

UHR-IonStar App User Manual



Version 1.5

01/2022

Outline

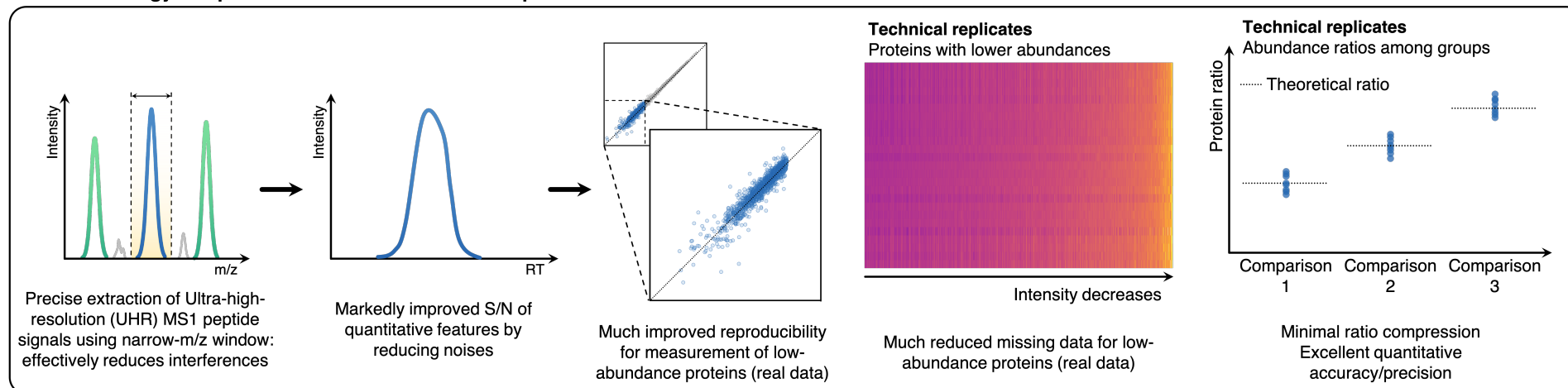
- **Introduction**
- **App setup**
- **Application**

Introduction

IonStar, recent updated to ultra-high-resolution- (UHR-) IonStar, is a MS1-based quantitative proteomics procedure which achieves accurate and robust proteomic quantification in large cohorts, even including low-abundant regulatory proteins.

IonStar effectively takes advantage of the high sensitivity/selectivity attainable by ultra-high-resolution-MS1 acquisition (*e.g.*, 120k-240k FWHM@ $m/z=200$) which is now widely available on Ultra-High-Field(UHF) Orbitrap instruments. By selectively and accurately procuring quantitative features of peptides within precisely-defined, very-narrow m/z windows appropriate to UHR-MS1-resolution, the method minimizes co-eluted interferences and substantially enhances S/N of low-abundance species.

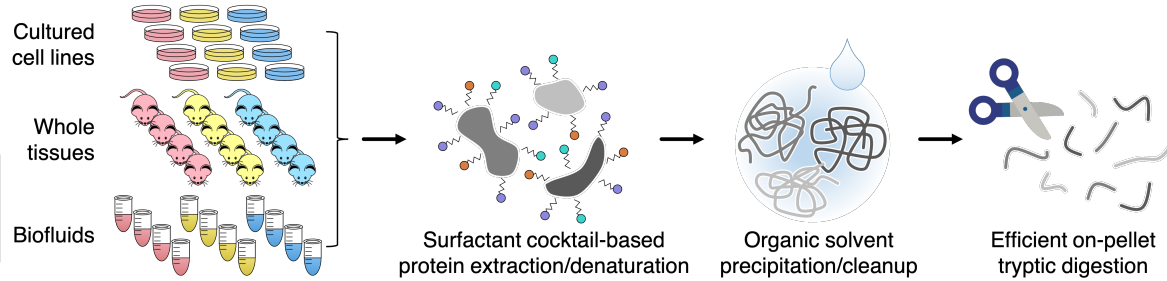
IonStar strategy for quantification of low-abundance proteins



UHR IONSTAR

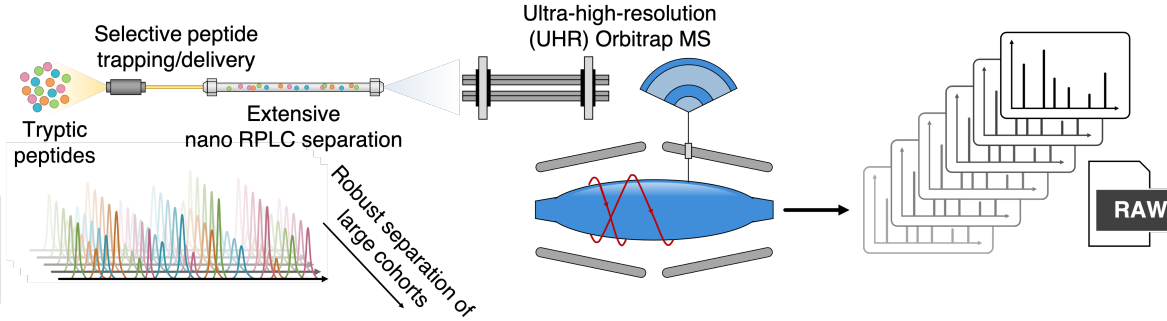
Sample preparation

SEPOD protocol for efficient and reproducible preparation of large sample cohorts



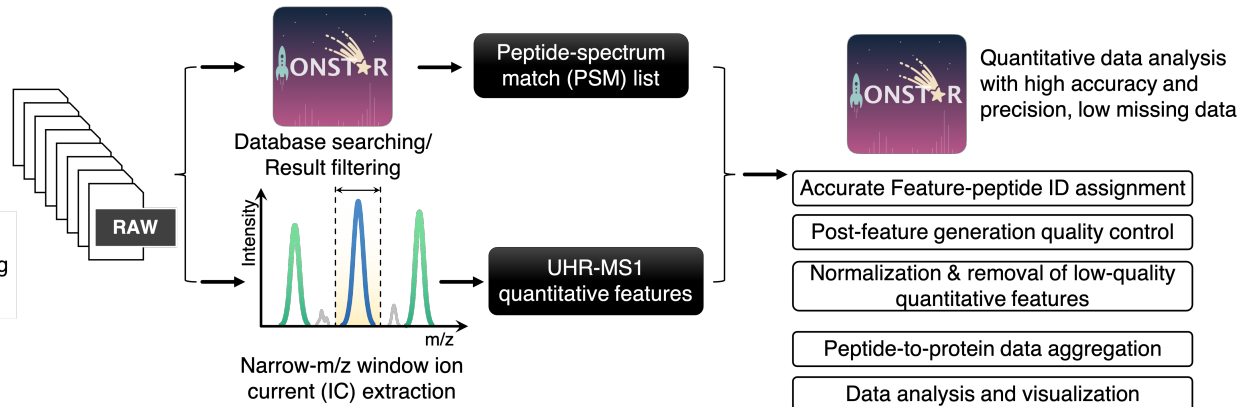
LC-MS analysis

A trapping nano LC-MS for extensive, sensitive analysis, robustly across large cohorts

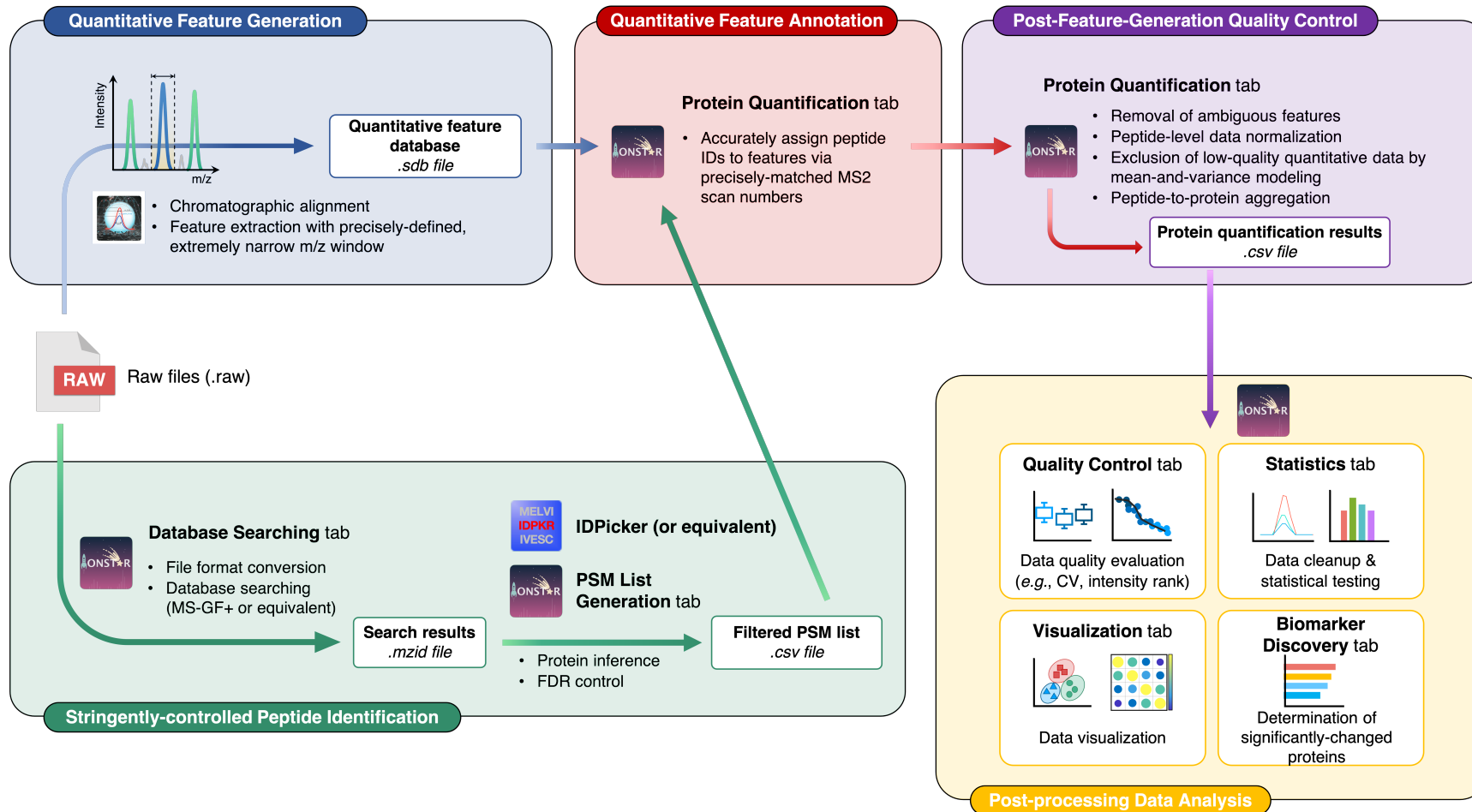


Data processing

A unique MS1-based strategy for reproducible quantification by taking full advantage of the UHR-MS1



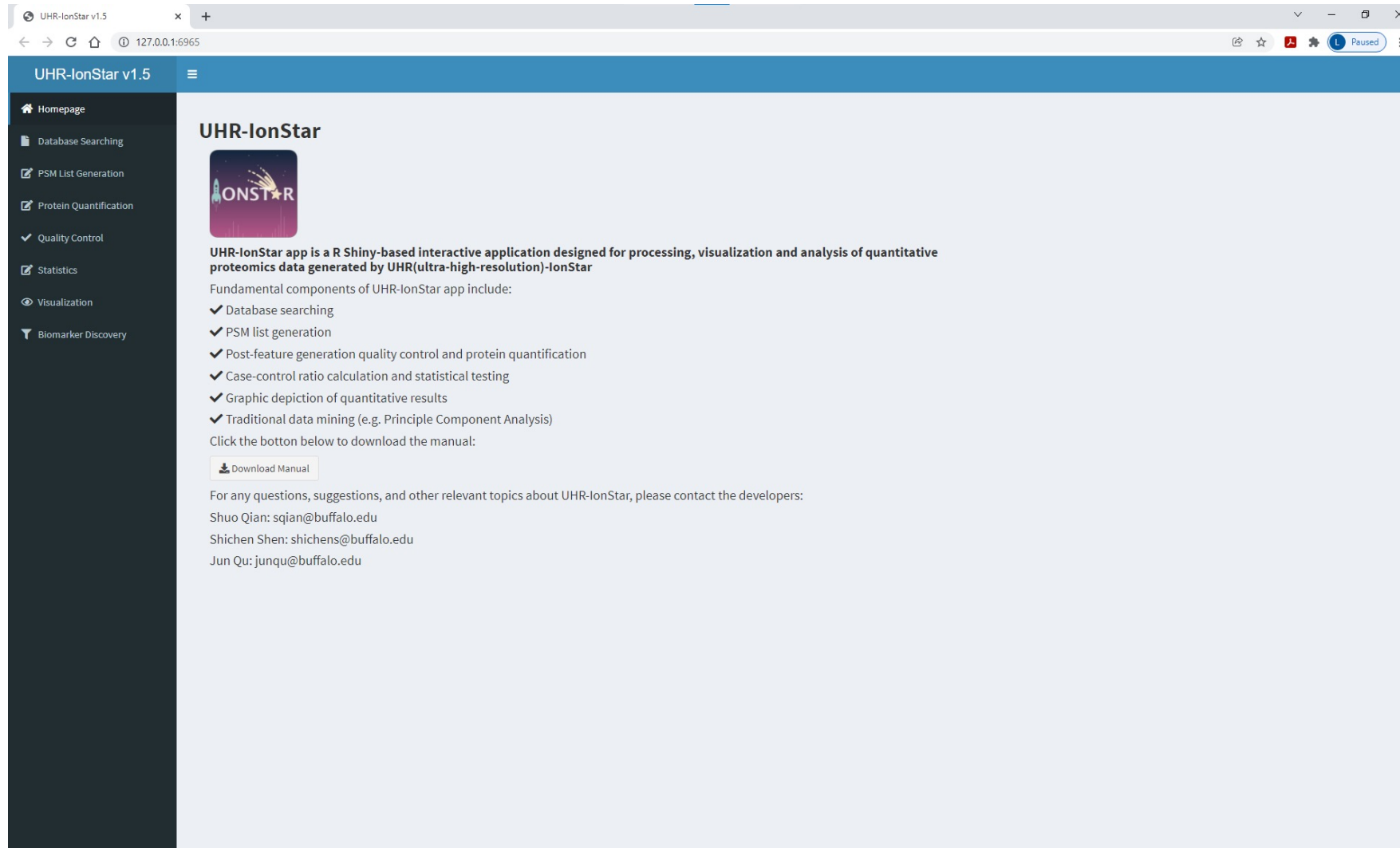
IonStar is not just a data processing protocol. Despite data acquisition and post-acquisition processing, IonStar contains a surfactant cocktail-aided extraction/precipitation/on-pellet digestion (SEPOD) protocol achieving efficient and reproducible sample preparation across large cohorts, and a trapping nano LC-UHR Orbitrap MS system for extensive and sensitive peptide analysis, robustly for large sample sets.



UHR-IonStar app is a R shiny-based interactive application designed for processing, visualization and analysis of quantitative proteomics data generated by UHR-IonStar.

<https://github.com/JunQu-Lab/UHRlonStarApp>

If everything is fine, the web interface would pop up after running the app starting code:



Application

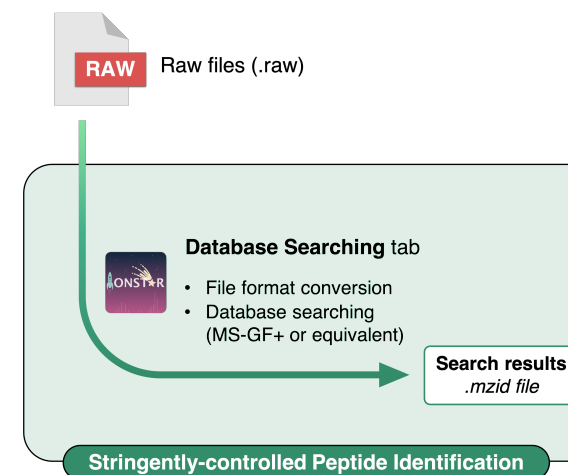
A carefully-designed multi-species spike-in sample set was constructed for evaluating the quantitative performance of label-free proteomics method(s). It was prepared by spiking small, variable amounts of E. Coli (for mimicking significant protein changes, i.e., true positives) and yeast (for balancing the variable levels of E. coli proteins) proteins into a large, constant background of human proteins (representing unchanged proteins). The sample set encompasses a total of 25 LC-MS sample runs (5 E.coli-level groups, 5 LC-MS replicate runs in each group).

	Group A	Group B	Group C	Group D	Group E
Human (%)	60	60	60	60	60
E. Coli (%)	5	7.5	10	15	20
Yeast (%)	35	32.5	30	25	20

The raw data are available for download on PRIDE (<https://www.ebi.ac.uk/pride/archive/projects/PXD030780>)

1. Raw file conversion and database searching

- Open the UHR-IonStar App and navigate to the “**Database Searching**” tab.
- Click buttons to locate the file folder containing raw files (.raw) and reference database of protein sequence (.fasta), and the folder of the MS-GF+ java file (.jar).
- Next, set the number of processing threads and MS-GF+ parameters.
- Finally, click **Start formatting and searching** to firstly convert raw files to a compatible format for searching engine (.mzXML), then perform database searching. The identification result files (.mzid) will be generated under the same file directory as the raw files.



- Homepage
- Database Searching
- PSM List Generation
- Protein Quantification
- Quality Control
- Statistics
- Visualization
- Biomarker Discovery

Raw file conversion and Database searching

E:/Data/TechAE25/

Select the folder containing raw files and the fasta database

E:/Data/

Select the folder containing the MSGF+ java file

Set the number of processing threads:

1 2 12

Start formatting and searching

MSGF+ Parameter Settings

Select if adding decoys or not.

Input precursor mass tolerance (ppm):

20

Input isotope error range (separated by comma):

-1,2

Choose the fragmentation method:

HCD

Choose the MS instrument:

QExactive

TOF

Select the enzyme:

Trypsin

Select the sample treatment protocol:

No protocol

Specify the cleavage specificity:

Fully tryptic peptides

Semitryptic peptides

unspecific peptides

Input peptide length range (separated by comma):

6,40

Input peptide charge range (separated by comma):

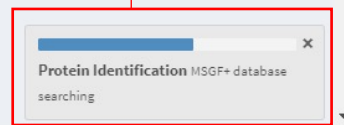
2,7

Input the number of matches to report per spectrum:

1

A small widget will pop up after clicking these buttons for guiding users selecting file directories.

After clicking **Start formatting and searching**, this widget will pop out for indicating which step the app is currently working.

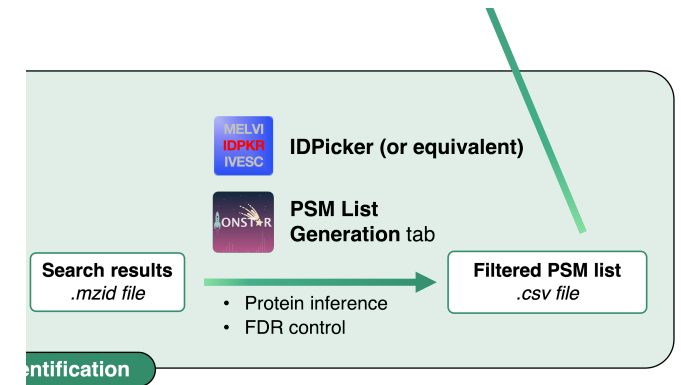


Recommended parameters for MS-GF+

Parameters	Recommended settings	Parameters	Recommended settings
Adding decoys	True	Cleavage specificity	Fully
Precursor mass tolerance (ppm)	20	Peptide length range	6, 40
Isotope error range	-1, 2	Peptide charge range	2, 7
Fragmentation method	HCD	The number of matches to report per spectrum	1
MS instrument	Q-Exactive	Maximum modification number per peptide	3
Enzyme	Trypsin	Fixed modifications	Carbamidomethyl (C)
Sample treatment protocol	No protocol	Variable modifications	Oxidation (M), Acetyl (N-term)

2. File conversion from IDPicker

The outputs from IDPicker lack several important variables which are necessary for frame annotation. In this part, UHR-IonStar app convert the original identification result files (.mzid) to directly readable format (.tsv) and assigns the additional information for filtered results from IDPicker to finally generate a PSM list.



- Open the UHR-IonStar App, then navigate to the “**PSM List Generation**” tab. Firstly, input the protein, peptide, and spectrum lists generated by IDPicker, and set the file directory containing all .mzid files. Then enter the file name of the filtered PSM list to be generated. Click the button **Generate PSM List** to start processing. The filtered PSM list is in the same file directory as .mzid files.

- Homepage
- Database Searching
- PSM List Generation**
- Protein Quantification
- Quality Control
- Statistics
- Visualization
- Biomarker Discovery

PSM List Generation

Upload Protein List from IDpicker:

Browse...

Proteins.tsv

Upload complete

Upload Peptide List from IDpicker:

Browse...

Peptides.tsv

Upload complete

Upload Spectrum List from IDpicker:

Browse...

Spectra.tsv

Upload complete

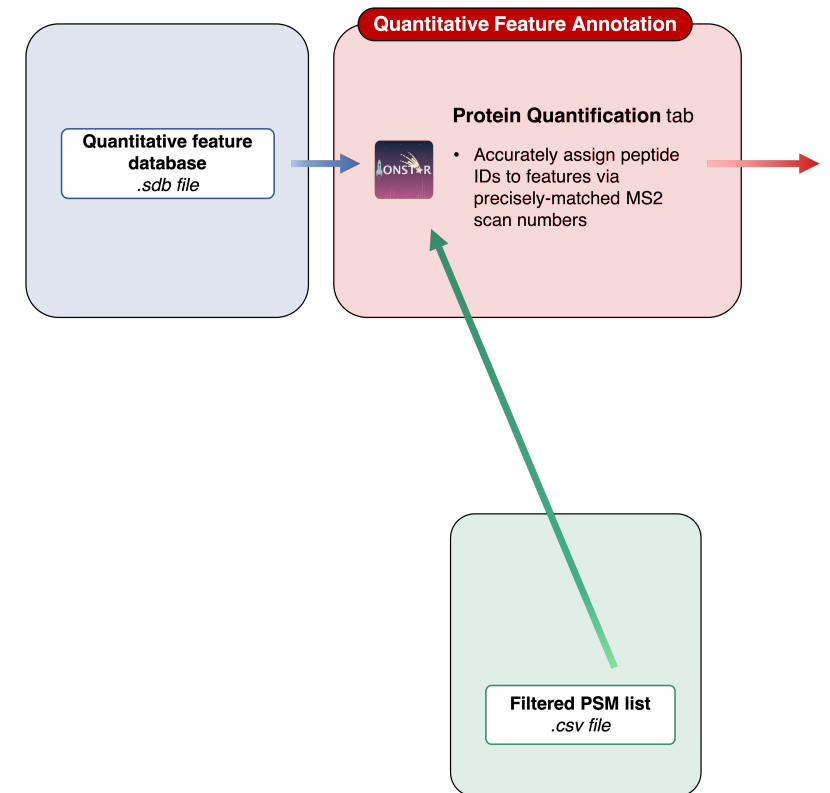
Name your PSM list ('.csv' is required):

PSM list will be generated at the same folder with .mzid files.

PSM List Generation

3. Frame annotation and quality control

- Navigate to the “**Protein Quantification**” tab, upload the feature database (.sdb) and the filtered PSM list (.csv). Then specify the ordinal numbers for the four designated columns. Click the button **Generate Frames** to start data processing and download the Annotated Frame List and the Sample ID file.



- Homepage
- Database Searching
- PSM List Generation
- Protein Quantification**
- Quality Control
- Statistics
- Visualization
- Biomarker Discovery

Protein Quantification

Frames Generation:

Upload SIEVE database (.sdb)

Browse... TechAE25.sdb

Upload complete

Please upload PSM list after you see "Upload complete" on the bar above!

Upload PSM list (.csv):

Browse... TechAE25_PSM_list.csv

Upload complete

Several settings in PSM list:

Specify the number of column which contains filename information:

1

Specify the number of column which contains MS2 scan number:

3

Specify the number of column which contains protein accessions:

16

Specify the number of column which contains peptide sequence:

15

Generate Frames

Download Annotated Frame List

Download Sample ID

Protein Quantification:

Upload annotated frame List(.csv)

Browse... No file selected

Upload sample id file (.csv)

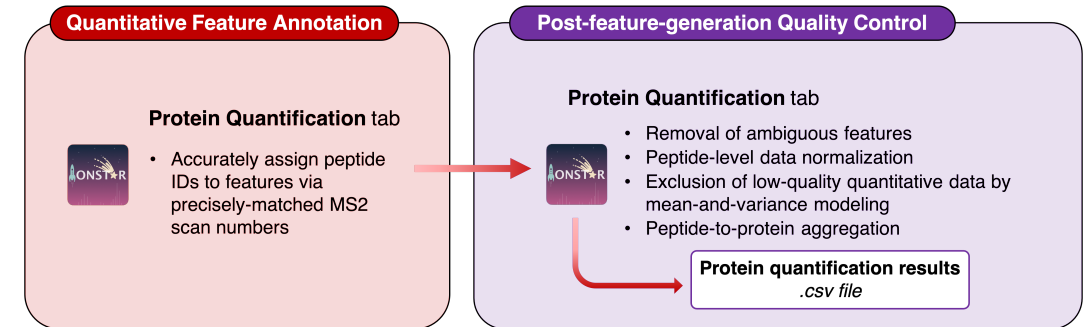
Browse... No file selected

IonStarStat

Please open the PSM list to make sure the column number of these requirements. Other columns will not be used for generating annotated frame dataset.

3. Frame annotation and quality control

- Open the Sample ID file and edit it by assigning group ID to all sample files. Re-upload the Annotated Frame List file and Sample ID file into the UHR-IonStar App.



- Select methods for normalization at the whole-dataset level and peptide-to-protein aggregation.
- For quality control of the quantitative features, we highly recommend selecting the options for detection and exclusion of low-quality quantitative data using mean-and-variance modeling at both feature (frame) level and peptide level to achieve high-quality protein quantification. Click **Perform Protein Quantification** to start data processing.
- Upon finishing, proteomic quantification results can be downloaded by clicking the **Download Quantification Data** button. Additional plots such as intra-group CV plot and protein abundance rank plot are generated under the “Quality Control” tab, which can be downloaded together.

1

Specify the number of column which contains MS2 scan number:


3



Specify the number of column which contains protein accessions:

16

Specify the number of column which contains peptide sequence:

15

 Generate Frames

 Download Annotated Frame List  Download Sample ID

Protein Quantification:

Upload annotated frame List(.csv)

Browse... TechAE_annotated_frame.csv

Upload complete

Upload sample id file (.csv)

Browse... TechAE_sampleid.csv

Upload complete

Select inter-group peptide normalization method:

Total Ion Currents (TIC)

Quantile

No normalization

Choose whether you would like to remove low-quality quantitative data at the frame level:

Frame-level Quality Control


Choose whether you would like to remove low-quality quantitative data at the peptide level:


Peptide-level Quality Control

Select the method that aggregates peptides to proteins

The Sum of Intensities

General Linear Mixed Model (GLMM)

 Perform Protein Quantification

 Download Quantification Data  Download Peptides

TIC normalization: normalize all samples to have the same total ion current.

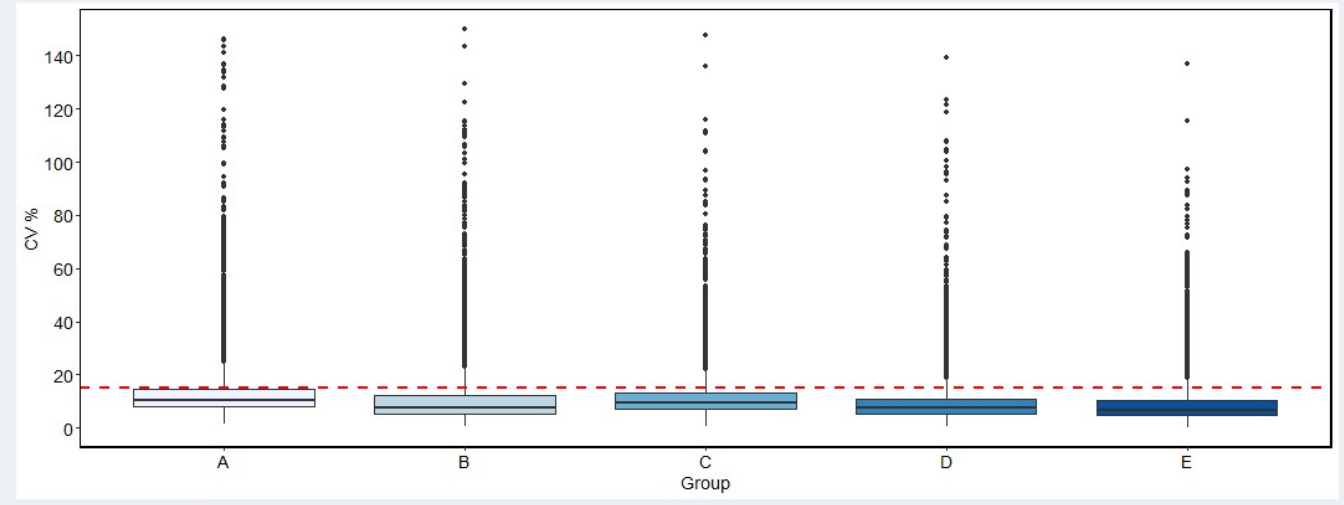
Quantile normalization: normalize all samples to have the same distribution of peptide abundance.

- Homepage
- Database Searching
- PSM List Generation
- Protein Quantification
- Quality Control**
- Statistics
- Visualization
- Biomarker Discovery

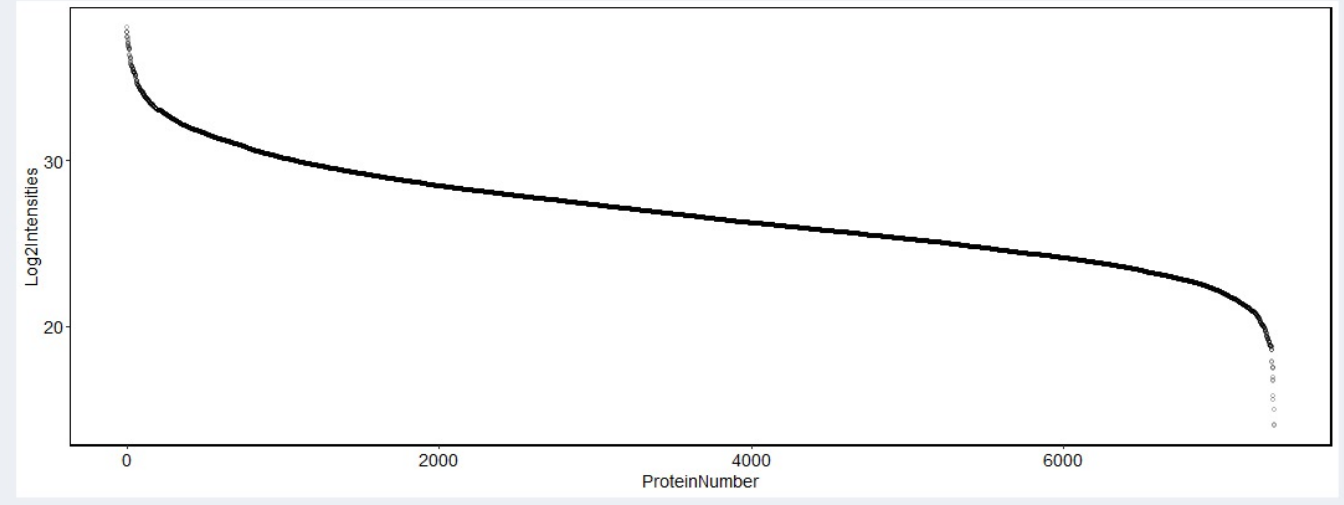
Quality Control

Download all quality plots

Intra-group CV plot



Protein Rank plot

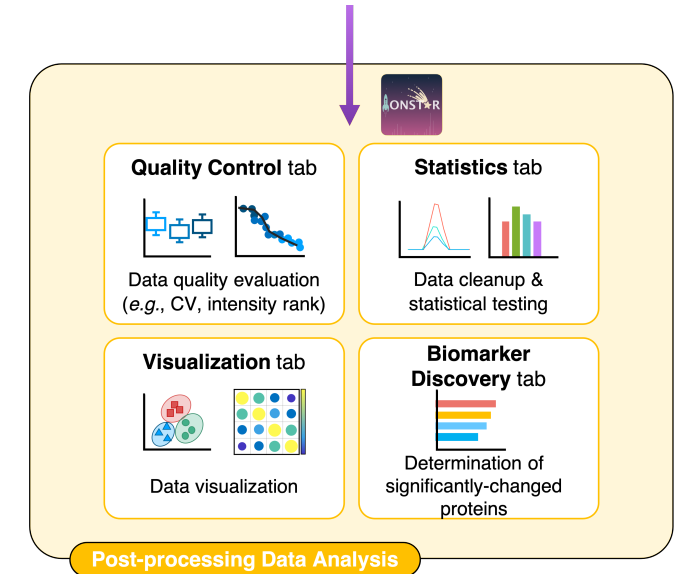


Mass Shift plot

4. Post-quantitation data analysis

(Optional, and still updating) Post-quantification data processing, statistical analysis and data visualization functions are also provided in the UHR-IonStar App.

- Under the “**Statistics**” tab, options are available to remove proteins containing missing data. Various statistical tests can also be performed.
- Under the “**Visualization**” tab, several types of data plots such as Inter-group correlation plot, Principal Component Analysis (PCA) plot, Ratio distribution plot can be graphed; under the “**Biomarker Discovery**” tab, significantly-changed proteins can be identified based on protein fold-of-change and p-value cutoff thresholds. We suggest employing the Experimental Null method to determine optimal cutoffs, while the users can also define their own values. Results can be visualized in volcano plots and exported as .csv files.



- Homepage
- Database Searching
- PSM List Generation
- Protein Quantification
- Quality Control
- Statistics**
- Visualization
- Biomarker Discovery

Protein Data Processing

Upload quantitative results (.csv):

Browse... TechAE25_proteins.csv

Upload complete

Upload group file (.csv):

Browse... TechAE_sampleid.csv

Upload complete

Please fill in your control group shown in the right, case sensitive (default is the first one):

Assign control group

X

Select method of statistical testing:

Original t-test

Variance equality:

 TRUE FALSE

Decoy Protein identifier

XXX

Start Data Processing

Processing completed.

Download Processed Quantity Results

Download Average Ratios

Download All Together

Note: Decoy entries and proteins with missing data will be removed automatically during the data processing step.

Quantitative Results

Show 10 entries

Search:

	Rawfiles	GroupID
1	A04	A
2	A05	A
3	B01	B
4	B02	B
5	B03	B
6	B04	B
7	B05	B
8	C01	C
9	C02	C
10	C03	C

Showing 1 to 10 of 25 entries

Previous 1 2 3 Next

- Homepage
- Database Searching
- PSM List Generation
- Protein Quantification
- Quality Control
- Statistics
- Visualization
- Biomarker Discovery

Data Visualization

Select plot type:

Inter-group correlation plot

***For Inter-group correlation plot ONLY:**

Please fill in the groups you want to see, case sensitive (Default is the first two groups):

Group 1

A

Group 2

E

Plot

Plot saving options:

Recommends larger width and height values when plots for Pearson correlation matrix and PCA.

Plot width



Plot height



Download Current Plot

Ratio distribution plots settings:

Choose group/groups you want to show in ratio distribution plot:

- B
- C
- D
- E

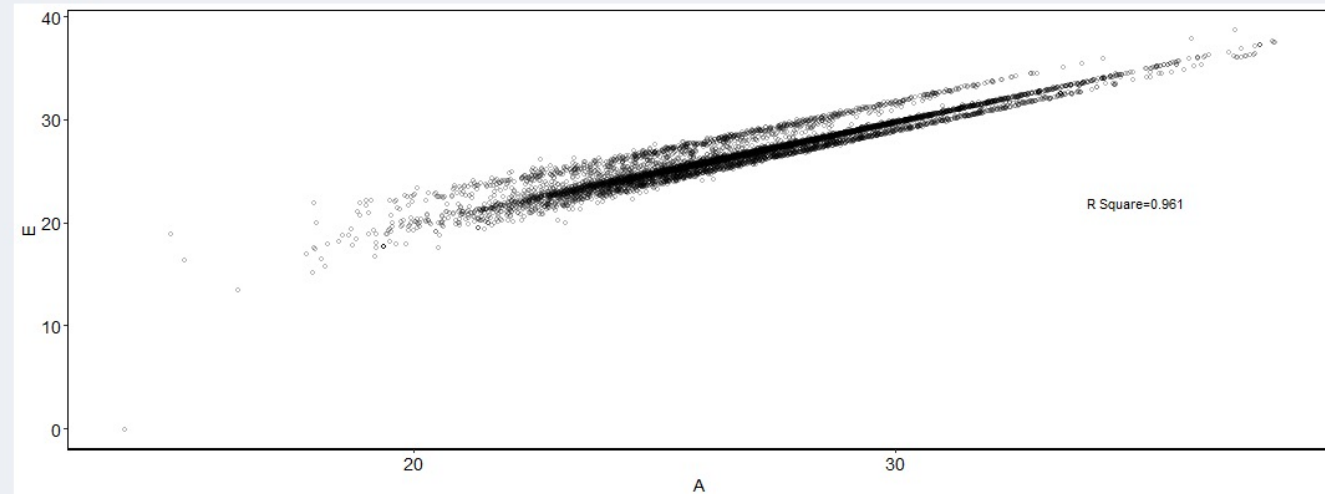
Choose whether you want to do distribution correction:

Correction

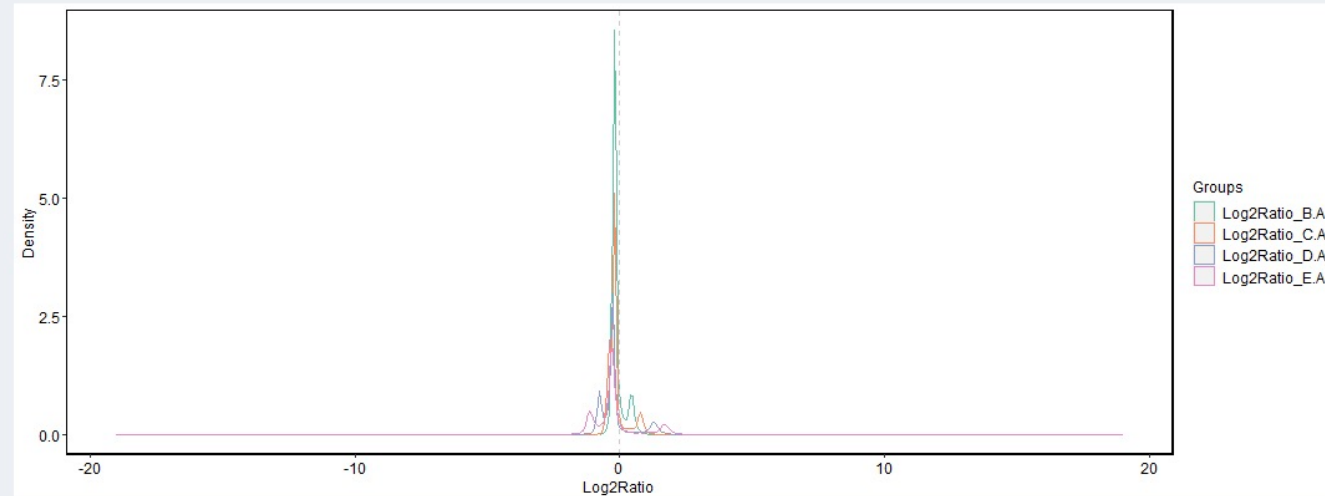
Plot for Ratio Distribution

Plot width

Plot Area



Area for Ratio Distribution Plots



Thank you for using UHR-IonStar!

For any questions, suggestions and other relevant topics about UHR-IonStar, please contact the developers:

Shuo Qian: sqian@buffalo.edu

Shichen shen: shichens@buffalo.edu

Jun Qu: junqu@buffalo.edu

Relevant articles:

- Shen, Xiaomeng, et al. "IonStar enables high-precision, low-missing-data proteomics quantification in large biological cohorts." *Proceedings of the National Academy of Sciences* 115.21 (2018): E4767-E4776.
- Wang, Xue, et al. "Ultra-High-Resolution IonStar Strategy Enhancing Accuracy and Precision of MS1-Based Proteomics and an Extensive Comparison with State-of-the-Art SWATH-MS in Large-Cohort Quantification." *Analytical chemistry* 93.11 (2021): 4884-4893.