Current Drug Targets, 2020, *21*, 34-54

REVIEW ARTICLE

Advances in Current Diabetes Proteomics: From the Perspectives of Label-free Quantification and Biomarker Selection

Jianbo Fu[1,†], Yongchao Luo[1,†], Minjie Mou[1], Hongning Zhang[1], Jing Tang[1,2], Yunxia Wang[1] and Feng Zhu[1,2,*]

[1]*College of Pharmaceutical Sciences, Zhejiang University, Hangzhou 310058, China;* [2]*School of Pharmaceutical Sciences and Innovative Drug Research Centre, Chongqing University, Chongqing 401331, China*

ARTICLE HISTORY

Received: May 06, 2019
Revised: July 17, 2019
Accepted: July 24, 2019

DOI:
10.2174/1389450120666190821160207

Abstract: ***Background*:** Due to its prevalence and negative impacts on both the economy and society, the diabetes mellitus (DM) has emerged as a worldwide concern. In light of this, the label-free quantification (LFQ) proteomics and diabetic marker selection methods have been applied to elucidate the underlying mechanisms associated with insulin resistance, explore novel protein biomarkers, and discover innovative therapeutic protein targets.

***Objective*:** The purpose of this manuscript is to review and analyze the recent computational advances and development of label-free quantification and diabetic marker selection in diabetes proteomics.

***Methods*:** Web of Science database, PubMed database and Google Scholar were utilized for searching label-free quantification, computational advances, feature selection and diabetes proteomics.

***Results*:** In this study, we systematically review the computational advances of label-free quantification and diabetic marker selection methods which were applied to get the understanding of DM pathological mechanisms. Firstly, different popular quantification measurements and proteomic quantification software tools which have been applied to the diabetes studies are comprehensively discussed. Secondly, a number of popular manipulation methods including transformation, pretreatment (centering, scaling, and normalization), missing value imputation methods and a variety of popular feature selection techniques applied to diabetes proteomic data are overviewed with objective evaluation on their advantages and disadvantages. Finally, the guidelines for the efficient use of the computation-based LFQ technology and feature selection methods in diabetes proteomics are proposed.

***Conclusion*:** In summary, this review provides guidelines for researchers who will engage in proteomics biomarker discovery and by properly applying these proteomic computational advances, more reliable therapeutic targets will be found in the field of diabetes mellitus.

Keywords: Label free quantification, diabetes proteomics, computation, target discovery, antidiabetic drug, mass spectrometry.

1. INTRODUCTION

Diabetes is a deadly and costly disease [1]. The WHO estimated that about 422 million adults are suffering from diabetes mellitus (DM) all over the globe as per latest data and each year people died directly from diabetes counts for nearly 1.6 million [2]. However, the curing of diabetes is still a very challenging task even with the great knowledge and successful development of effective treatments [3-5]. To overcome this problem, new drug targets for diabetes are clearly needed that will better address these unmet needs. The primary targets of the majority of the antidiabetic drugs (clinical trial/approved) are proteins [3, 6-10]. These kinds of changes of protein interactions, concentrations or functions might lead to the fundamental mechanism of various diseases including diabetes mellitus [11-13]. Thus, a comprehensive understanding of the molecular mechanisms underlying DM progression is of great significance for its diagnosis and treatment.

In the field of diabetes research, it is vital to understand the protein abundances and concentration, therefore the proteomics techniques are urgently needed [14]. So far, diabetes proteomics has grown rapidly in the intervening period. A search of the terms diabetes mellitus and proteomics or proteome shows that since 1995, publication has increased rapidly, with > 3500 each year for previous 5 years and the term diabetes mellitus publication has risen to over 20000 each year in the past five years (Fig. **1a**). Moreover, it is clear that proteomics has been gradually widely used in diabetes research for recent years, over 20 percent of all diabetes publication in 2018 as shown in (Fig. **1b**). In particular, it has

*Address correspondence to this author at the College of Pharmaceutical Sciences, Zhejiang University, Hangzhou 310058, China; Tel: 0086-571-8820-8444; Fax: 0086-571-8820-8444; E-mails: zhufeng@zju.edu.cn; prof.zhufeng@gmail.com

† *These authors contributed equally to the paper as first authors.*

1873-5592/20 $65.00+.00

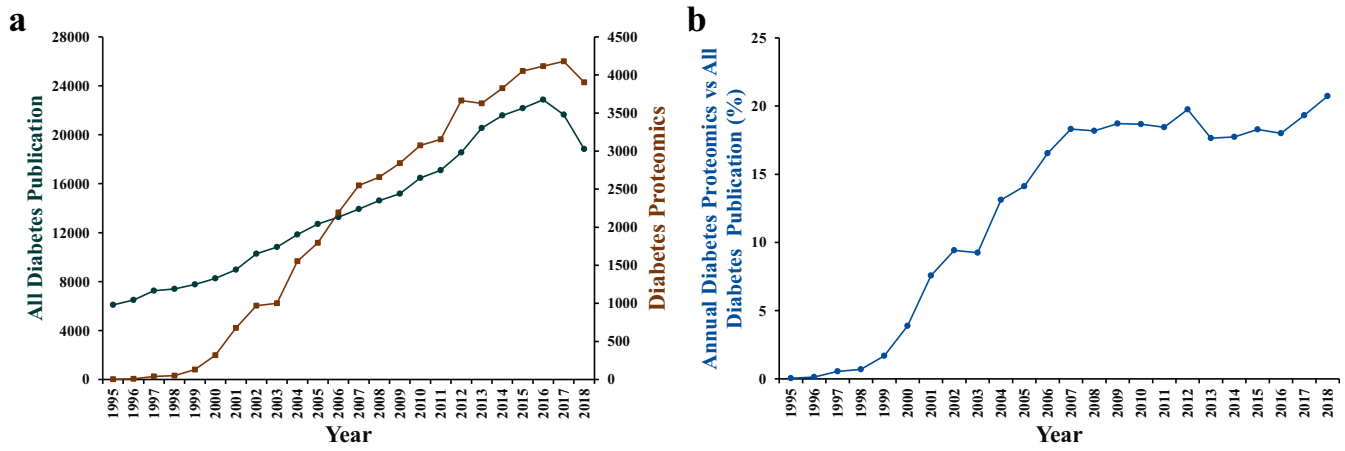

Fig. (1). Trends in diabetes proteomics. (**a**) Annual frequency, beginning with 1995, of publications using the terms *diabetes* (in green) compared with *diabetes proteomics* (in orange). (**b**) The number of diabetes proteomics studies has increased over 20% of all the diabetes publication in 2018, indicating increasingly popular application of the technologies.

been adopted to elucidate the underlying mechanisms associated with insulin resistance [15, 16], explore novel protein biomarkers [17-20], and discover innovative therapeutic protein targets [14, 21]. In addition, the development of antidiabetic drugs is greatly promoted by these remarkable techniques [22].

A convincing example to prove the importance of proteomics in diabetes is the study of glycated albumin (GA). GA was reported to be of higher concentration in the diabetes individual [23, 24], and had been commonly considered as a promising alternative of the traditional biomarkers of diabetes (such as glycated hemoglobin) [25]. However, traditional quantitative methods (such as fluorescence spectroscopy [26], phenylboronate affinity chromatography [27] and ELISA [28]) were found to be failed in the evaluation of the status of glycation [25]. Therefore, with the rapid development of MS technology, proteomics had been widely adopted for accurately identifying and quantifying GA in both healthy and diabetic individuals, so as to discover reliable diagnostic biomarkers used to predict prediabetes as well as secondary complications, respectively [29-34]. Moreover, a further proteomic study analyzing a large number of clinical samples showed that GA was indeed closely related to diabetic onsets [35], which demonstrated the feasibility of GA in the diabetes diagnosis. All these proteomic studies could substantially facilitate the diagnosis and treatment of diabetes [25].

Not only does the prevalence of diabetes need to be controlled, but also its serious complications need to be concerned, including neuropathy, retinopathy, cardiopathy, diabetic nephropathy and so on [36]. In a previous study, researchers identified peptides of differential abundance in serum between the controls and diabetic neuropathy patients, which were considered as biomarkers for detecting diabetic neuropathy [37]. In another proteomic research, the comprehensive alterations of vitreous proteome in proliferative diabetic retinopathy were observed and the abundances of proteins were quantified by LFQ methods [38]. The proteomic strategy was also adopted in the study of diabetic cardiomyopathy's pathogenesis and helped to demonstrate the effectiveness of antioxidant therapy in its treatment [39]. Additionally, in-depth analysis of urine proteome could provide reliable protein signature for identifying diabetic nephropathy (such as β2-microglobulin and ubiquitin) [40]. Therefore, the application of proteomics in diabetic complications is of great significance and can help reduce the harm of diabetes.

In recent years, with the advent of precision medicine [41], extensive attention and great efforts have been paid to the research of protein-based biomarkers to improve the success rate of clinical trials of corresponding drugs [42, 43], which has led to the rapid development of proteomics on the basis of MS [44]. The computational methods were widely used in the field of target and drug discovery [45-50] such as computer-aided drug design (CADD) [51-57], and the LFQ method demonstrates many advantages including allowing detection of proteomes without the expenditure of time and money due to introducing stable isotopes to prepare experimental samples, and could process large amount of samples from various sources [58-61]. These remarkable characteristics make LFQ the most commonly used quantification methods in diabetes proteomics [62-65]. For example, LFQ was applied to analyze the differential proteome of serum in type 1 diabetes [66] and discovered the insulin receptor substrate-1 in the insulin-stimulated interaction [64]. Furthermore, it is already used to discover a novel function of APMAP in gestational DM [67].

Although LFQ has been widely used in various aspects of current anti-diabetes research, there are still many technical challenges in this field [44, 68-71]. In particular, imprecision [72], inaccuracy [73, 74] and low reproducibility [5, 73] of the LFQ are regarded as pivotal "technical challenge" in the research of discovering diabetic marker [75]. In order to solve the above problems, mass spectrometry (MS) quantification measurements performance and some computational methods (like quantification software tools and data manipulation methods) were in progress [76, 77], and were widely applied for LFQ analyses [78] to discover therapeutic biomarkers for antidiabetic drug research [63-66]. Nowadays, it becomes more and more useful to overview the various quantification software tools [60, 79], various proteomics data manipulation methods [80, 81], and properly use of these methods for specific datasets [82, 83]. Moreover, no

such review of the LFQs has been conducted yet on diabetes proteomics.

Simultaneously, feature selection methods play a key role in identifying significant proteins/peptides (or features) between distinct groups from complex diabetes proteomic datasets [84-86]. So far, there are multiple feature selection techniques that have been successfully utilized in DM research to discover the hidden patterns and relationships [87-90]. Bagherzadeh-Khiabani *et al.* [91] made use of a clinical dataset comprising of 803 pre-diabetic females and predicted the likelihood of DM by comparing several common feature selection algorithms. In another work, Georga *et al.* [92] used random forest methods to predict the glucose concentrations in diabetic mellitus patients. However, due to the lack of robustness of feature selection methods, the consistency of the biomarkers has been discovered in most cases is ambiguous [93]. Therefore, to better understand which methods are more accurate when classifying data, some publicly available feature selection methods for proteomics data were recently compared [94-97]. But no such specific review of the feature selection has been conducted yet on diabetes proteomics.

In this study, we systematically review the computational advances of label-free quantification and diabetic marker selection which be applied to get the understanding of DM pathological mechanisms. First, 3 quantification measurements and 11 proteomic quantification tools frequently applied in diabetes proteomics were comprehensively reviewed. Second, a variety of manipulation methods including 4 transformation, 16 pretreatment (2 centering, 4 scaling & 10 normalization) and 3 missing value imputation methods, together with 12 popular feature selection techniques applied in diabetes proteomics were evaluated according to their reported advantages and disadvantages. Finally, the guidelines for the efficient use of the computation-based LFQ technology and feature selection methods in diabetes proteomics were proposed.

2. QUANTIFICATION MEASUREMENTS FOR DIABETES PROTEOMICS

There were three quantification measurements (QMs) used for proteomics-based diabetic biomarker discovery, which were (as shown in Fig. **2**) acquired by two different modes of acquisition: data-independent (DIA) and data-dependent (DDA) [44, 60]. For the QMs acquired by the mode of acquisition of DDA, individual precursors were selected for fragmentation in a semi stochastic manner, favoring the most intense peaks. In contrast, all precursors were fragmented and tandem mass spectrometry (MS/MS) data were acquired for all fragment ions in DIA acquisition [98, 99]. Under this circumstance, there were two main QMs: spectral counting and peak intensity [100]. The detail descriptions on these two QMs were frequently reported in previous studies [101, 102]. For the QM acquired by the mode of acquisition of DIA, the SWATH-MS was a measurement that enabled complete detection and quantification of almost all detectable peptide fragments in a sample [103]. Since it demonstrated the significantly enhanced quantification accuracy and precision [61, 104], it was known as one of the most advanced techniques in current MS-based proteomics.

As one of the most popular QMs acquired by the mode of acquisition of DDA, peak intensity offered more accurate quantitation and wider dynamic range than spectral counting [105]. In addition, the protein quantification applying peak intensity was reported to be better accuracy for higher resolution mass spectrometry [106]. But for low resolution machines, owing to great amounts of thermal noise, the precision of peak intensity was impaired [107]. The other popular QM was spectral counting, which was a very simple LFQ technique for taking into account the total number of spectra of identified proteins [108]. It demonstrated the best QM for quickly screening the differences between samples [109], and broad estimation of the identification of total proteins [110].

SWATH-MS, a newly developing QM, has gradually gained popularity due to it can comprehensively detect all ionized peptide fragments with the favorable sensitivity, reproducibility, accuracy and extended dynamic range for analyzing diabetes proteomics data [44, 103]. It has become an effective way to discover novel potential therapeutic targets for treating proliferative diabetic retinopathy [111], elucidate the underlying DM etiology associated with insulin resistance in gestational diabetes mellitus and diabetic cardiovascular complications [15, 112]. By now, although not many studies have used SWATH-MS for diabetic biomarker discovery, it is worth noting that this technology has the potential to address the limitations of many current diagnostic biomarkers or therapeutic targets [113].

3. QUANTIFICATION SOFTWARE TOOLS FOR DIABETES PROTEOMICS

So far, there are a variety of quantification software tools (QSTs) for diabetes proteomics data analysis. Some of them are freely accessible while others are commercial. These QSTs which have different algorithms could process diabetes proteome datasets generated by different QMs. Here we show 11 QSTs which are extensively applied in processing proteome raw data and divide them into three types according to the way of 3 different QMs. They are described as follow (Fig. **2**), and (Table **1**) provides explanations on each QST.

3.1. QSTs Used to Pre-process the Diabetes Proteomic Data Quantified by SWATH-MS

The DIA-Umpire is a freely accessible QST which can pre-process the raw data acquired by DIA [114], and it is extremely appropriate to the untargeted quantification employing proteomics dataset on the basis of SWATH-MS [115]. The advantage of this tool is that it can get robust protein quantification, so similar amount of proteins can be discovered among various sets of samples while traditional tools of DDA can't do this [116]. However, it is not suitable for quantifying the modified peptides after translation due to the design purpose of protein quantification [117]. The accuracy for protein quantification of this tool is improved to 93% by filling in the intensities of missing peptide ions [114]. Moreover, it has been frequently applied to analyze human proteome by targeted high-resolution mass spectrometry [118] and generate the results of pseudo-MS2 spectra protein identification [119].

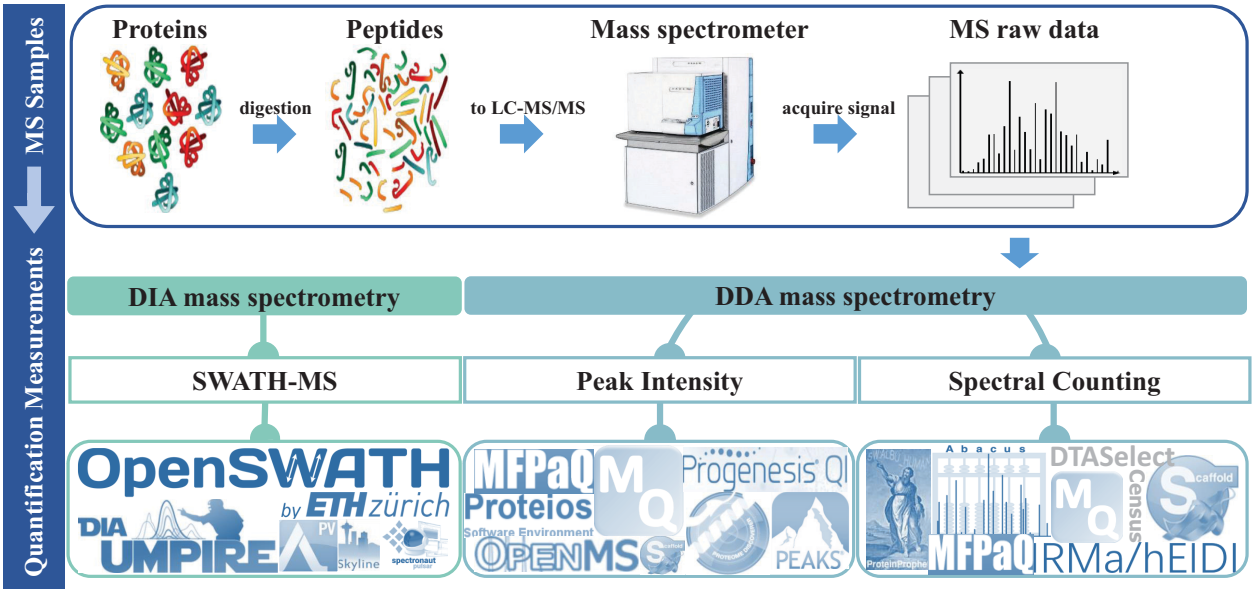

Fig. (2). Quantification measurements together with their representative quantification software tools used in diabetes proteomics.

Table 1. Thirteen popular quantification software tools for diabetes proteomic data.

Quantification Tool		Platform*	Version	Applied in Diabetes Proteomic Researches
SWATH-MS	DIA-Umpire	W; L; M	2.0 (Apr. 2016)	Applied to export the peptide identification results of pseudo-MS/MS spectra [119].
	OpenSWATH	W; L	2.4.0 (Oct. 2018)	Applied to generate a compendium of quantitative proteome data in dietary and metabolic perturbations mouse strains [121].
	PeakView	W	2.2 (Oct. 2014)	Applied to study the malleable nature of diet-induced cognitive dysfunction [110].
	Skyline	W	4.2.0 (Jan. 2019)	Applied to determine transitions for each target peptides in the study of protein changes within the diabetic vasculature [130].
	Spectronaut	W	11 (May 2016)	Applied to analyze the controversial omentin-regulated proteins in type 2 diabetes [133].
Peak Intensity	MaxQuant	W; L	1.6.5.0 (Oct. 2016)	Applied to establish an online 2D-LC-HCD-MS/MS platform for comprehensive glycated peptide quantification[163].
	Scaffold	W	4.8.9 (Dec. 2018)	Applied in an in-depth analysis of the urinary proteome based on different separation strategies [168].
	OpenMS	W; L; M	2.4.0 (Oct. 2018)	Applied for the proteomic analysis of corneal endothelial cell-descemet membrane tissues in type 2 diabetes mellitus [138].
	PEAKS	W	8.5 (Oct. 2017)	Applied to predict gestational diabetes through early second-trimester peptidomic serum peptides identification [141]
	Progenesis	W	4.1 (May 2018)	Applied to study the retinal proteome alterations in a mouse model of type 2 diabetes [145]
	Proteome Discoverer	W; L	2.2 (Aug. 2017)	Applied for determining changes in protein expressions in arterial tissue from patients with type 2 diabetes mellitus [151]
Spectral Counting	Census	W; L; M	2.3 (Mar. 2014)	Applied in a multiple proteomics study on liver mitochondria isolated from spontaneous diabetic rat model [155].
	MaxQuant	W; L	1.6.5.0 (Oct. 2016)	Applied to study the unique exocrine tissue proteomic profile of type 1 diabetes cadaveric human pancreata [164].
	Scaffold	W	4.8.9 (Dec. 2018)	Applied for identification of more specific biomarkers for prediction of Diabetic nephropathy [167]
	DTASelect	W	2.0.41 (Mar. 2010)	Applied to explain the relationship of Ces3 to obesity and diabetes [160].

*W = Windows OS, L= Linux OS, M= Mac OS.

As a QST designed with the ability to work under different platforms, OpenSWATH is an automated, open-accessible and high-throughput tool [120]. It is a cross-platform software, written in C++, which only depends on open-data formats, permitting the analysis of DIA data through multiple tools vendors and can deeply analyze diabetes proteomics data based on SWATH-MS [60, 103, 120]. The disadvantage of this tool is that it can't distinguish peptide ions with lots of the same MS/MS fragments [114]. In a previous study, the accuracy of the software can reach 87.5% with a high precision of 94.3% [120]. OpenSWATH has been applied to generate quantitative proteome data in an experiment of metabolic perturbations [121].

Compared with other tools of this type, the PeakView shows particular superiority of integrating major computational processing methods [119, 122]. The principle of this tool has been introduced in previous research [103]. The advantage of PeakView was the function that all the extracted ion chromatograms could be inspected visually [123], and the mass defect filtering (MDF) technique integrated into this tool provided a new method for rapid identification of various compounds [124]. However, it couldn't automatically compare the differential peaks between control and sample, so manual comparison is needed [123]. The accuracy of this software varies with sample size and parameters [125]. Nowadays, PeakView has rapidly grown into a powerful quantification tool for processing proteomic data in many fields. For example, in the study of the association between maternal metabolic changes and circulating exosomes in gestational diabetes, PeakView was used for the quantification calculations [15]. In addition, Johnson LA used PeakView for Mass spectrometry data analysis and highlighted the malleable nature of diet-induced cognitive dysfunction [110].

Skyline is also a freely accessible quantification tool which can process data acquired by different techniques of reaction monitoring and analyze both DDA and SWATH-MS data on the basis of MS1 information of quantification [126]. Besides its excellent file compatibility [127], it has strong capabilities of method editing and can respond to large amounts of complex data [128]. Moreover, it enables to quantify proteins, peptides as well as small molecules [129]. With the help of Skyline, an analysis of 572 diabetes patients introduced a biomarker panel with the potential to improve diagnosis of diabetic kidney disease [109]. Skyline was also applied to determine the conversion of each target peptides in the study of protein changes in the diabetic vascular system [130] and select the optimum peptides for multiple reaction monitoring in type 2 diabetes patients treated with Xiaoke Pill and Glibenclamide [131].

Another widely applied QST in this type is the Spectronaut, which can be used in DIA measurement analysis [116, 132]. The powerful functions of choosing peaks and interference auto-correction applying typical library of spectral make it widely used to sustain workflow that doesn't have a spectral library [60, 122], and this is the key strength of it. In addition, Spectronaut is more sensitive in the detection of differentially expressed proteins [125]. A proteomics analysis of omentin-regulated proteins which was reported controversial in type 2 diabetes has applied Spectronaut [133].

3.2. QSTs Used to Pre-process the Diabetes Proteomic Data Quantified by Peak Intensity

As a freely accessible QST, OpenMS can process MS-based raw proteomics data with the high-throughput and robust characteristics. It was widely used for analyzing diabetes proteomics data with enhanced reproducibility [134]. This tool is more flexible than other tools because it has smaller algorithmic components which can be rapidly combined for specific analysis [135]. In another study of OpenMS, the quantification coverage could reach 99%, and the obtained protein abundance was close to the known amounts, indicating high accuracy [136]. It is widely applied to blood transcriptomes and metabolomics for personalized medicine [110], identification of therapeutic targets in chronic kidney disease [137] and analyzing endothelial cell Descemet membrane of cornea proteomics in type 2 diabetes [138].

PEAKS has become a powerful QST for processing diabetes proteomic raw data to identify and quantify the proteins. It has developed into a comprehensive software of proteomics quantification that could handle both the label-free and labeling proteomics data. Besides, it was also a multi-functional tool which can be used for analyzing PTMs (Post-translational Modifications) and protein/peptide de novo sequencing [139]. Its advantage is that the generation of peaks makes the protein quantification have higher sensitivity and accuracy than other QSTs [140]. Moreover, PEAKS help to predict the identification of gestational diabetes by early second-trimester peptidomic serum peptides [141].

Progenesis which is a commercial tool, quantifies protein abundances by ion intensity from large scale proteomics data and ion detection according to a high sensitivity algorithm [142]. Operators can control each processing step when using this software [79]. It is limited to LFQ unless there are other programming interfaces [143]. The quantification accuracy of this tool was analyzed within various proteomic date sets, and it performed almost the best compared with other software [122]. Nowadays, Progenesis has been commonly used in diabetes proteomics study [14], including the human placental tissue impaired by gestational diabetes [144], retinal proteome alterations in type 2 diabetes mouse model [145], and analysis of vitreous body from proliferative diabetic retinopathy and type 2 diabetic patients [146].

Proteome Discoverer makes a large range of proteomic workflows easier, from the identification of proteins to isobaric mass tagging to PTM analysis and both LFQ and SILAC [110, 147]. It supports a variety of database search algorithms (Byonic, Mascot and Sequest, *et al.*) and diverse separation techniques (ETD, CID and HCD, *et al.*) for more detailed analysis [148]. Researchers could present MS-data directly from the instrument, which allows for the identifying and quantifying peptides and proteins by multiple seeking algorithms [148]. It strikes a good balance between usability and flexibility, which on the other hand may lead to functional impairment [149]. A Java library has been established to solve the problem of parsing and visualizing the output files of this software [147], which was caused by the lack of internal conversion tools [150]. Proteome Discoverer has been used in arterial tissue from patients with DM to detect differences of protein expressions [151] as well as in

the patient of diabetic nephropathy to determine changes in the proteome of human urinary exosomes [152].

3.3. QSTs Used to Pre-process the Diabetes Proteomic Data Quantified by Spectral Counting

Census is a commercial software based on spectral counting not only is accessible for different stable isotope labeling tests, but also can treat with the shotgun diabetes label-free proteomics data [153]. Multiple computational strategies for improving quality of quantification and its extensive coverage of quantification strategies makes it outstanding from other QSTs of spectral counting [154]. The quantitative accuracy of low abundance peptide was improved by using filtering strategy [154]. Census can be used for quantitative analysis and a multiple proteomic research on spontaneous diabetic rat model liver mitochondria was conducted [155].

Another Spectral Counting QSTs, DTASelect, can be used for analysis and effective identification of the proteins produced by the search engine of tandem MS database [56] which is a kind of most widely used [156]. The process of DTASelect contains filter, establishment and visualization of a great deal of a biosample tandem mass spectra [157]. This method takes typical proteins through elimination of the different identification and therefore improves quantification of proteins [157]. Users could decide to accept or reject the result of individual spectrum by setting complex criteria [157], which is a unique advantage of this software. It can make complicated tests possible by simplifying analysis of data [158] so that it could be adopted to various diabetes proteomic research with a low false positive [159] and linking Ces3 to obesity and diabetes [160].

3.4. The QSTs Used to Pre-process the Diabetes Proteomic Data Quantified by Multiple QMs

Two QSTs able to process raw data generated by various QMs are accessible, they are Scaffold and MaxQuant software packages. Specifically, both tools could process data obtained from spectral counting and peak intensity.

MaxQuant matches proteins across various samples since it integrates common used algorithms for quantifying proteins for MS-based apparatus with high resolution [161]. Nowadays, this tool is widely used for the analysis of tandem spectra produced by the high energy collisional, collision-induced and the electron-transfer [162] in the diabetes proteomics. A previous research showed that MaxQuant could quantify the largest number of proteins compared to other LFQ tools in the SGSDS, CPTAC and UPS1 data sets [122]. This tool has been applied to establish an online MS/MS platform for comprehensively quantifying glycated peptide [163]. In addition, it was also used to study the unique exocrine tissue proteomic profile of type 1 diabetes cadaveric human pancreas [164].

Scaffold is a kind of commercial QST offering good accuracy on identifying proteins through applying different computational methods and providing diversiform strategies for confirming the proteins identification accuracy from primary datasets [165]. It is a multifunctional software that can be used in complex LC-MS/MS experiment for analysis, quantitation, validation and other processes [165]. This tool was applied to study the alteration determined by exercise training of the skeletal muscle proteome in type 2 diabetes patients [166]. Since microalbuminuria had been reported to be limited in determining disease risk, Scaffold is used to identify biomarkers for diabetic nephropathy prediction more specifically [167]. A recent study where Scaffold is used presents a depth urinary proteome analysis according to different strategies of separation [168].

4. DATA MANIPULATION METHODS FOR DIABETES PROTEOMICS

Three manipulation methods were developed to process diabetes proteomic data, which included transformation, pretreatment (centering, scaling & normalization) and missing value imputation. More detailed description on those applied methods and the wide application in present diabetes mellitus research were then discussed in the next part and shown in Fig. (**3**). All these methods are summarized in Table **2**.

4.1. Methods for Transforming the Diabetes Proteomic Data

In most instances, diabetes proteomics data usually requires transformation as the first step [169]. In matrix of data, the abundances of proteins are sometimes discovered to distribute in a right-skewed way [169]. Therefore, a proper usage of the transformation seems to be quite important to generate data with much more symmetry-improved and normal distribution. At present, four transformation methods (Power, Log, Cube Root and Box-cox) were frequently applied to process the diabetes proteomics data.

Power transformation (POW) can be used to transform the normal linear model [170]. Usually, it deals with a round of functions and results in performing a monotonic transformation [170], and it has the capacity of the stability of variance [171]. POW has been applied to study progresses that were available in the nation-wide Finnish diabetic nephropathy study [172], and to indicate the significance of both subfascial and subcutaneous fat related to the metabolism of lipid and glucose [173]. Besides, another symmetric distribution before the dataset analysis usually comes from the log transformation (LOG). For the dataset where the residuals might be larger for values of the dependent variables, this method can be quite suitable [174]. Hence, the LOG was a popular method of transformation for microarray datasets [175] and has helped the analyses of the functional and proteomic alterations of high-density lipoproteins in plasma in diabetes patient and placenta in maternal obesity [176, 177]. The LOG method was usually performed by log2-scale, and the algorithm was programmed and implemented under the R environment.

Variance and mean of distribution using $N=1/3$ power (cube root) by substituting have been applied to process all kinds of diabetes mellitus proteomic dataset through the cube root (CUB), that is initially designed according to probability density function [178]. Yet this method is adopted to compare the proteomic analysis result from saliva of dogs with obesity-related metabolic dysfunction [179]. As a kind of method able for parametric power transformation, the box-cox (BOX) aims to break away from multiple anomalies [180], and has been applied in a robust and fast workflow

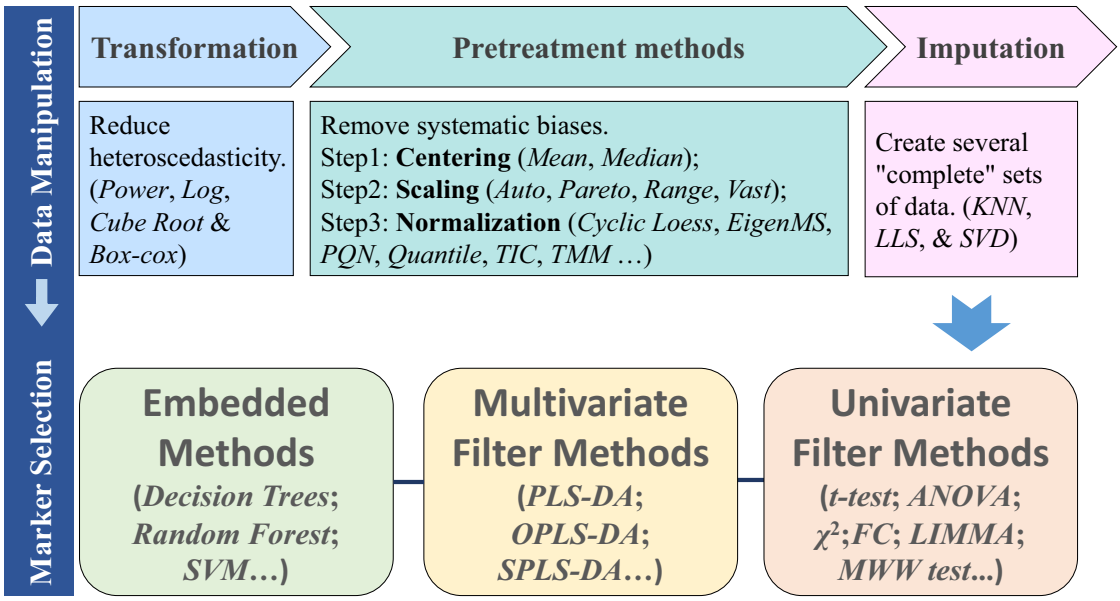

Fig. (3). Data manipulation methods sequentially applied in diabetes proteomics.

which shows potential in plasma biomarker identification [181]. The parameter λ in BOX was set to 0.3, and the algorithm was programmed and implemented under the R environment (version 3.5.1).

4.2. Methods for Pretreating the Diabetes Proteomic Data

The pretreatment methods which can remove systematic biases were taken as an inalienable sector of LFQ to improve the accuracy of relatively quantifying the peptides and proteins [182]. Currently, there are 2 centering, 4 scaling and 10 normalization methods which are commonly used in diabetes proteomics and have been overviewed in the following sections.

4.2.1. Methods for Centering the Diabetes Proteomic Data

Mean centering (MEC) adjusts the concentrations to fluctuate around 0 rather than the mean value. It is thus applied to concern the fluctuating section of the dataset, and then only remain the related differences (the variation among the samples) [183]. Some diabetes related proteomic analysis such as weight loss and maintenance [181], anti-diabetes lipid and bile acid markers [184], as well as biomarkers for neuroretinal degeneration in diabetic retinopathy [185] used the MEC. Its algorithm was programed by integrating the basic mean function (mean value) in the R-statistical programming and implemented under the R environment (version 3.5.1). Similarly, median centering (MDC) regulates changes in the dataset between proteins with low and high abundance beside median of the protein concentrations [183]. It was widely applied in exocrine tissue proteomic profile in type 1 diabetes [164] and diagnostic biomarker [186]. The algorithm of MDC was programmed by integrating the basic median function (median value) in the R-statistical programming and implemented under the R environment (version 3.5.1).

4.2.2. Methods for Scaling the Diabetes Proteomic Data

Before analyzing a multivariate data, a step of scaling might be essential to obtain accurate results, because protein concentration levels in samples can range by orders of magnitude [187]. The scaling can be crucial for avoiding the case in which the peak is the most influential in the multivariate dataset [187]. Some kinds of scaling methods exist, that is ATO, VAS, RAN and PAR.

The auto scaling (ATO, also named unit-variance scaling) method is the simplest way adjusting the proteomic variance and can scale the intensities into unit variances based on standard deviation of diabetes proteomic data [188, 189]. ATO has now also been used to identify miRNAs as predictive biomarkers of type 2 diabetes [190] and applied to several pregnancy conditions [177]. The algorithm of ATO was programmed by integrating basic *sd* function (*standard deviation*) in the R-statistical programming and implemented under the R environment (version 3.5.1). Vast scaling (VAS) is an extension of ATO [141]. VAS concern stable variables making the coefficient of variation and the standard deviation to be the factor of scaling [183]. Every peak is autoscaled and separated through the coefficient of variation. It is especially suitable for proteins bearing small fold changes [141]. VAS has been used to study the delayed storage influence on the metabolome and proteome of cerebrospinal fluid [191]. The algorithm of VAS was programmed by integrating two basic functions *var* and *mean* (*variance* and *mean value*) in R-statistical program and implemented under R environment.

In addition, pareto scaling (PAR) can utilize the standard deviation as a scaling factor [188] which is very similar to autoscaling. This approach is able to reduce the weight of huge intensities fold changes, that is much more important compared to ATO [188]. But as the major weight, the very large fold alteration might not change. So being sensitive to the large fold changes might be a disadvantage of PAR

Table 2. Manipulation methods available for LFQ-based diabetes proteomics.

Algorithm		Package (Function)	Applied in Diabetes Proteomic Study
Transformation	Power	car (bcPower)	Applied to indicate the significance of both subfascial and subcutaneous fat related to the metabolism of lipid and glucose [173].
	Log	metabolomic (LogTransform)	Applied to analyse the functional and proteomic alterations of high density lipoproteins in plasma in diabetes patient [176].
	Cube root	pamr (pamr.cube.roo)	Applied to compare the proteomic analysis result from saliva of dogs with obesity-related metabolic dysfunction [179].
	Box-cox	AID (boxcoxfr)	Applied in a robust and fast workflow which shows potential in plasma biomarker identification [181].
Centering	Mean centering	Mean	Applied in the discovery of biomarkers for neuroretinal degeneration in diabetic retinopathy [185].
	Median centering	Median	Applied to study unique exocrine tissue proteomic profile exhibited by type 1 diabetes cadaveric human pancreata [164].
Scaling	Auto scaling	Metabolomics	Identify miRNAs as predictive biomarkers of type 2 diabetes [190]
	Vast scaling	DiffCorr	Applied to study the delayed storage influence on the metabolome and proteome of cerebrospinal fluid [191].
	Pareto scaling	BioMark	Applied to process data on the basis the information of scan-level and help predict the risk of developing diabetes [192].
	Range scaling	DiffCorr	Applied in the study of omics data fusion integrated with an optimal data preprocessing strategy [171].
Normalization	EigenMS	DanteR	Applied to study the pathogenesis and pathophysiology of gestational diabetes mellitus [219].
	Lowess	LPE	Applied to the study of the influence on platelet micro-RNA expression from controlled type 2 diabetes mellitus [221].
	Mean	mixOmics; Normalyzer	Applied in the research of differential expression of proteins in different tissues of diabetic and non-diabetic rats [205].
	Median	Normalyzer mixOmics	Applied to normalize multiple spectra intensities in proteomic analysis of gestational diabetes [210].
	MAD	stats	Applied to establish a platform which can identify natural compounds regulating Pdx1 and insulin expression [213].
	PQN	KODAMA MALDIquant	Applied in a metabolomics study in gestational diabetes [215].
	RLR	Normalyzer	Applied in an early study of the relationship between visceral adipose tissues and early pathogenesis of type 2 diabetes applied the RLR [198].
	TIC	Normalyzer	Applied to explore the functional and proteomic changes of plasma high density lipoproteins in diabetes [203].
	TMM	edgeR	Applied to study the transcriptional reprogramming in human myocytes induced by type 2 diabetes [224].
	Z-score	mosaic	Applied to help calculate the significant changes in gene expression in different samples [226].
Imputation	KNN	imputation (knnImputation)	Applied to find out desired protein in proteomic study on diabetes mellitus patients with perimenopausal syndrome [233].
	LLS	pcaMethods (llsImpute)	Applied in oral glucose tolerance experiment in patients with gestational diabetes [236].
	SVD	pcaMethods (svdImpute)	Applied to study unsaturated plasma phospholipids level in the pregnancy patients diagnosed with gestational diabetes [235].

[183]. Processing data on the basis the information of scan-level with PAR could help predict the risk of developing diabetes [192]. The algorithm of PAR was programmed by integrating two basic functions *sd* and *sqrt* (*standard deviation* and *square root*) in the R statistical programming and implemented under the R environment (version 3.5.1). Differently, Range scaling (RAN) takes the biological range as the factor of scaling [193] which is maximal and minimal density attained by a typical protein in a series of tests [183]. Because there are 2 values used to assess the biological range, which makes RAN more susceptive to outliers [183]. It has been adopted to the study of omics data fusion integrated with an optimal data preprocessing strategy [171]. The algorithm of RAN was programmed by integrating two basic functions *max* and *min* (*maximum value* and *minimum value*) in the R-statistical programming and implemented under the R environment (version 3.5.1).

4.2.3. Methods for Normalizing the Diabetes Proteomic Data

Diabetes proteomics data retain innately biased because it might range from sample processing to differences issues of the instrumentation [194]. The normalization is treating process aiming to explain the bias as well as make these samples much more comparable while maintaining the biological variation [195]. Therefore, how to choose an optimal normalization method could be a vital issue to the downstream analysis reliability [195-197].

Robust linear regression (RLR) can be quite robust to outliers in datasets when compared with linear regression using least square estimation and it scales one referred range to another. RLR performs quite well when estimating the logFC of the spike-in proteins [182]. A study of the early pathogenesis in DM applied the RLR [198]. The median values over all the samples as the reference sample (to which all the other samples were normalized to) were first calculated, and then RLR method was programed by integrating the basic *rlm* function in R-statistical programming and implemented under the R environment, which was then programmed with default parameter settings. Different from other normalization methods, Total Ionic Current (TIC) could be used in proteomics dataset normalization on the basis of identifying same proteins to a specific sample and evaluating the sum of all the peak intensities of these proteins [199]. TIC was reported as one of the simplest and most common methods normalizing diabetes proteomics data [200-202]. The premise of using this method is the assumption that each protein is equally important among given sample. It has been used to explore the functional and proteomic changes of plasma high-density lipoproteins in diabetes [203]. The TIC method was programed by integrating the basic *sum* function (*summation*) in R-statistical programming and implemented under the R environment (version 3.5.1).

Various adjustments to the basic premise of TIC include mean normalization, median normalization, median absolute deviation, cyclic locally weighted regression and *etc*. [201]. The mean normalization (MEA) can normalize datasets using the mean of all proteins, which can remove the effect of background. This method regards the mean of intensity of all the variables in a specific specimen as each protein intensity in the same specimen [174]. The MEA has been widely applied in the research of the metabolomics related to cardiovascular disease in type 2 diabetes patients [204] and differential expression of proteins in different tissues of diabetic and non-diabetic rats [205]. The MEA method was implemented with the *Normalise* function using the metabolomics-package in R/Bioconductor, which was then programmed with the *mean* parameter settings under R environment. The median normalization (MED) is developed to make the median of samples in a dataset the same, based on the hypothesis that a constant separates the samples [206-208]. It scales the log of intensity values according to the global median value so that they have the same median [209]. So far, the intensities of multiple spectra have been normalized applying MED in proteomic analysis of gestational diabetes [210]. MED method was implemented with the *Normalise* function using the metabolomics-package in R/Bioconductor, which was then programmed with the *median* parameter settings under R environment.

Median Absolute Deviation (MAD) is a simple way for variation quantification, which could be used to estimate the sample standard deviation while just scaled through the 1.483 [211]. In contrast to standard deviation, MAD is more robust to outliers and expected to obtain fewer false negatives [212]. Furthermore, the method can improve the process of quality control of dataset on the basis of LC-MS. And MAD has helped to establish a platform which can identify natural compounds regulating Pdx1 and insulin expression [213]. MAD method was programed by integrating two basic functions *median* and *mad* (*median value* and *median absolute deviation*) in the R-statistical programming and implemented under the R environment (version 3.5.1). Probabilistic quotient normalization (PQN) operates in a similar way to MED, and it can transform proteomics data on the basis of the systematic assessment of most like dilutions [214], and it has significant accuracy and robustness compared to those on basis of the length of vector. The disadvantage of PQN is that, it operates assuming class of interest (such as species, sex) does not exhibit significant differences [207]. A study of metabolomics in gestational diabetes applied PQN, suggesting PQN is a reliable normalization technique as well [215]. The median values over all samples as the reference sample (to which all the other samples were normalized to) were first calculated, and then PQN method was implemented with the *median* function in the R-statistical programming and implemented under the R environment.

EigenMS (EIG) fits the variance model analysis to assess the treatment group effects, and next applies singular value decomposition to model residual matrix to remove the bias [216-218]. EIG has already been adopted to the study of the pathogenesis and pathophysiology of gestational diabetes mellitus [219]. The EIG method was implemented using its R-codes available for downloading from the *Sourceforge-repositories*. Locally weighted scatterplot smoothing (LOW) can normalize two-color expression data, in which the log-ratio for each sample can be adjusted by the lowess fitted value [220]. It's based on the assumption that spot intensity determines the appearance of dye bias [220], and can be used in both complete and incomplete datasets [220]. It has been adopted to the study of the influence on platelet micro-RNA expression from controlled type 2 diabetes mellitus [221], as well as amino acid and acyl-carnitine metabolism in infants

from dietary protein intake [222]. The LOW method was programed by the *preprocess* function using the software package of LPE in the R environment, which was then programmed with the *LOWESS* parameter settings under R environment.

Moreover, the trimmed mean of M-values (TMM) can also be very popular for its simplicity and high efficiency in processing RNA-sequence data [223]. This normalization can be applied to evaluate scaling factors in data, and has been used to study the transcriptional reprogramming in human myocytes induced by type 2 diabetes [224]. TMM method was implemented with the *tmm* function using the NOISeq-package in R/Bioconductor, and was then programmed with the default parameter settings under R environment. Z-score (ZSC) can normalize data acquired from a large scale of experiments on basis of the standard deviation and mean values then can be used in comparison of the microarray data [225]. Z-score has helped to calculate the significant changes in gene expression in different samples [226-228]. The formula of Z-score was the same as that of the Auto Scaling (ATO), and its algorithm was usually programed by integrating basic *sd* function (*standard deviation*) using the mosaic-package in the R-statistical programming and implemented under the R environment (version 3.5.1).

4.3. Methods for Imputing the Missing Values in Diabetes Proteomic Data

Diabetes proteomic dataset usually distributed sparsely [229], that implies a specific matrix of data contains lots of missing values among situations [122]. For instance, the detection instrument limitation is compared higher than the proteins concentration [216], different biological factors or analytical/technical lapse, the incorrect protein or peptide identification [230]. Therefore, the imputation approaches are usually existing for addressing these problems [216]. To the best of our knowledge, several imputation strategies applied to process the missing values are described as follows.

As an imputation method, K-nearest Neighbor Imputation (KNN) could identify K proteins similar to the protein that has values missed [231, 232]. And it has the ability to find out the most similar protein to the desired protein in proteomic study on diabetes mellitus patients with perimenopausal syndrome [233]. The KNN method was implemented with the *impute.knn* function using the impute-package in R/Bioconductor, and then programmed with the parameter settings of *k value equaling to 10* under R environment. Singular Value Decomposition (SVD) is on account of linear relation through all kinds of proteins of a typical sample [234]. When compared with KNN that uses the local pairwise information, SVD predicts missing value primarily by message from global matrix [234, 235]. The SVD method was implemented with the *pca* function using the pcaMethods-package in R/Bioconductor, and then programmed with the parameter settings of *svdImpute method* under R environment. What is more, Local Least Squares Imputation (LLS) can exploited dataset through local similar structures [231]. This method has been utilized in oral glucose tolerance experiment in patients with gestational diabetes [236]. LLS method was implemented with the *llsImpute* function using the pcaMethods-package in R/Bioconductor, and then programmed with the *k value equaling to 10* together with the default parameter settings under R environment.

5. DIABETES MARKER IDENTIFICATION USING PROTEOMIC DATA

In the study of identifying diabetes biomarker, the marker selection (also called feature selection) strategies are much-needed and quite popular [237]. Since several kinds of techniques were accessible for finding biomarker or disease related proteins imported through therapies for diabetes, it's quite a challenge using appropriate methods for doing any diabetes marker-related researches [238, 239]. All these feature selection strategies which commonly applied in the studies of DM divided into three categories, there are embedded methods, multivariate filter methods and univariate filter methods [240, 241]. All frequently-used marker selection methods are described as follows and shown in Table **3**.

Some filter methods could reflex the quality of every feature according to its discriminative ability in which feature can be reserved when the value of the metric were in a typical standard [242]. The function of t-test is to estimate if the mean values of the two sets that distribute normally have statistical difference [243, 244]. It was widely applied in the omics study [245] and has helped to exploit a new assay for rapid and multidimensional monitoring of diabetes [246] and also been adopted to study different content of proteins in arteries membrane between metformin users and patients with type 2 diabetes [151]. Analysis of Variance (ANOVA) [247] pays attention to analyze the dissimilitude in several groups variance or average value in a typical protein [248, 249]. ANOVA has been used to analyze a biomarker set of urinary, which could assess diabetic patients' early kidney risk [250], and explore the association of potential salivary biomarkers with diabetic retinopathy in type 2 diabetes [251]. Moreover, Chi-square (χ^2) may become a popular strategy in weighting divergence distribution if the class value has nothing to do with hypothetical features [252]. χ^2 has been applied to study the relationship between the severity of cardiovascular disease and proinflammatory as well as antioxidant proteins in type 2 diabetes [253], and the association between KCNQ1 mutations and hypertension in type 2 diabetes mellitus [254].

Mann-Whitney-Wilcoxon test (MWW) is a non-parametric alternative test with null hypothesis [255, 256]. MWW is normally applied while the assumption of the data or t-test is not met [257], so it has been used to explore individual glycation sites in blood plasma proteins, which are prospective biomarkers of type 2 diabetes mellitus [258]. By comparing the absolute value alteration among mean values of two groups, Fold Change (FC) calculated the ratio or log of the ratio levels between groups. Except for its application in genomics [259, 260], in plasma of type 1 diabetes patients, FC value has been applied in finding proteomic changes of high density lipoproteins [203], and monitoring the individual development of diabetic nephropathy patients with type 2 diabetes [261]. Linear Models for Microarray Data (LIMMA) pays attention to different protein expression analysis of raw-file produced by microarray [262]. LIMMA has been applied to explore significant genes associated with diabetic nephropathy [263].

Table 3. Twelve popular marker selection algorithms in diabetes proteomic researches.

Algorithm		Package (Function)	Applied in Diabetes Proteomic Study
Univariate Filter	t-test	stats (t.test)	Applied to exploit a new assay for rapid and multidimensional monitoring of diabetes [246].
	ANOVA	ANOVA.TFNs (fanova)	Applied to explore the association of potential salivary biomarkers with diabetic retinopathy in type-2 diabetes [251].
	Chi-square	stats (chisq.test)	Applied to study the relationship between the severity of cardiovascular disease and proinflammatory as well as antioxidant proteins in type-2 diabetes [253]
	MWW	stats (wilcox.test)	Applied to explore individual glycation sites in blood plasma proteins, which are prospective biomarkers of type-2 diabetes mellitus [258].
	Fold Change	metabolomics (FoldChange)	Applied to monitor the individual development of diabetic nephropathy patients with type-2 diabetes [261].
	LIMMA	limma (lmFit)	Applied to explore significant genes associated with diabetic nephropathy [263].
Multivariate Filter	PLS-DA	caret (plsda)	Applied for identify early urinary biomarkers of diabetic nephropathy identification in type 1 diabetes patient [266].
	OPLS-DA	ropls (opls)	Applied to investigate the proteomic alterations of human milk in women with gestational diabetes mellitus [268].
	SPLS-DA	mixOmics (splsda)	Applied to explore host-microbiota interactions in patients with type-1 diabetes [271].
Embedded	Decision Tree	dtree (pca)	Applied to improve the diagnostic accuracy for type 2 diabetes mellitus [225].
	Random Forest	randomForest (randomFores)	Applied in the study of the associations between maternal BMI and insulin resistance [277].
	SVM	e1071 (svm)	Applied to improve type 2 diabetes mellitusd diagnostic accuracy with the help of glycation sites in plasma proteins [225].

As for PLS-DA [264], the strategy belongs to linear two-class classifier [265]. Since the sample sizes are unequal, the PLS-DA would not produce a decision boundary with great accuracy [264]. In type 1 diabetes patient, it was also useful for early urinary biomarkers of diabetic nephropathy identification [266]. Orthogonal PLS-DA (OPLS-DA) method is a vigorous strategy for analyzing qualitative data structures, and the results predicted by it are similar to the results of standard PLS-DA [267]. The OPLS-DA has been applied to investigate the proteomic alterations of human milk in women with gestational diabetes mellitus [268]. What's more, similar to PLS-DA, the Sparse PLS-DA (SPLS-DA) may be inclined to ignore variables that only distinguish between small samples [269]. Variable selection and modeling in SPLS-DA strategy can be allowed in one step, and the interpretability is modified by valuable graphical output [270] therefore helps in exploring host-microbiota interactions in patients with type 1 diabetes [271].

Moreover, both the feature selection part and the learning part are of significance and could not be separated in embedded methods [272]. Take Decision Trees (DT) for example, it studies a battery of training examples, picks out a property and splits typical examples according to the values of that attribute by an iterative process [273]. This method has helped characterize drug targets [274] and has been applied in type 2 diabetes mellitus to improve diagnostic accuracy [225]. Through the idea of integrated learning, Random Forest (RF) integrates multiple trees [275]. This strategy helps clinical phenotypic discrimination and biomarker selection [276] as well as the associations between maternal BMI and insulin resistance study [277]. In addition, the Support Vector Machine (SVM) was reported to be a powerful classification tool, whose basic principle was to discover a multidimensional hyperplane by projecting the studied samples into the high dimensional space during the model construction [278]. It performs well in handling high-dimensional datasets with the help of a few training examples [279] although it has the limitation of relying on negative data sets [280]. SVM has been applied to the improvement of DM diagnostic accuracy with the help of glycated lysine-141 in haptoglobin combined with glycated hemoglobin HbA_{1c} [225].

CONCLUSION

Most current therapies for diabetes were developed in the absence of defined therapeutic targets or an understanding of molecular mechanisms of the disease. Proteomics helps to enhance our understanding of the pathogenesis of DM. Furthermore, as the computation methods of LFQ and diabetic marker selection are increasingly growing more specific and sensitive, the application of these advances is an important opportunity to enhance our knowledge of diabetes and identify new therapeutic targets for the discovery of antidiabetic drug. Different data analysis results differ in the choice of manipulation methods and the proper selection of methods is crucial to minimize the biological deviation of the omics datasets [141, 182, 195, 281]. For instance, as reported by the studies, LOG transforms data into a normal distribution, and the application of some parametric statistical tests (including MWW test and t-test) should be based on the assumption of such normal distribution [141]. All in all, these computational advances discussed above provided guidelines for researchers who will engage in proteomics biomarker discovery and by properly applying these proteomic advances, more stable therapeutic protein targets might be discovered in the research of DM.

LIST OF ABBREVIATIONS

ANOVA	=	Analysis of Variance
APMAP	=	Adipocyte Plasma Membrane-Associated Protein
ATO	=	Auto Scaling
BOX	=	Box-Cox
CADD	=	Computer Aided Drug Design
CUB	=	Cube Root
DDA	=	Data-dependent Acquisition
DIA	=	Data-independent Acquisition
DM	=	Diabetes Mellitus
DT	=	Decision Trees
EIG	=	EigenMS
FC	=	Fold Change
KNN	=	K-nearest Neighbor Imputation
LFQ	=	Label-free Quantification
LIMMA	=	Linear Models for Microarray Data
LLS	=	Local Least Squares Imputation
LOG	=	Log Transformation
LOW	=	Locally Weighted Scatterplot Smoothing
MAD	=	Median Absolute Deviation
MDC	=	Median Centering
MEA	=	Mean Normalization
MEC	=	Mean Centering
MED	=	Median Normalization
MWW	=	Mann-Whitney-Wilcoxon test
OPLS-DA	=	Orthogonal Partial Least Square Discriminant Analysis
PAR	=	Pareto Scaling
PLS-DA	=	Partial Least Square Discriminant Analysis
POW	=	Power Transformation
PQN	=	Probabilistic Quotient Normalization
PTM	=	Post-translational Modification
QMs	=	Quantification Measurements
QSTs	=	Quantification Software Tools
RAN	=	Range Scaling
RF	=	Random Forest
RLR	=	Robust Linear Regression
SPLS-DA	=	Sparse Partial Least Square Discriminant Analysis
SVD	=	Singular Value Decomposition
SVM	=	Support Vector Machine
SWATH	=	Sequential Window Acquisition of All Theoretical Mass Spectra
TIC	=	Total Ionic Current
TMM	=	Trimmed Mean of M-values
VAS	=	Vast Scaling
ZSC	=	Z-score
χ^2	=	Chi-square

CONSENT FOR PUBLICATION

Not applicable.

FUNDING

Funded by the National Key Research and Development Program of China (2018YFC0910500), 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 (2018QNA7023, 10611 CDJXZ238826, 2018CDQYSG0007 & CDJZR14468801).

CONFLICT OF INTEREST

The authors declare no conflict of interest, financial or otherwise.

ACKNOWLEDGEMENTS

Declared none.

REFERENCES

[1] Miller BR, Nguyen H, Hu CJ, Lin C, Nguyen QT. New and emerging drugs and targets for type 2 diabetes: reviewing the evidence. Am Health Drug Benefits 2014; 7(8): 452-63. [PMID: 25558307]

[2] Pearson-Stuttard J, Zhou B, Kontis V, Bentham J, Gunter MJ, Ezzati M. Worldwide burden of cancer attributable to diabetes and high body-mass index: a comparative risk assessment. Lancet Diabetes Endocrinol 2018; 6(2): 95-104. [http://dx.doi.org/10.1016/S2213-8587(17)30366-2] [PMID: 29195904]

[3] Kaur P, Mittal A, Nayak SK, Vyas M, Mishra V, Khatik GL. Current strategies and drug targets in the management of type 2 diabetes mellitus. Curr Drug Targets 2018; 19(15): 1738-66. [http://dx.doi.org/10.2174/1389450119666180727142902] [PMID: 30051787]

[4] Gurudeeban S, Satyavani K, Ramanathan T, Balasubramanian T. Antidiabetic effect of a black mangrove species Aegiceras corniculatum in alloxan-induced diabetic rats. J Adv Pharm Technol Res 2012; 3(1): 52-6. [PMID: 22470894]

[5] 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. [http://dx.doi.org/10.1016/j.phymed.2016.03.004] [PMID: 27161404]

[6] Mullard A. 2018 FDA drug approvals. Nat Rev Drug Discov 2019; 18(2): 85-9. [http://dx.doi.org/10.1038/d41573-019-00014-x] [PMID:

30710142]

[7] Grolmusz VI. Identifying diabetes-related important protein targets with few interacting partners with the PageRank algorithm. R Soc Open Sci 2015; 2(4)140252. [http://dx.doi.org/10.1098/rsos.140252] [PMID: 26064627]

[8] 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. [PMID: 29140520]

[9] Wei DQ, Selvaraj G, Kaushik AC. Computational perspective on the current state of the methods and new challenges in cancer drug discovery. Curr Pharm Des 2018; 24(32): 3725-6. [http://dx.doi.org/10.2174/1381612824321901091053 39] [PMID: 30675829]

[10] 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. [http://dx.doi.org/10.1093/nar/gkv1230] [PMID: 26578601]

[11] Nagaraj NS, Singh OV, Merchant NB. Proteomics: a strategy to understand the novel targets in protein misfolding and cancer therapy. Expert Rev Proteomics 2010; 7(4): 613-23. [http://dx.doi.org/10.1586/epr.10.70] [PMID: 20653514]

[12] 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. [http://dx.doi.org/10.1016/j.tips.2017.12.002] [PMID: 29295742]

[13] Kaliamurthi S, Selvaraj G, Chinnasamy S, *et al.* Exploring the papillomaviral proteome to identify potential candidates for a chimeric vaccine against cervix papilloma using immunomics and computational structural vaccinology. Viruses 2019; 11(1)E63. [http://dx.doi.org/10.3390/v11010063] [PMID: 30650527]

[14] López-Villar E, Martos-Moreno GA, Chowen JA, Okada S, Kopchick JJ, Argente J. A proteomic approach to obesity and type 2 diabetes. J Cell Mol Med 2015; 19(7): 1455-70. [http://dx.doi.org/10.1111/jcmm.12600] [PMID: 25960181]

[15] Jayabalan N, Lai A, Nair S, *et al.* Quantitative proteomics by SWATH-MS suggest an association between circulating exosomes and maternal metabolic changes in gestational diabetes mellitus. Proteomics 2019; 19(1-2)e1800164. [PMID: 30536821]

[16] Sylow L, Kleinert M, Richter EA, Jensen TE. Exercise-stimulated glucose uptake - regulation and implications for glycaemic control. Nat Rev Endocrinol 2017; 13(3): 133-48. [http://dx.doi.org/10.1038/nrendo.2016.162] [PMID: 27739515]

[17] Tuerxunyiming M, Xian F, Zi J, *et al.* Quantitative evaluation of serum proteins uncovers a protein signature related to maturity-onset diabetes of the young (MODY). J Proteome Res 2018; 17(1): 670-9. [http://dx.doi.org/10.1021/acs.jproteome.7b00727] [PMID: 29182332]

[18] Peters KE, Davis WA, Ito J, *et al.* Identification of novel circulating biomarkers predicting rapid decline in renal function in type 2 diabetes: the fremantle diabetes study phase II. Diabetes Care 2017; 40(11): 1548-55. [http://dx.doi.org/10.2337/dc17-0911] [PMID: 28851702]

[19] Zhang Q, Fillmore TL, Schepmoes AA, *et al.* Serum proteomics reveals systemic dysregulation of innate immunity in type 1 diabetes. J Exp Med 2013; 210(1): 191-203. [http://dx.doi.org/10.1084/jem.20111843] [PMID: 23277452]

[20] Selvaraj G, Kaliamurthi S, Lin S, Gu K, Wei DQ. Prognostic impact of tissue inhibitor of metalloproteinase-1 in non-small cell lung cancer: systematic review and meta-analysis. Curr Med Chem 2018. [http://dx.doi.org/10.2174/0929867325666180904114455] [PMID: 30182835]

[21] Maris M, Ferreira GB, D'Hertog W, *et al.* High glucose induces dysfunction in insulin secretory cells by different pathways: a proteomic approach. J Proteome Res 2010; 9(12): 6274-87. [http://dx.doi.org/10.1021/pr100557w] [PMID: 20942503]

[22] 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. [http://dx.doi.org/10.1007/s12539-015-0139-2] [PMID: 26750924]

[23] Rondeau P, Navarra G, Militello V, Bourdon E. On the aggregation of albumin: influences of the protein glycation. Protein Aggregation 2011; pp. 139-59.

[24] Toyokuni S, Yamada S, Kashima M, *et al.* Serum 4-hydroxy-2-nonenal-modified albumin is elevated in patients with type 2 diabetes mellitus. Antioxid Redox Signal 2000; 2(4): 681-5. [http://dx.doi.org/10.1089/ars.2000.2.4-681] [PMID: 11213473]

[25] Bhat S, Jagadeeshaprasad MG, Venkatasubramani V, Kulkarni MJ. Abundance matters: role of albumin in diabetes, a proteomics perspective. Expert Rev Proteomics 2017; 14(8): 677-89. [http://dx.doi.org/10.1080/14789450.2017.1352473] [PMID: 28689445]

[26] Westwood ME, Thornalley PJ. Molecular characteristics of methylglyoxal-modified bovine and human serum albumins. Comparison with glucose-derived advanced glycation endproduct-modified serum albumins. J Protein Chem 1995; 14(5): 359-72. [http://dx.doi.org/10.1007/BF01886793] [PMID: 8590604]

[27] Arif B, Ashraf JM, Moinuddin D, Ahmad J, Arif Z, Alam K. Structural and immunological characterization of Amadori-rich human serum albumin: role in diabetes mellitus. Arch Biochem Biophys 2012; 522(1): 17-25. [http://dx.doi.org/10.1016/j.abb.2012.04.005] [PMID: 22516656]

[28] Cohen MP, Sharma K, Jin Y, *et al.* Prevention of diabetic nephropathy in db/db mice with glycated albumin antagonists. A novel treatment strategy. J Clin Invest 1995; 95(5): 2338-45. [http://dx.doi.org/10.1172/JCI117926] [PMID: 7738197]

[29] Lapolla A, Gerhardinger C, Baldo L, *et al.* A study on *in vitro* glycation processes by matrix-assisted laser desorption ionization mass spectrometry. Biochim Biophys Acta 1993; 1225(1): 33-8. [http://dx.doi.org/10.1016/0925-4439(93)90118-K] [PMID: 8241287]

[30] Lapolla A, Fedele D, Reitano R, *et al.* Enzymatic digestion and mass spectrometry in the study of advanced glycation end products/peptides. J Am Soc Mass Spectrom 2004; 15(4): 496-509. [http://dx.doi.org/10.1016/j.jasms.2003.11.014] [PMID: 15047055]

[31] Stefanowicz P, Kijewska M, Kluczyk A, Szewczuk Z. Detection of glycation sites in proteins by high-resolution mass spectrometry combined with isotopic labeling. Anal Biochem 2010; 400(2): 237-43. [http://dx.doi.org/10.1016/j.ab.2010.02.011] [PMID: 20156417]

[32] Bhonsle HS, Korwar AM, Kote SS, *et al.* Low plasma albumin levels are associated with increased plasma protein glycation and HbA1c in diabetes. J Proteome Res 2012; 11(2): 1391-6. [http://dx.doi.org/10.1021/pr201030m] [PMID: 22181049]

[33] Korwar AM, Vannuruswamy G, Jagadeeshaprasad MG, *et al.* Development of diagnostic fragment ion library for glycated peptides of human serum albumin: targeted quantification in prediabetic, diabetic, and microalbuminuria plasma by parallel reaction monitoring, SWATH, and MSE. Mol Cell Proteomics 2015; 14(8): 2150-9. [http://dx.doi.org/10.1074/mcp.M115.050518] [PMID: 26023067]

[34] Soboleva A, Modzel M, Didio A, *et al.* Quantification of prospective type 2 diabetes mellitus biomarkers by stable isotope dilution with bi-labeled standard glycated peptides. Anal Methods 2017; 9: 409-18. [http://dx.doi.org/10.1039/C6AY02483A]

[35] Selvin E, Rawlings AM, Grams M, *et al.* Fructosamine and glycated albumin for risk stratification and prediction of incident diabetes and microvascular complications: a prospective cohort analysis of the Atherosclerosis Risk in Communities (ARIC) study. Lancet Diabetes Endocrinol 2014; 2(4): 279-88. [http://dx.doi.org/10.1016/S2213-8587(13)70199-2] [PMID: 24703046]

[36] Forbes JM, Cooper ME. Mechanisms of diabetic complications. Physiol Rev 2013; 93(1): 137-88. [http://dx.doi.org/10.1152/physrev.00045.2011] [PMID: 23303908]

[37] Tang W, Shi YQ, Zou JJ, *et al.* Serum biomarker of diabetic peripheral neuropathy indentified by differential proteomics. Front Biosci 2011; 16: 2671-81. [http://dx.doi.org/10.2741/3879] [PMID: 21622202]

[38] Loukovaara S, Nurkkala H, Tamene F, *et al.* Quantitative proteomics analysis of vitreous humor from diabetic retinopathy patients. J Proteome Res 2015; 14(12): 5131-43. [http://dx.doi.org/10.1021/acs.jproteome.5b00900] [PMID: 26490944]

[39] Trinh VM, Le VD, Nguyen DV. [The anatomical basis and sources of error of selective and supraselective gastric vagotomies (author's transl)]. J Chir (Paris) 1981; 118(12): 725-38.

[PMID: 7320100]

[40] Papale M, Di Paolo S, Magistroni R, *et al.* Urine proteome analysis may allow noninvasive differential diagnosis of diabetic nephropathy. Diabetes Care 2010; 33(11): 2409-15. [http://dx.doi.org/10.2337/dc10-0345] [PMID: 20671095]

[41] Shameer K, Badgeley MA, Miotto R, Glicksberg BS, Morgan JW, Dudley JT. Translational bioinformatics in the era of real-time biomedical, health care and wellness data streams. Brief Bioinform 2017; 18(1): 105-24. [http://dx.doi.org/10.1093/bib/bbv118] [PMID: 26876889]

[42] Li YH, Li XX, Hong JJ, *et al.* Clinical trials, progression-speed differentiating features and swiftness rule of the innovative targets of first-in-class drugs. Brief Bioinform 2019. [http://dx.doi.org/10.1093/bib/bby130] [PMID: 30689717]

[43] Kaliamurthi S, Demir-Korkmaz A, Selvaraj G, *et al.* Viewing the emphasis on state-of-the-art magnetic nanoparticles: synthesis, physical oroperties, and applications in cancer theranostics. Curr Pharm Des 2019. [http://dx.doi.org/10.2174/1381612825666190523105004] [PMID: 31119998]

[44] Shao S, Guo T, Aebersold R. Mass spectrometry-based proteomic quest for diabetes biomarkers. Biochim Biophys Acta 2015; 1854(6): 519-27. [http://dx.doi.org/10.1016/j.bbapap.2014.12.012] [PMID: 25556002]

[45] Dhanda SK, Usmani SS, Agrawal P, Nagpal G, Gautam A, Raghava GPS. Novel *in silico* tools for designing peptide-based subunit vaccines and immunotherapeutics. Brief Bioinform 2017; 18(3): 467-78. [PMID: 27016393]

[46] Ma W, Zhang L, Zeng P, *et al.* An analysis of human microbe-disease associations. Brief Bioinform 2017; 18(1): 85-97. [http://dx.doi.org/10.1093/bib/bbw005] [PMID: 26883326]

[47] Azuaje F. Computational models for predicting drug responses in cancer research. Brief Bioinform 2017; 18(5): 820-9. [PMID: 27444372]

[48] Nobile MS, Cazzaniga P, Tangherloni A, Besozzi D. Graphics processing units in bioinformatics, computational biology and systems biology. Brief Bioinform 2017; 18(5): 870-85. [PMID: 27402792]

[49] Kaliamurthi S, Selvaraj G, Junaid M, Khan A, Gu K, Wei DQ. Cancer immunoinformatics: a promising era in the development of peptide vaccines for human papillomavirus-induced cervical cancer. Curr Pharm Des 2018; 24(32): 3791-817. [http://dx.doi.org/10.2174/1381612824666181106094133] [PMID: 30398106]

[50] Selvaraj G, Kaliamurthi S, Thirungnasambandam R, Vivekanandan L, Balasubramanian T. Anti-nociceptive effect in mice of thillai flavonoid rutin. Biomed Environ Sci 2014; 27(4): 295-9. [PMID: 24758758]

[51] 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. [http://dx.doi.org/10.1111/cns.13051] [PMID: 30106219]

[52] Li XX, Yin J, Tang J, *et al.* Determining the balance between drug efficacy and safety by the network and biological system profile of its therapeutic target. Front Pharmacol 2018; 9: 1245. [http://dx.doi.org/10.3389/fphar.2018.01245] [PMID: 30429792]

[53] 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. [http://dx.doi.org/10.1039/C7CP07869B] [PMID: 29451287]

[54] Zheng G, Yang F, Fu T, *et al.* Computational characterization of the selective inhibition of human norepinephrine and serotonin transporters by an escitalopram scaffold. Phys Chem Chem Phys 2018; 20(46): 29513-27. [http://dx.doi.org/10.1039/C8CP06232C] [PMID: 30457616]

[55] Zhang Y, Ying JB, Hong JJ, *et al.* How does chirality determine the selective inhibition of histone deacetylase 6? A lesson from trichostatin A enantiomers based on molecular dynamics. ACS Chem Neurosci 2019; 10(5): 2467-80. [http://dx.doi.org/10.1021/acschemneuro.8b00729] [PMID: 30784262]

[56] 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. [http://dx.doi.org/10.1021/acschemneuro.7b00173] [PMID: 28557437]

[57] 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. [http://dx.doi.org/10.1016/j.bbagen.2017.07.022] [PMID: 28757337]

[58] Zhu W, Smith JW, Huang CM. Mass spectrometry-based label-free quantitative proteomics. J Biomed Biotechnol 2010; 2010840518 [http://dx.doi.org/10.1155/2010/840518] [PMID: 19911078]

[59] 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. [http://dx.doi.org/10.1186/1559-0275-11-27] [PMID: 25097465]

[60] 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. [http://dx.doi.org/10.1038/nbt.3685] [PMID: 27701404]

[61] 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. [http://dx.doi.org/10.3389/fphar.2018.00681] [PMID: 29997509]

[62] Yang Y, Gu Y, Wan B, Ren X, Guo LH. Label-free electrochemical biosensing of small-molecule inhibition on O-GlcNAc glycosylation. Biosens Bioelectron 2017; 95: 94-9. [http://dx.doi.org/10.1016/j.bios.2017.04.009] [PMID: 28427012]

[63] Zhang M, Xu W, Deng Y. A new strategy for early diagnosis of type 2 diabetes by standard-free, label-free LC-MS/MS quantification of glycated peptides. Diabetes 2013; 62(11): 3936-42. [http://dx.doi.org/10.2337/db13-0347] [PMID: 23894188]

[64] Geetha T, Langlais P, Luo M, *et al.* Label-free proteomic identification of endogenous, insulin-stimulated interaction partners of insulin receptor substrate-1. J Am Soc Mass Spectrom 2011; 22(3): 457-66. [http://dx.doi.org/10.1007/s13361-010-0051-2] [PMID: 21472564]

[65] do Nascimento de Oliveira V, Lima-Neto ABM, van Tilburg MF, *et al.* Proteomic analysis to identify candidate biomarkers associated with type 1 diabetes. Diabetes Metab Syndr Obes 2018; 11: 289-301. [http://dx.doi.org/10.2147/DMSO.S162008] [PMID: 29942143]

[66] Moulder R, Bhosale SD, Erkkilä T, *et al.* Serum proteomes distinguish children developing type 1 diabetes in a cohort with HLA-conferred susceptibility. Diabetes 2015; 64(6): 2265-78. [http://dx.doi.org/10.2337/db14-0983] [PMID: 25616278]

[67] Ma Y, Gao J, Yin J, *et al.* Identification of a novel function of adipocyte plasma membrane-associated protein (APMAP) in gestational diabetes mellitus by proteomic analysis of omental adipose tissue. J Proteome Res 2016; 15(2): 628-37. [http://dx.doi.org/10.1021/acs.jproteome.5b01030] [PMID: 26767403]

[68] Sandin M, Teleman J, Malmström J, Levander F. Data processing methods and quality control strategies for label-free LC-MS protein quantification. Biochim Biophys Acta 2014; 1844(1 Pt A): 29-41. [http://dx.doi.org/10.1016/j.bbapap.2013.03.026] [PMID: 23567904]

[69] Neilson KA, Ali NA, Muralidharan S, *et al.* Less label, more free: approaches in label-free quantitative mass spectrometry. Proteomics 2011; 11(4): 535-53. [http://dx.doi.org/10.1002/pmic.201000553] [PMID: 21243637]

[70] Xie F, Liu T, Qian WJ, Petyuk VA, Smith RD. Liquid chromatography-mass spectrometry-based quantitative proteomics. J Biol Chem 2011; 286(29): 25443-9. [http://dx.doi.org/10.1074/jbc.R110.199703] [PMID: 21632532]

[71] Tang J, Fu J, Wang Y, *et al.* Simultaneous improvement in the precision, accuracy and robustness of label-free proteome quantification by optimizing data manipulation chains. Mol Cell Proteomics 2019; 18(8): 1683-99. [http://dx.doi.org/10.1074/mcp.RA118.001169] [PMID: 31097671]

[72] Lyutvinskiy Y, Yang H, Rutishauser D, Zubarev RA. *In silico* instrumental response correction improves precision of label-free proteomics and accuracy of proteomics-based predictive models. Mol Cell Proteomics 2013; 12(8): 2324-31.

[http://dx.doi.org/10.1074/mcp.O112.023804] [PMID: 23589346]

[73] Cox J, Hein MY, Luber CA, Paron I, Nagaraj N, Mann M. Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. Mol Cell Proteomics 2014; 13(9): 2513-26.
[http://dx.doi.org/10.1074/mcp.M113.031591] [PMID: 24942700]

[74] 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.
[http://dx.doi.org/10.1080/14789450.2017.1365602] [PMID: 28799854]

[75] Hong JJ, Luo YC, Zhang Y, *et al.* Protein functional annotation of simultaneously improved stability, accuracy and false discovery rate achieved by a sequence-based deep learning. Brief Bioinform 2019.
[http://dx.doi.org/10.1093/bib/bbz081]

[76] Tang J, Fu J, Wang Y, *et al.* ANPELA: analysis and performance assessment of the label-free quantification workflow for metaproteomic studies. Brief Bioinform 2019.
[http://dx.doi.org/10.1093/bib/bby127] [PMID: 30649171]

[77] Leipzig J. A review of bioinformatic pipeline frameworks. Brief Bioinform 2017; 18(3): 530-6.
[PMID: 27013646]

[78] Kuharev J, Navarro P, Distler U, Jahn O, Tenzer S. In-depth evaluation of software tools for data-independent acquisition based label-free quantification. Proteomics 2015; 15(18): 3140-51.
[http://dx.doi.org/10.1002/pmic.201400396] [PMID: 25545627]

[79] 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 ariability of protein abundance. J Proteome Res 2017; 16(4): 1410-24.
[http://dx.doi.org/10.1021/acs.jproteome.6b00645] [PMID: 28217993]

[80] Bubis JA, Levitsky LI, Ivanov MV, Tarasova IA, Gorshkov MV. Comparative evaluation of label-free quantification methods for shotgun proteomics. Rapid Commun Mass Spectrom 2017; 31(7): 606-12.
[http://dx.doi.org/10.1002/rcm.7829] [PMID: 28097710]

[81] Ning K, Fermin D, Nesvizhskii AI. Comparative analysis of different label-free mass spectrometry based protein abundance estimates and their correlation with RNA-Seq gene expression data. J Proteome Res 2012; 11(4): 2261-71.
[http://dx.doi.org/10.1021/pr201052x] [PMID: 22329341]

[82] Chawade A, Sandin M, Teleman J, Malmström J, Levander F. Data processing has major impact on the outcome of quantitative label-free LC-MS analysis. J Proteome Res 2015; 14(2): 676-87.
[http://dx.doi.org/10.1021/pr500665j] [PMID: 25407311]

[83] Nakato R, Shirahige K. Recent advances in ChIP-seq analysis: from quality management to whole-genome annotation. Brief Bioinform 2017; 18(2): 279-90.
[PMID: 26979602]

[84] Goh WW, Wong L. Evaluating feature-selection stability in next-generation proteomics. J Bioinform Comput Biol 2016; 14(5)1650029
[http://dx.doi.org/10.1142/S0219720016500293] [PMID: 27640811]

[85] Xia J, Sinelnikov IV, Han B, Wishart DS. MetaboAnalyst 3.0--making metabolomics more meaningful. Nucleic Acids Res 2015; 43(W1)W251-7
[http://dx.doi.org/10.1093/nar/gkv380] [PMID: 25897128]

[86] Hoekman B, Breitling R, Suits F, Bischoff R, Horvatovich P. msCompare: a framework for quantitative analysis of label-free LC-MS data for comparative candidate biomarker studies. Mol Cell Proteomics 2012; 11(6)015974
[http://dx.doi.org/10.1074/mcp.M111.015974] [PMID: 22318370]

[87] Worachartcheewan A, Nantasenamat C, Isarankura-Na-Ayudhya C, Pidetcha P, Prachayasittikul V. Identification of metabolic syndrome using decision tree analysis. Diabetes Res Clin Pract 2010; 90(1): e15-8.
[http://dx.doi.org/10.1016/j.diabres.2010.06.009] [PMID: 20619912]

[88] Worachartcheewan A, Nantasenamat C, Isarankura-Na-Ayudhya C, Prachayasittikul V. Quantitative population-health relationship (QPHR) for assessing metabolic syndrome. EXCLI J 2013; 12: 569-83.
[PMID: 26622213]

[89] Worachartcheewan A, Nantasenamat C, Prasertsrithong P, *et al.* Machine learning approaches for discerning intercorrelation of hematological parameters and glucose level for identification of diabetes mellitus. EXCLI J 2013; 12: 885-93.
[PMID: 27092034]

[90] Worachartcheewan A, Shoombuatong W, Pidetcha P, Nopnithipat W, Prachayasittikul V, Nantasenamat C. Predicting metabolic syndrome using the random forest method. ScientificWorldJournal 2015; 2015581501
[http://dx.doi.org/10.1155/2015/581501] [PMID: 26290899]

[91] Bagherzadeh-Khiabani F, Ramezankhani A, Azizi F, Hadaegh F, Steyerberg EW, Khalili D. A tutorial on variable selection for clinical prediction models: feature selection methods in data mining could improve the results. J Clin Epidemiol 2016; 71: 76-85.
[http://dx.doi.org/10.1016/j.jclinepi.2015.10.002] [PMID: 26475568]

[92] Georga EI, Protopappas VC, Polyzos D, Fotiadis DI. Evaluation of short-term predictors of glucose concentration in type 1 diabetes combining feature ranking with regression models. Med Biol Eng Comput 2015; 53(12): 1305-18.
[http://dx.doi.org/10.1007/s11517-015-1263-1] [PMID: 25773366]

[93] Spratt HM, Ju H. Statistical approaches to candidate biomarker panel selection. Adv Exp Med Biol 2016; 919: 463-92.
[http://dx.doi.org/10.1007/978-3-319-41448-5_22] [PMID: 27975231]

[94] Liu Q, Sung AH, Qiao M, *et al.* Comparison of feature selection and classification for MALDI-MS data. BMC Genomics 2009; 10(Suppl. 1): S3.
[http://dx.doi.org/10.1186/1471-2164-10-S1-S3] [PMID: 19594880]

[95] Christin C, Hoefsloot HC, Smilde AK, *et al.* A critical assessment of feature selection methods for biomarker discovery in clinical proteomics. Mol Cell Proteomics 2013; 12(1): 263-76.
[http://dx.doi.org/10.1074/mcp.M112.022566] [PMID: 23115301]

[96] Lualdi M, Fasano M. Statistical analysis of proteomics data: A review on feature selection. J Proteomics 2019; 198: 18-26.
[http://dx.doi.org/10.1016/j.jprot.2018.12.004] [PMID: 30529743]

[97] Wang W, Sue AC, Goh WWB. Feature selection in clinical proteomics: with great power comes great reproducibility. Drug Discov Today 2017; 22(6): 912-8.
[http://dx.doi.org/10.1016/j.drudis.2016.12.006] [PMID: 27988358]

[98] Silva JC, Denny R, Dorschel CA, *et al.* Quantitative proteomic analysis by accurate mass retention time pairs. Anal Chem 2005; 77(7): 2187-200.
[http://dx.doi.org/10.1021/ac048455k] [PMID: 15801753]

[99] Hu A, Noble WS, Wolf-Yadlin A. Technical advances in proteomics: new developments in data-independent acquisition. F1000 Res 2016; 5: 5.
[http://dx.doi.org/10.12688/f1000research.7042.1] [PMID: 27092249]

[100] Argentini A, Goeminne LJ, Verheggen K, *et al.* moFF: a robust and automated approach to extract peptide ion intensities. Nat Methods 2016; 13(12): 964-6.
[http://dx.doi.org/10.1038/nmeth.4075] [PMID: 27898063]

[101] Nikolov M, Schmidt C, Urlaub H. Quantitative mass spectrometry-based proteomics: an overview. Methods Mol Biol 2012; 893: 85-100.
[http://dx.doi.org/10.1007/978-1-61779-885-6_7] [PMID: 22665296]

[102] Arike L, Peil L. Spectral counting label-free proteomics. Methods Mol Biol 2014; 1156: 213-22.
[http://dx.doi.org/10.1007/978-1-4939-0685-7_14] [PMID: 24791991]

[103] 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): 17.
[http://dx.doi.org/10.1002/pmic.201600278] [PMID: 28127880]

[104] Aebersold R, Mann M. Mass-spectrometric exploration of proteome structure and function. Nature 2016; 537(7620): 347-55.
[http://dx.doi.org/10.1038/nature19949] [PMID: 27629641]

[105] 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.
[http://dx.doi.org/10.1002/pmic.200700426] [PMID: 18324724]
[106] America AH, Cordewener JH. Comparative LC-MS: a landscape of peaks and valleys. Proteomics 2008; 8(4): 731-49.
[http://dx.doi.org/10.1002/pmic.200700694] [PMID: 18297651]
[107] 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.
[http://dx.doi.org/10.1007/s00216-007-1486-6] [PMID: 17668192]
[108] 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.
[http://dx.doi.org/10.1021/pr700758r] [PMID: 18173218]
[109] Bringans SD, Ito J, Stoll T, *et al.* Comprehensive mass spectrometry based biomarker discovery and validation platform as applied to diabetic kidney disease. EuPA Open Proteom 2017; 14: 1-10.
[http://dx.doi.org/10.1016/j.euprot.2016.12.001] [PMID: 29900119]
[110] Johnson LA, Zuloaga KL, Kugelman TL, *et al.* Amelioration of metabolic syndrome-associated cognitive impairments in mice *via* a reduction in dietary fat content or infusion of non-diabetic plasma. EBioMedicine 2015; 3: 26-42.
[http://dx.doi.org/10.1016/j.ebiom.2015.12.008] [PMID: 26870815]
[111] Vähätupa M, Nättinen J, Jylhä A, *et al.* SWATH-MS proteomic analysis of oxygen-induced retinopathy reveals novel potential therapeutic targets. Invest Ophthalmol Vis Sci 2018; 59(8): 3294-306.
[http://dx.doi.org/10.1167/iovs.18-23831] [PMID: 30025079]
[112] Banarjee R, Sharma A, Bai S, Deshmukh A, Kulkarni M. Proteomic study of endothelial dysfunction induced by AGEs and its possible role in diabetic cardiovascular complications. J Proteomics 2018; 187: 69-79.
[http://dx.doi.org/10.1016/j.jprot.2018.06.009] [PMID: 29935336]
[113] Ludwig C, Gillet L, Rosenberger G, Amon S, Collins BC, Aebersold R. Data-independent acquisition-based SWATH-MS for quantitative proteomics: a tutorial. Mol Syst Biol 2018; 14(8)e8126
[http://dx.doi.org/10.15252/msb.20178126] [PMID: 30104418]
[114] 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.
[http://dx.doi.org/10.1038/nmeth.3255] [PMID: 25599550]
[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.
[http://dx.doi.org/10.1002/pmic.201500526] [PMID: 27246681]
[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.
[http://dx.doi.org/10.1074/mcp.M114.044305] [PMID: 25724911]
[117] Meyer JG, Mukkamalla S, Steen H, Nesvizhskii AI, Gibson BW, Schilling B. PIQED: automated identification and quantification of protein modifications from DIA-MS data. Nat Methods 2017; 14(7): 646-7.
[http://dx.doi.org/10.1038/nmeth.4334] [PMID: 28661500]
[118] Lawrence RT, Searle BC, Llovet A, Villén J. Plug-and-play analysis of the human phosphoproteome by targeted high-resolution mass spectrometry. Nat Methods 2016; 13(5): 431-4.
[http://dx.doi.org/10.1038/nmeth.3811] [PMID: 27018578]
[119] Wu L, Amon S, Lam H. A hybrid retention time alignment algorithm for SWATH-MS data. Proteomics 2016; 16(15-16): 2272-83.
[http://dx.doi.org/10.1002/pmic.201500511] [PMID: 27302277]
[120] 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.
[http://dx.doi.org/10.1038/nbt.2841] [PMID: 24727770]
[121] Terfve C, Sabidó E, Wu Y, *et al.* System-wide quantitative proteomics of the metabolic syndrome in mice: genotypic and dietary effects. J Proteome Res 2017; 16(2): 831-41.
[http://dx.doi.org/10.1021/acs.jproteome.6b00815] [PMID: 27936760]
[122] Välikangas T, Suomi T, Elo LL. A comprehensive evaluation of popular proteomics software workflows for label-free proteome quantification and imputation. Brief Bioinform 2018; 19(6): 1344-55.
[PMID: 28575146]
[123] Rafiei A, Sleno L. Comparison of peak-picking workflows for untargeted liquid chromatography/high-resolution mass spectrometry metabolomics data analysis. Rapid Commun Mass Spectrom 2015; 29(1): 119-27.
[http://dx.doi.org/10.1002/rcm.7094] [PMID: 25462372]
[124] Li CY, Song HT, Liu SJ, *et al.* Systematic screening and characterization of astragalosides in an oral solution of Radix Astragali by liquid chromatography with quadrupole time-of-flight mass spectrometry and Peakview software. J Sep Sci 2016; 39(6): 1099-109.
[http://dx.doi.org/10.1002/jssc.201501278] [PMID: 27027590]
[125] 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.
[http://dx.doi.org/10.1021/acs.jproteome.6b00767] [PMID: 27995803]
[126] 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.
[http://dx.doi.org/10.1002/elps.201500527] [PMID: 26935773]
[127] Pino LK, Searle BC, Bollinger JG, Nunn B, MacLean B, MacCoss MJ. The Skyline ecosystem: Informatics for quantitative mass spectrometry proteomics. Mass Spectrom Rev 2017.
[http://dx.doi.org/10.1002/mas.21540] [PMID: 28691345]
[128] 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.
[http://dx.doi.org/10.1093/bioinformatics/btq054] [PMID: 20147306]
[129] Henderson CM, Shulman NJ, MacLean B, MacCoss MJ, Hoofnagle AN. Skyline performs as well as Vendor software in the quantitative analysis of serum 25-hydroxy vitamin D and vitamin D binding globulin. Clin Chem 2018; 64(2): 408-10.
[http://dx.doi.org/10.1373/clinchem.2017.282293] [PMID: 29203474]
[130] Dwinovan J, Colella AD, Chegeni N, Chataway TK, Sokoya EM. Proteomic analysis reveals downregulation of housekeeping proteins in the diabetic vascular proteome. Acta Diabetol 2017; 54(2): 171-90.
[http://dx.doi.org/10.1007/s00592-016-0929-y] [PMID: 27796656]
[131] Zhang X, Sun H, Paul SK, *et al.* The serum protein responses to treatment with Xiaoke Pill and Glibenclamide in type 2 diabetes patients. Clin Proteomics 2017; 14: 19.
[http://dx.doi.org/10.1186/s12014-017-9154-0] [PMID: 28529466]
[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.
[http://dx.doi.org/10.1002/pmic.201500488] [PMID: 27213465]
[133] Niersmann C, Hauck SM, Kannenberg JM, *et al.* Omentin-regulated proteins combine a pro-inflammatory phenotype with an anti-inflammatory counterregulation in human adipocytes: A proteomics analysis. Diabetes Metab Res Rev 2019; 35(1)e3074
[http://dx.doi.org/10.1002/dmrr.3074] [PMID: 30198166]
[134] 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.
[http://dx.doi.org/10.1021/pr300992u] [PMID: 23391308]
[135] Sturm M, Bertsch A, Gröpl C, *et al.* OpenMS - an open-source software framework for mass spectrometry. BMC Bioinformatics 2008; 9: 163.
[http://dx.doi.org/10.1186/1471-2105-9-163] [PMID: 18366760]
[136] Weisser H, Choudhary JS. Targeted feature detection for data-dependent shotgun proteomics. J Proteome Res 2017; 16(8): 2964-74.
[http://dx.doi.org/10.1021/acs.jproteome.7b00248] [PMID: 28673088]
[137] Cisek K, Krochmal M, Klein J, Mischak H. The application of multi-omics and systems biology to identify therapeutic targets in chronic kidney disease. Nephrol Dial Transplant 2016; 31(12): 2003-11.
[http://dx.doi.org/10.1093/ndt/gfv364] [PMID: 26487673]
[138] Skeie JM, Aldrich BT, Goldstein AS, Schmidt GA, Reed CR, Greiner MA. Proteomic analysis of corneal endothelial cell-

descemet membrane tissues reveals influence of insulin dependence and disease severity in type 2 diabetes mellitus. PLoS One 2018; 13(3)e0192287
[http://dx.doi.org/10.1371/journal.pone.0192287] [PMID: 29529022]

[139] Tran NH, Qiao R, Xin L, *et al.* Deep learning enables de novo peptide sequencing from data-independent-acquisition mass spectrometry. Nat Methods 2019; 16(1): 63-6.
[http://dx.doi.org/10.1038/s41592-018-0260-3] [PMID: 30573815]

[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(4)010587
[http://dx.doi.org/10.1074/mcp.M111.010587] [PMID: 22186715]

[141] Di Guida R, Engel J, Allwood JW, *et al.* Non-targeted UHPLC-MS metabolomic data processing methods: a comparative investigation of normalisation, missing value imputation, transformation and scaling. Metabolomics 2016; 12: 93.
[http://dx.doi.org/10.1007/s11306-016-1030-9] [PMID: 27123000]

[142] 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.
[http://dx.doi.org/10.1007/s00216-016-9482-3] [PMID: 27055884]

[143] Qi D, Brownridge P, Xia D, *et al.* A software toolkit and interface for performing stable isotope labeling and top3 quantification using Progenesis LC-MS. OMICS 2012; 16(9): 489-95.
[http://dx.doi.org/10.1089/omi.2012.0042] [PMID: 22888986]

[144] Roverso M, Brioschi M, Banfi C, *et al.* A preliminary study on human placental tissue impaired by gestational diabetes: a comparison of gel-based *versus* gel-free proteomics approaches. Eur J Mass Spectrom (Chichester) 2016; 22(2): 71-82.
[http://dx.doi.org/10.1255/ejms.1412] [PMID: 27419900]

[145] Ly A, Scheerer MF, Zukunft S, *et al.* Retinal proteome alterations in a mouse model of type 2 diabetes. Diabetologia 2014; 57(1): 192-203.
[http://dx.doi.org/10.1007/s00125-013-3070-2] [PMID: 24078137]

[146] Li J, Lu Q, Lu P. Quantitative proteomics analysis of vitreous body from type 2 diabetic patients with proliferative diabetic retinopathy. BMC Ophthalmol 2018; 18(1): 151.
[http://dx.doi.org/10.1186/s12886-018-0821-3] [PMID: 29940965]

[147] 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.
[http://dx.doi.org/10.1021/pr2005154] [PMID: 21714566]

[148] Tao L, Zhu F, Qin C, *et al.* Nature's contribution to today's pharmacopeia. Nat Biotechnol 2014; 32(10): 979-80.
[http://dx.doi.org/10.1038/nbt.3034] [PMID: 25299914]

[149] 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.
[http://dx.doi.org/10.1021/acs.jproteome.6b00407] [PMID: 27476824]

[150] Aiyetan P, Zhang B, Chen L, Zhang Z, Zhang H. M2Lite: An open-source, light-weight, pluggable and fast proteome discoverer MSF to mzIdentML tool. J Bioinform 2014; 1(2): 40-9.
[PMID: 25346941]

[151] Preil SA, Kristensen LP, Beck HC, *et al.* Quantitative proteome analysis reveals increased content of basement membrane proteins in arteries from patients with type 2 diabetes mellitus and lower levels among metformin users. Circ Cardiovasc Genet 2015; 8(5): 727-35.
[http://dx.doi.org/10.1161/CIRCGENETICS.115.001165] [PMID: 26371159]

[152] Zubiri I, Posada-Ayala M, Sanz-Maroto A, *et al.* Diabetic nephropathy induces changes in the proteome of human urinary exosomes as revealed by label-free comparative analysis. J Proteomics 2014; 96: 92-102.
[http://dx.doi.org/10.1016/j.jprot.2013.10.037] [PMID: 24211404]

[153] Park SK, Yates JR III. Census for proteome quantification. Curr Protoc Bioinformatics 2010; Chapter 13: 1-11.
[http://dx.doi.org/10.1002/0471250953.bi1312s29] [PMID: 20205189]

[154] Park SK, Venable JD, Xu T, Yates JR III. A quantitative analysis software tool for mass spectrometry-based proteomics. Nat Methods 2008; 5(4): 319-22.
[http://dx.doi.org/10.1038/nmeth.1195] [PMID: 18345006]

[155] Deng WJ, Nie S, Dai J, Wu JR, Zeng R. Proteome, phosphoproteome, and hydroxyproteome of liver mitochondria in diabetic rats at early pathogenic stages. Mol Cell Proteomics 2010; 9(1): 100-16.
[http://dx.doi.org/10.1074/mcp.M900020-MCP200] [PMID: 19700791]

[156] 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.
[http://dx.doi.org/10.1080/08927022.2017.1309653]

[157] Cociorva D, L Tabb D, Yates JR. Validation of tandem mass spectrometry database search results using DTASelect. Curr Protoc Bioinformatics 2007; Chapter 13: 4.
[http://dx.doi.org/10.1002/0471250953.bi1304s16] [PMID: 18428785]

[158] 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.
[http://dx.doi.org/10.1021/pr015504q] [PMID: 12643522]

[159] 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.
[http://dx.doi.org/10.1021/acs.jproteome.6b00376] [PMID: 27537616]

[160] Dominguez E, Galmozzi A, Chang JW, *et al.* Integrated phenotypic and activity-based profiling links Ces3 to obesity and diabetes. Nat Chem Biol 2014; 10(2): 113-21.
[http://dx.doi.org/10.1038/nchembio.1429] [PMID: 24362705]

[161] Tyanova S, Temu T, Cox J. The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. Nat Protoc 2016; 11(12): 2301-19.
[http://dx.doi.org/10.1038/nprot.2016.136] [PMID: 27809316]

[162] 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.
[http://dx.doi.org/10.1002/pmic.201400449] [PMID: 25644178]

[163] Zhang L, Liu CW, Zhang Q. Online 2D-LC-MS/MS platform for analysis of glycated proteome. Anal Chem 2018; 90(2): 1081-6.
[http://dx.doi.org/10.1021/acs.analchem.7b03342] [PMID: 29281256]

[164] Liu CW, Atkinson MA, Zhang Q. Type 1 diabetes cadaveric human pancreata exhibit a unique exocrine tissue proteomic profile. Proteomics 2016; 16(9): 1432-46.
[http://dx.doi.org/10.1002/pmic.201500333] [PMID: 26935967]

[165] Codrea MC, Nahnsen S. Platforms and pipelines for proteomics data analysis and management. Adv Exp Med Biol 2016; 919: 203-15.
[http://dx.doi.org/10.1007/978-3-319-41448-5_9] [PMID: 27975218]

[166] Hussey SE, Sharoff CG, Garnham A, *et al.* Effect of exercise on the skeletal muscle proteome in patients with type 2 diabetes. Med Sci Sports Exerc 2013; 45(6): 1069-76.
[http://dx.doi.org/10.1249/MSS.0b013e3182814917] [PMID: 23274603]

[167] Ahn JM, Kim BG, Yu MH, Lee IK, Cho JY. Identification of diabetic nephropathy-selective proteins in human plasma by multilectin affinity chromatography and LC-MS/MS. Proteomics Clin Appl 2010; 4(6-7): 644-53.
[http://dx.doi.org/10.1002/prca.200900196] [PMID: 21137083]

[168] Zhao M, Li M, Yang Y, *et al.* A comprehensive analysis and annotation of human normal urinary proteome. Sci Rep 2017; 7(1): 3024.
[http://dx.doi.org/10.1038/s41598-017-03226-6] [PMID: 28596590]

[169] 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.
[http://dx.doi.org/10.1586/14789450.5.6.819] [PMID: 19086861]

[170] 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.
[http://dx.doi.org/10.1016/j.atmosenv.2013.01.038] [PMID:

23606806]
[171] 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.
[http://dx.doi.org/10.1039/C7CP05688E] [PMID: 29057413]
[172] Mäkinen VP, Tynkkynen T, Soininen P, *et al.* Metabolic diversity of progressive kidney disease in 325 patients with type 1 diabetes (the FinnDiane Study). J Proteome Res 2012; 11(3): 1782-90.
[http://dx.doi.org/10.1021/pr201036j] [PMID: 22204613]
[173] Kahleova H, Klementova M, Herynek V, *et al.* The effect of a vegetarian *vs* conventional hypocaloric diabetic diet on thigh adipose tissue distribution in subjects with type 2 diabetes: a randomized study. J Am Coll Nutr 2017; 36(5): 364-9.
[http://dx.doi.org/10.1080/07315724.2017.1302367] [PMID: 28604251]
[174] De Livera AM, Dias DA, De Souza D, *et al.* Normalizing and integrating metabolomics data. Anal Chem 2012; 84(24): 10768-76.
[http://dx.doi.org/10.1021/ac302748b] [PMID: 23150939]
[175] Selvaraj G, Kaliamurthi S, Kaushik AC, *et al.* Identification of target gene and prognostic evaluation for lung adenocarcinoma using gene expression meta-analysis, network analysis and neural network algorithms. J Biomed Inform 2018; 86: 120-34.
[http://dx.doi.org/10.1016/j.jbi.2018.09.004] [PMID: 30195659]
[176] Lau CH, Muniandy S. Adiponectin and resistin gene polymorphisms in association with their respective adipokine levels. Ann Hum Genet 2011; 75(3): 370-82.
[http://dx.doi.org/10.1111/j.1469-1809.2010.00635.x] [PMID: 21323646]
[177] Fattuoni C, Mandò C, Palmas F, *et al.* Preliminary metabolomics analysis of placenta in maternal obesity. Placenta 2018; 61: 89-95.
[http://dx.doi.org/10.1016/j.placenta.2017.11.014] [PMID: 29277276]
[178] 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.
[http://dx.doi.org/10.1002/chem.201503496] [PMID: 26836577]
[179] Lucena S, Varela Coelho A, Anjo SI, *et al.* Comparative proteomic analysis of saliva from dogs with and without obesity-related metabolic dysfuntion. J Proteomics 2019; 201: 65-72.
[http://dx.doi.org/10.1016/j.jprot.2019.04.010] [PMID: 30991157]
[180] Sakia RM. The Box-Cox transformation technique - a review. J R Stat Soc 1992; 41: 169-78.
[181] Bruderer R, Muntel J, Müller S, *et al.* Analysis of 1508 plasma samples by capillary flow data-independent acquisition profiles proteomics of weight loss and maintenance. Mol Cell Proteomics 2019; 18(6): 1242-54.
[http://dx.doi.org/10.1074/mcp.RA118.001288] [PMID: 30948622]
[182] Välikangas T, Suomi T, Elo LL. A systematic evaluation of normalization methods in quantitative label-free proteomics. Brief Bioinform 2018; 19(1): 1-11.
[PMID: 27694351]
[183] 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.
[http://dx.doi.org/10.1186/1471-2164-7-142] [PMID: 16762068]
[184] Qi Y, Jiang C, Cheng J, *et al.* Bile acid signaling in lipid metabolism: metabolomic and lipidomic analysis of lipid and bile acid markers linked to anti-obesity and anti-diabetes in mice. Biochim Biophys Acta 2015; 1851(1): 19-29.
[http://dx.doi.org/10.1016/j.bbalip.2014.04.008] [PMID: 24796972]
[185] Guha Mazumder A, Chatterjee S, Chatterjee S, *et al.* Spectropathology-corroborated multimodal quantitative imaging biomarkers for neuroretinal degeneration in diabetic retinopathy. Clin Ophthalmol 2017; 11: 2073-89.
[http://dx.doi.org/10.2147/OPTH.S140110] [PMID: 29200821]
[186] Mundt F, Johansson HJ, Forshed J, *et al.* Proteome screening of pleural effusions identifies galectin 1 as a diagnostic biomarker and highlights several prognostic biomarkers for malignant mesothelioma. Mol Cell Proteomics 2014; 13(3): 701-15.
[http://dx.doi.org/10.1074/mcp.M113.030775] [PMID: 24361865]
[187] Brennan L. NMR-based metabolomics: from sample preparation to applications in nutrition research. Prog Nucl Magn Reson Spectrosc 2014; 83: 42-9.
[http://dx.doi.org/10.1016/j.pnmrs.2014.09.001] [PMID: 25456316]
[188] 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.
[http://dx.doi.org/10.1007/s11306-011-0350-z] [PMID: 22593726]
[189] 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.
[http://dx.doi.org/10.1007/s11306-014-0738-7]
[190] Jiménez-Lucena R, Rangel-Zúñiga OA, Alcalá-Díaz JF, *et al.* Circulating miRNAs as predictive biomarkers of type 2 diabetes mellitus development in coronary heart disease patients from the CORDIOPREV study. Mol Ther Nucleic Acids 2018; 12: 146-57.
[http://dx.doi.org/10.1016/j.omtn.2018.05.002] [PMID: 30195754]
[191] Rosenling T, Stoop MP, Smolinska A, *et al.* The impact of delayed storage on the measured proteome and metabolome of human cerebrospinal fluid. Clin Chem 2011; 57(12): 1703-11.
[http://dx.doi.org/10.1373/clinchem.2011.167601] [PMID: 21998343]
[192] Marco-Ramell A, Tulipani S, Palau-Rodriguez M, *et al.* Untargeted profiling of concordant/discordant phenotypes of high insulin resistance and obesity to predict the risk of developing diabetes. J Proteome Res 2018; 17(7): 2307-17.
[http://dx.doi.org/10.1021/acs.jproteome.7b00855] [PMID: 29905079]
[193] Smilde AK, van der Werf MJ, Bijlsma S, van der Werff-van der Vat BJ, Jellema RH. Fusion of mass spectrometry-based metabolomics data. Anal Chem 2005; 77(20): 6729-36.
[http://dx.doi.org/10.1021/ac051080y] [PMID: 16223263]
[194] Klont F, Bras L, Wolters JC, *et al.* Assessment of sample preparation bias in mass spectrometry-based proteomics. Anal Chem 2018; 90(8): 5405-13.
[http://dx.doi.org/10.1021/acs.analchem.8b00600] [PMID: 29608294]
[195] Chawade A, Alexandersson E, Levander F. Normalyzer: a tool for rapid evaluation of normalization methods for omics data sets. J Proteome Res 2014; 13(6): 3114-20.
[http://dx.doi.org/10.1021/pr401264n] [PMID: 24766612]
[196] 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.
[http://dx.doi.org/10.1093/nar/gkx449] [PMID: 28525573]
[197] 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.
[http://dx.doi.org/10.1038/srep38881] [PMID: 27958387]
[198] Kim SJ, Chae S, Kim H, *et al.* A protein profile of visceral adipose tissues linked to early pathogenesis of type 2 diabetes mellitus. Mol Cell Proteomics 2014; 13(3): 811-22.
[http://dx.doi.org/10.1074/mcp.M113.035501] [PMID: 24403596]
[199] 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.
[http://dx.doi.org/10.1021/acs.analchem.6b02848] [PMID: 27792874]
[200] Noonan MJ, Tinnesand HV, Buesching CD. Normalizing gas-chromatography-mass spectrometry data: method choice can alter biological inference. BioEssays 2018; 40(6)e1700210
[http://dx.doi.org/10.1002/bies.201700210] [PMID: 29709068]
[201] Wulff JE, Mitchell MW. A comparison of various normalization methods for LC/MS metabolomics data. Adv Biosci Biotechnol 2018; 9: 339.
[http://dx.doi.org/10.4236/abb.2018.98022]
[202] Meng X, Li X, Zhang P, Wang J, Zhou Y, Chen M. Circular RNA: an emerging key player in RNA world. Brief Bioinform 2017; 18(4): 547-57.
[PMID: 27255916]
[203] Manjunatha S, Distelmaier K, Dasari S, Carter RE, Kudva YC, Nair KS. Functional and proteomic alterations of plasma high density lipoproteins in type 1 diabetes mellitus. Metabolism 2016; 65(9): 1421-31.
[http://dx.doi.org/10.1016/j.metabol.2016.06.008] [PMID: 27506748]

[204] García-Fontana B, Morales-Santana S, Díaz Navarro C, *et al.* Metabolomic profile related to cardiovascular disease in patients with type 2 diabetes mellitus: A pilot study. Talanta 2016; 148: 135-43. [http://dx.doi.org/10.1016/j.talanta.2015.10.070] [PMID: 26653434]

[205] Cho WC, Yip TT, Chung WS, Leung AW, Cheng CH, Yue KK. Differential expression of proteins in kidney, eye, aorta, and serum of diabetic and non-diabetic rats. J Cell Biochem 2006; 99(1): 256-68. [http://dx.doi.org/10.1002/jcb.20923] [PMID: 16598775]

[206] Rao Y, Lee Y, Jarjoura D, *et al.* A comparison of normalization techniques for microRNA microarray data. Stat Appl Genet Mol Biol 2008; 7(1)e22 [http://dx.doi.org/10.2202/1544-6115.1287] [PMID: 18673291]

[207] Wang W, Zhou H, Lin H, *et al.* Quantification of proteins and metabolites by mass spectrometry without isotopic labeling or spiked standards. Anal Chem 2003; 75(18): 4818-26. [http://dx.doi.org/10.1021/ac026468x] [PMID: 14674459]

[208] Xu J, Bai J, Zhang X, *et al.* A comprehensive overview of lncRNA annotation resources. Brief Bioinform 2017; 18(2): 236-49. [PMID: 26944085]

[209] 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. [http://dx.doi.org/10.1074/mcp.M800514-MCP200] [PMID: 19596695]

[210] Hajduk J, Klupczynska A, Dereziński P, *et al.* A combined metabolomic and proteomic analysis of gestational diabetes mellitus. Int J Mol Sci 2015; 16(12): 30034-45. [http://dx.doi.org/10.3390/ijms161226133] [PMID: 26694367]

[211] 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. [http://dx.doi.org/10.1093/bioinformatics/btr479] [PMID: 21852304]

[212] Chung N, Zhang XD, Kreamer A, *et al.* Median absolute deviation to improve hit selection for genome-scale RNAi screens. J Biomol Screen 2008; 13(2): 149-58. [http://dx.doi.org/10.1177/1087057107312035] [PMID: 18216396]

[213] Hill JA, Szabat M, Hoesli CA, *et al.* A multi-parameter, high-content, high-throughput screening platform to identify natural compounds that modulate insulin and Pdx1 expression. PLoS One 2010; 5(9)e12958 [http://dx.doi.org/10.1371/journal.pone.0012958] [PMID: 20886041]

[214] 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. [http://dx.doi.org/10.1021/ac051632c] [PMID: 16808434]

[215] Mao X, Chen X, Chen C, Zhang H, Law KP. Metabolomics in gestational diabetes. Clin Chim Acta 2017; 475: 116-27. [http://dx.doi.org/10.1016/j.cca.2017.10.019] [PMID: 29066210]

[216] 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. [http://dx.doi.org/10.1093/bioinformatics/btp426] [PMID: 19602524]

[217] Karpievitch YV, Nikolic SB, Wilson R, Sharman JE, Edwards LM. Metabolomics data normalization with EigenMS. PLoS One 2014; 9(12)e116221 [http://dx.doi.org/10.1371/journal.pone.0116221] [PMID: 25549083]

[218] Xia S, Feng J, Lei L, *et al.* Comprehensive characterization of tissue-specific circular RNAs in the human and mouse genomes. Brief Bioinform 2017; 18(6): 984-92. [PMID: 27543790]

[219] Law KP, Zhang H. The pathogenesis and pathophysiology of gestational diabetes mellitus: Deductions from a three-part longitudinal metabolomics study in China. Clin Chim Acta 2017; 468: 60-70. [http://dx.doi.org/10.1016/j.cca.2017.02.008] [PMID: 28213010]

[220] Yang YH, Dudoit S, Luu P, Speed TP. Normalization for cDNA microarray data. Microarrays: optical Technologies And Informatics 2001; 4266: 141-52.

[221] Stratz C, Nührenberg T, Fiebich BL, *et al.* Controlled type II diabetes mellitus has no major influence on platelet micro-RNA expression. Results from micro-array profiling in a cohort of 60 patients. Thromb Haemost 2014; 111(5): 902-11. [http://dx.doi.org/10.1160/TH13-06-0476] [PMID: 24352417]

[222] Kirchberg FF, Harder U, Weber M, *et al.* European Childhood Obesity Trial Study Group. Dietary protein intake affects amino acid and acylcarnitine metabolism in infants aged 6 months. J Clin Endocrinol Metab 2015; 100(1): 149-58. [http://dx.doi.org/10.1210/jc.2014-3157] [PMID: 25368978]

[223] 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. [http://dx.doi.org/10.1186/s12864-015-2353-z] [PMID: 26732976]

[224] Väremo L, Henriksen TI, Scheele C, *et al.* Type 2 diabetes and obesity induce similar transcriptional reprogramming in human myocytes. Genome Med 2017; 9(1): 47. [http://dx.doi.org/10.1186/s13073-017-0432-2] [PMID: 28545587]

[225] Spiller S, Li Y, Blüher M, Welch L, Hoffmann R. Glycated lysine-141 in haptoglobin improves the diagnostic accuracy for type 2 diabetes mellitus in combination with glycated hemoglobin HbA_{1c} and fasting plasma glucose. Clin Proteomics 2017; 14: 10. [http://dx.doi.org/10.1186/s12014-017-9145-1] [PMID: 28360826]

[226] Cheadle C, Vawter MP, Freed WJ, Becker KG. Analysis of microarray data using Z score transformation. J Mol Diagn 2003; 5(2): 73-81. [http://dx.doi.org/10.1016/S1525-1578(10)60455-2] [PMID: 12707371]

[227] Yang Q, Li B, Tang J, *et al.* Consistent gene signature of schizophrenia identified by a novel feature selection strategy from comprehensive sets of transcriptomic data. Brief Bioinform 2019.bbz049 [http://dx.doi.org/10.1093/bib/bbz049] [PMID: 31157371]

[228] Kryuchkova-Mostacci N, Robinson-Rechavi M. A benchmark of gene expression tissue-specificity metrics. Brief Bioinform 2017; 18(2): 205-14. [PMID: 26891983]

[229] 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. [http://dx.doi.org/10.1021/pr300227y] [PMID: 22845868]

[230] Karpievitch YV, Dabney AR, Smith RD. Normalization and missing value imputation for label-free LC-MS analysis. BMC Bioinformatics 2012; 13(Suppl. 16): S5. [http://dx.doi.org/10.1186/1471-2105-13-S16-S5] [PMID: 23176322]

[231] 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. [PMID: 24876803]

[232] Cui X, Yang Q, Li B, *et al.* Assessing the effectiveness of direct data merging strategy in long-term and large-scale pharmacometabonomics. Front Pharmacol 2019; 10: 127. [http://dx.doi.org/10.3389/fphar.2019.00127] [PMID: 30842738]

[233] Jiang N, Liu HF, Li SD, *et al.* An integrated metabonomic and proteomic study on Kidney-Yin Deficiency Syndrome patients with diabetes mellitus in China. Acta Pharmacol Sin 2015; 36(6): 689-98. [http://dx.doi.org/10.1038/aps.2014.169] [PMID: 25937635]

[234] 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. [http://dx.doi.org/10.1073/pnas.97.18.10101] [PMID: 10963673]

[235] Law KP, Mao X, Han TL, Zhang H. Unsaturated plasma phospholipids are consistently lower in the patients diagnosed with gestational diabetes mellitus throughout pregnancy: A longitudinal metabolomics study of Chinese pregnant women part 1. Clin Chim Acta 2017; 465: 53-71. [http://dx.doi.org/10.1016/j.cca.2016.12.010] [PMID: 27988319]

[236] Lehmann R, Friedrich T, Krebiehl G, *et al.* Metabolic profiles during an oral glucose tolerance test in pregnant women with and without gestational diabetes. Exp Clin Endocrinol Diabetes 2015; 123(7): 483-38. [http://dx.doi.org/10.1055/s-0035-1549887] [PMID: 26171623]

[237] Song X, Waitman LR, Hu Y, Yu ASL, Robins D, Liu M. Robust

clinical marker identification for diabetic kidney disease with ensemble feature selection. J Am Med Inform Assoc 2019; 26(3): 242-53.
[http://dx.doi.org/10.1093/jamia/ocy165] [PMID: 30602020]

[238] Peddinti G, Cobb J, Yengo L, *et al.* Early metabolic markers identify potential targets for the prevention of type 2 diabetes. Diabetologia 2017; 60(9): 1740-50.
[http://dx.doi.org/10.1007/s00125-017-4325-0] [PMID: 28597074]

[239] Min S, Lee B, Yoon S. Deep learning in bioinformatics. Brief Bioinform 2017; 18(5): 851-69.
[PMID: 27473064]

[240] Labani M, Moradi P, Ahmadizar F, Jalili M. A novel multivariate filter method for feature selection in text classification problems. Eng Appl Artif Intell 2018; 70: 25-37.
[http://dx.doi.org/10.1016/j.engappai.2017.12.014]

[241] Tang J, Wang Y, Fu J, *et al.* A critical assessment of the feature selection methods used for biomarker discovery in current metaproteomics studies. Brief Bioinform 2019.bbz061
[http://dx.doi.org/10.1093/bib/bbz061] [PMID: 31197323]

[242] Sánchez-Maroño N, Alonso-Betanzos A, Tombilla-Sanromán M. Filter methods for feature selection–a comparative study. Inter Conf Intel Data Eng and Auto Learning 2007; pp. 178-87.
[http://dx.doi.org/10.1007/978-3-540-77226-2_19]

[243] Cibrik DM, Warner RL, Kommareddi M, Song P, Luan FL, Johnson KJ. Identification of a protein signature in renal allograft rejection. Proteomics Clin Appl 2013; 7(11-12): 839-49.
[http://dx.doi.org/10.1002/prca.201200036] [PMID: 24323459]

[244] Wright DB. Comparing groups in a before-after design: when t test and ANCOVA produce different results. Br J Educ Psychol 2006; 76(Pt 3): 663-75.
[http://dx.doi.org/10.1348/000709905X52210] [PMID: 16953968]

[245] Militello G, Weirick T, John D, Döring C, Dimmeler S, Uchida S. Screening and validation of lncRNAs and circRNAs as miRNA sponges. Brief Bioinform 2017; 18(5): 780-8.
[PMID: 27373735]

[246] Mao P, Wang D. Top-down proteomics of a drop of blood for diabetes monitoring. J Proteome Res 2014; 13(3): 1560-9.
[http://dx.doi.org/10.1021/pr401074t] [PMID: 24533899]

[247] Kempthorne O. The correlation between relatives on the supposition of mendelian inheritance. Sci T R So 1919; 52: 399-433.

[248] McHugh ML. Multiple comparison analysis testing in ANOVA. Biochem Med (Zagreb) 2011; 21(3): 203-9.
[http://dx.doi.org/10.11613/BM.2011.029] [PMID: 22420233]

[249] Pritchard CC, Hsu L, Delrow J, Nelson PS. Project normal: defining normal variance in mouse gene expression. Proc Natl Acad Sci USA 2001; 98(23): 13266-71.
[http://dx.doi.org/10.1073/pnas.221465998] [PMID: 11698685]

[250] Roscioni SS, de Zeeuw D, Hellemons ME, *et al.* A urinary peptide biomarker set predicts worsening of albuminuria in type 2 diabetes mellitus. Diabetologia 2013; 56(2): 259-67.
[http://dx.doi.org/10.1007/s00125-012-2755-2] [PMID: 23086559]

[251] Chee CS, Chang KM, Loke MF, Angela Loo VP, Subrayan V. Association of potential salivary biomarkers with diabetic retinopathy and its severity in type-2 diabetes mellitus: a proteomic analysis by mass spectrometry. PeerJ 2016; 4e2022.
[http://dx.doi.org/10.7717/peerj.2022] [PMID: 27280065]

[252] Koletsi D, Pandis N. The chi-square test for trend. Am J Orthod Dentofacial Orthop 2016; 150(6): 1066-7.
[http://dx.doi.org/10.1016/j.ajodo.2016.10.001] [PMID: 27894529]

[253] García-Fontana B, Morales-Santana S, Longobardo V, *et al.* Relationship between proinflammatory and antioxidant proteins with the severity of cardiovascular disease in type 2 diabetes mellitus. Int J Mol Sci 2015; 16(5): 9469-83.
[http://dx.doi.org/10.3390/ijms16059469] [PMID: 25923078]

[254] Zhang W, Wang H, Guan X, Niu Q, Li W. Variant rs2237892 of KCNQ1 is potentially associated with hypertension and macrovascular complications in type 2 diabetes mellitus in a Chinese Han population. Genomics Proteomics Bioinformatics 2015; 13(6): 364-70.
[http://dx.doi.org/10.1016/j.gpb.2015.05.004] [PMID: 26678516]

[255] Whitney J. Testing for differences with the nonparametric Mann-Whitney U test. J Wound Ostomy Continence Nurs 1997; 24(1): 12.
[http://dx.doi.org/10.1016/S1071-5754(97)90044-9] [PMID: 9213501]

[256] Marx A, Backes C, Meese E, Lenhof HP, Keller A. EDISON-WMW: exact dynamic programing solution of the Wilcoxon-Mann-Whitney test. Genomics Proteomics Bioinformatics 2016; 14(1): 55-61.
[http://dx.doi.org/10.1016/j.gpb.2015.11.004] [PMID: 26829645]

[257] Tang Y. Size and power estimation for the Wilcoxon-Mann-Whitney test for ordered categorical data. Stat Med 2011; 30(29): 3461-70.
[http://dx.doi.org/10.1002/sim.4407] [PMID: 22086799]

[258] Soboleva A, Mavropulo-Stolyarenko G, Grishina T, *et al.* Individual glycation sites in blood plasma proteins–prospective biomarkers of type 2 diabetes mellitus. Young AGErs Symposium. 2017: 24.

[259] Han Z, Xue W, Tao L, Zhu F. Identification of key long noncoding RNAs in the pathology of Alzheimer's disease and their functions based on genome-wide associations study, microarray, and RNA-seq data. J Alzheimers Dis 2019; 68(1): 339-55.
[http://dx.doi.org/10.3233/JAD-181051] [PMID: 30776002]

[260] Han Z, Xue W, Tao L, Lou Y, Qiu Y, Zhu F. Genome-wide identification and analysis of the eQTL lncRNAs in multiple sclerosis based on RNA-seq data. Brief Bioinform 2019.bbz036
[http://dx.doi.org/10.1093/bib/bbz036] [PMID: 31323688]

[261] Pena MJ, Lambers Heerspink HJ, Hellemons ME, *et al.* Urine and plasma metabolites predict the development of diabetic nephropathy in individuals with Type 2 diabetes mellitus. Diabet Med 2014; 31(9): 1138-47.
[http://dx.doi.org/10.1111/dme.12447] [PMID: 24661264]

[262] Ritchie ME, Phipson B, Wu D, *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res 2015; 43(7)e47
[http://dx.doi.org/10.1093/nar/gkv007] [PMID: 25605792]

[263] Wang Z, Wang Z, Zhou Z, Ren Y. Crucial genes associated with diabetic nephropathy explored by microarray analysis. BMC Nephrol 2016; 17(1): 128.
[http://dx.doi.org/10.1186/s12882-016-0343-2] [PMID: 27613243]

[264] Wold S, Sjostrom M, Eriksson L. PLS-regression: a basic tool of chemometrics. Chemometr Intel Lab 2001; 58: 109-30.
[http://dx.doi.org/10.1016/S0169-7439(01)00155-1]

[265] Wong KH, Razmovski-Naumovski V, Li KM, Li GQ, Chan K. Differentiation of Pueraria lobata and Pueraria thomsonii using partial least square discriminant analysis (PLS-DA). J Pharm Biomed Anal 2013; 84: 5-13.
[http://dx.doi.org/10.1016/j.jpba.2013.05.040] [PMID: 23777642]

[266] Vitova L, Tuma Z, Moravec J, Kvapil M, Matejovic M, Mares J. Early urinary biomarkers of diabetic nephropathy in type 1 diabetes mellitus show involvement of kallikrein-kinin system. BMC Nephrol 2017; 18(1): 112.
[http://dx.doi.org/10.1186/s12882-017-0519-4] [PMID: 28359252]

[267] Bylesjö M, Rantalainen M, Cloarec O, *et al.* OPLS discriminant analysis: combining the strengths of PLS-DA and SIMCA classification. J Chemometr 2006; 20: 341-51.
[http://dx.doi.org/10.1002/cem.1006]

[268] Smilowitz JT, Totten SM, Huang J, *et al.* Human milk secretory immunoglobulin a and lactoferrin N-glycans are altered in women with gestational diabetes mellitus. J Nutr 2013; 143(12): 1906-12.
[http://dx.doi.org/10.3945/jn.113.180695] [PMID: 24047700]

[269] Jiang M, Wang C, Zhang Y, Feng Y, Wang Y, Zhu Y. Sparse partial-least-squares discriminant analysis for different geographical origins of Salvia miltiorrhiza by (1) H-NMR-based metabolomics. Phytochem Anal 2014; 25(1): 50-8.
[http://dx.doi.org/10.1002/pca.2461] [PMID: 23868756]

[270] Lê Cao KA, Boitard S, Besse P. Sparse PLS discriminant analysis: biologically relevant feature selection and graphical displays for multiclass problems. BMC Bioinformatics 2011; 12: 253.
[http://dx.doi.org/10.1186/1471-2105-12-253] [PMID: 21693065]

[271] Gavin PG, Mullaney JA, Loo D, *et al.* Intestinal metaproteomics reveals host-microbiota interactions in subjects at risk for type 1 diabetes. Diabetes Care 2018; 41(10): 2178-86.
[http://dx.doi.org/10.2337/dc18-0777] [PMID: 30100563]

[272] Sun Y, Todorovic S, Goodison S. Local-learning-based feature selection for high-dimensional data analysis. IEEE Trans Pattern Anal Mach Intell 2010; 32(9): 1610-26.
[http://dx.doi.org/10.1109/TPAMI.2009.190] [PMID: 20634556]

[273] Wang Y, Makedon FS, Ford JC, Pearlman J. HykGene: a hybrid approach for selecting marker genes for phenotype classification using microarray gene expression data. Bioinformatics 2005; 21(8): 1530-7.

[http://dx.doi.org/10.1093/bioinformatics/bti192] [PMID: 15585531]

[274] 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. [http://dx.doi.org/10.1021/acschemneuro.7b00490] [PMID: 29300091]

[275] Zhu F, Han B, Kumar P, *et al.* Update of TTD: therapeutic target database. Nucleic Acids Res 2010; 38(Database issue): D787-91. [http://dx.doi.org/10.1093/nar/gkp1014] [PMID: 19933260]

[276] Chen T, Cao Y, Zhang Y, *et al.* Random forest in clinical metabolomics for phenotypic discrimination and biomarker selection. Evid Based Complement Alternat Med 2013; 2013298183 [http://dx.doi.org/10.1155/2013/298183] [PMID: 23573122]

[277] Sandler V, Reisetter AC, Bain JR, *et al.* HAPO Study Cooperative Research Group. Associations of maternal BMI and insulin resistance with the maternal metabolome and newborn outcomes. Diabetologia 2017; 60(3): 518-30. [http://dx.doi.org/10.1007/s00125-016-4182-2] [PMID: 27981358]

[278] Yu W, Liu T, Valdez R, Gwinn M, Khoury MJ. Application of support vector machine modeling for prediction of common diseases: the case of diabetes and pre-diabetes. BMC Med Inform Decis Mak 2010; 10: 16. [http://dx.doi.org/10.1186/1472-6947-10-16] [PMID: 20307319]

[279] Bottou L, Vapnik V. Local Learning Algorithms. Neural Comput 1992; 4: 888-900. [http://dx.doi.org/10.1162/neco.1992.4.6.888]

[280] Chen X, Yan CC, Zhang X, You ZH. Long non-coding RNAs and complex diseases: from experimental results to computational models. Brief Bioinform 2017; 18(4): 558-76. [PMID: 27345524]

[281] Rausch TK, Schillert A, Ziegler A, Lüking A, Zucht HD, Schulz-Knappe P. Comparison of pre-processing methods for multiplex bead-based immunoassays. BMC Genomics 2016; 17(1): 601. [http://dx.doi.org/10.1186/s12864-016-2888-7] [PMID: 27515389]