

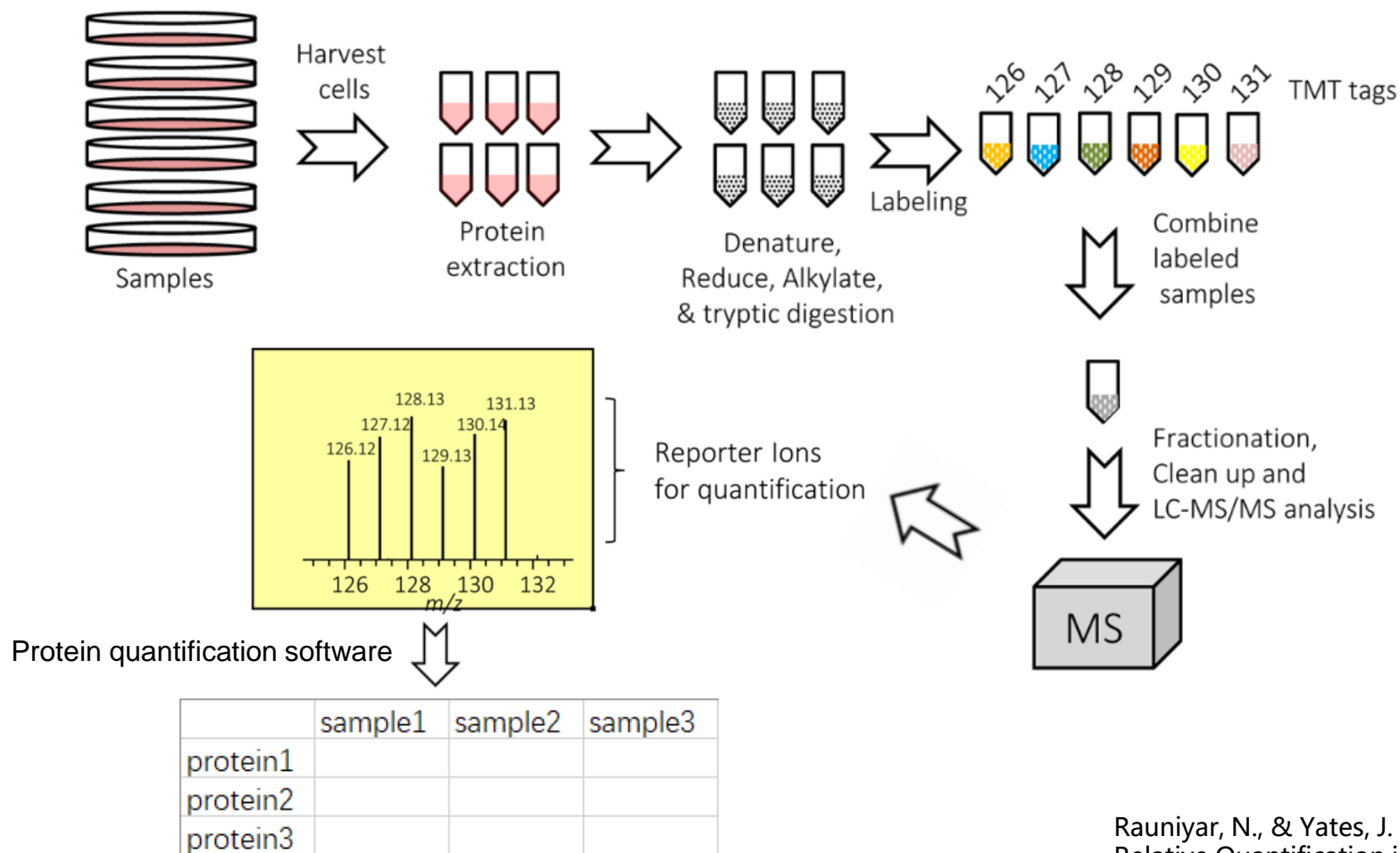
Comparative Evaluation of Proteome Discoverer and FragPipe for the TMT-based Proteome Quantification

蛋白质组数据分析软件Proteome Discoverer和
FragPipe基于串联质谱标签进行蛋白质定量的比较评估

Tianen He
何天恩

蛋白质组大数据实验室
www.guomics.com

Background: Tandem mass tags (TMT)

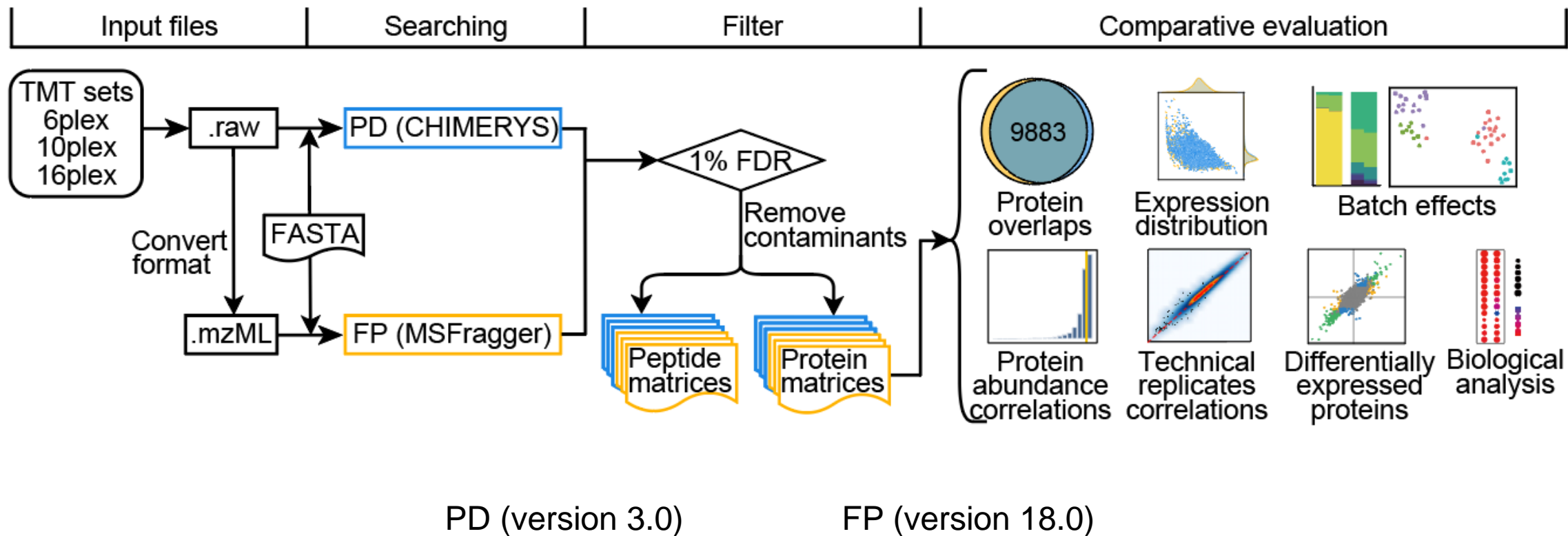


Rauniyar, N., & Yates, J. R.. (2014). Isobaric Labeling-Based Relative Quantification in Shotgun Proteomics. *Journal of Proteome Research*, 13(12), 5293–5309. (figure adapted)

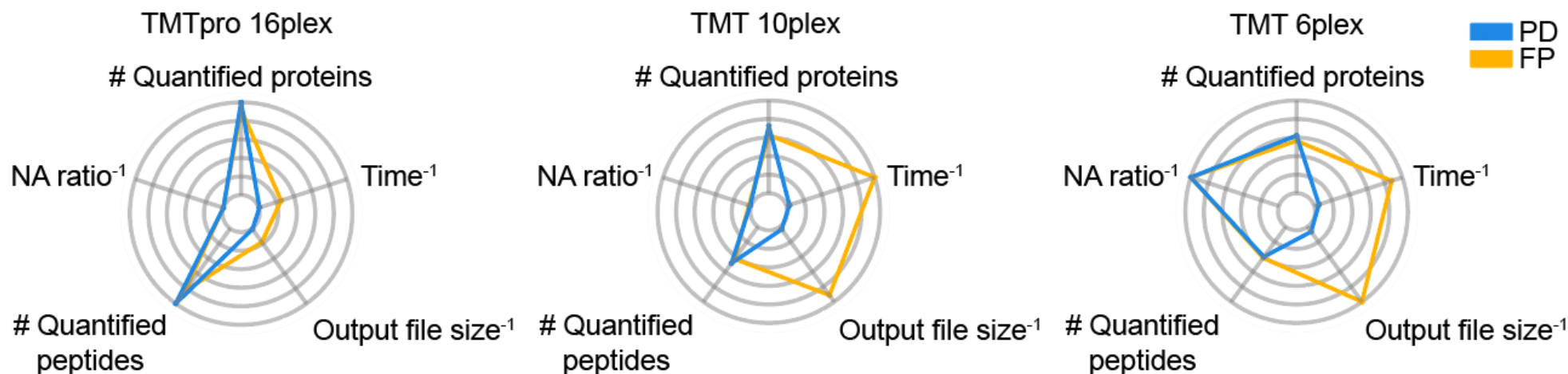
Objectives

- To compare the performance of Proteome Discoverer (PD) and FragPipe (FP)
 - PD: commercial software, with search engine CHIMERYS
 - FP: free for academic use, with search engine MSFragger
- To provide reference for researchers to choose the suitable software

Study Design

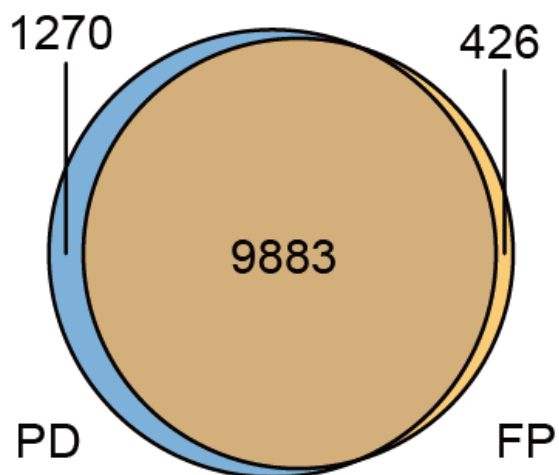


FP obtained similar quantification with PD while saving time and space

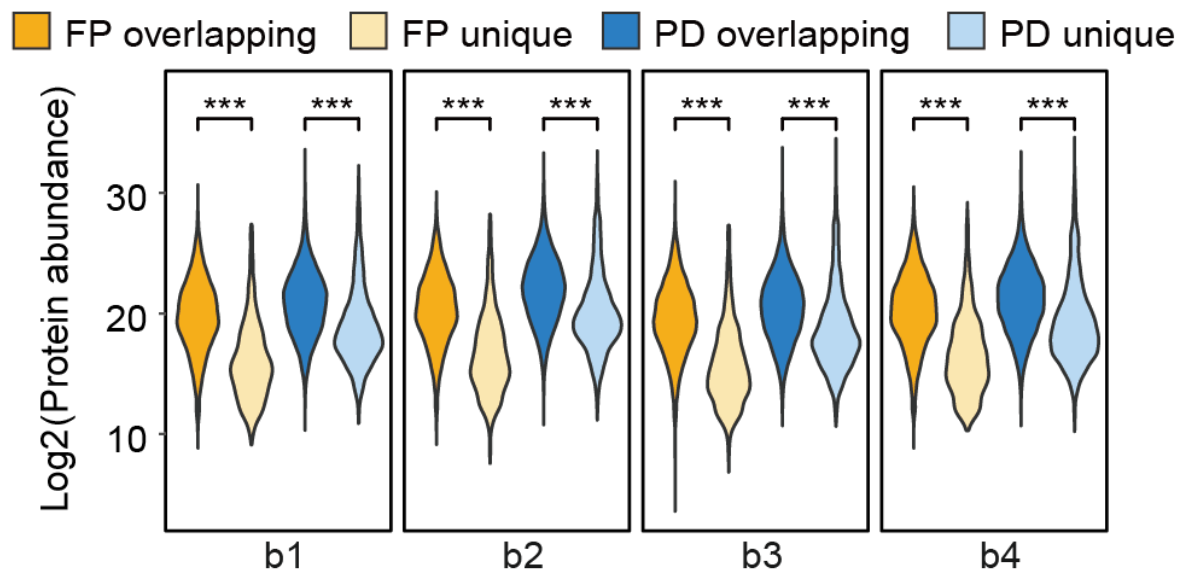


Dataset	TMTpro 16plex (4 batches, 120 files, 29.3 G)		TMT 10plex (1 batch, 40 files, 9.28 G)		TMT 6plex (1 batch, 10 files, 10.0 G)	
Software	PD	FP	PD	FP	PD	FP
Time (min)	3960	24+118	1074	5+31	675	7+34
# Quantified proteins	11,153	10,309	8088	7006	6938	6306
NA ratio of quantified proteins (%)	11.74	10.52	0.85	0.38	0.00	0.00
# Quantified peptides	140,402	117,309	68,439	60,193	56,252	58,221
Output file size (G)	120	5.77	29.5	1.14	15.7	1.04

PD and FP quantified highly overlapping proteomes



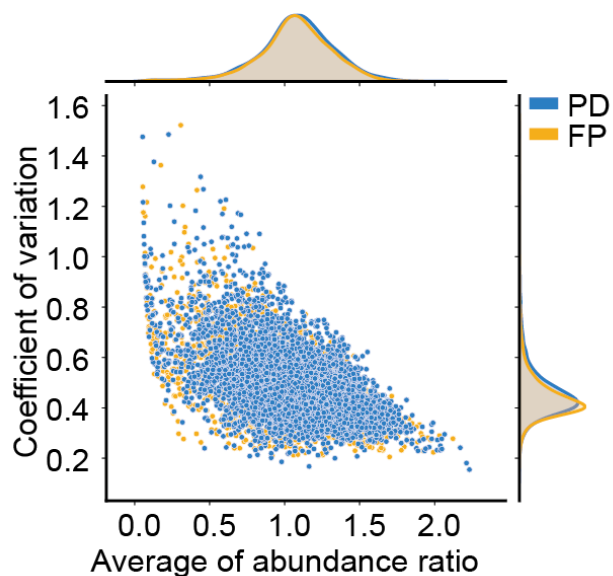
Overlap of quantified proteins



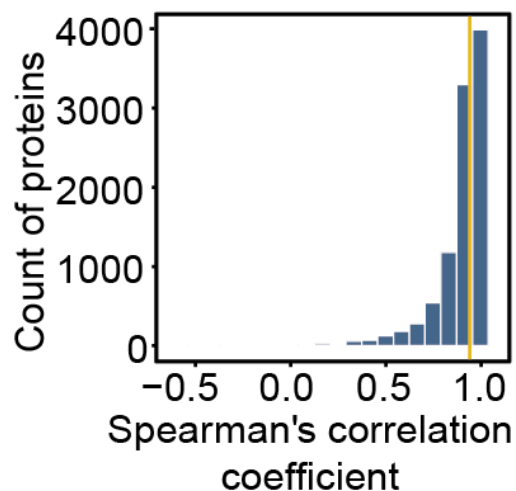
The uniquely-quantified proteins were less abundant

Using TMTpro 16plex dataset as the example. Results from the other two datasets are similar.

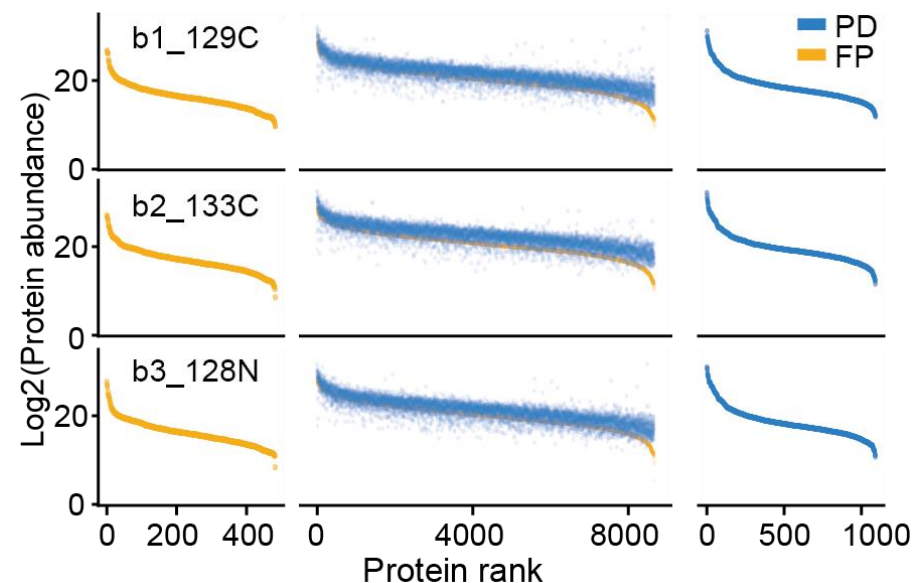
PD and FP generated highly correlated quantification results



Average values and coefficients of variation (CVs) of the abundance ratios of each protein across all samples

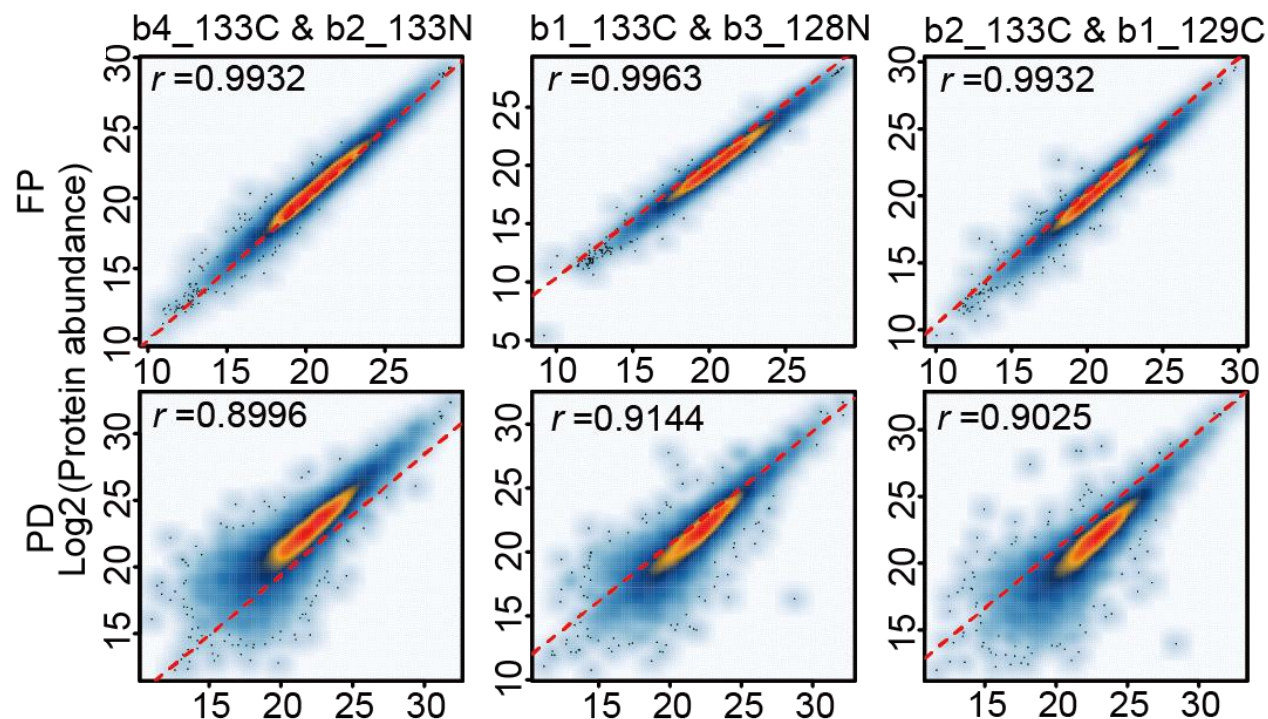


Spearman's correlation coefficients (SCCs) between the abundance ratios of the same protein quantified by PD and FP



Log2 abundances of the same proteins quantified by PD and FP

FP quantified the technical replicates with higher consistency

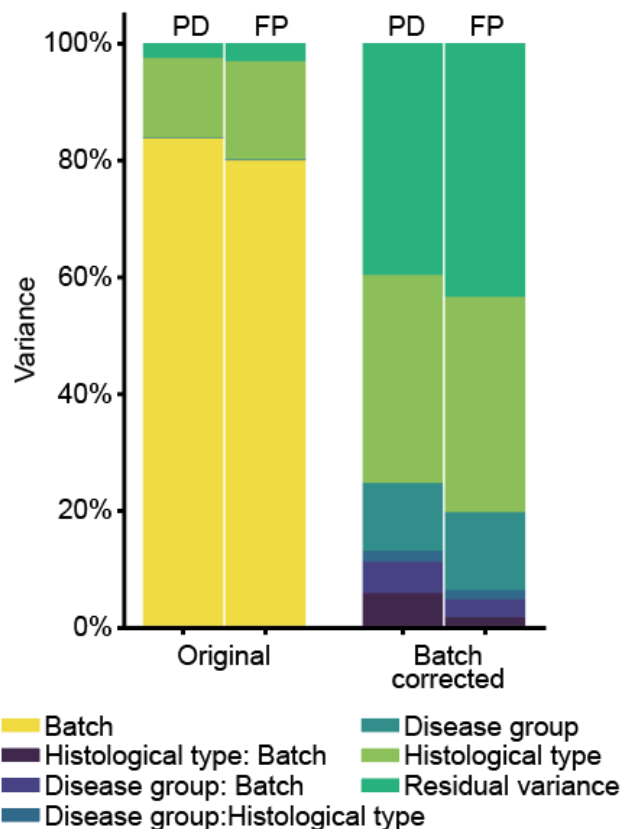


Spearman's correlation coefficient:

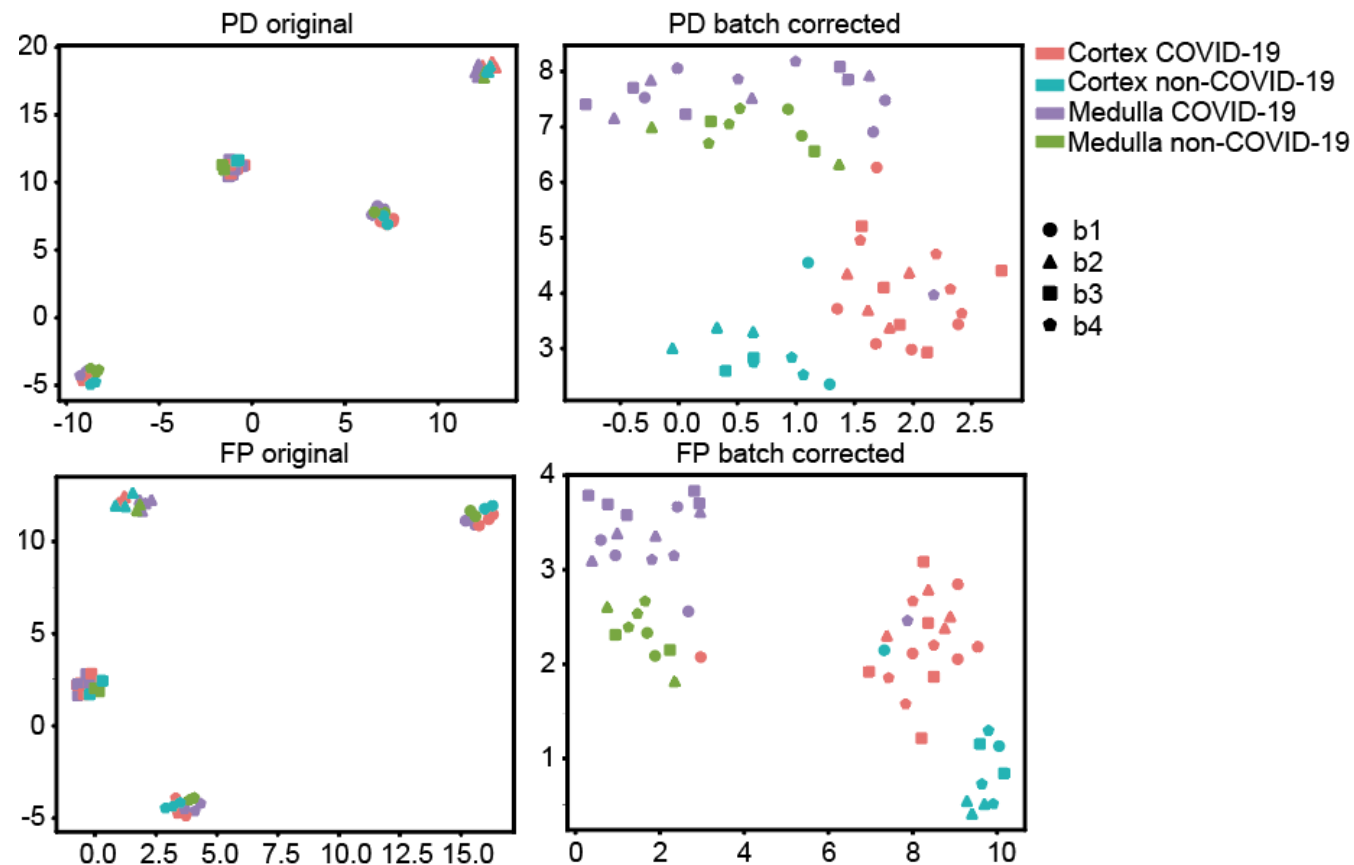
FP (0.9932, 0.9963, 0.9932) significantly higher than PD (0.8996, 0.9144, 0.9025)

(Welch two-sample t-test, P-value = 0.0009)

PD and FP outputs had similar batch effect before and after correction

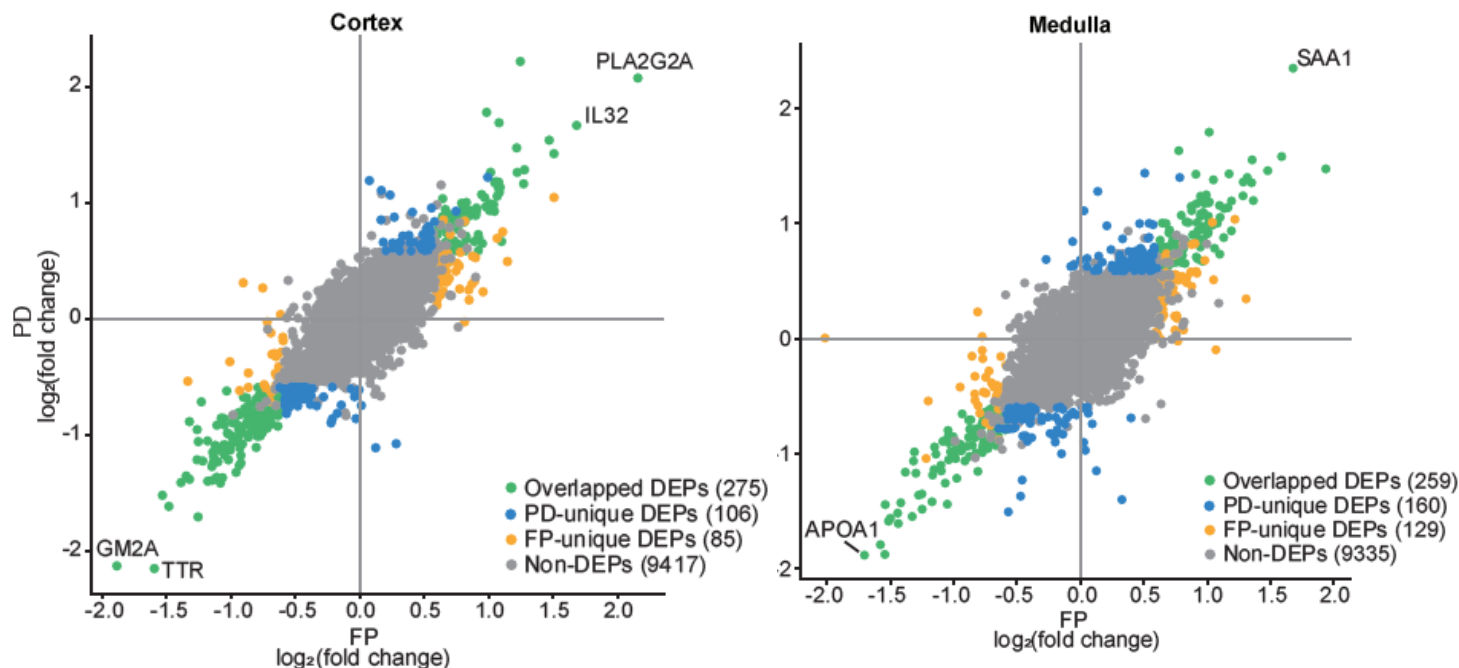
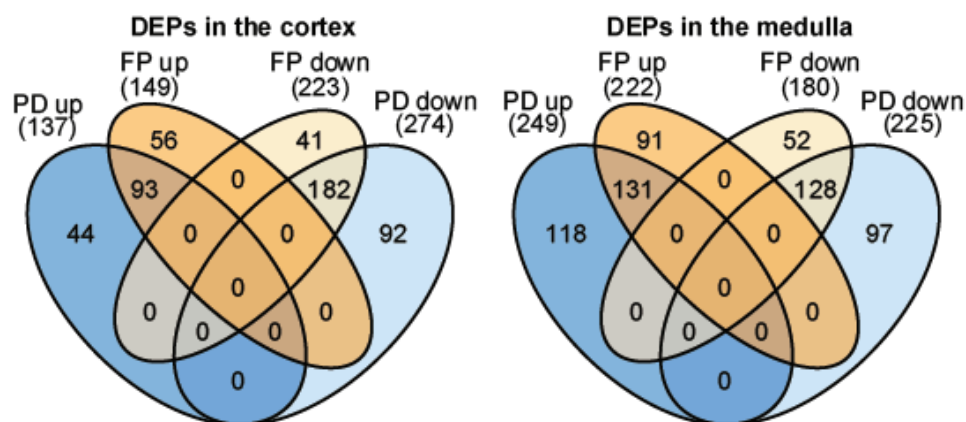


Contribution of different grouping variables and their interaction terms on the variance of the expression data



Visualization using uniform manifold approximation and projection (UMAP)

Differentially expressed proteins identified from PD and FP outputs were partially overlapping



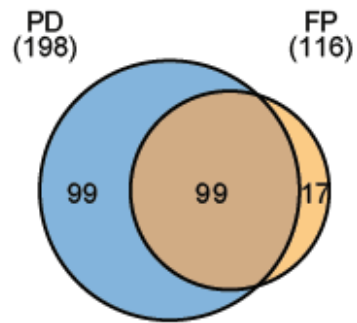
The cortex and medulla samples were analysed separately because the histological type was the main contributing factor to the data variance.

Differential expression analysis: COVID-19 vs. non-COVID-19

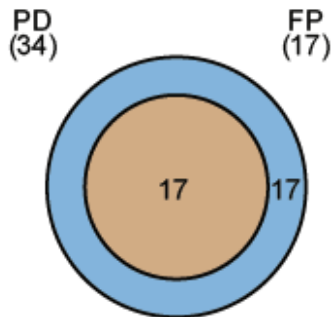
Criteria: fold change > 1.50 and Benjamini-Hochberg adjusted P-value < 0.05

The most significantly enriched functions and pathways were robust in PD and FP outputs

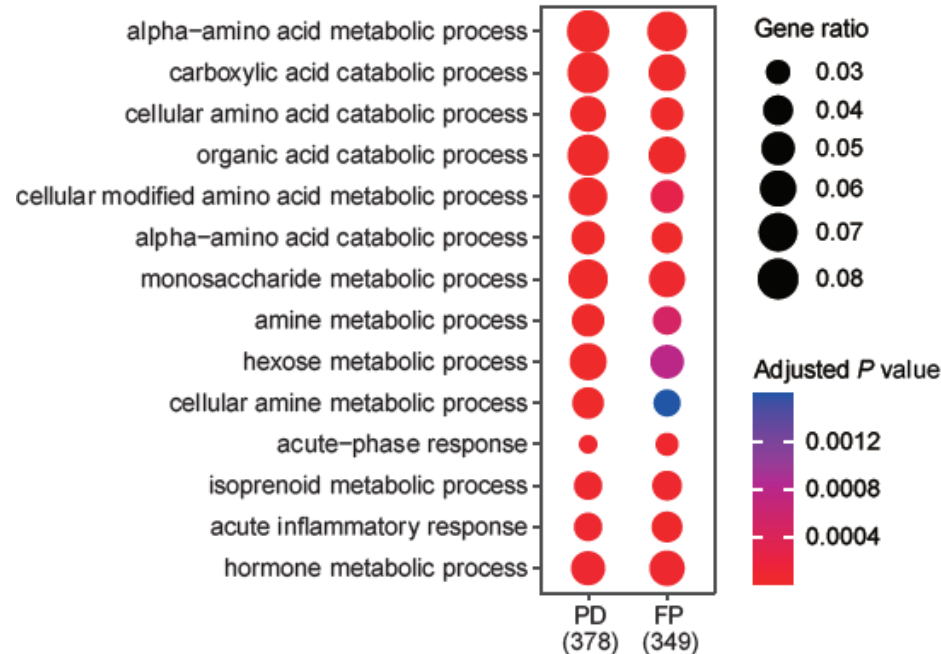
Enriched biological processes in the cortex



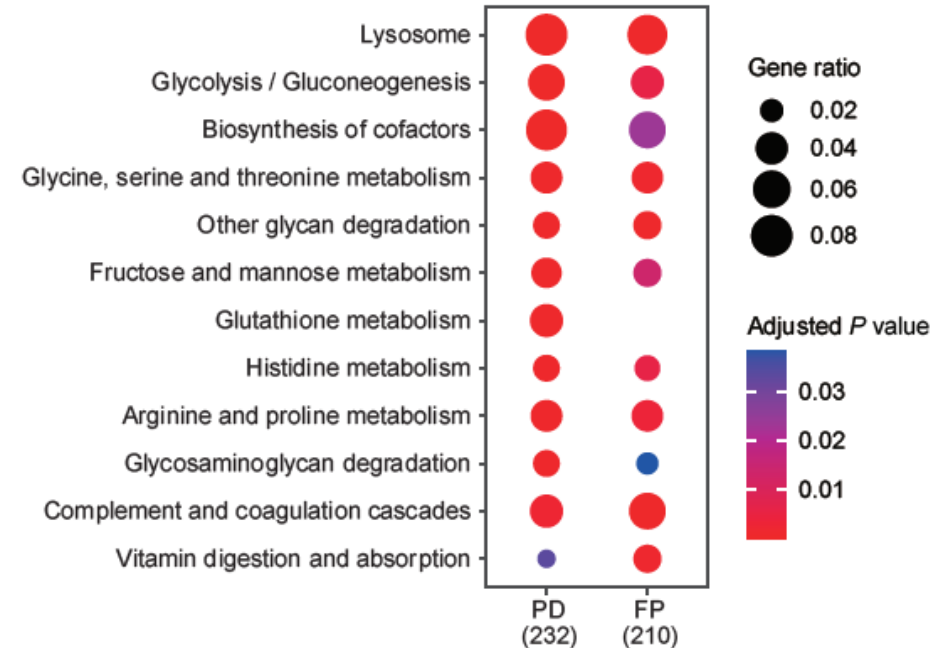
Enriched KEGG pathways in the cortex



Top enriched biological processes in the cortex



Top enriched KEGG pathways in the cortex



The ten most significantly enriched GO biological processes and KEGG pathways in the cortex DEPs identified by PD and FP

Summary

PD and FP quantification results are similar in many aspects:

- protein identification
- protein abundance
- batch effect
- differential expression
- functional enrichment

PD identified more proteins; FP required shorter computational time.

Note: Conclusions may change if there are major updates of the two software.

Publication

Journal of
proteome
research

pubs.acs.org/jpr

Technical Note


Comparative Evaluation of Proteome Discoverer and FragPipe for the TMT-Based Proteome Quantification


Tianen He,[#] Youqi Liu,[#] Yan Zhou, Lu Li, He Wang, Shanjun Chen, Jinlong Gao, Wenhao Jiang, Yi Yu, Weigang Ge, Hui-Yin Chang, Ziquan Fan, Alexey I. Nesvizhskii,* Tiannan Guo,* and Yaoting Sun*


 Cite This: *J. Proteome Res.* 2022, 21, 3007–3015

 Read Online

ACCESS |

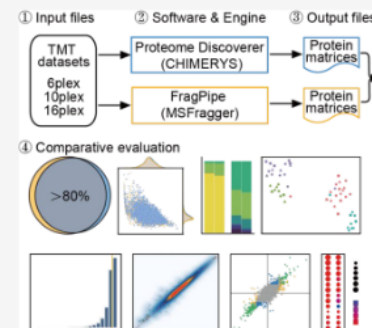
 Metrics & More

 Article Recommendations

 Supporting Information

ABSTRACT: Isobaric labeling-based proteomics is widely applied in deep proteome quantification. Among the platforms for isobaric labeled proteomic data analysis, the commercial software Proteome Discoverer (PD) is widely used, incorporating the search engine CHIMERYS, while FragPipe (FP) is relatively new, free for noncommercial purposes, and integrates the engine MSFragger. Here, we compared PD and FP over three public proteomic data sets labeled using 6plex, 10plex, and 16plex tandem mass tags. Our results showed the protein abundances generated by the two software are highly correlated. PD quantified more proteins (10.02%, 15.44%, 8.19%) than FP with comparable NA ratios (0.00% vs. 0.00%, 0.85% vs. 0.38%, and 11.74% vs. 10.52%) in the three data sets. Using the 16plex data set, PD and FP outputs showed high consistency in quantifying technical replicates, batch effects, and functional enrichment in differentially expressed proteins. However, FP saved 93.93%, 96.65%, and 96.41% of processing time compared to PD for analyzing the three data sets, respectively. In conclusion, while PD is a well-maintained commercial software integrating various additional functions and can quantify more proteins, FP is freely available and achieves similar output with a shorter computational time. Our results will guide users in choosing the most suitable quantification software for their needs.

KEYWORDS: *FragPipe, Proteome Discoverer, tandem mass tag, labeled quantitative proteomics, mass spectrometry*



Acknowledgements





THANK YOU

西湖大學
WESTLAKE UNIVERSITY