

A streamlined workflow for the quality control, interpretation and analysis of timsTOF Pro data: towards large-scale proteomics

USHUPO 2021, P08.17

Chengpin Shen¹, Xianming Liu², Xiaoxian Du², Xiaoqing Wang¹, Hebin Liu¹ and Ruoyu Wu²

¹ Omicsolution Co., Ltd, Shanghai, China
² Bruker Daltonics, Shanghai, China

Introduction

The importance of quality control (QC) has been long acknowledged, especially in the large-scale proteomics researches, but most of them only utilized protein quantitation sheet produced by search engine. Such kind of QC lacks comprehensive LC and MS level information that may influence dataset quality in all aspects. Here, we introduce a web based proteomics data interpretation platform, OmicsCloud, to provide a streamlined MS data processing solution. All the factors that affect the data quality like missed cleavage, contaminants, chromatographic tailing, contaminants and etc. are evaluated. Furthermore, large-scale DIA/DDA data quantitation together with missing value imputation, batch effect correction, project-wide normalization and etc. are provided. Several species and organ specific libraries were generated by Bruker timsTOF Pro platform, which are set as reference libraries to evaluate user's proteomics data coverage depth. We also established an automatic proteome sequence updating and annotation pipeline to provide an up-to-date bioinformatics analysis, including gene ontology, KEGG, interpro, COG/KOG, STRING and etc. In conclusion, a standardized MS data QC, large-scale DIA quantitation and bioinformatics platform called OmicsCloud can tremendously facilitates large-scale proteomics research.

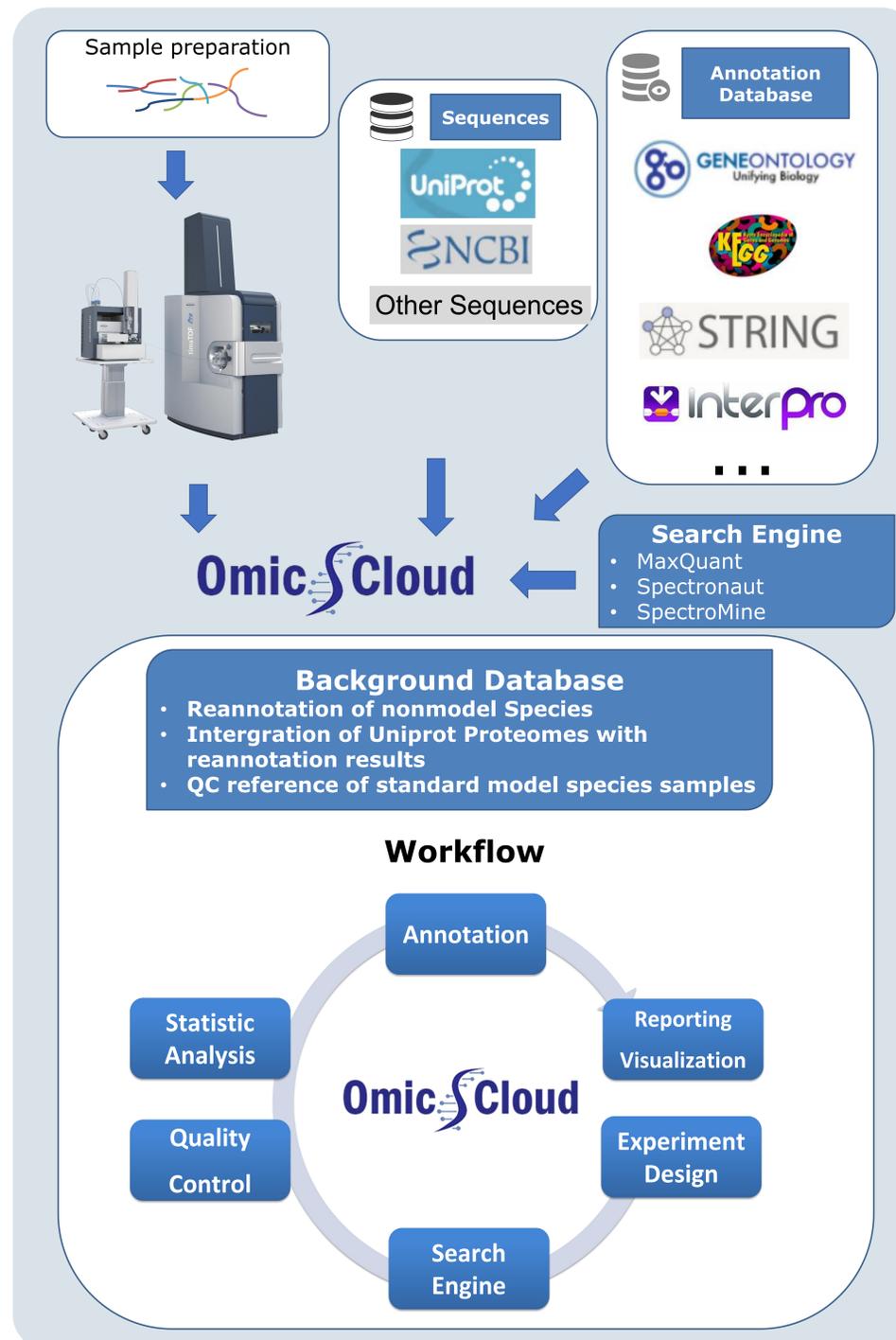


Figure 1: Analysis workflow of OmicsCloud:

The complete workflow can be fully automated after uploading of MS raw files and set up experiment design.

LC & Quantitative Quality Control				
Mass Accuracy	Missed Cleavage	FWHM	Dynamic Range	LC tailing
ID frequency	ID redundancy	Contaminants	<u>Seq database Optimization</u>	<u>Unintended PTM</u>
<u>QC Reference compare</u>	<u>Labeling efficiency</u>			
Quantitative Quality Control				
DPPP	Peaks Capacity	Normalization	Delta RT	Absolute Deviation
Quantity Per Sample Distribution	Data Completeness	HCA	PCA	Batch effect correction

Table 1&2: QC items

We designed QC workflow for single sample ID, multiple sample quantitative analysis. Items with underline will be available soon.

Results

- We developed an automated workflow including Uniprot proteome sequence fetching, sequence annotation updating based on blast algorithm.
- We are continuously collecting standard model species samples as QC references.
- A demo server with several published data is available on project.omicsolution.com/omicscloud/index
- Demo projects include both DDA/DIA PASEF data, HeLa and large scale clinical data from published literature. More types of data will soon available online.

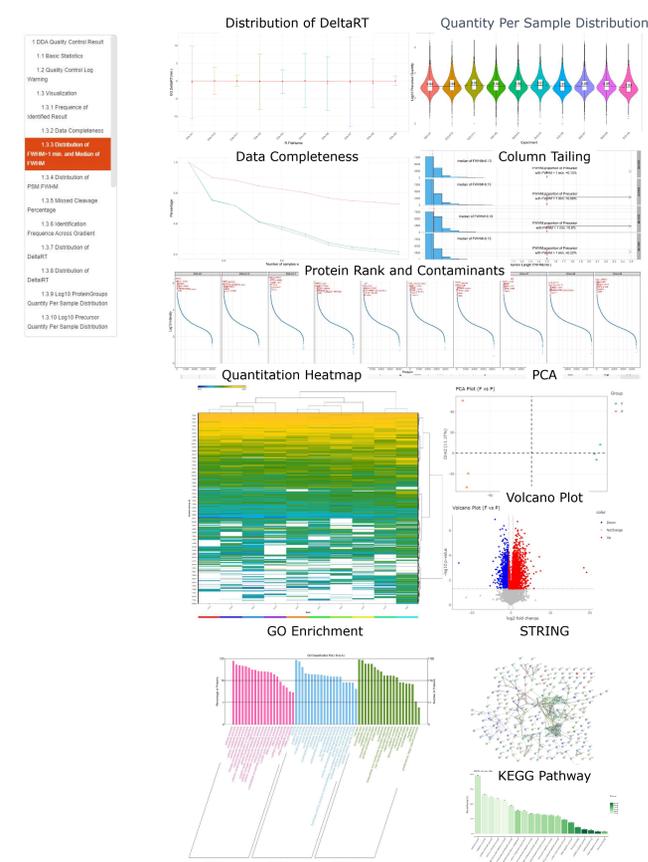


Figure 2: Screenshot of OmicsCloud result: A typical workflow of quantitative proteomics results in OmicsCloud include all information you need from QC, statistical analysis and functional annotations. Some QC figures are generated by Spectronaut and intergrated into the report.

Methods

Sample preparation was performed using iST Kit (PreOmics GmbH) and optimized for different types of samples. A nanoElute (Bruker Daltonics) nano-flow LC was coupled to a high-resolution TIMS-QTOF (timsTOF Pro, Bruker Daltonics) with a CaptiveSpray ion source (Bruker Daltonics). The peptide mixtures were loaded onto a 250 mm pulled emitter column (IonOpticks, Australia). DDA/DIA PASEF data was analyzed by SpectroMine/Spectronaut (Biognosys AG). QC and downstream annotation were all analyzed on OmicsCloud(Omicsolution Co., Ltd).

Conclusions

- QC analysis is extremely important. We developed a streamlined workflow of all aspects of proteomics.
- Our workflow can be distributed locally or used online.
- A specific QC algorithm containing database search, de novo sequencing, PTM open search is under development.
- Multi-omics analysis like RNA, protein, PTM and metabolite will be available in next version.
- More statistical and machine learning tools are available on our WuKong Omics platform. www.omicsolution.com/wkomics/main English version is under translation.