 9(1): 70-95.

[92] Aratyn-Schaus Y, Ramanathan R. Advances in high-resolution MS and hepatocyte models solve a long-standing metabolism challenge: the loratadine story. Bioanalysis 2016; 8(16): 1645-62.

[93] Jia J, Zhu F, Ma X, *et al.* Mechanisms of drug combinations: interaction and network perspectives. Nat Rev Drug Discov 2009; 8(2): 111-28.

[94] Krasny L, Bland P, Kogata N, *et al.* SWATH mass spectrometry as a tool for quantitative profiling of the matrisome. J Proteomics 2018; pii: S1874-3919(18): 30083-6.

[95] Roemmelt AT, Steuer AE, Kraemer T. Liquid chromatography, in combination with a quadrupole time-of-flight instrument, with sequential window acquisition of all theoretical fragment-ion spectra acquisition: validated quantification of 39 antidepressants in whole blood as part of a simultaneous screening and quantification procedure. Anal Chem 2015; 87(18): 9294-301.

[96] Xue W, Wang P, Li B, *et al.* Identification of the inhibitory mechanism of FDA approved selective serotonin reuptake inhibitors: An insight from molecular dynamics simulation study. Phys Chem Chem Phys 2016; 18(4): 3260-71.

[97] Shi T, Song E, Nie S, *et al.* Advances in targeted proteomics and applications to biomedical research. Proteomics 2016; 16(15-16): 2160-82.

[98] Gillet LC, Navarro P, Tate S, *et al.* Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: A new concept for consistent and accurate proteome analysis. Mol Cell Proteomics 2012; 11 O111.016717

[99] Jamwal R, Barlock BJ, Adusumalli S, Ogasawara K, Simons BL, Akhlaghi F. Multiplex and Label-Free Relative Quantification Approach for Studying Protein Abundance of Drug Metabolizing Enzymes in Human Liver Microsomes Using SWATH-MS. J Proteome Res 2017; 16(11): 4134-43.

[100] Zhu F, Han L, Zheng C, *et al.* What are next generation innovative therapeutic targets? Clues from genetic, structural, physicochemical, and systems profiles of successful targets. J Pharmacol Exp Ther 2009; 330(1): 304-15.

[101] Cox J, Mann M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat Biotechnol 2008; 26(12): 1367-72.

[102] Tyanova S, Temu T, Cox J. The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. Nat Protoc 2016; 11(12): 2301-19.
[103] Tyanova S, Temu T, Carlson A, Sinitcyn P, Mann M, Cox J. Visualization of LC-MS/MS proteomics data in MaxQuant. Proteomics 2015; 15(8): 1453-6.
[104] Weisser H, Nahnsen S, Grossmann J, *et al.* An automated pipeline for high-throughput label-free quantitative proteomics. J Proteome Res 2013; 12(4): 1628-44.
[105] Iglesias-Gato D, Wikström P, Tyanova S, *et al.* The Proteome of Primary Prostate Cancer. Eur Urol 2016; 69(5): 942-52.
[106] Zhu F, Han B, Kumar P, *et al.* Update of TTD: Therapeutic Target Database. Nucleic Acids Res 2010; 38(Database issue): D787-91.
[107] Gautier V, Mouton-Barbosa E, Bouyssié D, *et al.* Label-free quantification and shotgun analysis of complex proteomes by one-dimensional SDS-PAGE/NanoLC-MS: evaluation for the large scale analysis of inflammatory human endothelial cells. Mol Cell Proteomics 2012; 11(8): 527-39.
[108] Bouyssié D, Gonzalez de Peredo A, Mouton E, *et al.* Mascot file parsing and quantification (MFPaQ), a new software to parse, validate, and quantify proteomics data generated by ICAT and SILAC mass spectrometric analyses: Application to the proteomics study of membrane proteins from primary human endothelial cells. Mol Cell Proteomics 2007; 6(9): 1621-37.
[109] Hoedt E, Chaoui K, Huvent I, *et al.* SILAC-based proteomic profiling of the human MDA-MB-231 metastatic breast cancer cell line in response to the two antitumoral lactoferrin isoforms: the secreted lactoferrin and the intracellular delta-lactoferrin. PLoS One 2014; 9(8): e104563.
[110] Zhu F, Ma XH, Qin C, *et al.* Drug discovery prospect from untapped species: indications from approved natural product drugs. PLoS One 2012; 7(7): e39782.
[111] Codrea MC, Nahnsen S. Platforms and Pipelines for Proteomics Data Analysis and Management. Adv Exp Med Biol 2016; 919: 203-15.
[112] Vidotto A, Morais AT, Ribeiro MR, *et al.* Systems Biology Reveals NS4B-Cyclophilin A Interaction: A New Target to Inhibit YFV Replication. J Proteome Res 2017; 16(4): 1542-55.
[113] Twigt JM, Bezstarosti K, Demmers J, Lindemans J, Laven JS, Steegers-Theunissen RP. Preconception folic acid use influences the follicle fluid proteome. Eur J Clin Invest 2015; 45(8): 833-41.
[114] Dorts J, Kestemont P, Thézenas ML, Raes M, Silvestre F. Effects of cadmium exposure on the gill proteome of Cottus gobio: modulatory effects of prior thermal acclimation. Aquat Toxicol 2014; 154: 87-96.
[115] Tsou CC, Tsai CF, Teo GC, Chen YJ, Nesvizhskii AI. Untargeted, spectral library-free analysis of data-independent acquisition proteomics data generated using Orbitrap mass spectrometers. Proteomics 2016; 16(15-16): 2257-71.
[116] Bruderer R, Bernhardt OM, Gandhi T, *et al.* Extending the limits of quantitative proteome profiling with data-independent acquisition and application to acetaminophen-treated three-dimensional liver microtissues. Mol Cell Proteomics 2015; 14(5): 1400-10.
[117] Kreimer S, Gao Y, Ray S, *et al.* Host Cell Protein Profiling by Targeted and Untargeted Analysis of Data Independent Acquisition Mass Spectrometry Data with Parallel Reaction Monitoring Verification. Anal Chem 2017; 89(10): 5294-302.
[118] Wu L, Amon S, Lam H. A hybrid retention time alignment algorithm for SWATH-MS data. Proteomics 2016; 16(15-16): 2272-83.
[119] Röst HL, Rosenberger G, Navarro P, *et al.* OpenSWATH enables automated, targeted analysis of data-independent acquisition MS data. Nat Biotechnol 2014; 32(3): 219-23.
[120] Rosenberger G, Liu Y, Röst HL, *et al.* Inference and quantification of peptidoforms in large sample cohorts by SWATH-MS. Nat Biotechnol 2017; 35(8): 781-8.
[121] Wu JX, Song X, Pascovici D, *et al.* SWATH Mass Spectrometry Performance Using Extended Peptide MS/MS Assay Libraries. Mol Cell Proteomics 2016; 15(7): 2501-14.
[122] Liu Y, Hüttenhain R, Surinova S, *et al.* Quantitative measurements of N-linked glycoproteins in human plasma by SWATH-MS. Proteomics 2013; 13(8): 1247-56.
[123] Shao S, Guo T, Koh CC, *et al.* Minimal sample requirement for highly multiplexed protein quantification in cell lines and tissues by PCT-SWATH mass spectrometry. Proteomics 2015; 15(21): 3711-21.
[124] Gao Y, Lim TK, Lin Q, Li SF. Evaluation of sample extraction methods for proteomics analysis of green algae Chlorella vulgaris. Electrophoresis 2016; 37(10): 1270-6.
[125] MacLean B, Tomazela DM, Shulman N, *et al.* Skyline: An open source document editor for creating and analyzing targeted proteomics experiments. Bioinformatics 2010; 26(7): 966-8.
[126] Escher C, Reiter L, MacLean B, *et al.* Using iRT, a normalized retention time for more targeted measurement of peptides. Proteomics 2012; 12(8): 1111-21.
[127] Zhang Y, Bilbao A, Bruderer T, *et al.* The Use of Variable Q1 Isolation Windows Improves Selectivity in LC-SWATH-MS Acquisition. J Proteome Res 2015; 14(10): 4359-71.
[128] Han ZJ, Xue WW, Tao L, Zhu F. Identification of novel immune-relevant drug target genes for Alzheimer's Disease by combining ontology inference with network analysis. CNS Neurosci Ther 2018; 24(12): 1253-63.
[129] Schilling B, Rardin MJ, MacLean BX, *et al.* Platform-independent and label-free quantitation of proteomic data using MS1 extracted ion chromatograms in skyline: Application to protein acetylation and phosphorylation. Mol Cell Proteomics 2012; 11(5): 202-14.
[130] Chen R, Pan S, Yi EC, *et al.* Quantitative proteomic profiling of pancreatic cancer juice. Proteomics 2006; 6(13): 3871-9.
[131] Lin F, Li Z, Hua Y, Lim YP. Proteomic profiling predicts drug response to novel targeted anticancer therapeutics. Expert Rev Proteomics 2016; 13(4): 411-20.
[132] Bruderer R, Bernhardt OM, Gandhi T, Reiter L. High-precision iRT prediction in the targeted analysis of data-independent acquisition and its impact on identification and quantitation. Proteomics 2016; 16(15-16): 2246-56.
[133] Röst HL, Sachsenberg T, Aiche S, *et al.* OpenMS: A flexible open-source software platform for mass spectrometry data analysis. Nat Methods 2016; 13(9): 741-8.
[134] Schlaffner CN, Pirklbauer GJ, Bender A, Choudhary JS. Fast, Quantitative and Variant Enabled Mapping of Peptides to Genomes. Cell Syst 2017; 5(2): 152-156.e4.
[135] Khoonsari PE, Häggmark A, Lönnberg M, *et al.* Analysis of the Cerebrospinal Fluid Proteome in Alzheimer's Disease. PLoS One 2016; 11(3): e0150672.
[136] Weston AJ, Dunlap WC, Shick JM, *et al.* A profile of an endosymbiont-enriched fraction of the coral Stylophora pistillata reveals proteins relevant to microbial-host interactions. Mol Cell Proteomics 2012; 11 M111.015487
[137] Surinova S, Choi M, Tao S, *et al.* Prediction of colorectal cancer diagnosis based on circulating plasma proteins. EMBO Mol Med 2015; 7(9): 1166-78.
[138] Ma B, Zhang K, Hendrie C, *et al.* PEAKS: powerful software for peptide de novo sequencing by tandem mass spectrometry. Rapid Commun Mass Spectrom 2003; 17(20): 2337-42.
[139] Rao HB, Zhu F, Yang GB, Li ZR, Chen YZ. Update of PROFEAT: A web server for computing structural and physicochemical features of proteins and peptides from amino acid sequence. Nucleic Acids Res 2011; 39(Web Server issue): W385-90.
[140] Zhang J, Xin L, Shan B, *et al.* PEAKS DB: de novo sequencing assisted database search for sensitive and accurate peptide identification. Mol Cell Proteomics 2012; 11 M111.010587
[141] Zhang J, Yang W, Li S, *et al.* An intelligentized strategy for endogenous small molecules characterization and quality evaluation of earthworm from two geographic origins by ultra-high performance HILIC/QTOF MS(E) and Progenesis QI. Anal Bioanal Chem 2016; 408(14): 3881-90.
[142] Al Shweiki MR, Mönchgesang S, Majovsky P, Thieme D, Trutschel D, Hoehenwarter W. Assessment of Label-Free Quantification in Discovery Proteomics and Impact of Technological Factors and Natural Variability of Protein Abundance. J Proteome Res 2017; 16(4): 1410-24.
[143] Almeida AM, Nanni P, Ferreira AM, *et al.* The longissimus thoracis muscle proteome in Alentejana bulls as affected by growth path. J Proteomics 2017; 152: 206-15.
[144] Timms JF, Arslan-Low E, Kabir M, *et al.* Discovery of serum biomarkers of ovarian cancer using complementary proteomic profiling strategies. Proteomics Clin Appl 2014; 8(11-12): 982-93.

[145] MacKeigan JP, Clements CM, Lich JD, Pope RM, Hod Y, Ting JP. Proteomic profiling drug-induced apoptosis in non-small cell lung carcinoma: identification of RS/DJ-1 and RhoGDIalpha. Cancer Res 2003; 63(20): 6928-34.

[146] Zhu F, Qin C, Tao L, *et al.* Clustered patterns of species origins of nature-derived drugs and clues for future bioprospecting. Proc Natl Acad Sci USA 2011; 108(31): 12943-8.

[147] Gärdén P, Alm R, Häkkinen J. PROTEIOS: An open source proteomics initiative. Bioinformatics 2005; 21(9): 2085-7.

[148] Häkkinen J, Vincic G, Månsson O, Wårell K, Levander F. The proteios software environment: An extensible multiuser platform for management and analysis of proteomics data. J Proteome Res 2009; 8(6): 3037-43.

[149] Levander F, Krogh M, Wårell K, Gärdén P, James P, Häkkinen J. Automated reporting from gel-based proteomics experiments using the open source Proteios database application. Proteomics 2007; 7(5): 668-74.

[150] Olsson N, Carlsson P, James P, *et al.* Grading breast cancer tissues using molecular portraits. Mol Cell Proteomics 2013; 12(12): 3612-23.

[151] Li YH, Xu JY, Tao L, *et al.* SVM-Prot 2016: A Web-Server for Machine Learning Prediction of Protein Functional Families from Sequence Irrespective of Similarity. PLoS One 2016; 11(8): e0155290.

[152] Colaert N, Barsnes H, Vaudel M, *et al.* Thermo-msf-parser: An open source Java library to parse and visualize Thermo Proteome Discoverer msf files. J Proteome Res 2011; 10(8): 3840-3.

[153] Wang P, Yang F, Yang H, *et al.* Identification of dual active agents targeting 5-HT1A and SERT by combinatorial virtual screening methods. Biomed Mater Eng 2015; 26(Suppl. 1): S2233-9.

[154] Tao L, Zhu F, Qin C, *et al.* Nature's contribution to today's pharmacopeia. Nat Biotechnol 2014; 32(10): 979-80.

[155] Veit J, Sachsenberg T, Chernev A, Aicheler F, Urlaub H, Kohlbacher O. LFQProfiler and RNP(xl): Open-Source Tools for Label-Free Quantification and Protein-RNA Cross-Linking Integrated into Proteome Discoverer. J Proteome Res 2016; 15(9): 3441-8.

[156] Casado-Vela J, Martínez-Esteso MJ, Rodriguez E, Borrás E, Elortza F, Bru-Martínez R. iTRAQ-based quantitative analysis of protein mixtures with large fold change and dynamic range. Proteomics 2010; 10(2): 343-7.

[157] Yan JF, Kim H, Jeong SK, *et al.* Integrated Proteomic and Genomic Analysis of Gastric Cancer Patient Tissues. J Proteome Res 2015; 14(12): 4995-5006.

[158] Fermin D, Basrur V, Yocum AK, Nesvizhskii AI. Abacus: A computational tool for extracting and pre-processing spectral count data for label-free quantitative proteomic analysis. Proteomics 2011; 11(7): 1340-5.

[159] Zheng G, Xue W, Wang P, *et al.* Exploring the Inhibitory Mechanism of Approved Selective Norepinephrine Reuptake Inhibitors and Reboxetine Enantiomers by Molecular Dynamics Study. Sci Rep 2016; 6: 26883.

[160] Zhao L, Lee BY, Brown DA, *et al.* Identification of candidate biomarkers of therapeutic response to docetaxel by proteomic profiling. Cancer Res 2009; 69(19): 7696-703.

[161] Park SK, Yates JR III. Census for proteome quantification. Curr Protoc Bioinformatics 2010; Chapter 13: 1-11.

[162] Prieto JH, Koncarevic S, Park SK, Yates J III, Becker K. Large-scale differential proteome analysis in Plasmodium falciparum under drug treatment. PLoS One 2008; 3(12): e4098.

[163] Savas JN, Park SK, Yates JR III. Proteomic Analysis of Protein Turnover by Metabolic Whole Rodent Pulse-Chase Isotopic Labeling and Shotgun Mass Spectrometry Analysis. Methods Mol Biol 2016; 1410: 293-304.

[164] Wang P, Zhang X, Fu T, *et al.* Differentiating Physicochemical Properties between Addictive and Nonaddictive ADHD Drugs Revealed by Molecular Dynamics Simulation Studies. ACS Chem Neurosci 2017; 8(6): 1416-28.

[165] Yang FY, Fu TT, Zhang XY, *et al.* Comparison of computational model and X-ray crystal structure of human serotonin transporter: potential application for the pharmacology of human monoamine transporters. Mol Simul 2017; 43: 1089-98.

[166] Cociorva D, L Tabb D, Yates JR. Validation of tandem mass spectrometry database search results using DTASelect. Curr Protoc Bioinformatics 2007; Chapter 13: 4.

[167] Tabb DL, McDonald WH, Yates JR III. DTASelect and Contrast: tools for assembling and comparing protein identifications from shotgun proteomics. J Proteome Res 2002; 1(1): 21-6.

[168] Park GW, Hwang H, Kim KH, *et al.* Integrated Proteomic Pipeline Using Multiple Search Engines for a Proteogenomic Study with a Controlled Protein False Discovery Rate. J Proteome Res 2016; 15(11): 4082-90.

[169] Wan J, Roth AF, Bailey AO, Davis NG. Palmitoylated proteins: purification and identification. Nat Protoc 2007; 2(7): 1573-84.

[170] Yang H, Qin C, Li YH, *et al.* Therapeutic target database update 2016: enriched resource for bench to clinical drug target and targeted pathway information. Nucleic Acids Res 2016; 44(D1): D1069-74.

[171] Dupierris V, Masselon C, Court M, Kieffer-Jaquinod S, Bruley C. A toolbox for validation of mass spectrometry peptides identification and generation of database: IRMa. Bioinformatics 2009; 25(15): 1980-1.

[172] Ramus C, Hovasse A, Marcellin M, *et al.* Spiked proteomic standard dataset for testing label-free quantitative software and statistical methods. Data Brief 2015; 6: 286-94.

[173] Osseni A, Sébastien M, Sarrault O, *et al.* Triadin and CLIMP-63 form a link between triads and microtubules in muscle cells. J Cell Sci 2016; 129(20): 3744-55.

[174] Nesvizhskii AI, Keller A, Kolker E, Aebersold R. A statistical model for identifying proteins by tandem mass spectrometry. Anal Chem 2003; 75(17): 4646-58.

[175] Keller A, Nesvizhskii AI, Kolker E, Aebersold R. Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. Anal Chem 2002; 74(20): 5383-92.

[176] Shraibman B, Barnea E, Melamed Kadosh D, *et al.* Identification of tumor antigens among the HLA peptidomes of Glioblastoma tumors and plasma. Mol Cell Proteomics 2018. pii: mcp.RA118.000792

[177] Duhamel M, Rodet F, Delhem N, *et al.* Molecular Consequences of Proprotein Convertase 1/3 (PC1/3) Inhibition in Macrophages for Application to Cancer Immunotherapy: A Proteomic Study. Mol Cell Proteomics 2015; 14(11): 2857-77.

[178] Johnson PE, Sayers RL, Gethings LA, *et al.* Quantitative Proteomic Profiling of Peanut Allergens in Food Ingredients Used for Oral Food Challenges. Anal Chem 2016; 88(11): 5689-95.

[179] Boffa DJ, Graf RP, Salazar MC, *et al.* Cellular Expression of PD-L1 in the Peripheral Blood of Lung Cancer Patients is Associated with Worse Survival. Cancer Epidemiol Biomarkers Prev 2017; 26(7): 1139-45.

[180] de Roos B. Proteomic analysis of human plasma and blood cells in nutritional studies: development of biomarkers to aid disease prevention. Expert Rev Proteomics 2008; 5(6): 819-26.

[181] Xia J, Wishart DS. Web-based inference of biological patterns, functions and pathways from metabolomic data using MetaboAnalyst. Nat Protoc 2011; 6(6): 743-60.

[182] Field EA, Price CJ, Sleet RB, Marr MC, Schwetz BA, Morrissey RE. Developmental toxicity evaluation of acrylamide in rats and mice. Fundam Appl Toxicol 1990; 14(3): 502-12.

[183] Warton DI, Hui FK. The arcsine is asinine: the analysis of proportions in ecology. Ecology 2011; 92(1): 3-10.

[184] Chen K, Cheng Y, Berkout O, Lindhiem O. Analyzing Proportion Scores as Outcomes for Prevention Trials: A Statistical Primer. Prev Sci 2017; 18(3): 312-21.

[185] Wang P, Fu T, Zhang X, *et al.* Differentiating physicochemical properties between NDRIs and sNRIs clinically important for the treatment of ADHD. Biochim Biophys Acta, Gen Subj 2017; 1861(11 Pt A): 2766-77.

[186] Sarhan AR, Patel TR, Cowell AR, *et al.* LAR protein tyrosine phosphatase regulates focal adhesions through CDK1. J Cell Sci 2016; 129(15): 2962-71.

[187] Zilli L, Beirão J, Schiavone R, Herraez MP, Gnoni A, Vilella S. Comparative proteome analysis of cryopreserved flagella and head plasma membrane proteins from sea bream spermatozoa: effect of antifreeze proteins. PLoS One 2014; 9(6): e99992.

[188] Sakia RM. The Box-Cox Transformation Technique - a Review. Statistician 1992; 41: 169-78.
[189] Kohl M, Megger DA, Trippler M, *et al.* A practical data processing workflow for multi-OMICS projects. Biochim Biophys Acta 2014; 1844(1 Pt A): 52-62.
[190] Raji Reddy C, Rani Valleti R, Dilipkumar U. One-Pot Sequential Propargylation/Cycloisomerization: A Facile [4+2]-Benzannulation Approach to Carbazoles. Chemistry 2016; 22(7): 2501-6.
[191] Coombes KR, Tsavachidis S, Morris JS, Baggerly KA, Hung MC, Kuerer HM. Improved peak detection and quantification of mass spectrometry data acquired from surface-enhanced laser desorption and ionization by denoising spectra with the undecimated discrete wavelet transform. Proteomics 2005; 5(16): 4107-17.
[192] De Livera AM, Dias DA, De Souza D, *et al.* Normalizing and integrating metabolomics data. Anal Chem 2012; 84(24): 10768-76.
[193] Wiśniewski JR, Ostasiewicz P, Duś K, Zielińska DF, Gnad F, Mann M. Extensive quantitative remodeling of the proteome between normal colon tissue and adenocarcinoma. Mol Syst Biol 2012; 8: 611.
[194] Li L, Wu J, Ghosh JK, Ritz B. Estimating Spatiotemporal Variability of Ambient Air Pollutant Concentrations with A Hierarchical Model. Atmos Environ (1994) 2013; 71(71): 54-63.
[195] Zheng G, Xue W, Yang F, *et al.* Revealing vilazodone's binding mechanism underlying its partial agonism to the 5-HT_{1A} receptor in the treatment of major depressive disorder. Phys Chem Chem Phys 2017; 19(42): 28885-96.
[196] Liu Y, Hu H, Wang K, *et al.* Multidimensional analysis of gene expression reveals TGFB1I1-induced EMT contributes to malignant progression of astrocytomas. Oncotarget 2014; 5(24): 12593-606.
[197] Dicker L, Lin X, Ivanov AR. Increased power for the analysis of label-free LC-MS/MS proteomics data by combining spectral counts and peptide peak attributes. Mol Cell Proteomics 2010; 9(12): 2704-18.
[198] Välikangas T, Suomi T, Elo LL. A systematic evaluation of normalization methods in quantitative label-free proteomics. Brief Bioinform 2018; 19(1): 1-11.
[199] Kohl SM, Klein MS, Hochrein J, Oefner PJ, Spang R, Gronwald W. State-of-the art data normalization methods improve NMR-based metabolomic analysis. Metabolomics 2012; 8(Suppl. 1): 146-60.
[200] Gromski PS, Xu Y, Hollywood KA, Turner ML, Goodacre R. The influence of scaling metabolomics data on model classification accuracy. Metabolomics 2015; 11: 684-95.
[201] Reindl J, Pesek J, Krüger T, *et al.* Proteomic biomarkers for psoriasis and psoriasis arthritis. J Proteomics 2016; 140: 55-61.
[202] Nezami Ranjbar MR, Zhao Y, Tadesse MG, Wang Y, Ressom HW. Gaussian process regression model for normalization of LC-MS data using scan-level information. Proteome Sci 2013; 11(Suppl. 1): S13.
[203] Webb-Robertson BJ, Kim YM, Zink EM, *et al.* A Statistical Analysis of the Effects of Urease Pre-treatment on the Measurement of the Urinary Metabolome by Gas Chromatography-Mass Spectrometry. Metabolomics 2014; 10(5): 897-908.
[204] Ballman KV, Grill DE, Oberg AL, Therneau TM. Faster cyclic loess: normalizing RNA arrays via linear models. Bioinformatics 2004; 20(16): 2778-86.
[205] Keeping AJ, Collins RA. Data variance and statistical significance in 2D-gel electrophoresis and DIGE experiments: comparison of the effects of normalization methods. J Proteome Res 2011; 10(3): 1353-60.
[206] Xue W, Wang P, Tu G, *et al.* Computational identification of the binding mechanism of a triple reuptake inhibitor amitifadine for the treatment of major depressive disorder. Phys Chem Chem Phys 2018; 20(9): 6606-16.
[207] Valikangas T, Suomi T, Elo LL. A systematic evaluation of normalization methods in quantitative label-free proteomics. Brief Bioinform 2016.
[208] Karpievitch YV, Nikolic SB, Wilson R, Sharman JE, Edwards LM. Metabolomics data normalization with EigenMS. PLoS One 2014; 9(12): e116221.
[209] Karpievitch YV, Dabney AR, Smith RD. Normalization and missing value imputation for label-free LC-MS analysis. BMC Bioinformatics 2012; 13(Suppl. 16): S5.
[210] Karpievitch YV, Polpitiya AD, Anderson GA, Smith RD, Dabney AR. Liquid Chromatography Mass Spectrometry-Based Proteomics: Biological and Technological Aspects. Ann Appl Stat 2010; 4(4): 1797-823.
[211] Yang YH, Dudoit S, Luu P, Speed TP. Normalization for cDNA microarray data. Microarrays. Optical Technologies And Informatics 2001; 4266: 141-52.
[212] Andjelkovic V, Thompson R. Changes in gene expression in maize kernel in response to water and salt stress. Plant Cell Rep 2006; 25(1): 71-9.
[213] Ejigu BA, Valkenborg D, Baggerman G, *et al.* Evaluation of normalization methods to pave the way towards large-scale LC-MS-based metabolomics profiling experiments. OMICS 2013; 17(9): 473-85.
[214] Padoan A, Basso D, La Malfa M, *et al.* Reproducibility in urine peptidome profiling using MALDI-TOF. Proteomics 2015; 15(9): 1476-85.
[215] Yu CY, Li XX, Yang H, *et al.* Assessing the Performances of Protein Function Prediction Algorithms from the Perspectives of Identification Accuracy and False Discovery Rate. Int J Mol Sci 2018; 19(1): 19.
[216] Bolstad BM, Irizarry RA, Astrand M, Speed TP. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. Bioinformatics 2003; 19(2): 185-93.
[217] Callister SJ, Barry RC, Adkins JN, *et al.* Normalization approaches for removing systematic biases associated with mass spectrometry and label-free proteomics. J Proteome Res 2006; 5(2): 277-86.
[218] Matzke MM, Waters KM, Metz TO, *et al.* Improved quality control processing of peptide-centric LC-MS proteomics data. Bioinformatics 2011; 27(20): 2866-72.
[219] van den Berg RA, Hoefsloot HC, Westerhuis JA, Smilde AK, van der Werf MJ. Centering, scaling, and transformations: improving the biological information content of metabolomics data. BMC Genomics 2006; 7: 142.
[220] Dieterle F, Ross A, Schlotterbeck G, Senn H. Probabilistic quotient normalization as robust method to account for dilution of complex biological mixtures. Application in 1H NMR metabonomics. Anal Chem 2006; 78(13): 4281-90.
[221] López-Fernández H, Santos HM, Capelo JL, Fdez-Riverola F, Glez-Peña D, Reboiro-Jato M. Mass-Up: An all-in-one open software application for MALDI-TOF mass spectrometry knowledge discovery. BMC Bioinformatics 2015; 16: 318.
[222] Hong MG, Lee W, Nilsson P, Pawitan Y, Schwenk JM. Multidimensional normalization to minimize plate effects of suspension bead array data. J Proteome Res 2016; 15(10): 3473-80.
[223] Gaspari M, Chiesa L, Nicastri A, *et al.* Proteome Speciation by Mass Spectrometry: Characterization of Composite Protein Mixtures in Milk Replacers. Anal Chem 2016; 88(23): 11568-74.
[224] Borgaonkar SP, Hocker H, Shin H, Markey MK. Comparison of normalization methods for the identification of biomarkers using MALDI-TOF and SELDI-TOF mass spectra. OMICS 2010; 14(1): 115-26.
[225] Lin Y, Golovnina K, Chen ZX, *et al.* Comparison of normalization and differential expression analyses using RNA-Seq data from 726 individual Drosophila melanogaster. BMC Genomics 2016; 17: 28.
[226] Huber W, von Heydebreck A, Sültmann H, Poustka A, Vingron M. Variance stabilization applied to microarray data calibration and to the quantification of differential expression. Bioinformatics 2002; 18(Suppl. 1): S96-S104.
[227] Kultima K, Nilsson A, Scholz B, Rossbach UL, Fälth M, Andrén PE. Development and evaluation of normalization methods for label-free relative quantification of endogenous peptides. Mol Cell Proteomics 2009; 8(10): 2285-95.
[228] Ghule PN, Xie RL, Colby JL, *et al.* p53 checkpoint ablation exacerbates the phenotype of Hinfp dependent histone H4 deficiency. Cell Cycle 2015; 14(15): 2501-8.
[229] Cheadle C, Vawter MP, Freed WJ, Becker KG. Analysis of microarray data using Z score transformation. J Mol Diagn 2003; 5(2): 73-81.

[230] Webb-Robertson BJ, Matzke MM, Jacobs JM, Pounds JG, Waters KM. A statistical selection strategy for normalization procedures in LC-MS proteomics experiments through dataset-dependent ranking of normalization scaling factors. Proteomics 2011; 11(24): 4736-41.

[231] Saha S, Dazard JE, Xu H, Ewing RM. Computational framework for analysis of prey-prey associations in interaction proteomics identifies novel human protein-protein interactions and networks. J Proteome Res 2012; 11(9): 4476-87.

[232] Karpievitch YV, Taverner T, Adkins JN, *et al.* Normalization of peak intensities in bottom-up MS-based proteomics using singular value decomposition. Bioinformatics 2009; 25(19): 2573-80.

[233] Chai LE, Law CK, Mohamad MS, *et al.* Investigating the effects of imputation methods for modelling gene networks using a dynamic bayesian network from gene expression data. Malays J Med Sci 2014; 21(2): 20-7.

[234] Pedreschi R, Hertog ML, Carpentier SC, *et al.* Treatment of missing values for multivariate statistical analysis of gel-based proteomics data. Proteomics 2008; 8(7): 1371-83.

[235] Koopmans F, Cornelisse LN, Heskes T, Dijkstra TM. Empirical Bayesian random censoring threshold model improves detection of differentially abundant proteins. J Proteome Res 2014; 13(9): 3871-80.

[236] Lin D, Zhang J, Li J, Xu C, Deng HW, Wang YP. An integrative imputation method based on multi-omics datasets. BMC Bioinformatics 2016; 17: 247.

[237] Kim H, Golub GH, Park H. Missing value estimation for DNA microarray gene expression data: local least squares imputation. Bioinformatics 2005; 21(2): 187-98.

[238] Wu WS, Jhou MJ. MVIAeval: A web tool for comprehensively evaluating the performance of a new missing value imputation algorithm. BMC Bioinformatics 2017; 18(1): 31.

[239] Alter O, Brown PO, Botstein D. Singular value decomposition for genome-wide expression data processing and modeling. Proc Natl Acad Sci USA 2000; 97(18): 10101-6.

[240] Gan X, Liew AW, Yan H. Microarray missing data imputation based on a set theoretic framework and biological knowledge. Nucleic Acids Res 2006; 34(5): 1608-19.

[241] Gatto L, Lilley KS. MSnbase-an R/Bioconductor package for isobaric tagged mass spectrometry data visualization, processing and quantitation. Bioinformatics 2012; 28(2): 288-9.