



# Discovery of the Consistently Well-Performed Analysis Chain for SWATH-MS Based Pharmacoproteomic Quantification

Jianbo Fu<sup>1</sup>, Jing Tang<sup>1,2</sup>, Yunxia Wang<sup>1</sup>, Xuejiao Cui<sup>1,2</sup>, Qingxia Yang<sup>1,2</sup>, Jiajun Hong<sup>1</sup>, Xiaoxu Li<sup>1,2</sup>, Shuang Li<sup>1,2</sup>, Yuzong Chen<sup>3</sup>, Weiwei Xue<sup>2</sup> and Feng Zhu<sup>1,2\*</sup>

<sup>1</sup> College of Pharmaceutical Sciences, Zhejiang University, Hangzhou, China, <sup>2</sup> School of Pharmaceutical Sciences and Collaborative Innovation Center for Brain Science, Chongqing University, Chongqing, China, <sup>3</sup> Bioinformatics and Drug Design Group, Department of Pharmacy, Center for Computational Science and Engineering, National University of Singapore, Singapore, Singapore

## OPEN ACCESS

### Edited by:

Zhi-Liang Ji,  
Xiamen University, China

### Reviewed by:

Dariusz Mrozek,  
Silesian University of Technology,  
Poland

Qing-Chuan Zheng,  
Jilin University, China

### \*Correspondence:

Feng Zhu  
zhufeng@zju.edu.cn;  
zhufeng.ns@gmail.com

### Specialty section:

This article was submitted to  
Translational Pharmacology,  
a section of the journal  
Frontiers in Pharmacology

**Received:** 27 April 2018

**Accepted:** 05 June 2018

**Published:** 26 June 2018

### Citation:

Fu J, Tang J, Wang Y, Cui X, Yang Q, Hong J, Li X, Li S, Chen Y, Xue W and Zhu F (2018) Discovery of the Consistently Well-Performed Analysis Chain for SWATH-MS Based Pharmacoproteomic Quantification. *Front. Pharmacol.* 9:681. doi: 10.3389/fphar.2018.00681

Sequential windowed acquisition of all theoretical fragment ion mass spectra (SWATH-MS) has emerged as one of the most popular techniques for label-free proteome quantification in current pharmacoproteomic research. It provides more comprehensive detection and more accurate quantitation of proteins comparing with the traditional techniques. The performance of SWATH-MS is highly susceptible to the selection of processing method. Till now,  $\geq 27$  methods (transformation, normalization, and missing-value imputation) are sequentially applied to construct numerous analysis chains for SWATH-MS, but it is still not clear which analysis chain gives the optimal quantification performance. Herein, the performances of 560 analysis chains for quantifying pharmacoproteomic data were comprehensively assessed. Firstly, the most complete set of the publicly available SWATH-MS based pharmacoproteomic data were collected by comprehensive literature review. Secondly, substantial variations among the performances of various analysis chains were observed, and the consistently well-performed analysis chains (CWPACs) across various datasets were for the first time generalized. Finally, the log and power transformations sequentially followed by the total ion current normalization were discovered as one of the best performed analysis chains for the quantification of SWATH-MS based pharmacoproteomic data. In sum, the CWPACs identified here provided important guidance to the quantification of proteomic data and could therefore facilitate the cutting-edge research in any pharmacoproteomic studies requiring SWATH-MS technique.

**Keywords:** pharmacoproteomics, SWATH-MS, processing method, transformation, normalization

## INTRODUCTION

The pharmacoproteomics has been widely applied to various aspects of current pharmaceutical researches by discovering disease-related genes (Mrozek et al., 2013; Quiros et al., 2017; Zeng et al., 2017) or new drug targets (Li et al., 2018; Saei et al., 2018), constructing pharmacology screening model (Hauser et al., 2005), and revealing the drug mechanism of action (Yue et al., 2016; Zhu et al., 2018), resistance (Paul et al., 2016), and toxicity (Tan et al., 2017; Wang et al., 2017b). Recent findings uncover its potentials to fulfill the promise that the pharmacogenomics has not accomplished yet (D'Alessandro and Zolla, 2010; Chambliss and Chan, 2016; Yang et al., 2016).

As a newly emerging technique (Anjo et al., 2017), the *sequential windowed acquisition of all theoretical fragment ion mass spectra* (SWATH-MS) has been reported to provide much more comprehensive detection and accurate quantitation of proteins compared to the traditional techniques used in pharmacoproteomic analyses (Zhu et al., 2008b; Tao et al., 2015; Aebersold and Mann, 2016; Li et al., 2016a; Anjo et al., 2017), and it thus becomes one of the most popular techniques for target discovery (Li et al., 2016b; Xu et al., 2016; Anjo et al., 2017), drug/lead quantification (Roemmelt et al., 2015) and identification (Scheidweiler et al., 2015; Wang et al., 2015; Aratyn-Schaus and Ramanathan, 2016; Li B. et al., 2017), construction of assay library for targeted proteomic analysis (Schubert et al., 2015), and quantitative protein profiling (Krasny et al., 2018) for recognizing drug-induced alterations (Roemmelt et al., 2015; Xue et al., 2016).

However, due to the interdependent nature among multiple acquisition parameters (dwell time, duty cycle, precursor isolation window width, and mass range), the protein quantification based on SWATH-MS is reported to be limited in dynamic range (Anjo et al., 2017) and in turn low in accuracy (Gillet et al., 2012; Huang et al., 2015; Shi et al., 2016; Yang et al., 2017; Xue et al., 2018b). The problems above can be even worse considering the innate complexity of clinical samples (Jamwal et al., 2017), small amount of proteins (Sajic et al., 2015), and low abundance of drug-metabolizing enzymes (Jamwal et al., 2017). To cope with these problems, a variety of popular quantification tools, including *DIA-Umpire* (Sajic et al., 2015), *OpenSWATH* (Rost et al., 2014), *Skyline* (MacLean et al., 2010), *Spectronaut* (Bruederer et al., 2015), and *SWATH2.0* (Li S. et al., 2017), and dozens of subsequent processing methods (transformation, normalization, and missing-value imputation) are developed to enhance the accuracy of SWATH-MS (Navarro et al., 2016). Recent reports further reveal that SWATH-MS' accuracies depend heavily on the specific quantification tool/processing method used in a particular study (Navarro et al., 2016), and the protein quantification can significantly benefit from comparative benchmarking of the performance of these tools and methods (Gatto et al., 2016; Zheng et al., 2016). Therefore, it is urgently needed to assess the performances of tools/methods for discovering the optimal one(s) for SWATH-MS based pharmacoproteomic studies.

The performance of various quantification tools has already been systematically evaluated by benchmark SWATH-MS data (Navarro et al., 2016). Among those tools, only 2 (*OpenSWATH* and *Skyline*) are non-commercial ones, and the *OpenSWATH* (Rost et al., 2014) is of the most popular one used to quantify SWATH-MS based pharmacoproteomic data (Rost et al., 2014; Parker et al., 2015; Weisser and Choudhary, 2017). So far,  $\geq 4$  transformation,  $\geq 15$  normalization, and  $\geq 6$  missing-value imputation algorithms (Guo et al., 2015; Li et al., 2016c; Ori et al., 2016; Wu et al., 2016; Tan et al., 2017; Wang et al., 2017a) have been sequentially applied to process pharmacoproteomic data. Among these algorithms, four for normalizing label-free proteomic data have been assessed to identify the best performed one (Callister et al., 2006) and six for missing-value imputation have been evaluated to discover the one enhancing proteomic

quantifications in the differential expression analysis (Valikangas et al., 2017). Appropriate integrations of the processing methods into a sequential analysis chain are reported to improve the quantification accuracies (Karpievitch et al., 2012; Chawade et al., 2015; Valikangas et al., 2017) with some chains identified as highly accurate in particular pharmacoproteomic studies (Guo et al., 2015; Ori et al., 2016; Tan et al., 2017; Zheng et al., 2017). For example, log transformation followed by median normalization performs well in identifying the therapeutic target/pathway for *Down syndrome* (Sullivan et al., 2017), endogenous toxins inducing the haploinsufficiency of tumor suppressor (Tan et al., 2017) and biological mechanism underlying the role of proteins played in *Alzheimer's disease* (Khoonsari et al., 2016). Since the processing methods are sequentially used to form the integrated analysis chain (Guo et al., 2015; Ori et al., 2016; Tan et al., 2017), any performance assessment aiming solely at transformation, normalization, or imputation may not be able to reflect the overall performance of the whole analysis chain. Considering the huge amount of possible analysis chains [560 in total, taking non-transformation, non-normalization, and non-imputation into account adopted by previous studies (Guo et al., 2015; Liu et al., 2015; Wu et al., 2016)] by randomly integrating those processing methods, it is therefore essential to comprehensively evaluate the performance of all analysis chains to identify the optimal one for specific pharmacoproteomic dataset. However, no such analysis has been conducted yet.

In this study, the performances of all possible analysis chains integrating 4 transformation, 15 normalization, and 6 imputation algorithms were comprehensively assessed by their precisions based on the proteomes among replicates (Kuharev et al., 2015; Navarro et al., 2016; Chignell et al., 2018; Muller et al., 2018). Systematic literature review on the popular quantification tool *OpenSWATH* firstly yielded seven SWATH-MS based benchmark pharmacoproteomic datasets of varied sample sizes (from 6 to 116). To the best of our knowledge, these seven provided the most complete set of the publicly available pharmacoproteomic data based on the SWATH-MS technique. Secondly, the performance of analysis chains was assessed by each dataset. Thirdly, the analysis chains consistently performed well across all datasets were identified for the first time and compared with those popular chains frequently applied in current pharmacoproteomic studies. Finally, the consistently well-performed analysis chains were further discussed based on their performances. The analysis chains identified in and the corresponding findings of this study provided important guidance to current pharmacoproteomic studies.

## MATERIALS AND METHODS

### Collection of SWATH-MS Based Benchmark Pharmacoproteomic Datasets

A systematic literature review on the popular quantification tool *OpenSWATH* and the analysis on the datasets provided in the PRIDE database (Navarro et al., 2016) were

collectively conducted to find SWATH-MS based benchmark pharmacoproteomic datasets. Firstly, PRIDE database was searched against by keyword “SWATH-MS.” Together with the literature review on the resulting projects, 85 projects were identified as based on SWATH-MS, among which 76 and 9 projects were acquired by TripleTOF instruments 5600 and 6600, respectively. Secondly, several criteria were used to guarantee the availability and processability of the raw proteomic data, which included (1) complete set of raw data files, (2) well-defined parameters (isolation scheme, range of retention time, and transition settings), (3) availability of spectral library and protein database to search against, and (4) clear description on sample groups. The application of these criteria on the resulting PRIDE projects yielded seven SWATH-MS based benchmark pharmacoproteomic datasets of varied sample sizes (Table 1), which covered both TripleTOF instruments (5600 and 6600) of all 85 projects. Therefore, these datasets can be recognized as representatives of SWATH-MS based pharmacoproteomic data. To the best of our knowledge, these datasets provided the most complete set of SWATH-MS based pharmacoproteomic data.

## Processing Methods for Data Transformation, Normalization, and Imputation

So far,  $\geq 4$  transformation,  $\geq 15$  normalization, and  $\geq 6$  missing-value imputation algorithms (Guo et al., 2015; Li et al., 2016c; Ori et al., 2016; Wu et al., 2016; Tan et al., 2017; Wang et al., 2017a) have been reported to be sequentially and frequently used to process pharmacoproteomic data. Based on our comprehensive literature review, their corresponding applications to current proteomic research were discussed in Supplementary Method S1. These 25 methods include 4 **transformation**: *Box-cox* (Sakia, 1992), *Cube Root* (Wen et al., 2017), *Log* (De Livera et al., 2012), and *Power* (Zhang, 2014), 15 **normalization**: *Auto Scaling* (Kohl et al., 2012), *Cyclic Loess* (Zhu et al., 2012b), *EigenMS*

(Zhu et al., 2009), *Locally Weighted Scatterplot Smoothing* (Wilson et al., 2003), *Mean* (Andjelkovic and Thompson, 2006), *Median* (Bolstad et al., 2003), *Median Absolute Deviation* (Matzke et al., 2011), *Pareto* (Zhu et al., 2010), *Probabilistic Quotient* (Dieterle et al., 2006), *Quantile* (Callister et al., 2006), *Robust Linear Regression* (Hong et al., 2016), *Total Ion Current* (Gaspari et al., 2016), *Trimmed Mean of M Values* (Lin et al., 2016), *VSN* (Huber et al., 2002), and *Z-score* (Cheadle et al., 2003), and 6 **imputation**: *Background* (Chai et al., 2014), *Bayesian Principal* (Chai et al., 2014), *Censored* (Valikangas et al., 2017), *K-nearest Neighbor* (Zhu et al., 2008a), *Singular Value Decomposition* (Alter et al., 2000), and *Zero Imputation* (Gan et al., 2006). As shown in the Supplementary Method S1, due to their popularity in current pharmacoproteomic studies, these 25 methods were included, sequentially applied, and analyzed in this study. Each method was abbreviated by a three-letter code which was demonstrated in Supplementary Table S1.

## Assessing Analysis Chain Using the Precision Based on Proteomes Among Replicates

Diverse methods for proteomic data processing (transformation, normalization, and imputation) profoundly affected the precision of protein quantification which was frequently assessed using the value of pooled intragroup median absolute deviation (PMAD) of reported protein intensity among replicates (Chawade et al., 2014; Kuharev et al., 2015; Valikangas et al., 2018; Yu et al., 2018). Particularly, the PMAD was designed to demonstrate the capacity of each analysis chain to reduce the variation among replicates, and therefore to enhance the technical reproducibility (Chawade et al., 2014). The lower value of PMAD denoted the more thorough removal of the experimentally induced noise and indicated better precision of the corresponding analysis chain (Valikangas et al., 2018). So far, PMAD value within the range of  $\leq 0.3$ ,  $> 0.3$  &  $\leq 0.7$ , and  $> 0.7$  was generally accepted as with

**TABLE 1** | Seven SWATH-MS based benchmark pharmacoproteomic datasets collected for the analysis of this study.

Datasets	PRIDE ID	Sample size and Dataset description	Analysis Chain	Instrument
<i>Nat. Biotechnol.</i> 34:1130-6, 2016	PXD002952	3 samples of 65% human, 30% yeast, and 5% <i>E. coli</i> proteins 3 samples of 65% human, 15% yeast, and 20% <i>E. coli</i> proteins	LOG-MED-???	TripleTOF 6600
<i>Cell Rep.</i> 20:1229-41, 2017	PXD003278	6 siRNA-treated Cal51 cell samples 6 PRPF8-depleted Cal51 cell samples	LOG-QUA-NON	TripleTOF 5600
<i>Cell.</i> 169:1105-18, 2017	PXD006106	10 formaldehyde treated HeLa Kyoto cell samples 10 formaldehyde untreated HeLa Kyoto cell samples	LOG-MED-NON	TripleTOF 5600
<i>Nat Med.</i> 21:407-13, 2015	PXD000672	18 tumorous kidney tissue biopsies 18 non-tumorous kidney tissue biopsies	LOG-QUA-NON	TripleTOF 5600
<i>Sci Rep.</i> 7:14818, 2017	PXD004880	18 plasma samples from individuals with <i>Down syndrome</i> 18 plasma samples from healthy controls	LOG-MED-NON	TripleTOF 5600
<i>Cell Rep.</i> 18:3219-26, 2017	PXD003972	20 wild type mouse samples 20 knock-in mouse samples expressing endogenous GRB2	LOG-??-???	TripleTOF 5600
<i>Mol Syst. Biol.</i> 11:786, 2015	PXD001064	72 blood samples of monozygotic twins 44 blood samples of dizygotic twins	??-RLR-BAK	TripleTOF 5600

All datasets were from PRIDE database (Navarro et al., 2016). Each method in the analysis chain was abbreviated by a three-letter code as demonstrated in Supplementary Table S1, and ??? indicated that the corresponding method was not specified in the corresponding study of the dataset.

superior, good, and poor precision, respectively (Chawade et al., 2014; Valikangas et al., 2018), which had gradually become a popular metric for assessing the precision of processing methods in OMICs (Chawade et al., 2014; Valikangas et al., 2018).

## Performance Assessment Among Various Analysis Chains by Hierarchical Clustering

Pooled intragroup median absolute deviation values of 560 possible analysis chains across the seven benchmark datasets were firstly calculated. Fifty-one out of these 560 analysis chains reported error for processing at least one of the benchmark datasets. Therefore, the hierarchical clustering of the remaining 509 analysis chains with calculatable results of all seven PMADs was conducted to identify the relationship among the performances of various analysis chains. Particularly, PMAD values of a specific analysis chain among 7 datasets were used to form a 7-dimensional vector. Then, hierarchical clustering was applied to investigate the relationship among those 509 vectors, and therefore among the corresponding analysis chains. To measure the distance between any 2 vectors, the *Euclidean distance* was adopted, which could be demonstrated as below:

$$\text{Euclidean distance } (a, b) = \sqrt{\sum_{i=1}^n (a_i - b_i)^2}$$

where  $i$  denoted each dimension of the analysis chain  $a$  and  $b$ . The clustering algorithm applied here was Ward's minimum variance algorithm (Barer and Harwood, 1999), which was designed to minimize the total within-cluster variance. Ward's minimum variance module in R package (Tippmann, 2015) was used. To visualize the hierarchical tree graph among those 509 analysis chains, the tree generator *iTOL* was used to generate and display the hierarchical tree structure (Letunic and Bork, 2016).

## RESULTS AND DISCUSSION

### Ranking the Analysis Chains Based on Their Performances on Each Benchmark

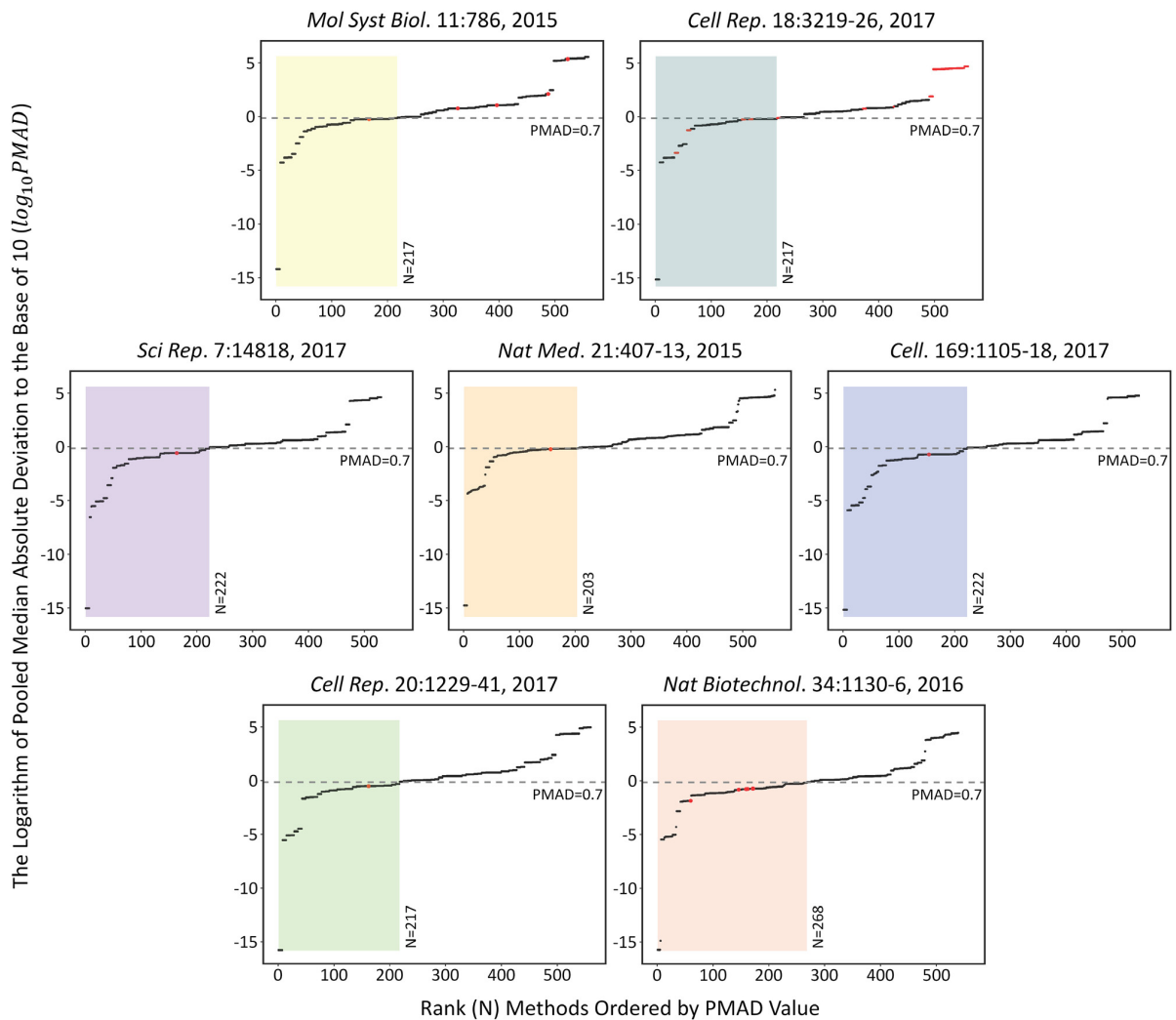
The performances of each analysis chain on the seven SWATH-MS based benchmark datasets (Table 1) were assessed by measuring the corresponding PMAD values. As shown in Figure 1, the performances of 509 analysis chains ( $\log_{10}$  PMAD, Y-axis) with calculatable PMAD values were measured and ranked (X-axis). Because some analysis chains may not be able to result in a PMAD value, there were slight variations among the number of analysis chains for different benchmark datasets (from 530 to 560). Taking the dataset shown in the center of Figure 1 as an example (Nat Med. 21:407-13, 2015), a total of 558 analysis chains were assessed and ranked, and the performance of different analysis chains varied significantly (PMAD from  $1.8 \times 10^{-15}$  to  $2.0 \times 10^5$ ). With reference to the frequently adopted cutoff (PMAD = 0.7) for differentiating the analysis chains of good and poor precision (Chawade et al., 2014; Valikangas et al., 2018), 203 (36.4%) out of these 558 analysis chains were ranked as well-performed. Similar to this dataset

(Nat Med. 21:407-13, 2015), the performance of different analysis chains for the other datasets also differentiated substantially (PMAD from  $1.7 \times 10^{-16}$  to  $3.4 \times 10^5$ ) with 38.8%~49.7% of the analysis chains ranked as well-performed.

The specific analysis chains for each benchmark dataset adopted in the corresponding original studies were identified by literature review (Table 1). Particularly, 4 out of these datasets were with the clearly defined analysis chain (LOG-QUA-NON, LOG-MED-NON, LOG-QUA-NON, and LOG-MED-NON for PXD003278, PXD006106, PXD000672, and PXD004880, respectively), while the remaining 3 datasets were with incomplete information of the adopted analysis chain (LOG-MED-???, LOG-??-???, and ???-RLR-BAK for the datasets of PXD002952, PXD003972, and PXD001064, respectively). Taking the same dataset in the middle of Figure 1 as an example (Nat Med. 21:407-13, 2015), the red dot indicated the PMAD of the analysis chain adopted by this study and its corresponding ranking among all 558 analysis chains. As shown, the adopted chain (LOG-QUA-NON) in this study was ranked to be the 156th well-performed one (PMAD = 0.598) showing its capacity to reduce variations among replicates and thus enhance technical reproducibility (Chawade et al., 2014). However, there were 155 chains performed better than the adopted one (PMAD from  $1.8 \times 10^{-15}$  to 0.595) with POW-TMM-ZER chain performed the best. Similar to this example dataset, the analysis chains adopted by the corresponding studies of PXD003278, PXD006106, and PXD004880 were ranked 162nd, 154th, and 164th well-performed ones, which demonstrated appropriate selection of analysis chain in previous studies. However, there were still more than a hundred chains performed better than the adopted ones, which may further enhance the accuracy of SWATH-MS based protein quantification. For the studies with incomplete information of the adopted chain (PXD002952, PXD003972, and PXD001064), the possible integrations based on the known information were highlighted by multiple red dots. 1 (20%) out of 5, 28 (25%) out of 112, and 7 (100%) out of 7 integrations were within the ranges of well-performance for PXD002952, PXD003972, and PXD001064, respectively.

### Analysis Chains Consistently Well-Performed Across All Benchmark Datasets

The performances of 20 representative analysis chains across different datasets were illustrated in Figure 2. PMAD within the ranges of  $\leq 0.3$ ,  $> 0.3$  &  $\leq 0.7$ , and  $> 0.7$  was generally accepted as with superior, good, and poor performance, respectively (Chawade et al., 2014; Valikangas et al., 2018), which was illustrated by a circle of various diameters (the smaller diameter denoted the lower PMAD value). As shown, the performances of specific chain among various datasets varied significantly. Particularly, the LOG-PQN-BPC performed superior, good, and poor in 3, 3, and 1 datasets, respectively, and POW-ZSC-ZER performed superior, good, and poor in 1, 5, and 1 datasets, respectively. These results demonstrated a certain level of variations among the seven datasets for each analysis chain. However, as shown in Figure 2, there were some chains

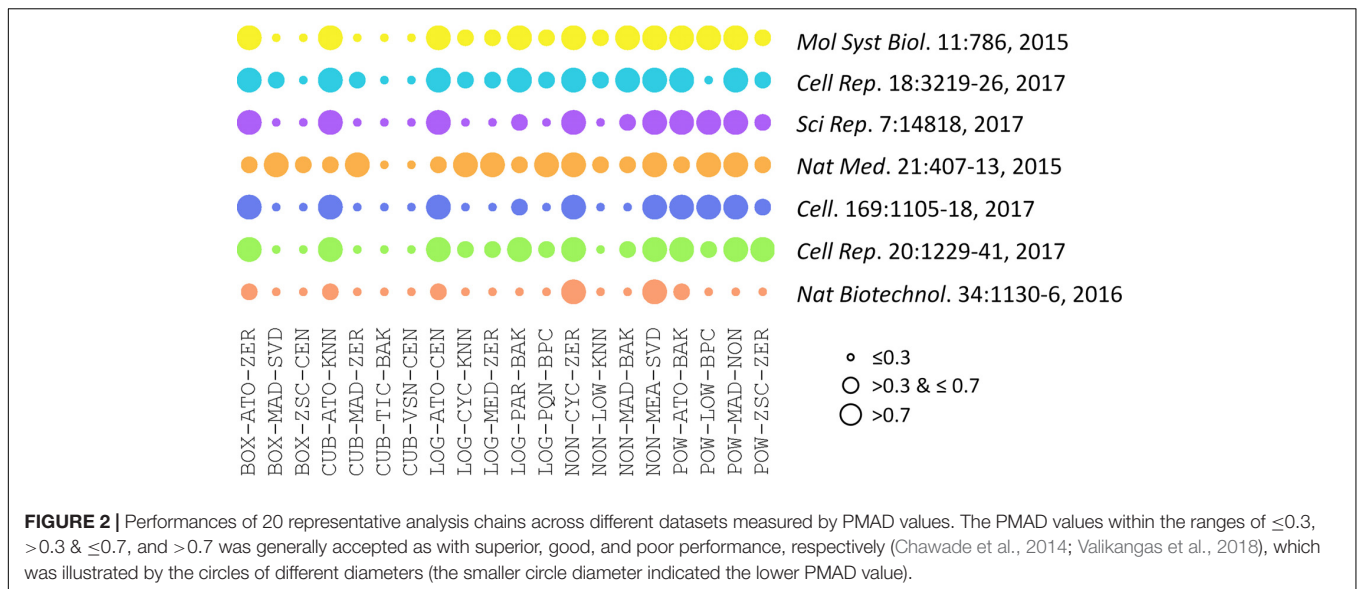


**FIGURE 1** | The performances of each analysis chain on those seven SWATH-MS based benchmark datasets assessed by measuring the corresponding PMAD values [ $>500$  analysis chains ( $\log_{10}$  PMAD, Y-axis) were measured and ranked (X-axis)]. Since some analysis chains may not be able to result in a specific PMAD value, there were slight variations among the number of analysis chains for different benchmark datasets (from 530 to 560). Detail information on these seven datasets were provided in **Table 1**.

performed consistently across different benchmark datasets. For instance, CUB-TIC-BAK and CUB-VSN-CEN performed superior in all datasets, while 2 other chains (NON-CYC-ZER and NON-MEA-SVD) performed poor in all seven benchmarks. It was of great interests to explore dataset-independent properties underlying the consistency across datasets, which thus inspired us to further investigate the similarity among performances of different analysis chains.

Since the type of instrument (TripleTOF 5600 and 6600) covered by seven benchmark datasets were the same as that of 85 SWATH-MS based projects, those datasets could be recognized as representative datasets of SWATH-MS based pharmacoproteomic data. Thus, the discovery of analysis chain performed consistently well across the various datasets might give great insights into the selection of the most appropriate analysis chain in SWATH-MS based proteomic study. To

identify such chains performed consistently well across datasets, the hierarchical clustering with the ward algorithm (Barer and Harwood, 1999; Zhu et al., 2011; Fu et al., 2018; Xue et al., 2018a) was used to identify the “consistently well-performed” analysis chains (CWPACs) based on their PMAD values across different datasets. Theoretically, there were 560 possible analysis chains by randomly integrating 5 transformation, 16 normalization, and 7 imputation algorithms (including non-transformation, non-normalization, and non-imputation). 51 (9.1%) out of these 560 were with at least one PMAD value of the seven datasets unavailable due to the calculation error. Then, the PMAD values of the remaining 509 analysis chains were applied for clustering analysis. As illustrated in **Figure 3**, six partitions of the analysis chains ( $A_1$ ,  $A_2$ ,  $A_3$ , B, C, and D) were identified. The PMADs meeting the “well-performed” criterion ( $\leq 0.7$ ) were displayed



by blue color, with the  $\log_{10}$  PMAD  $\leq -5$  set as exact blue and the larger  $\log_{10}$  PMAD gradually fading toward white (PMAD = 0.7). Meanwhile, those “poor-performed” PMADs ( $>0.7$ ) were colored by orange, with  $\log_{10}$  PMAD  $\geq 5$  set as exact orange and the smaller PMAD gradually fading toward white (PMAD = 0.7).

The analysis chains in the partition  $A_1$ ,  $A_2$ , and  $A_3$  were “consistently well-performed” across all datasets (Figure 3). For partition  $A_1$ , 320 (99.4%) out of 322 PMAD values were  $\leq 0.1$ , and the remaining PMADs were  $\leq 0.7$  (Supplementary Figure S1). For partition  $A_2$ , 288 (52.7%), 209 (38.3%), and 40 (7.3%) out of those 546 PMAD values were  $\leq 0.1$ ,  $\leq 0.3$ , and  $\leq 0.7$ , respectively (Supplementary Figure S2). In partition  $A_3$ , 187 (46.1%) and 183 (45.1%) out of 406 PMADs were  $\leq 0.3$  and  $\leq 0.7$ , respectively (Supplementary Figure S3). In summary, 608 (47.7%), 396 (31.1%), and 225 (17.7%) out of all 1,274 PMADs in the partition combined by  $A_1$ ,  $A_2$ , and  $A_3$  were  $\leq 0.1$ ,  $\leq 0.3$ , and  $\leq 0.7$ , respectively, indicating an extremely high percentage (96.5%) of the PMAD values meeting the widely adopted cutoff (PMAD = 0.7) for differentiating the chain of good and poor performances (Chawade et al., 2014; Valikangas et al., 2018). Comprehensive literature review on the 85 SWATH-MS based proteomic projects further identified the analysis chains adopted by their corresponding studies (Supplementary Table S2). In total, there were 55 analysis chains previously applied in proteomic studies, which were mapped to and labeled on Figure 3 (pink triangles). As illustrated, 7 (12.7%), 9 (16.4%), and 21 (38.2%) out of the 55 analysis chains previously adopted were within the partition  $A_1$ ,  $A_2$ , and  $A_3$ , respectively, which indicated that the majority (67.3%) of these analysis chains were the CWPACs.

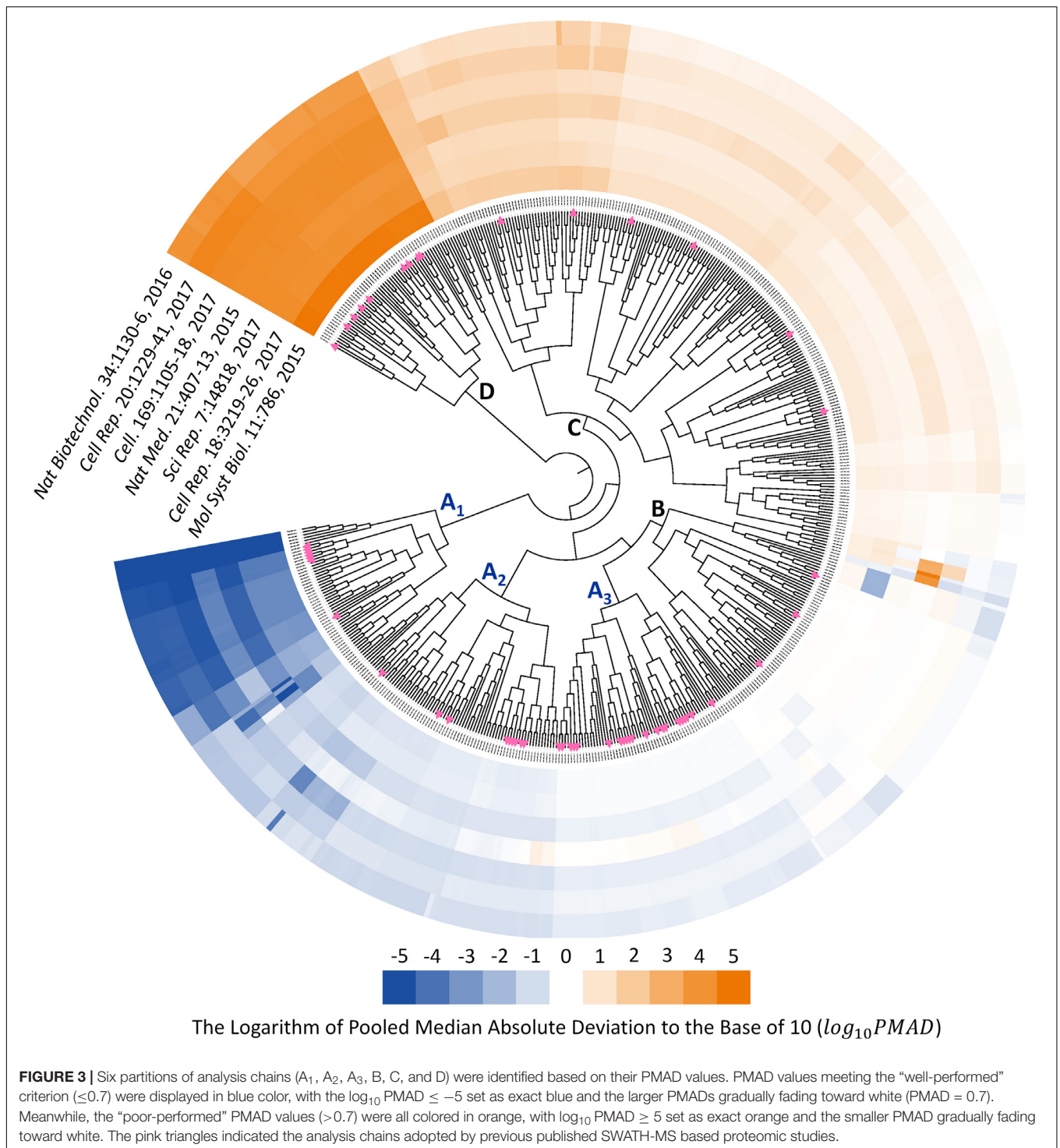
As shown in Supplementary Figure S4, the percentage of each processing method adopted by the previous proteomic studies were analyzed. *Log Transformation* was the only transformation method used in SWATH-MS based proteomic studies, and was widely recognized as powerful in quantifying thousands of proteins (Rao et al., 2011; De Livera et al., 2012;

Wisniewski et al., 2012; Zhu et al., 2012a; Feng et al., 2014). For normalizations, *Median Normalization*, *Total Ion Current*, and *Quantile Normalization* were the top-3 ranked methods in their popularity. The *Median* and *Quantile Normalization* were frequently adopted in MS-based label-free proteomic analyses (Callister et al., 2006), while the *Total Ion Current* was reported to be preferably used in the proteomic profiling based on MALDI- and SELDI-TOF mass spectra (Borgaonkar et al., 2010). For imputation, *K-nearest Neighbor* and *Background Imputation* accounted for  $>80\%$  of the SWATH-MS based proteomic studies adopting imputation methods. Among those methods used in proteomic studies (4 transformation, 15 normalization, and 6 missing-value imputation), Supplementary Figure S4 showed that some methods were adopted seldomly in SWATH-MS based proteomic studies (such as *Box-Cox Transformation*, *Pareto Scaling*, and *Singular Value Decomposition*). Therefore, it is of great interests to discover whether there are other methods suitable or demonstrating enhanced performance in SWATH-MS based proteomic analysis.

Fifty-three analysis chains consistently performed poor among datasets were also discovered by Figure 3 (partition D), all of which did not adopt any transformation method in their analysis. In total, 101 out of the 509 analysis chains (Figure 3) adopted non-transformation, and 53 (52.5%), 10 (9.9%), 11 (10.9%), 14 (13.9%), 6 (5.9%), and 7 (6.9%) out of these 101 chains were within the partition D, C, B,  $A_3$ ,  $A_2$ , and  $A_1$ , respectively. These results demonstrated the important roles played by transformation methods in the quantification performance of analysis chains.

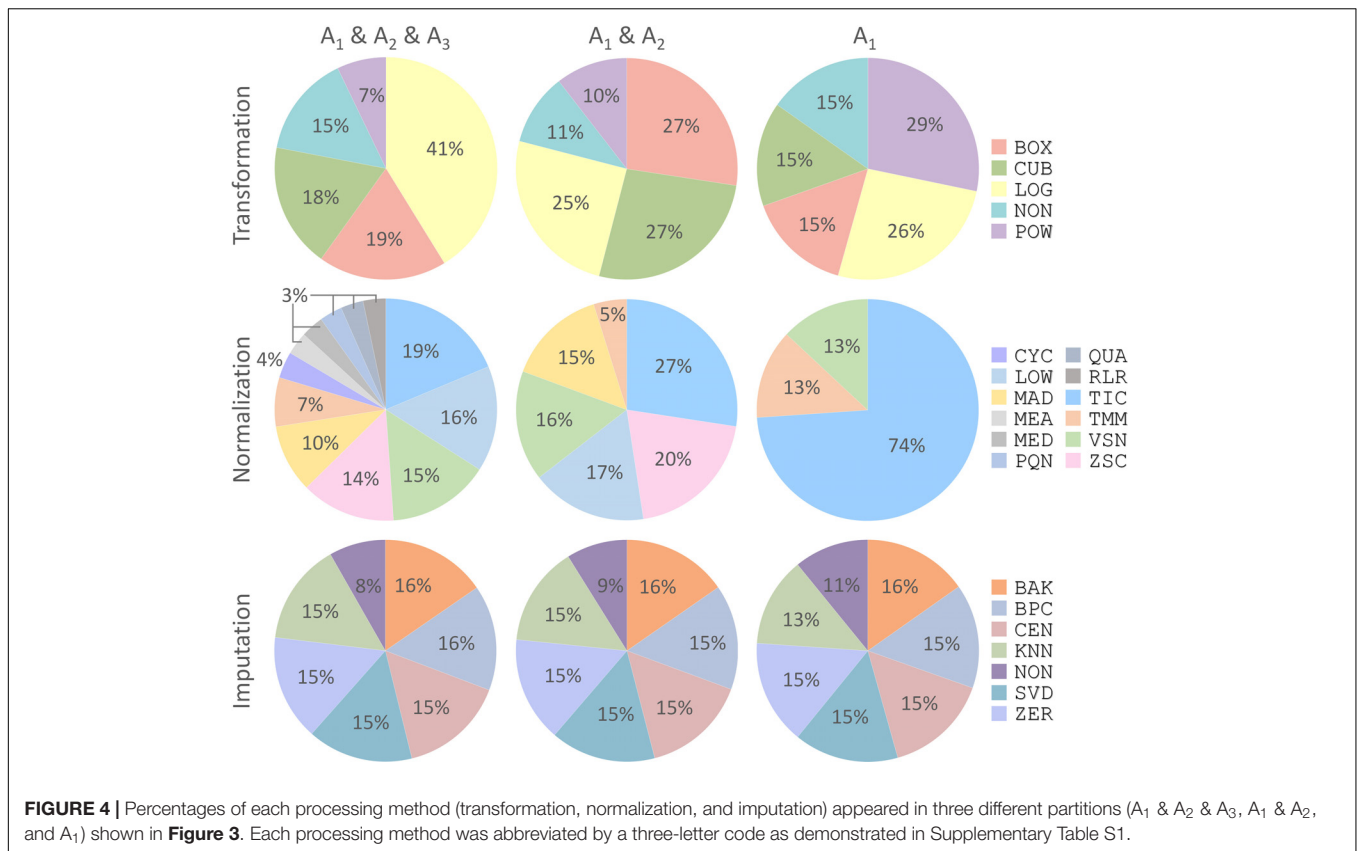
## Contribution of Each Processing Method to the Performance of Analysis Chain

With the discovery of a variety of CWPACs based on those independent benchmark datasets, it was interesting to go back



to each processing method used to integrate these CWPACs, which might be able to discover processing methods with significant contributions to the performance of CWPACs. Therefore, all CWPACs listed in Supplementary Figures S1–S3 were investigated by analyzing their corresponding processing methods. As shown in **Figure 4**, the percentage of each method appeared in 3 different partitions ( $A_1$  &  $A_2$  &  $A_3$ ,  $A_1$  &  $A_2$ ,

and  $A_1$ ) were analyzed. For transformation, the percentage of *Power Transformation* significantly increased from 7% to 10% to 29% with the gradual narrow down of partitions (from  $A_1$  &  $A_2$  &  $A_3$  to  $A_1$  &  $A_2$  to  $A_1$ ), which showed significantly enhanced role played by this transformation to achieve good performance in protein quantifications. However, *Log Transformation* decreased greatly from 41% to 25% to



26%. This indicated that *Log Transformation* contributed most to the CWPACs compared to other transformations. But when it came to the superior performance (partition  $A_1$  with  $PMAD \leq 0.1$ ), its contribution decreased and ranked as the second. For normalization, the *Total Ion Current* method stood out among all methods as the one with the highest contribution to CWPAC. With gradual narrow down of partitions (from  $A_1$  &  $A_2$  &  $A_3$  to  $A_1$  &  $A_2$  to  $A_1$ ), the importance of *Total Ion Current* method was enhanced significantly from 19% to 27% to 74%. For imputation, methods were almost evenly distributed with no clear change among different partitions. This indicated that each imputation method contributed equally to CWPACs, and the selection of any of those methods could not make statistical difference in protein quantification. Due to the equal contribution of imputation methods, it was essential to focus on selecting the appropriate combinations of transformation and normalization methods to achieve the optimal performance of analysis chains, which included POW-TMM, LOG-TIC, BOX-TIC, CUB-TIC, NON-TIC, POW-TIC, and LOG-VSN (Supplementary Figure S1).

## CONCLUSION

Based on the most complete set of the publicly available pharmacoproteomic data generated by SWATH-MS technique, this study revealed a substantial variation among

the performances of various analysis chains applied for pharmacoproteomic quantification, and the analysis chains performed consistently well across a diverse set of publicly available pharmacoproteomic data were discovered. As a result, log and power transformations sequentially followed by total ion current normalization were discovered as one of the best performed analysis chains applied for the SWATH-MS based pharmacoproteomic quantification. In summary, the identified analysis chains provided important guidance to current proteomic research and could thus facilitate the cutting-edge research in any proteomic studies requiring SWATH-MS technique.

## AUTHOR CONTRIBUTIONS

FZ conceived the idea and supervised the work. JF, JT, and YW performed the research. JF, XC, QY, JH, XL, SL, YC, and WX prepared and analyzed the data. FZ and JF wrote the manuscript. All authors have read and approved this manuscript.

## FUNDING

Funded by the research support of the Precision Medicine Project of the National Key Research and Development Plan of China



(2016YFC0902200); Innovation Project on Industrial Generic Key Technologies of Chongqing (cstc2015zdcy-ztxx120003); and Fundamental Research Funds for the Central Universities (10611CDJXZ238826, CDJZR14468801, and CDJKXB14011).

## REFERENCES

- Aebersold, R., and Mann, M. (2016). Mass-spectrometric exploration of proteome structure and function. *Nature* 537, 347–355. doi: 10.1038/nature19949
- Alter, O., Brown, P. O., and Botstein, D. (2000). Singular value decomposition for genome-wide expression data processing and modeling. *Proc. Natl. Acad. Sci. U.S.A.* 97, 10101–10106. doi: 10.1073/pnas.97.18.10101
- Andjelkovic, V., and Thompson, R. (2006). Changes in gene expression in maize kernel in response to water and salt stress. *Plant Cell Rep.* 25, 71–79. doi: 10.1007/s00299-005-0037-x
- Anjo, S. I., Santa, C., and Manadas, B. (2017). SWATH-MS as a tool for biomarker discovery: from basic research to clinical applications. *Proteomics* 17:1600278. doi: 10.1002/pmic.201600278
- Aratyn-Schaus, Y., and Ramanathan, R. (2016). Advances in high-resolution MS and hepatocyte models solve a long-standing metabolism challenge: the loratadine story. *Bioanalysis* 8, 1645–1662. doi: 10.4155/bio-2016-0094
- Barer, M. R., and Harwood, C. R. (1999). Bacterial viability and culturability. *Adv. Microb. Physiol.* 41, 93–137. doi: 10.1016/S0065-2911(08)60166-6
- Bolstad, B. M., Irizarry, R. A., Astrand, M., and Speed, T. P. (2003). A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics* 19, 185–193. doi: 10.1093/bioinformatics/19.2.185
- Borgaonkar, S. P., Hocker, H., Shin, H., and Markey, M. K. (2010). Comparison of normalization methods for the identification of biomarkers using MALDI-TOF and SELDI-TOF mass spectra. *OMICS* 14, 115–126. doi: 10.1089/omi.2009.0082
- Bruderer, R., Bernhardt, O. M., Gandhi, T., Miladinovic, S. M., Cheng, L. Y., Messner, S., et al. (2015). Extending the limits of quantitative proteome profiling with data-independent acquisition and application to acetaminophen-treated three-dimensional liver microtissues. *Mol. Cell. Proteomics* 14, 1400–1410. doi: 10.1074/mcp.M114.044305
- Callister, S. J., Barry, R. C., Adkins, J. N., Johnson, E. T., Qian, W. J., Webb-Robertson, B. J., et al. (2006). Normalization approaches for removing systematic biases associated with mass spectrometry and label-free proteomics. *J. Proteome Res.* 5, 277–286. doi: 10.1021/pr050300l
- Chai, L. E., Law, C. K., Mohamad, M. S., Chong, C. K., Choon, Y. W., Deris, S., et al. (2014). Investigating the effects of imputation methods for modelling gene networks using a dynamic bayesian network from gene expression data. *Malays. J. Med. Sci.* 21, 20–27.
- Chambliss, A. B., and Chan, D. W. (2016). Precision medicine: from pharmacogenomics to pharmacoproteomics. *Clin. Proteomics* 13:25. doi: 10.1186/s12014-016-9127-8
- Chawade, A., Alexandersson, E., and Levander, F. (2014). Normalyzer: a tool for rapid evaluation of normalization methods for omics data sets. *J. Proteome Res.* 13, 3114–3120. doi: 10.1021/pr401264n
- Chawade, A., Sandin, M., Teleman, J., Malmstrom, J., and Levander, F. (2015). Data processing has major impact on the outcome of quantitative label-free LC-MS analysis. *J. Proteome Res.* 14, 676–687. doi: 10.1021/pr500665j
- Cheadle, C., Vawter, M. P., Freed, W. J., and Becker, K. G. (2003). Analysis of microarray data using Z score transformation. *J. Mol. Diagn.* 5, 73–81. doi: 10.1016/S1525-1578(10)60455-2
- Chignell, J. F., Park, S., Lacerda, C. M. R., De Long, S. K., and Reardon, K. F. (2018). Label-free proteomics of a defined, binary co-culture reveals diversity of competitive responses between members of a model soil microbial system. *Microb. Ecol.* 75, 701–719. doi: 10.1007/s00248-017-1072-1
- D'Alessandro, A., and Zolla, L. (2010). Pharmacoproteomics: a chess game on a protein field. *Drug Discov. Today* 15, 1015–1023. doi: 10.1016/j.drudis.2010.10.002
- De Livera, A. M., Dias, D. A., De Souza, D., Rupasinghe, T., Pyke, J., Tull, D., et al. (2012). Normalizing and integrating metabolomics data. *Anal. Chem.* 84, 10768–10776. doi: 10.1021/ac302748b
- Dieterle, F., Ross, A., Schlotterbeck, G., and Senn, H. (2006). Probabilistic quotient normalization as robust method to account for dilution of complex biological mixtures. Application in 1H NMR metabolomics. *Anal. Chem.* 78, 4281–4290. doi: 10.1021/ac051632c
- Feng, C., Wang, H., Lu, N., Chen, T., He, H., Lu, Y., et al. (2014). Log-transformation and its implications for data analysis. *Shanghai Arch. Psychiatry* 26, 105–109. doi: 10.3969/j.issn.1002-0829.2014.02.009
- Fu, T., Zheng, G., Tu, G., Yang, F., Chen, Y., Yao, X., et al. (2018). Exploring the binding mechanism of metabotropic glutamate receptor 5 negative allosteric modulators in clinical trials by molecular dynamics simulations. *ACS Chem. Neurosci.* doi: 10.1021/acscchemneuro.8b00059 [Epub ahead of print].
- Gan, X., Liew, A. W., and Yan, H. (2006). Microarray missing data imputation based on a set theoretic framework and biological knowledge. *Nucleic Acids Res.* 34, 1608–1619. doi: 10.1093/nar/gkl047
- Gaspari, M., Chiesa, L., Nicastrì, A., Gabriele, C., Harper, V., Britti, D., et al. (2016). Proteome speciation by mass spectrometry: characterization of composite protein mixtures in milk replacers. *Anal. Chem.* 88, 11568–11574. doi: 10.1021/acs.analchem.6b02848
- Gatto, L., Hansen, K. D., Hoopmann, M. R., Hermjakob, H., Kohlbacher, O., and Beyer, A. (2016). Testing and validation of computational methods for mass spectrometry. *J. Proteome Res.* 15, 809–814. doi: 10.1021/acs.jproteome.5b00852
- Gillet, L. C., Navarro, P., Tate, S., Rost, H., Selevsek, N., Reiter, L., et al. (2012). Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis. *Mol. Cell. Proteomics* 11:O111.016717. doi: 10.1074/mcp.O111.016717
- Guo, T., Kouvonen, P., Koh, C. C., Gillet, L. C., Wolski, W. E., Rost, H. L., et al. (2015). Rapid mass spectrometric conversion of tissue biopsy samples into permanent quantitative digital proteome maps. *Nat. Med.* 21, 407–413. doi: 10.1038/nm.3807
- Hauser, D. S., Stade, M., Schmidt, A., and Hanauer, G. (2005). Cardiovascular parameters in anaesthetized guinea pigs: a safety pharmacology screening model. *J. Pharmacol. Toxicol. Methods* 52, 106–114. doi: 10.1016/j.vascn.2005.03.003
- Hong, M. G., Lee, W., Nilsson, P., Pawitan, Y., and Schwenk, J. M. (2016). Multidimensional normalization to minimize plate effects of suspension bead array data. *J. Proteome Res.* 15, 3473–3480. doi: 10.1021/acs.jproteome.5b01131
- Huang, Q., Yang, L., Luo, J., Guo, L., Wang, Z., Yang, X., et al. (2015). SWATH enables precise label-free quantification on proteome scale. *Proteomics* 15, 1215–1223. doi: 10.1002/pmic.201400270
- Huber, W., von Heydebreck, A., Sultmann, H., Poustka, A., and Vingron, M. (2002). Variance stabilization applied to microarray data calibration and to the quantification of differential expression. *Bioinformatics* 18, S96–S104. doi: 10.1093/bioinformatics/18.suppl\_1.S96
- Jamwal, R., Barlock, B. J., Adusumalli, S., Ogasawara, K., Simons, B. L., and Akhlaghi, F. (2017). Multiplex and label-free relative quantification approach for studying protein abundance of drug metabolizing enzymes in human liver microsomes using SWATH-MS. *J. Proteome Res.* 16, 4134–4143. doi: 10.1021/acs.jproteome.7b00505
- Karpievitch, Y. V., Dabney, A. R., and Smith, R. D. (2012). Normalization and missing value imputation for label-free LC-MS analysis. *BMC Bioinformatics* 13:S5. doi: 10.1186/1471-2105-13-S16-S5
- Khoonsari, P. E., Haggmark, A., Lonnberg, M., Mikus, M., Kilander, L., Lannfelt, L., et al. (2016). Analysis of the cerebrospinal fluid proteome in Alzheimer's Disease. *PLoS One* 11:e0150672. doi: 10.1371/journal.pone.0150672
- Kohl, S. M., Klein, M. S., Hochrein, J., Oefner, P. J., Spang, R., and Gronwald, W. (2012). State-of-the art data normalization methods improve NMR-based metabolomic analysis. *Metabolomics* 8, 146–160. doi: 10.1007/s11306-011-0350-z

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fphar.2018.00681/full#supplementary-material>

- Krasny, L., Bland, P., Kogata, N., Wai, P., Howard, B. A., Natrajan, R. C., et al. (2018). SWATH mass spectrometry as a tool for quantitative profiling of the matrisome. *J. Proteomics* doi: 10.1016/j.jprot.2018.02.026 [Epub ahead of print].
- Kuharev, J., Navarro, P., Distler, U., Jahn, O., and Tenzer, S. (2015). In-depth evaluation of software tools for data-independent acquisition based label-free quantification. *Proteomics* 15, 3140–3151. doi: 10.1002/pmic.201400396
- Letunic, I., and Bork, P. (2016). Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res.* 44, W242–W245. doi: 10.1093/nar/gkw290
- Li, B., Tang, J., Yang, Q., Cui, X., Li, S., Chen, S., et al. (2016a). Performance evaluation and online realization of data-driven normalization methods used in LC/MS based untargeted metabolomics analysis. *Sci. Rep.* 6:38881. doi: 10.1038/srep38881
- Li, B., Tang, J., Yang, Q., Li, S., Cui, X., Li, Y., et al. (2017). NOREVA: normalization and evaluation of MS-based metabolomics data. *Nucleic Acids Res.* 45, W162–W170. doi: 10.1093/nar/gkx449
- Li, S., Cao, Q., Xiao, W., Guo, Y., Yang, Y., Duan, X., et al. (2017). 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.* 16, 738–747. doi: 10.1021/acs.jproteome.6b00767
- Li, Y. H., Wang, P. P., Li, X. X., Yu, C. Y., Yang, H., Zhou, J., et al. (2016b). The human kinome targeted by fda approved multi-target drugs and combination products: a comparative study from the drug-target interaction network perspective. *PLoS One* 11:e0165737. doi: 10.1371/journal.pone.0165737
- Li, Y. H., Xu, J. Y., Tao, L., Li, X. F., Li, S., Zeng, X., et al. (2016c). SVM-Prot 2016: a web-server for machine learning prediction of protein functional families from sequence irrespective of similarity. *PLoS One* 11:e0155290. doi: 10.1371/journal.pone.0155290
- Li, Y. H., Yu, C. Y., Li, X. X., Zhang, P., Tang, J., Yang, Q., et al. (2018). Therapeutic target database update 2018: enriched resource for facilitating bench-to-clinic research of targeted therapeutics. *Nucleic Acids Res.* 46, D1121–D1127. doi: 10.1093/nar/gkx1076
- Lin, Y., Golovkina, K., Chen, Z. X., Lee, H. N., Negron, Y. L., Sultana, H., et al. (2016). Comparison of normalization and differential expression analyses using RNA-Seq data from 726 individual *Drosophila melanogaster*. *BMC Genomics* 17:28. doi: 10.1186/s12864-015-2353-z
- Liu, Y., Buil, A., Collins, B. C., Gillet, L. C., Blum, L. C., Cheng, L. Y., et al. (2015). Quantitative variability of 342 plasma proteins in a human twin population. *Mol. Syst. Biol.* 11:786. doi: 10.15252/msb.20145728
- MacLean, B., Tomazela, D. M., Shulman, N., Chambers, M., Finney, G. L., Frewen, B., et al. (2010). Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics* 26, 966–968. doi: 10.1093/bioinformatics/btq054
- Matzke, M. M., Waters, K. M., Metz, T. O., Jacobs, J. M., Sims, A. C., Baric, R. S., et al. (2011). Improved quality control processing of peptide-centric LC-MS proteomics data. *Bioinformatics* 27, 2866–2872. doi: 10.1093/bioinformatics/btr479
- Mrozek, D., Malysiak-Mrozek, B., and Siaznik, A. (2013). search GenBank: interactive orchestration and ad-hoc choreography of Web services in the exploration of the biomedical resources of the National Center For Biotechnology Information. *BMC Bioinformatics* 14:73. doi: 10.1186/1471-2105-14-73
- Muller, F., Fischer, L., Chen, Z. A., Auchynnikava, T., and Rappsilber, J. (2018). On the reproducibility of label-free quantitative cross-linking/mass spectrometry. *J. Am. Soc. Mass Spectrom.* 29, 405–412. doi: 10.1007/s13361-017-1837-2
- Navarro, P., Kuharev, J., Gillet, L. C., Bernhardt, O. M., MacLean, B., Rost, H. L., et al. (2016). A multicenter study benchmarks software tools for label-free proteome quantification. *Nat. Biotechnol.* 34, 1130–1136. doi: 10.1038/nbt.3685
- Ori, A., Iskar, M., Buczak, K., Kastiris, P., Parca, L., Andres-Pons, A., et al. (2016). Spatiotemporal variation of mammalian protein complex stoichiometries. *Genome Biol.* 17:47. doi: 10.1186/s13059-016-0912-5
- Parker, S. J., Rost, H., Rosenberger, G., Collins, B. C., Malmstrom, L., Amodei, D., et al. (2015). Identification of a set of conserved eukaryotic internal retention time standards for data-independent acquisition mass spectrometry. *Mol. Cell. Proteomics* 14, 2800–2813. doi: 10.1074/mcp.O114.042267
- Paul, D., Chanukuppa, V., Reddy, P. J., Taunk, K., Adhav, R., Srivastava, S., et al. (2016). Global proteomic profiling identifies etoposide chemoresistance markers in non-small cell lung carcinoma. *J. Proteomics* 138, 95–105. doi: 10.1016/j.jprot.2016.02.008
- Quiros, P. M., Prado, M. A., Zamboni, N., D'Amico, D., Williams, R. W., Finley, D., et al. (2017). Multi-omics analysis identifies ATF4 as a key regulator of the mitochondrial stress response in mammals. *J. Cell Biol.* 216, 2027–2045. doi: 10.1083/jcb.201702058
- Rao, H. B., Zhu, F., Yang, G. B., Li, Z. R., and Chen, Y. Z. (2011). Update of PROFEAT: a web server for computing structural and physicochemical features of proteins and peptides from amino acid sequence. *Nucleic Acids Res.* 39, W385–W390. doi: 10.1093/nar/gkr284
- Roemmel, A. T., Steuer, A. E., and Kraemer, T. (2015). Liquid chromatography, in combination with a quadrupole time-of-flight instrument, with sequential window acquisition of all theoretical fragment-ion spectra acquisition: validated quantification of 39 antidepressants in whole blood as part of a simultaneous screening and quantification procedure. *Anal. Chem.* 87, 9294–9301. doi: 10.1021/acs.analchem.5b02031
- Rost, H. L., Rosenberger, G., Navarro, P., Gillet, L., Miladinovic, S. M., Schubert, O. T., et al. (2014). OpenSWATH enables automated, targeted analysis of data-independent acquisition MS data. *Nat. Biotechnol.* 32, 219–223. doi: 10.1038/nbt.2841
- Saei, A. A., Sabatier, P., Guler Tokat, U., Chernobrovkin, A., Pirmoradian, M., and Zubarev, R. A. (2018). Comparative proteomics of dying and surviving cancer cells improves the identification of drug targets and sheds light on cell life/death decisions. *Mol. Cell. Proteomics* 17, 1144–1155. doi: 10.1074/mcp.RA118.000610
- Sajic, T., Liu, Y., and Aebersold, R. (2015). Using data-independent, high-resolution mass spectrometry in protein biomarker research: perspectives and clinical applications. *Proteomics Clin. Appl.* 9, 307–321. doi: 10.1002/prca.201400117
- Sakia, R. M. (1992). The box-cox transformation technique - a review. *Statistician* 41, 169–178. doi: 10.2307/2348250
- Scheidweiler, K. B., Jarvis, M. J., and Huestis, M. A. (2015). Nontargeted SWATH acquisition for identifying 47 synthetic cannabinoid metabolites in human urine by liquid chromatography-high-resolution tandem mass spectrometry. *Anal. Bioanal. Chem.* 407, 883–897. doi: 10.1007/s00216-014-8118-8
- Schubert, O. T., Gillet, L. C., Collins, B. C., Navarro, P., Rosenberger, G., Wolski, W. E., et al. (2015). Building high-quality assay libraries for targeted analysis of SWATH MS data. *Nat. Protoc.* 10, 426–441. doi: 10.1038/nprot.2015.015
- Shi, T., Song, E., Nie, S., Rodland, K. D., Liu, T., Qian, W. J., et al. (2016). Advances in targeted proteomics and applications to biomedical research. *Proteomics* 16, 2160–2182. doi: 10.1002/pmic.201500449
- Sullivan, K. D., Evans, D., Pandey, A., Hraha, T. H., Smith, K. P., Markham, N., et al. (2017). Trisomy 21 causes changes in the circulating proteome indicative of chronic autoinflammation. *Sci. Rep.* 7:14818. doi: 10.1038/s41598-017-13858-3
- Tan, S. L. W., Chadha, S., Liu, Y., Gabasova, E., Perera, D., Ahmed, K., et al. (2017). A class of environmental and endogenous toxins induces BRCA2 haploinsufficiency and genome instability. *Cell* 169, 1105–1118. doi: 10.1016/j.cell.2017.05.010
- Tao, L., Zhu, F., Xu, F., Chen, Z., Jiang, Y. Y., and Chen, Y. Z. (2015). Co-targeting cancer drug escape pathways confers clinical advantage for multi-target anticancer drugs. *Pharmacol. Res.* 102, 123–131. doi: 10.1016/j.phrs.2015.09.019
- Tippmann, S. (2015). Programming tools: adventures with R. *Nature* 517, 109–110. doi: 10.1038/517109a
- Valikangas, T., Suomi, T., and Elo, L. L. (2017). A comprehensive evaluation of popular proteomics software workflows for label-free proteome quantification and imputation. *Brief. Bioinform.* doi: 10.1093/bib/bbx054 [Epub ahead of print].
- Valikangas, T., Suomi, T., and Elo, L. L. (2018). A systematic evaluation of normalization methods in quantitative label-free proteomics. *Brief. Bioinform.* 19, 1–11. doi: 10.1093/bib/bbw095
- Wang, P., Fu, T., Zhang, X., Yang, F., Zheng, G., Xue, W., et al. (2017a). Differentiating physicochemical properties between NDRIs and sNRIs clinically important for the treatment of ADHD. *Biochim. Biophys. Acta* 1861, 2766–2777. doi: 10.1016/j.bbagen.2017.07.022

- Wang, P., Yang, F., Yang, H., Xu, X., Liu, D., Xue, W., et al. (2015). Identification of dual active agents targeting 5-HT1A and SERT by combinatorial virtual screening methods. *Biomed. Mater. Eng.* 26, S2233–S2239. doi: 10.3233/BME-151529
- Wang, P., Zhang, X., Fu, T., Li, S., Li, B., Xue, W., et al. (2017b). Differentiating physicochemical properties between addictive and nonaddictive ADHD drugs revealed by molecular dynamics simulation studies. *ACS Chem. Neurosci.* 8, 1416–1428. doi: 10.1021/acscchemneuro.7b00173
- Weisser, H., and Choudhary, J. S. (2017). Targeted feature detection for data-dependent shotgun proteomics. *J. Proteome Res.* 16, 2964–2974. doi: 10.1021/acs.jproteome.7b00248
- Wen, B., Mei, Z., Zeng, C., and Liu, S. (2017). metaX: a flexible and comprehensive software for processing metabolomics data. *BMC Bioinformatics* 18:183. doi: 10.1186/s12859-017-1579-y
- Wilson, D. L., Buckley, M. J., Helliwell, C. A., and Wilson, I. W. (2003). New normalization methods for cDNA microarray data. *Bioinformatics* 19, 1325–1332. doi: 10.1093/bioinformatics/btg146
- Wisniewski, J. R., Ostasiewicz, P., Dus, K., Zielinska, D. F., Gnad, F., and Mann, M. (2012). Extensive quantitative remodeling of the proteome between normal colon tissue and adenocarcinoma. *Mol. Syst. Biol.* 8:611. doi: 10.1038/msb.2012.44
- Wu, J. X., Song, X., Pascovici, D., Zaw, T., Care, N., Krisp, C., et al. (2016). SWATH mass spectrometry performance using extended peptide MS/MS assay libraries. *Mol. Cell. Proteomics* 15, 2501–2514. doi: 10.1074/mcp.M115.055558
- Xu, J., Wang, P., Yang, H., Zhou, J., Li, Y., Li, X., et al. (2016). Comparison of FDA approved kinase targets to clinical trial ones: insights from their system profiles and drug-target interaction networks. *Biomed. Res. Int.* 2016:2509385. doi: 10.1155/2016/2509385
- Xue, W., Wang, P., Li, B., Li, Y., Xu, X., Yang, F., et al. (2016). Identification of the inhibitory mechanism of FDA approved selective serotonin reuptake inhibitors: an insight from molecular dynamics simulation study. *Phys. Chem. Chem. Phys.* 18, 3260–3271. doi: 10.1039/c5cp05771j
- Xue, W., Wang, P., Tu, G., Yang, F., Zheng, G., Li, X., et al. (2018a). Computational identification of the binding mechanism of a triple reuptake inhibitor amitifadine for the treatment of major depressive disorder. *Phys. Chem. Chem. Phys.* 20, 6606–6616. doi: 10.1039/c7cp07869b
- Xue, W., Yang, F., Wang, P., Zheng, G., Chen, Y., Yao, X., et al. (2018b). 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.* 9, 1128–1140. doi: 10.1021/acscchemneuro.7b00490
- Yang, F. Y., Fu, T. T., Zhang, X. Y., Hu, J., Xue, W. W., Zheng, G. X., et al. (2017). Comparison of computational model and X-ray crystal structure of human serotonin transporter: potential application for the pharmacology of human monoamine transporters. *Mol. Simul.* 43, 1089–1098. doi: 10.1080/08927022.2017.1309653
- Yang, H., Qin, C., Li, Y. H., Tao, L., Zhou, J., Yu, C. Y., et al. (2016). Therapeutic target database update 2016: enriched resource for bench to clinical drug target and targeted pathway information. *Nucleic Acids Res.* 44, D1069–D1074. doi: 10.1093/nar/gkv1230
- Yu, C. Y., Li, X. X., Yang, H., Li, Y. H., Xue, W. W., Chen, Y. Z., et al. (2018). Assessing the performances of protein function prediction algorithms from the perspectives of identification accuracy and false discovery rate. *Int. J. Mol. Sci.* 19:E183 doi: 10.3390/ijms19010183
- Yue, Q., Feng, L., Cao, B., Liu, M., Zhang, D., Wu, W., et al. (2016). Proteomic analysis revealed the important role of vimentin in human cervical carcinoma hela cells treated with gambogic acid. *Mol. Cell. Proteomics* 15, 26–44. doi: 10.1074/mcp.M115.053272
- Zeng, X., Ding, N., Rodriguez-Paton, A., and Zou, Q. (2017). Probability-based collaborative filtering model for predicting gene-disease associations. *BMC Med. Genomics* 10:76. doi: 10.1186/s12920-017-0313-y
- Zhang, Z. (2014). Recombinant human activated protein C for the treatment of severe sepsis and septic shock: a study protocol for incorporating observational evidence using a Bayesian approach. *BMJ Open* 4:e005622. doi: 10.1136/bmjopen-2014-005622
- Zheng, G., Xue, W., Wang, P., Yang, F., Li, B., Li, X., et al. (2016). Exploring the inhibitory mechanism of approved selective norepinephrine reuptake inhibitors and reboxetine enantiomers by molecular dynamics study. *Sci. Rep.* 6:26883. doi: 10.1038/srep26883
- Zheng, G., Xue, W., Yang, F., Zhang, Y., Chen, Y., Yao, X., et al. (2017). Revealing vilazodone's binding mechanism underlying its partial agonism to the 5-HT1A receptor in the treatment of major depressive disorder. *Phys. Chem. Chem. Phys.* 19, 28885–28896. doi: 10.1039/c7cp05688e
- Zhu, F., Han, B., Kumar, P., Liu, X., Ma, X., Wei, X., et al. (2010). Update of TTD: therapeutic target database. *Nucleic Acids Res.* 38, D787–D791. doi: 10.1093/nar/gkp1014
- Zhu, F., Han, L., Zheng, C., Xie, B., Tammi, M. T., Yang, S., et al. (2009). What are next generation innovative therapeutic targets? Clues from genetic, structural, physicochemical, and systems profiles of successful targets. *J. Pharmacol. Exp. Ther.* 330, 304–315. doi: 10.1124/jpet.108.149955
- Zhu, F., Han, L. Y., Chen, X., Lin, H. H., Ong, S., Xie, B., et al. (2008a). Homology-free prediction of functional class of proteins and peptides by support vector machines. *Curr. Protein Pept. Sci.* 9, 70–95. doi: 10.2174/138920308783565697
- Zhu, F., Li, X. X., Yang, S. Y., and Chen, Y. Z. (2018). Clinical success of drug targets prospectively predicted by in silico study. *Trends Pharmacol. Sci.* 39, 229–231. doi: 10.1016/j.tips.2017.12.002
- Zhu, F., Ma, X. H., Qin, C., Tao, L., Liu, X., Shi, Z., et al. (2012a). Drug discovery prospect from untapped species: indications from approved natural product drugs. *PLoS One* 7:e39782. doi: 10.1371/journal.pone.0039782
- Zhu, F., Qin, C., Tao, L., Liu, X., Shi, Z., Ma, X., et al. (2011). Clustered patterns of species origins of nature-derived drugs and clues for future bioprospecting. *Proc. Natl. Acad. Sci. U.S.A.* 108, 12943–12948. doi: 10.1073/pnas.1107336108
- Zhu, F., Shi, Z., Qin, C., Tao, L., Liu, X., Xu, F., et al. (2012b). Therapeutic target database update 2012: a resource for facilitating target-oriented drug discovery. *Nucleic Acids Res.* 40, D1128–D1136. doi: 10.1093/nar/gkr797
- Zhu, F., Zheng, C. J., Han, L. Y., Xie, B., Jia, J., Liu, X., et al. (2008b). Trends in the exploration of anticancer targets and strategies in enhancing the efficacy of drug targeting. *Curr. Mol. Pharmacol.* 1, 213–232. doi: 10.2174/1874467210801030213

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Fu, Tang, Wang, Cui, Yang, Hong, Li, Li, Chen, Xue and Zhu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.