

User Manual

[Update Log](#)

[Demo Website](#)

[Url](#)

[Demo Screenshot](#)

1 General Information

[1.1 Scope of ProPro Software](#)

[1.2 Computer System Requirements](#)

[1.3 Acquisition Method and Vendor File Conversion](#)

[1.4 Code Available](#)

[1.5 License](#)

2 Getting Started

[2.1 Associated software](#)

[2.2 Deployment](#)

[2.2.1 Single-Node Deployment](#)

[For windows users](#)

[For mac/linux users](#)

[2.2.2 Multi-Nodes Deployment](#)

[MongoDB Nodes](#)

[Redis Node](#)

[2.3 Configuration](#)

[2.3.1 ProPro](#)

[2.4 Demo](#)

3 How to use ProPro

[3.1 Data Prepare](#)

[3.2 Library](#)

[3.2.1 Import Library](#)

[3.2.2 Generate Decoy](#)

[3.2.3 Peptide List](#)

[3.2.4 Peptide Detail](#)

[3.3 Method](#)

[3.3.1 Method List](#)

[3.3.2 Method Detail](#)

[3.4 Project](#)

[3.4.1 Init Project](#)

[3.4.2 Scan Project](#)

[3.4.3 Update Project](#)

[3.4.4 Runs & Overviews](#)

[3.4.5 Run List](#)

[3.4.6 Run Detail](#)

[3.4.7 Overview List](#)

[3.4.8 Analysis Task](#)

[3.5 Protein Clinic](#)

[3.5.1 Protein&Peptide List](#)

[3.5.2 EIC](#)

[3.5.2.1 Base Information](#)

[3.5.2.2 Quick Tools](#)

[3.5.2.3 Runs name and Fragments](#)

[3.5.2.4 Marked Peak Group](#)

[3.5.2.5 Spectrum Viewer](#)

[3.5.2.6 Manual Peak Picker](#)

[3.5.2.7 Tags](#)

[3.5.3 Scores Result](#)

[3.5.3.1 Score Type List](#)

[3.5.4 LFQBench](#)

[3.5.5 iRT Results](#)

[3.5.6 RT Distribution](#)

[3.5.7 Target Decoy Distribution](#)

Update Log

Version	Update Data	Log	Editor
1.0.0	2021.12.15	First version for user manual	James Lu

Demo Website

Url

Demo Screenshot



ProPro Protein Clinic

1 General Information

1.1 Scope of ProPro Software

ProPro is an open-source web-based platform for **DIA/SWATH targeted MS data analysis**.

ProPro contains the core modules for DIA data analysis: iRT, data extractor, noise estimator, peak finder, peak group picker, peak group scoring, peptide quantification, semi-supervise algorithm with **LDA or XGBoost**. In addition, ProPro also provides novel functions such as **multi runs comparison and alignment, library management, multi strategy decoy peptides generator, manual peak calibration** and so on.

ProPro is an web-based platform which is written by Java. It is built using **SpringBoot + MongoDB + Redis + React**. We can use One-Click mode to install ProPro on a single personal computer, or deploying ProPro on multiple Linux servers to handle larger scale data.

1.2 Computer System Requirements

ProPro is available for Windows 7(or above), Linux and MacOS. The minimum and recommended system specifications for **single-node deployment** are described in Table 1.

Specifications	Minimum	Recommended
CPU	Intel Core i5 10-gen or AMD R5	Intel Core i9 12-gen or AMD R9 5900
Hard drive	50GB free space	2x data set size
RAM	16GB	64GB(Core i9–10gen), 128GB(Core i9–12gen)

Table 1. **single-node deployment requirements**

the **multi-nodes deployment** requirements are described in Table 2.

Server Requirements	Deploy software
Core i7–10gen, 32GB RAM, 5TB SSD	MongoDB(Main Node), Redis
Core i7–10gen, 32GB RAM, 5TB SSD	MongoDB(Child Node, optional)
NAS(such as Synology RS2821), 100TB HDD	Data Repository
Core i9–12gen, 128GB RAM, 500GB SSD	ProPro(Core computing node)
Core i7–10gen, 16GB RAM, 100GB HDD	AirdPro(Data Conversion)

Table 2. **multi-nodes deployment requirements recommended**

1.3 Acquisition Method and Vendor File Conversion

ProPro now support for DIA/SWATH data acquisition. Due to the need to support the whole platform of mass spectrometry file processing, we use a self-developed Computing-Oriented high compression data format—Aird. That means users should convert the vendor file to the aird format files using AirdPro. AirdPro supports the vendor files including .raw, .wiff and .d formats.

1.4 Code Available

ProPro is divided into two projects. One is the back-end project, which is written in Java. The other is front – End project, which is based on React.

AirdPro: <https://github.com/CSi-Studio/AirdPro>

ProPro-Server: <https://github.com/CSi-Studio/ProPro-Server>

ProPro-React: <https://github.com/CSi-Studio/ProPro-React>

1.5 License

ProPro-Server and ProPro-React are both licensed under the Apache License 2.0

 CSi-Studio/ProPro is licensed under the Apache License 2.0	Permissions ✓ Commercial use ✓ Modification ✓ Distribution ✓ Patent use ✓ Private use	Limitations ✗ Trademark use ✗ Liability ✗ Warranty	Conditions ① License and copyright notice ② State changes
--	---	--	--

ProPro-Server License

 CSi-Studio/ProPro-V2-Front is licensed under the Apache License 2.0	Permissions ✓ Commercial use ✓ Modification ✓ Distribution ✓ Patent use ✓ Private use	Limitations ✗ Trademark use ✗ Liability ✗ Warranty	Conditions ① License and copyright notice ② State changes
---	---	--	--

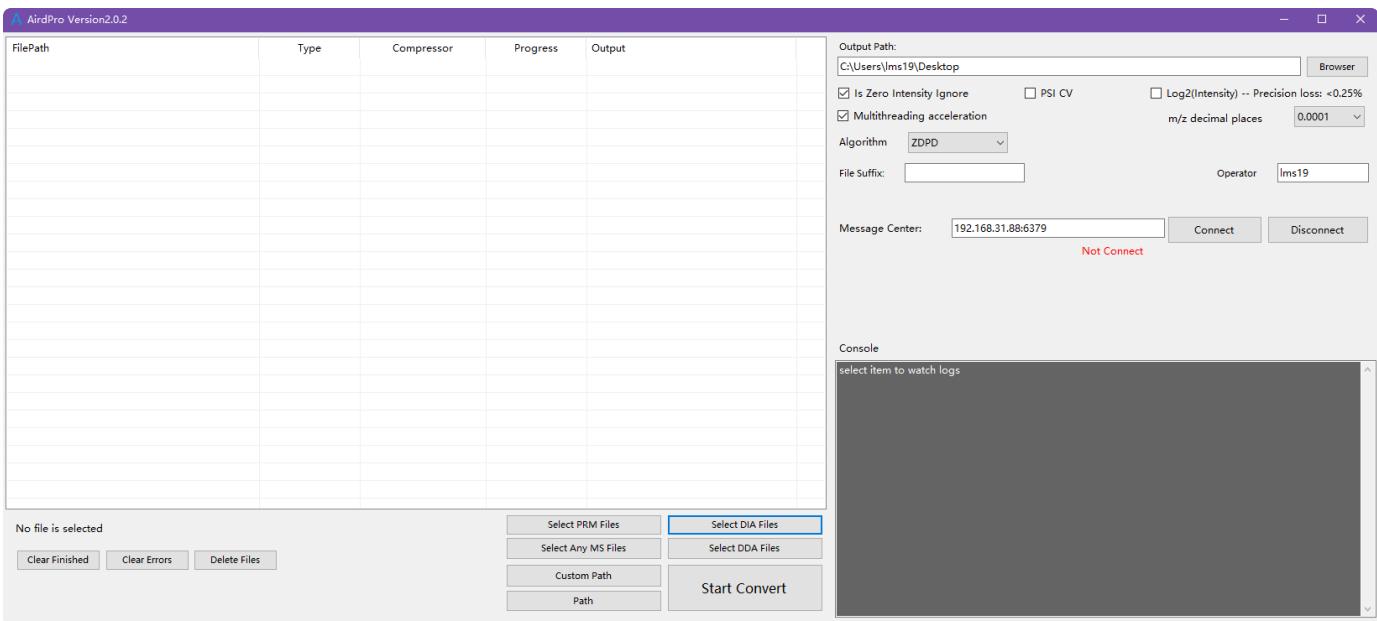
ProPro-React License

2 Getting Started

2.1 Associated software

ProPro depend on the database and basic software below. All dependencies are free to use.

Software	Version	Download Link
Java(or OpenJDK)	17.0.1	OracleJDK: https://www.oracle.com/java/technologies/downloads/ OpenJDK: http://openjdk.java.net/projects/jdk/17/
Redis	6.2.6	Windows: https://github.com/MSOpenTech/redis/releases Others: https://redis.io/download
MongoDB	5.0.5	https://www.mongodb.com/try/download/community
AirdPro	2.0.0	https://github.com/CSi-Studio/AirdPro/releases



AirdPro

2.2 Deployment

2.2.1 Single–Node Deployment

No need for download the software listed in Chapter 2.1. Users can download the zip/exe package which contains all the need software.

For windows users

Download the EXE/Zip setup file and double click the EXE file to install the ProPro on your computer, you'd better install the ProPro on the recommended disk which has an enough disk space. If your computer already install the redis server or the mongodb server. You'd better use the Zip setup file and do not install the existed software on your computer. Unzip the Zip file to the recommended disk. All the need software are in the package.

If you used the EXE setup to install ProPro, you can see the package like this:

名称	修改日期	类型	大小
bin	2021/12/15 15:49	文件夹	
dbfile	2021/12/15 15:54	文件夹	
jdk	2021/12/15 15:49	文件夹	
libs	2021/12/15 15:49	文件夹	
mongodb	2021/12/15 15:50	文件夹	
Redis	2021/12/15 15:50	文件夹	
unins000.dat	2021/12/15 15:50	DAT 文件	103 KB
unins000.exe	2021/12/15 15:49	应用程序	3,144 KB

You can uninstall the ProPro using the `unins000.exe`

For mac/linux users

Download the Zip setup file and unzip the file to the recommended disk which has an enough disk space.

2.2.2 Multi–Nodes Deployment

Multi–Node deployment is a solution that the team need to analyzing more than 500 DIA runs per week. Multi–Node distributed deployment uses multiple storage nodes and computing nodes to greatly cluster computing throughput. However, this deployment mode requires more complex cluster configuration.

MongoDB Nodes

The first try is to download the MongoDB on a single node computer with a much high configuration computer(Intel Core i9–12gen, 128GB RAM, 5TB SSD).

If this mode can not meet your requirement. You need to learn how to deploy the mongodb as a cluster.

The mongodb deployment guide: <https://github.com/mongodb/mongodb-kubernetes-operator/blob/master/README.md>

Redis Node

Since the usage of ProPro for redis is relatively small, redis can be deployed on the primary node server of mongodb on a same node.

ProPro Nodes

The computation nodes can just deploy as the single–node solution but without the redis and mongodb installation.

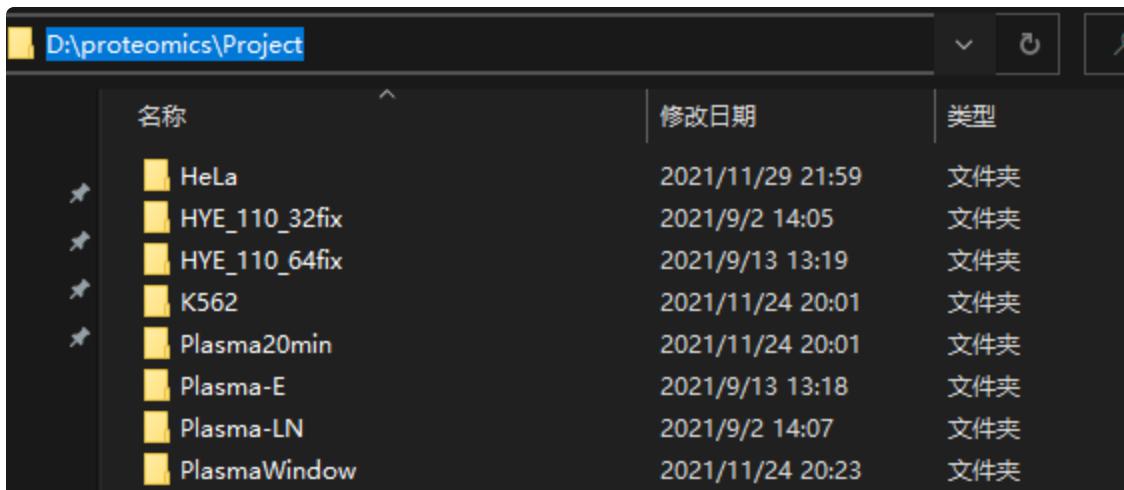
2.3 Configuration

2.3.1 ProPro

User can edit the configuration file "application.properties" under the "libs" directory

Properties	Default	Comments
spring.data.mongo.db.port	27017	default port for mongodb
spring.data.mongo.db.database	propro	default collection name for mongodb
spring.data.mongo.db.host	127.0.0.1	default IP for mongodb
repository	D:/proteomics	set your aird data repository location.Like D:/proteomics
multiple	3	depend on your computer performance. ≈RAM/10GB
spring.redis.host	127.0.0.1	default IP for redis
spring.redis.port	6379	default Port for redis
spring.redis.database	0	default db for redis
spring.redis.password		no password as default

repository: User must change the **repository** before running ProPro. Select a proper disk which contains enough disk space to store the DIA MS data file(here means Aird format data). Suppose you choosed D:/Proteomics as the repository. Then you need to add a empty directory "Project" under the repository. Every Project Folders need to be placed under the "Project" directory. ProPro will scan the "Project" folder automatically.



multiple: multiple is a variable that controls how many analysis tasks ProPro can run in parallel.

Here is an easy formula. Using the JVM memory divide 10. For example. Your single node computer has 64GB RAM, and you use 40GB for the JVM to start ProPro. Then you need to set multiple=4. If any

`java.lang.OutOfMemoryError: Java heap space`

error is display on the console. You should lower the value of this variable.

2.4 Demo

We provide a demo website for ProPro

3 How to use ProPro

3.1 Data Prepare

An targeted DIA data analysis needs at least three files

1. iRT library(tsv or TraML)
2. Standard library(tsv or TraML)
3. Raw files(Aird, vendor files need to be converted using AirdPro)

In ProPro, We have built-in two most common Internal standard libraries. The commonly used iRT library and a CiRT library

Protein:	Please enter	length:	Please enter	PeptideRef:	Please enter	m/z	RT	Cutinfo	m/z	Intensity	Charge	Annotations
Library / PeptideList_BelongLibRT_SGS												
enable	Proteins	length	PeptideRef									
Effective	AQRT_08	13	ADVTPDFSSWSC_2 @PSDAESTWIVKA		726.836	54.969	y9 y8+2 y10	1056.47 333.74 1167.54	16000 2013.3 1824	1 2 1	y9<0.01 y8>2>0.01 y10<0.01	
Effective	AQRT_07	13	DGGLADASYVAPVK_2 @SAYVEDAPAKLDR		699.336	43.2819	y7 y8 y10 y5 y3	853.432 936.466 1112.53 605.338 354.209	19000 8214.4 5916.2 5073.5 3383.4	1 1 1 1 1	y7>0.00 y8<0.01 y10<0.01 y5<0.00 y3>17>0.01	
Effective	AQRT_02	14	IGASSESPYVLDQAC_2 @TOSLASDAPEGVK		644.823	0.227424	y8 y10 y9 b6	800.449 1016.52 929.492 471.181	10000 1559.3 996.7 950	1 1 1 1	y8<0.00 y10<0.01 y9<0.00 b6>18>0.00	
Effective	AQRT_10	12	GTHIDPAAVIK_2 @TGDAPIVIAK		636.869	86.7152	y8 y2 y3 y9	854.507 741.45 598.152 967.584	6542.3 4649.2 3996.5 1527.7	1 1 1 1	y8<0.00 y7>0.00 y3>17>0.00 y9<0.01	
Effective	AQRT_09	12	GTHIDPGSOV_2 @DODGSRHTR		622.854	71.3819	y6 y7 y8 y9	598.36 713.395 826.494 939.572	10000 6151.3 4786.2 2763.2	1 1 1 1	y6<0.01 y7>0.00 y8<0.02 y9<0.01	
Effective	AQRT_01	9	LGSGNEQVTR_2 @SQVLEGTR		487.257	-28.3083	y8 y5 y2 y3	860.417 632.329 276.167 375.237	7886.9 2112.5 1904 1641	1 1 1 1	y8<0.01 y5<0.01 y2<0.00 y3<0.00	
Effective	AQRT_05	12	TPVSGSGPVY_2 @YEPGVPRSYTR		669.836	28.9999	y8 y9 y10 y2 b3 b4 a2	928.416 1041.5 1140.55 338.182 290.175 411.26 171.11	10000 6723.3 2296.2 1742.7 4085.7 1456.3 1279.4 887.5	1 1 1 1 1 1 1 1	y8<0.00 y9<0.00 y10<0.01 y2<0.00 b3>10>0.00 b4<0.00 a2<0.00	
Effective	AQRT_06	12	TPVITGAPVYR_2 @GYTPVWAITER		683.854	33.6311	y8 y9 y10 y7 b2 b4 a2	956.453 1098.54 1168.58 727.336 1199.104 411.26 411.257	10000 5695.2 5695.2 4992.7 1456.3 1279.4 2133.2 887.5	1 1 1 1 1 1 1 1	y8<0.01 y9<0.00 y10<0.02 y7<0.00 b2<0.00 b4<0.00 a2<0.00	
Effective	AQRT_03	13	VEATFOVDESNAK_2 @EDSAVEGFTVK		683.828	13.1078	y11 y10 y5 b2 y9 y7 y12	1138.54 1097.5 582.559 229.117 966.45 727.336 1168.58 411.257	5759.2 3652.2 3161 8313.6 6726.6 2133.2 1097.5 5695.2	1 1 1 1 1 1 1 1	y11<0.00 y10<0.00 y5<0.00 b2<0.00 y9<0.00 y7<0.00 y12<0.00	

Peptide List in iRT Library

Peptide List in CiRT library

The repository folder structure will be based on *Project- > [specific project name] -> [Aird File List]*

For example: Here we have two projects call HYE110-var64 and Plasma-2021. We need to create two folders with the same name under the Project Folder, and put the converted Aird mass spectrometry files into the corresponding project folder.

-Proteomics_Repository

-Project

-HYE110-var64

-project1-dia1.aird
-project1-dia1.json
-project1-dia2.aird
-project1-dia2.json

-Plasma-2021

-project2-dia1.aird
-project2-dia1.json
-project2-dia2.aird
-project2-dia2.json

3.2 Library

There are two types of library files in ProPro—The **INS(Internal Standard)** library and the **ANA(Analytes to analyze)** library.

3.2.1 Import Library

tsv,csv and TraML format are supported in ProPro. Click the library button in the left menu. And click the Add button to import a library. Once the library is successfully imported. User can see the library basic statistic information in the list page. Like peptides count, proteins count, fragments count.

The screenshot shows the CSI ProPro software interface. On the left, there is a navigation sidebar with options: Project, Protein, Library (selected), Method, Dictionary, and Task. The main area displays a 'Library' list with columns: LibraryName, Type, Algorithm, Organism, ProteinCount, PeptideCount, fragmentCount, Description, and Option. A red arrow points to the 'fragmentCount' column header. Below this is a 'Library List' search bar and a 'Type' dropdown. A modal dialog box titled 'Create Library' is open in the center. It has tabs for 'ManualUpload' (selected) and 'AutoImport'. It requires input for 'LibraryName' (with a placeholder 'Please enter') and 'Type' (with a placeholder 'Please select'). A file upload area is shown with a cloud icon and the text 'Click or drag the file to this area', supporting 'tsv', 'csv', and 'TraML' formats. A 'Detail description' text area with a placeholder 'Please enter' is also present. At the bottom of the dialog are 'Cancel' and 'OK' buttons.

Import Library Form

User needs to set the library name, library type and the library file before uploading. **tsv**, **csv** and **TraML** formats are supported now.

If your library contains less than 50000 precursors. The import process is usually completed in a few seconds. User can refresh the library list page or jump to the Task list to check the progress.

TaskTemplate	TaskStatus	SpendTime	CreatTime
UPLOAD_LIBRARY_FILE	RUNNING	Not started	2021-12-16 03:23:57
EXTRACT_PEAKPICK_SCORE-Analyze-EPPS-HYE_110_64var	SUCCESS	1265.841m	2021-12-10 06:13:39
EXTRACT_PEAKPICK_SCORE-Analyze-EPPS-HYE_110_64var	SUCCESS	1342.574m	2021-12-10 06:13:39
EXTRACT_PEAKPICK_SCORE-Analyze-EPPS-HYE_110_64var	SUCCESS	1278.044m	2021-12-10 06:13:39
EXTRACT_PEAKPICK_SCORE-Analyze-EPPS-HYE_110_64var	SUCCESS	1287.277m	2021-12-10 06:13:39
EXTRACT_PEAKPICK_SCORE-Analyze-EPPS-HYE_110_64var	SUCCESS	1281.208m	2021-12-10 06:13:39
EXTRACT_PEAKPICK_SCORE-Analyze-EPPS-HYE_110_64var	SUCCESS	1315.794m	2021-12-10 06:13:39
EXTRACT_PEAKPICK_SCORE-Analyze-EPPS-HYE_110_64var	SUCCESS	244.791m	2021-12-10 03:56:20
EXTRACT_PEAKPICK_SCORE-Analyze-EPPS-HYE_110_64var	SUCCESS	240.184m	2021-12-10 03:56:20
EXTRACT_PEAKPICK_SCORE-Analyze-EPPS-HYE_110_64var	SUCCESS	248.651m	2021-12-10 03:56:20
EXTRACT_PEAKPICK_SCORE-Analyze-EPPS-HYE_110_64var	SUCCESS	243.547m	2021-12-10 03:56:20
EXTRACT_PEAKPICK_SCORE-Analyze-EPPS-HYE_110_64var	SUCCESS	246.200m	2021-12-10 03:56:20
EXTRACT_PEAKPICK_SCORE-Analyze-EPPS-HYE_110_64var	SUCCESS	245.298m	2021-12-10 03:56:20
EXTRACT_PEAKPICK_SCORE-Analyze-EPPS-HYE_110_64var	SUCCESS	258.648m	2021-12-09 14:11:12
EXTRACT_PEAKPICK_SCORE-Analyze-EPPS-HYE_110_64var	SUCCESS	252.207m	2021-12-09 14:11:12
EXTRACT_PEAKPICK_SCORE-Analyze-EPPS-HYE_110_64var	SUCCESS	263.613m	2021-12-09 14:11:12
EXTRACT_PEAKPICK_SCORE-Analyze-EPPS-HYE_110_64var	SUCCESS	260.11m	2021-12-09 14:11:12
EXTRACT_PEAKPICK_SCORE-Analyze-EPPS-HYE_110_64var	SUCCESS	254.964m	2021-12-09 14:11:12
EXTRACT_PEAKPICK_SCORE-Analyze-EPPS-HYE_110_64var	SUCCESS	258.969m	2021-12-09 14:11:12
EXTRACT_PEAKPICK_SCORE-Analyze-EPPS-HYE_110_64var	SUCCESS	249.794m	2021-12-09 13:24:36

The upload progress in the Task

HYE_64_Var

LibraryName: [HYE_64_Var](#)

LibraryId: [61319ce33111e944c73feba6](#)

Type: [ANA](#)

Algorithm: [replace](#)

Organism: [NotSet](#)

ProteinCount: [6793](#)

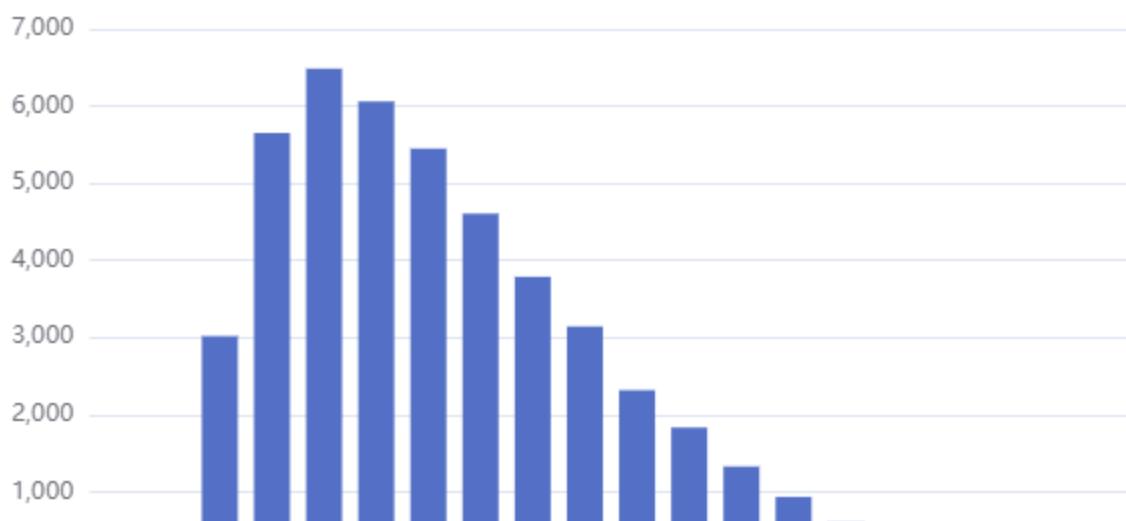
PeptideCount: [46052](#) [Check](#)

fragmentCount: [276312](#)

Description:

Peptide_Dist_On_Mz_5

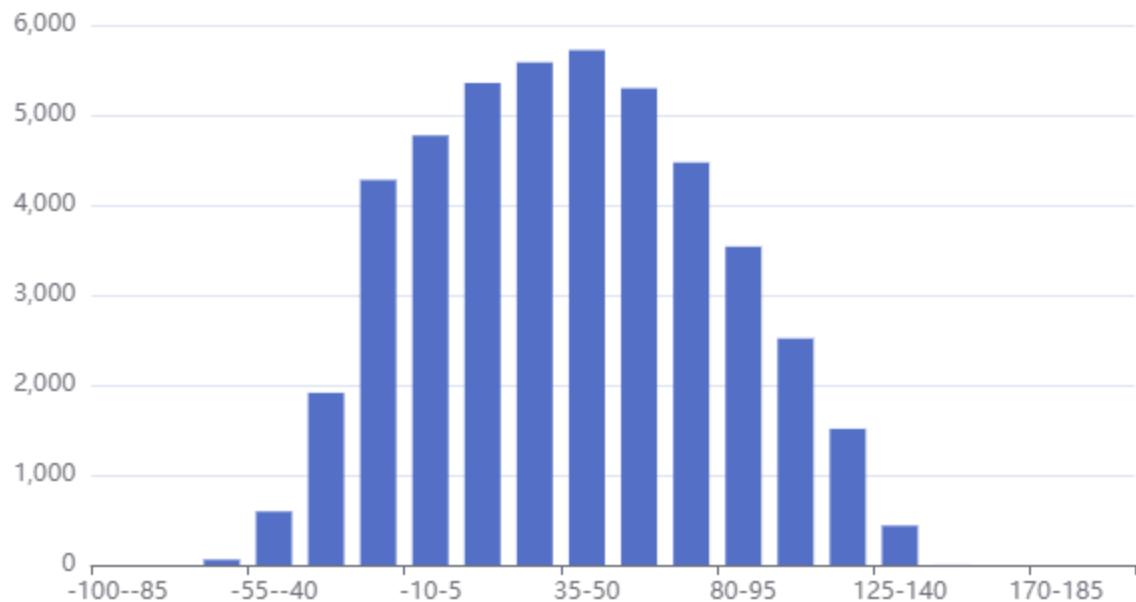
Peptide_Dist_On_Mz_5





Peptide_Dist_On_RT_5

Peptide_Dist_On_RT_5



Statistic for library

3.2.2 Generate Decoy

After imported library. User can generate decoy for the ANA type library.

LibraryName	Type	Algorithm	Organism	ProteinCount	PeptideCount	fragmentCount	Description
HYT_32_Fix	ANA	shuffle	NotSet	6740	41194 Check	271104	Shuffle
HYT_32_Var	ANA	shuffle	NotSet	6778	45933 Check	275160	Nico
HYT_64_Fix	ANA	shuffle	NotSet	6791	45970 Check	275862	Replace
HYT_64_Var	ANA	replace	NotSet	6793	46052 Check	276312	
CRT	INS	shuffle	NotSet	298	124 Check	619	
PanHuman	ANA	replace	NotSet	12046	211370 Check	2832208	
Plasma	ANA	shuffle	NotSet	404	4853 Check	67911	
DPLH	ANA	replace	NotSet	11848	342748 Check	2056470	

Generate decoy for ANA library

ProPro provides three decoy generator methods.

Shuffle: Shuffle algorithm for each peptide's amino acids for 10 times. And get the most unsimilar sequence as the decoy peptide.

Nico: Algorithm to find a most unsimilar sequence for each peptide's amino acids.

Replace: Replace the amino acids "GAVLIFMPWSCTYHKRQEND" to "LLLTVLLLTSSSSLLNDQE".

3.2.3 Peptide List

Click the check button on the *peptide count* column in library list page to jump to the peptide list page.

Peptide List

3.2.4 Peptide Detail

Select one peptide and click the detail button to check the peptide detail information.

AAAAAAAAAPAAAATPTTAATTAAATAAQ_3

ID: 6128abd49aef1e6785de36f2

LibId: 6128abd49aef1e6785de01ee

ProteinName: sp|P37108|SRP14_HUMAN

PeptideRef: AAAAAAAAAPAAAATPTTAATTAAATAAQ_3

m/z: 790.0749487

RT: 38.9

Charge: 3

peptideFullName: AAAAAAAAAAPAAAATPTTAATTAAATAAQ

Sequence: AAAAAAAAAPAAAATPTTAATTAAATAAQ

DecoySequence: AATQAAPTAAPAATATAAATAAAAAAAAT

Fragments

CutInfo	m/z	Intensity	Charge	Annotations
b5	356.1928455	7798.6	1	b5/-0.000,m2:6/-0.000
b6	427.2299592	7586.1	1	b6/-0.001,m2:7/-0.001
b7	498.267073	5788.8	1	b7/-0.001,b19^3/0.003,m2:8/-0.001
b8	569.3041868	4657.9	1	b8/-0.001,m2:9/-0.001
b9	640.3413006	2198	1	b9/0.001
b10	737.3940645	717.2	1	b10/-0.003,m2:11/-0.003

pseudolonFra

CutInfo	m/z	Intensity	Charge	Annotations
b5	443.22542403190005	7798.6	1	b5/-0.000,m2:6/-0.000
b6	514.2615380319	7586.1	1	b6/-0.001,m2:7/-0.001
b7	611.3143020319001	5788.8	1	b7/-0.001,b19^3/0.003,m2:8/-0.001
b8	712.3619810319001	4657.9	1	b8/-0.001,m2:9/-0.001
b9	783.4010950319	2198	1	b9/0.001
b10	854.4342090319	717.2	1	b10/-0.003,m2:11/-0.003

Peptide Detail

All the fragments of the selected peptide are shown in the detail panel.

3.3 Method

Method is an algorithm parameter group for DIA data analysis. User need to create a method before data analysis.

3.3.1 Method List

Method List

Click the Method button on the menu to jump to the method list page. On this page. Click the Add button to create an method.

3.3.2 Method Detail

Creat Method

basic Params	xic Params	table.IRtParams	peakFindingParams	scoreParams
MethodName name	mzWindow rtWindow extraRtWindow maxlons	pickedNumbers wantedNumber ionsLow ionsHigh smoothMethod smoothPoints baselineMethod Baseline RT Window	peakFindingMethod IONS_COUNT	scoreTypes Please select
Description description	300 200 maxlons	50	50	RegressionParams Algorithm LDA fdr 0.01

Cancel **OK**

All the important parameters have the description on their help tag. If you don't know anything about how to set the parameter. Just input the method name and leave all the other parameters as their default value.

3.4 Project

Project is the most important concept in ProPro. You'd better put the runs under a same Mass Spectrometer with same running parameters in a same project so that they can be analyzed

with same libraries and method.

3.4.1 Init Project

Press the Add button in project list page. You can see the create form.

The screenshot shows the 'Create project' dialog box. At the top left is the title 'Create project'. On the right side is a close button (X). Below the title are several input fields arranged in a grid:

- * ProjectName**: A dropdown menu with a placeholder 'Please select'. A red arrow points to this field.
- ProjectAlias**: An input field with a placeholder 'Please enter the project ali...'. It contains the text 'HeLa-MSG'.
- RunType**: A dropdown menu with a placeholder 'Please select'. It contains the text 'DIA'.
- Group**: An input field with a placeholder 'Please enter'. It contains the text 'Please enter the person in ...'.
- Tags**: An input field with a placeholder 'Please select'.
- Library**: A dropdown menu with a placeholder 'Please select'.
- InnerLibrary**: A dropdown menu with a placeholder 'Please select'.
- MethodPackage**: A dropdown menu with a placeholder 'Please select'.

Below these fields is a section labeled 'Description:' with a text input field containing the placeholder 'Please enter'.

At the bottom right of the dialog are two buttons: 'Cancel' and 'OK'.

Create Project Form

ProPro would scan the repository in the setting file. If a new folder under the project folder has not been added in the ProPro. The folder name will appear in the dropdown box. Then you need to set the INS library, ANA library and the method.

3.4.2 Scan Project

If you add some new aird files to an existed project. You need to select the existed project and click the scan button in the project page list. The ProPro will index the new files into the system.

The screenshot shows a table titled "ProjectList" with the following columns:

- Group**: Contains entries like "DIA", "LMS", "HPLC", etc.
- ProName**: Contains entries like "Hela-MHRM", "K562", "Plasma20min", "DIANN", "HYE", "HYE_110_32tx_backup", "HYE_110_32tx", "HYE_124_32tx-6000", "HYE_124_32tx-5600", "HYE_110_64var", "HYE_110_32var", "HYE_124_64var-6600", "HYE_124_64var-5600", "HYE_110_32tx", "Plasma-LN", "Plasma-E", and "SGS".
- Alias**: Contains entries like "5_Check", "4_Check", "3_Check", "7_Check", "6_Check", "42_Check", "18_Check", and "90_Check".
- RunNum**: Contains entries like "5", "4", "3", "7", "6", "6", "6", "6", "6", "6", "6", "6", "6", "6", "42", "18", and "90".
- OvNum**: Contains entries like "5", "4", "3", "7", "6", "6", "6", "6", "6", "6", "6", "6", "6", "6", "42", "18", and "90".
- Type**: Contains entries like "DIA", "LMS", "CRT", "RT_library", "Plasma", "NotSet", "ParHuman", "IonShape", and "SGS".
- Director**: Contains entries like "LMS", "HPLC", "RT_library", "Plasma", "NotSet", "ParHuman", and "IonShape".
- InnerLibrary**: Contains entries like "CRT", "RT_library", "Plasma", "NotSet", "ParHuman", and "IonShape".
- Library**: Contains entries like "ParHuman", "IonShape", "NotSet", "Plasma", "NotSet", and "SGS".
- MethodPackage**: Contains entries like "IonShape-Igboot", "NotSet", and "IonShape".
- Tags**: Contains entries like "6600", "5600", "5600", "5600", "5600", "5600", "5600", "5600", "5600", "5600", "5600", "5600", "5600", "5600", "5600", "5600", "5600", "5600", and "6600".
- CreateTime**: Contains dates and times like "2021-11-30 09:06:19", "2021-11-29 00:57:28", "2021-11-25 01:29:57", "2021-11-25 01:30:36", "2021-10-13 09:25:17", "2021-08-23 08:49:28", "2021-08-23 08:49:28", "2021-09-14 08:42:24", "2021-09-14 08:40:40", "2021-09-13 14:17:38", "2021-09-13 14:17:51", "2021-09-14 08:43:49", "2021-09-14 08:44:54", "2021-08-23 07:12:22", "2021-08-23 09:09:48", "2021-08-23 09:10:36", and "2021-09-08 03:55:35".

Project list page

3.4.3 Update Project

Project configuration below can be modified. Project name is not permitted to change.

The dialog box contains the following fields:

- * LibName**: HYE_110_64var
- ProjectAlias**: Please enter
- RunType**: DIA
- Director**: LMS
- Group**: HYE
- Tags**: 6600
- Library**: HYE_64_Var
- InnerLibrary**: iRT_library
- MethodPackage**: IonsCount
- Description**: Human:Yeast:Ecoli
A:67:30:3
B:67:3:30
- Buttons**: Cancel (grayed out) and OK

3.4.4 Runs & Overviews

A run corresponds to an Aird file. One project can contain multiple runs. The overview is the analysis result for each run. If we analyze for one run twice. ProPro will build two overviews.

Group	ProName	Alias	RunNum	OvNum	Type	Director	InnerLibrary
	HeLa-MHRM		5	Check	5	CIR	
	K562		4	Check	0	CIR	
DIANN	Plasma20min		3	Check	3	iRT_library	
DIANN	PlasmaWindow		7	Check	0	CIR	
HYE	HYE_110_32fix_backup		6	Check	0	iRT_library	
HYE	HYE_110_64fix		6	Check	0	iRT_library	
HYE	HYE_124_32fix-6600		6	Check	6	iRT_library	
HYE	HYE_124_32fix-5600		6	Check	0	iRT_library	
<input checked="" type="checkbox"/> HYE	HYE_110_64var		6	Check	18	Check	iRT_library
HYE	HYE_110_32var		6	Check	0	iRT_library	
HYE	HYE_124_64var-6600		6	Check	6	Check	iRT_library
HYE	HYE_124_64var-5600		6	Check	0	iRT_library	
HYE	HYE_110_32fix		6	Check	0	iRT_library	
Plasma	Plasma-LN		42	Check	0	iRT_library	
Plasma	Plasma-E		18	Check	0	CIR	
SGS	SGS		90	Check	90	Check	iRT_SGS

Runs and Overviews

On the project list page. Each project row has a RunNum column and an OvNum(Overview Num) column. Click the Check button can jump to the Run/Overview list page.

3.4.5 Run List

Project	Group	RunName	RunAlias	Tags	Aird: Vendor(MB)	SWATHWindow	IRTVerRes	FragmentMode	Manufacturer	FragmentMode	Option	Interfactor	Starting	Batchsize	NewRIT	ViewRIT	Reset	Query
Project	Group	RunName	RunAlias	Tags	Aird: Vendor(MB)	SWATHWindow	IRTVerRes	FragmentMode	Manufacturer	FragmentMode	Option	Interfactor	Starting	Batchsize	NewRIT	ViewRIT	C	X
Project	A	HYE110_TTOF6600_64var_lgillet_I160305_003	A-2	DIA	2993 [330] 9.4%	64 Check	y=0.026x-63.885	CID	SCIEK	TripleTOF 6600								
Project	A	HYE110_TTOF6600_64var_lgillet_I160305_001	A-1	DIA	2993 [330] 9.4%	64 Check	y=0.026x-63.454	CID	SCIEK	TripleTOF 6600								
Project	B	HYE110_TTOF6600_64var_lgillet_I160305_002	B-1	DIA	3096 [339] 9.8%	64 Check	y=0.026x-63.956	CID	SCIEK	TripleTOF 6600								
Project	A	HYE110_TTOF6600_64var_lgillet_I160305_005	A-3	DIA	2921 [327] 9.5%	64 Check	y=0.026x-63.626	CID	SCIEK	TripleTOF 6600								
Project	B	HYE110_TTOF6600_64var_lgillet_I160305_004	B-2	DIA	3083 [342] 9.9%	64 Check	y=0.026x-64.377	CID	SCIEK	TripleTOF 6600								
Project	B	HYE110_TTOF6600_64var_lgillet_I160305_006	B-3	DIA	3043 [339] 9.9%	64 Check	y=0.026x-63.576	CID	SCIEK	TripleTOF 6600								

Run List Page

After scanning the project. ProPro will load the aird file's metadata(in the json file) into the database automatically. So that users can have a visualization page for each run.

3.4.6 Run Detail

HYE110_TTOF6600_64var_lgillet_I160305_003

Group: A

RunName: HYE110_TTOF6600_64var_lgillet_I160305_003

RunAlias: A-2

Tags:

RunId: 61935bc3293468189b74cabc

Type: DIA

Aird : Vendor(MB): 2993 3305 9.4%

SWATHWindow: 64 Check

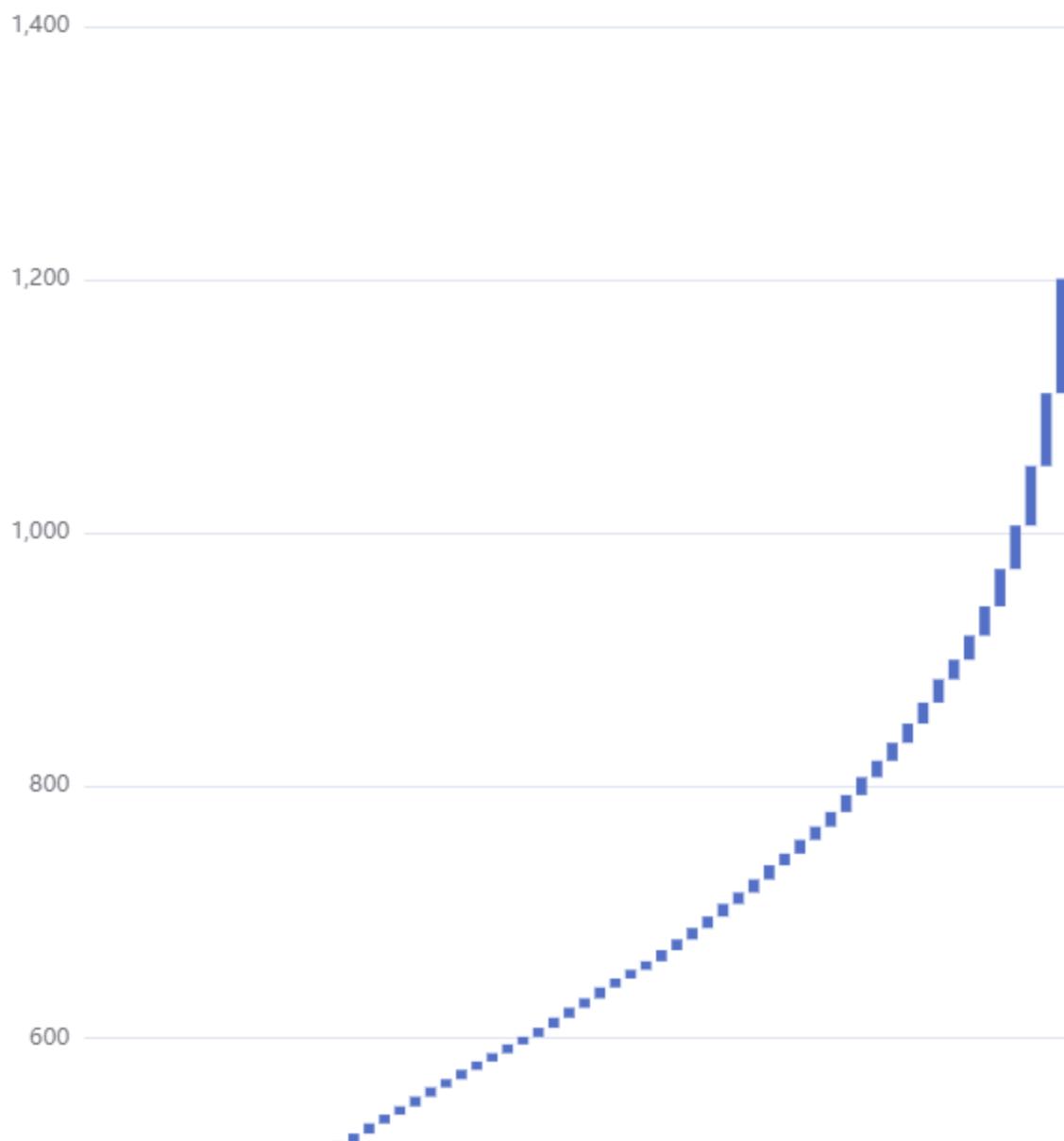
IRTVerRes: $y=0.026x-63.885$

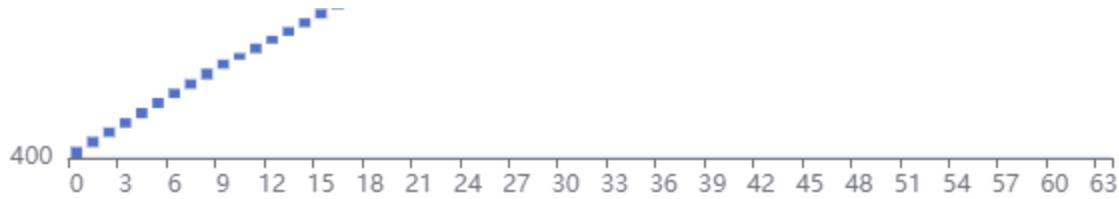
FragmentMode: CID

Manufacturer: SCIEX

FragmentMode: TripleTOF 6600

Swath Chart

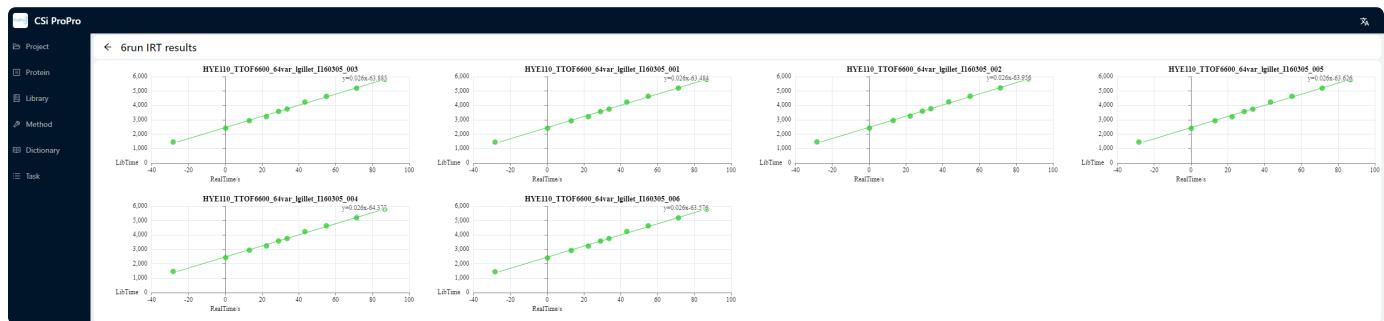




Run Detail

ProPro can show the DIA windows for each Run. If the runs have already do the iRT analysis.

User can select multiple runs and click the "view irt" button to see the irt linear regression charts.



iRT linear regression chart

3.4.7 Overview List

RunName	OnlyDef	reSelect	MinTotal	peakCount	PeptideNum	PeptideUni	PeptideAll	ProteinUni	ProteinAll	Tags	Remark	CreatTime	Option
HYE110_TTOF6600_64var_Iglier_I160305_006	No	No	-1.325	484962	92104	30679	33010	40337	5815	NewParameters	2021-12-10 08:02:05		
HYE110_TTOF6600_64var_Iglier_I160305_004	No	No	-1.521	491903	92104	30625	33171	40552	5840	NewParameters	2021-12-10 07:18:23		
HYE110_TTOF6600_64var_Iglier_I160305_003	No	No	-1.380	479519	92104	31996	34584	5088	6087	NewParameters	2021-12-10 06:56:56		
HYE110_TTOF6600_64var_Iglier_I160305_002	No	No	-1.549	489416	92104	30905	33229	40555	5834	NewParameters	2021-12-10 06:53:35		
HYE110_TTOF6600_64var_Iglier_I160305_001	No	No	-1.575	480294	92104	32308	34882	5127	6128	NewParameters	2021-12-10 06:13:39		
HYE110_TTOF6600_64var_Iglier_I160305_006	Yes	No	0.358	484962	92104	29057	31296	40227	5806	调整参数	2021-12-10 06:16:44		
HYE110_TTOF6600_64var_Iglier_I160305_005	Yes	No	0.362	478018	92104	30165	32663	40232	5904	调整参数	2021-12-10 04:12:44		
HYE110_TTOF6600_64var_Iglier_I160305_004	Yes	No	0.353	491903	92104	29065	31321	40668	5842	调整参数	2021-12-10 04:08:35		
HYE110_TTOF6600_64var_Iglier_I160305_003	Yes	No	0.347	479519	92104	30451	32990	40934	5925	调整参数	2021-12-10 04:04:32		
HYE110_TTOF6600_64var_Iglier_I160305_002	Yes	No	0.350	489416	92104	29387	31645	40444	5827	调整参数	2021-12-10 04:00:26		
HYE110_TTOF6600_64var_Iglier_I160305_001	Yes	No	0.337	480294	92104	30793	33322	40895	5976	调整参数	2021-12-10 03:56:20		
HYE110_TTOF6600_64var_Iglier_I160305_005	No	No	-1.541	478018	92104	32100	34690	5114	6119	NewParameters	2021-12-10 03:59:42		
HYE110_TTOF6600_64var_Iglier_I160305_006	No	No	0.453	484962	92104	28981	31215	40113	5589	xgboost-ionshape	2021-12-09 14:32:41		
HYE110_TTOF6600_64var_Iglier_I160305_005	No	No	0.453	478018	92104	30250	32771	40204	5910	xgboost-ionshape	2021-12-09 14:28:29		
HYE110_TTOF6600_64var_Iglier_I160305_004	No	No	0.459	491903	92104	29003	31454	40440	5620	xgboost-ionshape	2021-12-09 14:24:00		
HYE110_TTOF6600_64var_Iglier_I160305_003	No	No	0.459	479519	92104	30997	32958	40951	5937	xgboost-ionshape	2021-12-09 14:19:51		
HYE110_TTOF6600_64var_Iglier_I160305_002	No	No	0.444	489416	92104	29073	31537	40339	5619	xgboost-ionshape	2021-12-09 14:15:31		
HYE110_TTOF6600_64var_Iglier_I160305_001	No	No	0.459	480294	92104	30995	33118	40555	5946	xgboost-ionshape	2021-12-09 14:11:12		

Overview List Page

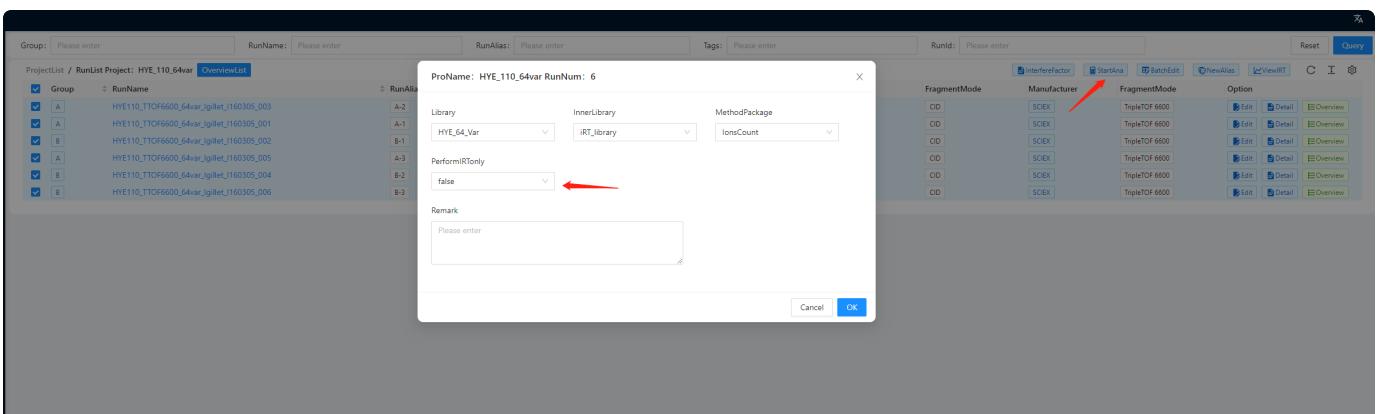
After do the DIA analysis. Each run will get an overview result. User can click the detail button to check all the identification results for each peptides.

Protein	Decoy	PeptideRef	FDR	Status	RealRt/LibrRt/ΔT	Sum
HYE110_TTOF6600_64var_Igillet_JI60305_006-iRT_Library-HYE_64_Var-20211210160205						
sp C82975 C82975_HUMAN	True	ELANIEPKK_2	0.0000	Success	3459 NaN	175339
sp P04817 CHCLU_ECOLI	True	MOMQNLHRR_2	0.0000	Success	2718 NaN	17275
sp P04817 CHCLU_HUMAN	True	LDPYFLSLVHNHC_3	0.0000	Success	5490 NaN	29162
sp P04817 CHCLU_ECOLI	True	LUEEEAAK_2	0.0000	Success	3076 NaN	87274
sp C22240 B2224_YEAS8	True	KALEEVVK_2	0.0000	Success	2341 NaN	8533
sp P21796 D0471_HUMAN	True	LTDGFSFPNTQK_3	0.0000	Success	3480 NaN	40072
sp P03949 RPUTA_ECOLI	True	LASSLALLNSLSQK_3	0.0000	Success	5433 NaN	4829
sp P39883 ACO02_ECOLI	True	DVAEDRSRFLSQQK_3	0.0000	Success	3046 NaN	33401
sp P43886 PR56_HUMAN	True	LIFSTITSK_3	0.0000	Success	4170 NaN	76596
sp P09999 KPF1_ECOLI	True	AACFVNNSK_3	0.0000	Success	1473 NaN	49490
sp P02133 ANXAA5_HUMAN	True	AMEGAGDERK_2	0.0000	Success	1074 NaN	6544
sp P04849 LEFTS_ECOLI	True	AGNVAADGIVK_2	0.0000	Success	2263 NaN	272389
sp P23847 CP98_ECOLI	True	GALTYWAK_2	0.0000	Success	2434 NaN	17733
sp P22087 FBRL_HUMAN	True	VSSEDGDKKRYK_3	0.0000	Success	3096 NaN	53637
sp P31467 DNUF7_YEAS8	True	LDSPKLYHSSGDKH_3	0.0007	Success	3611 NaN	4021
sp P04849 RPUTA_ECOLI	True	AGPLAVYVIVMGR_3	0.0000	Success	3580 NaN	57873
sp P36683 ACO02_ECOLI	True	VTVTLODK_2	0.0000	Success	3404 NaN	66312
sp Q9Q0Q9 MEB1A_HUMAN	False	DVAEDRSRFLSQQK_3	0.0059	Success	2781 NaN	4818
sp P28737 G881_HUMAN	True	EVVIGQQLK_2	0.0000	Success	3209 NaN	26482
sp P028737 G881_HUMAN	True	ELAUGHTDQSCUnMod4CUnMod4R_3	0.0000	Success	2496 NaN	30400

Peptides Identification Result

3.4.8 Analysis Task

This is the most important step for analyzing. On the run list page. Selecting the runs that user need to analyze. Then click the StatAna button to preparing. User can select the INS and ANA libraries, Method for the comming analysis. User can set the OnlyIrt option true if user just want to run the iRT analysis not the whole analysis. If OnlyIrt is set to false. ProPro will check if the run had already got a iRT result. If it exists. ProPro would reuse the iRT result for the next analysis. If users want a new iRT analysis. Users need to remove the irt result first.



3.5 Protein Clinic

Protein Clinic is a novel function for DIA analysis overview. Users can comprehensively review every detail of the DIA analysis process.



Protein Clinic Panel

Left-Top part is the protein list for ANA library used in the Project. Once select a protein. The peptides belongs to the protein will list in the Left-Bottom part as peptide list. The main area is the EIC for each run. On the top of the main area is the fragments information of the selected peptide.

3.5.1 Protein&Peptide List

Protein 

sp P0AC86 PHSG_ECOLI	Search for peptides		
tr C8ZBF9 C8ZBF9_YEAS8	<input type="button" value="Search"/>	<input type="button" value="Reset"/>	
sp P0A705 IF2_ECOLI			
sp P16403 H12_HUMAN			
sp P33368 YOHF_ECOLI			
sp O43776 SYNC_HUMAN			
tr C8Z9W7 C8Z9W7_YEAS8			
sp Q8TF65 GIPC2_HUMAN			
sp Q08722 CD47_HUMAN			
sp Q96CP2 FWCH2_HUMAN			
sp P08590 MYL3_HUMAN			
tr C8Z7W8 C8Z7W8_YEAS8			
sp P25522 MNME_ECOLI			

< 2 3 4 5 ... 523 > Go to Page

Uni	Peptide	m/z	rt
T	KPAAATVTK_2	443.771	-32.300
F	KAASGEAKPK_2	493.785	-53.300
F	SGVSLAALKK_2	487.306	7.100
F	ALAAAAGYDVEKNNSR_3	526.934	-2.400
F	SGVSLAALK_2	423.258	25.400
F	KALAAAAGYDVEK_3	412.559	-3.400
T	KAAGGGATPK_2	400.735	-51.800
F	ASGPPVSELITK_2	599.838	34.200
F	ALAAAAGYDVEK_2	554.288	7.400

< 2 >

Protein & Peptide List Panel

You can go through all the proteins identification result one by one. Search function is provided when thousands of proteins are identified.

Here we also provide the Keyboard shortcuts:

ArrowUp and ArrowDown for Peptide Switch;

Shift+ArrowUp and Shift+ArrowDown for Protein Switch;

3.5.2 EIC

EIC is the core function of the Protein Clinic.



EIC tag

3.5.2.1 Base Information



EIC detail

The first line on the top show the protein and peptide information. The fragments with intensity of the selected peptide.

3.5.2.2 Quick Tools

The second line provides several functions.

RT Align: Rerange the x-Axis of each eic-charts to the same RT range so that users can views the technical repeatability runs more friendly

Smooth: Using the gauss smooth for all the EIC data.



3.5.2.3 Runs name and Fragments

The third line list the runs in current view. ProPro will show the run's alias firstly.

The fourth line displays the fragments with mz. Here comes 5 specific tags

H: High Ions Count

L: Low Ions Count

HS: High Smoothed Ions Count

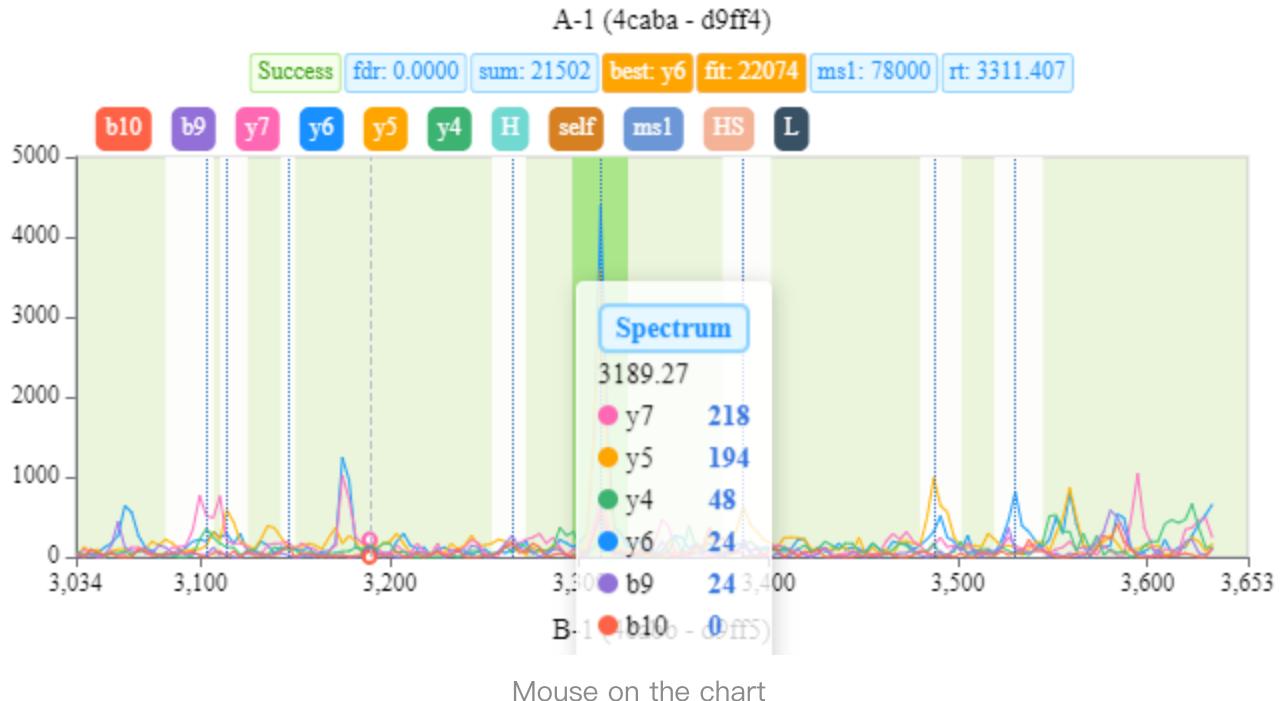
self: precursor not fragment in the MS2

ms1: precursor in the MS1

These five tags are not selected as default. User also can select or unselect all the fragment to display or hidden the fragment.

3.5.2.4 Marked Peak Group

If ProPro found some peak groups in the EIC. The peak groups will be marked as the white color background. The dotted line in the peak group range stands for the Apex RT.



3.5.2.5 Spectrum Viewer

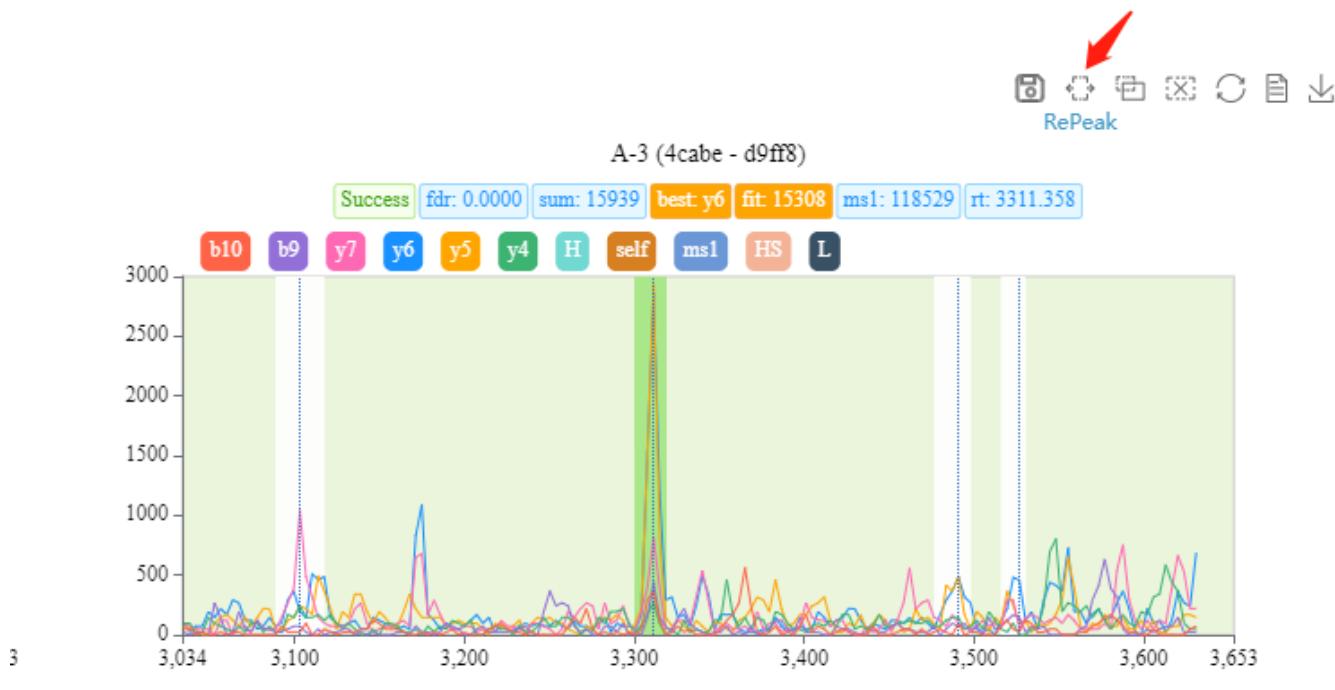
If user move the mouse on the EIC chart. User can see the intensity of each fragments in the selected RT. User can click the Spectrum button to see the Spectrum data.



Spectrum Data Chart

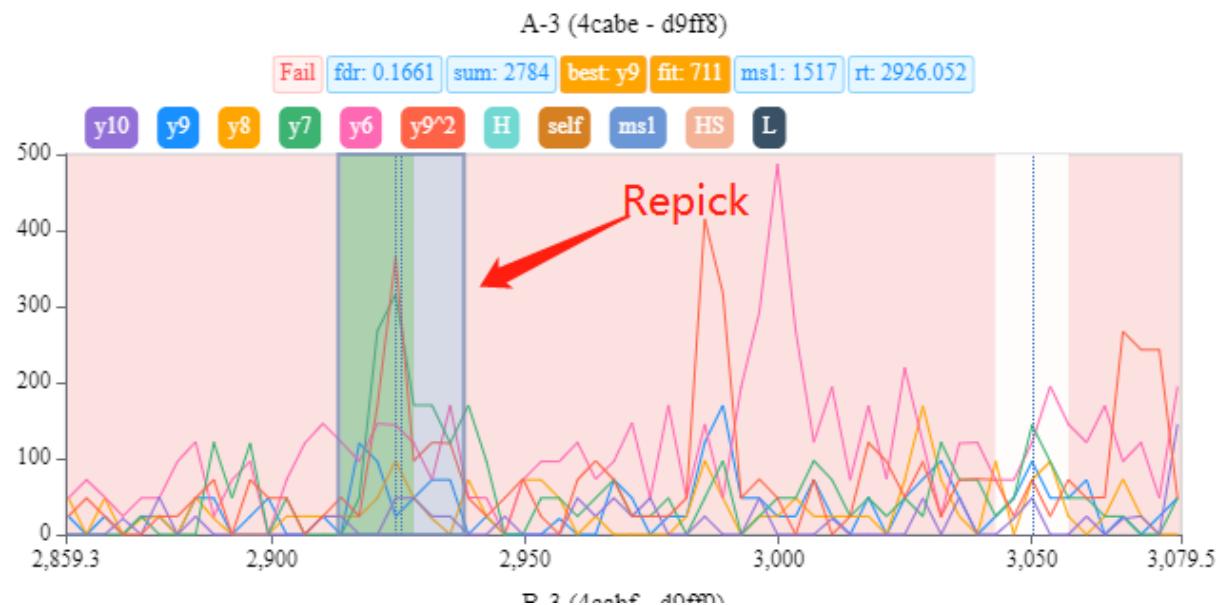
The red part is the theoretically superimposed value range for target fragments.

3.5.2.6 Manual Peak Picker



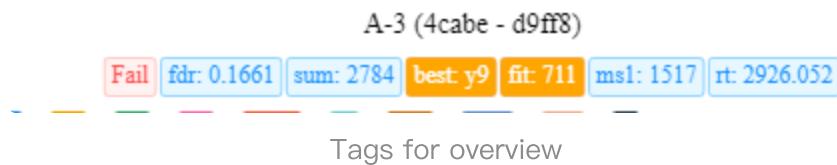
Repick for Peak Group 1

Click the PeakPicker button on the right-top toolbar. Then user can select the peak group manually.



Select a new peak range and click the save button to commit your manual pick result.

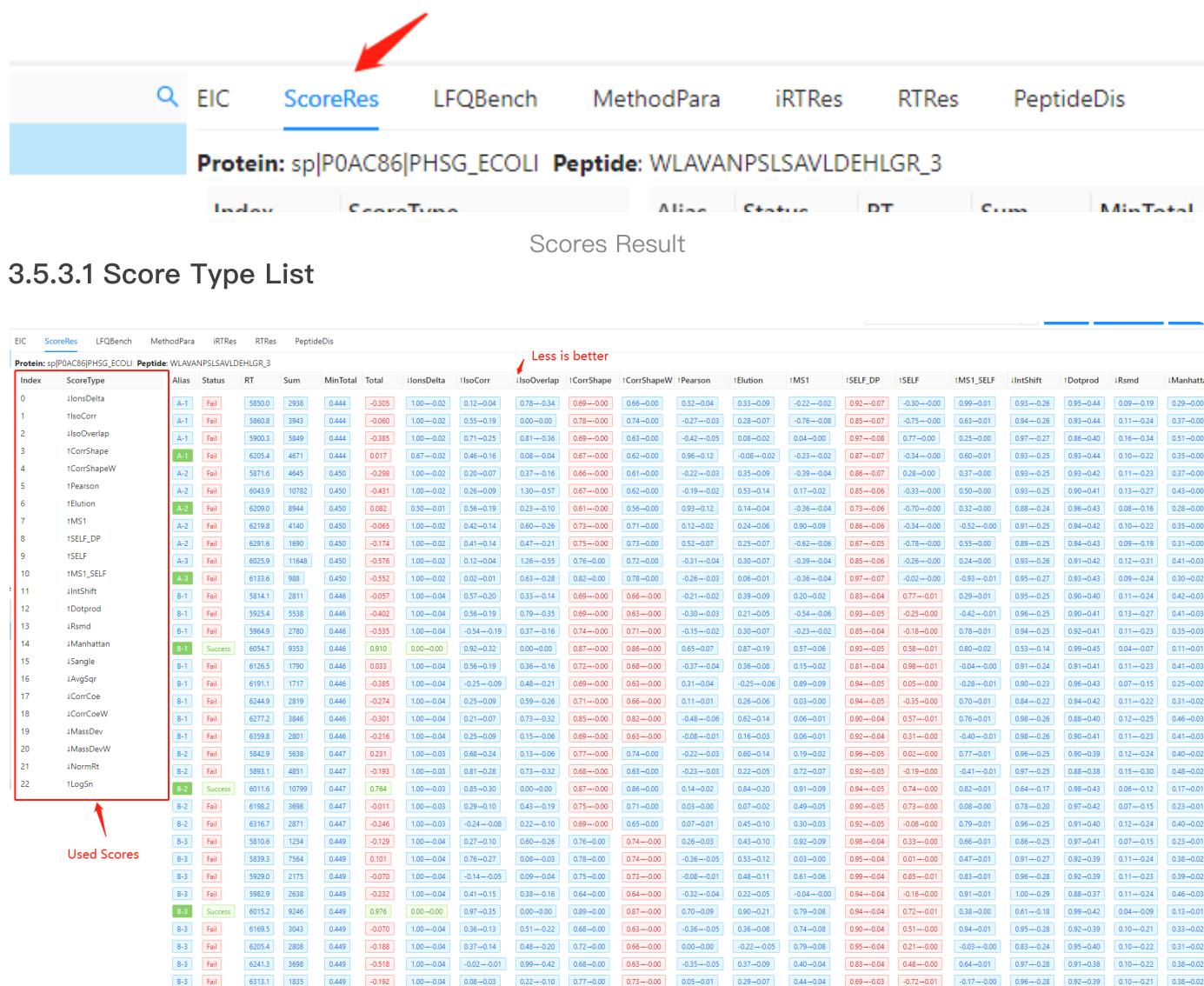
3.5.2.7 Tags



Tag	Comments
Status(First Tag)	Identification result. All the status are shown below: <i>WAIT, SUCCESS, FAILED, NO_ENOUGH_FRAGMENTS, NO_PEAK_GROUP_FIND, NO_EIC_FIND</i>
FDR	the final fdr for the peptide
sum	quantification by MS2 intensity sum
best	the best fragment marked by algorithm
fit	quantification fit by best fragment
ms1	quantification by MS1 intensity sum
rt	the apex retention time

3.5.3 Scores Result

Scores result shows all the scores for each peak group in each overview.



Score List

Left panel is the score types used in current overviews.

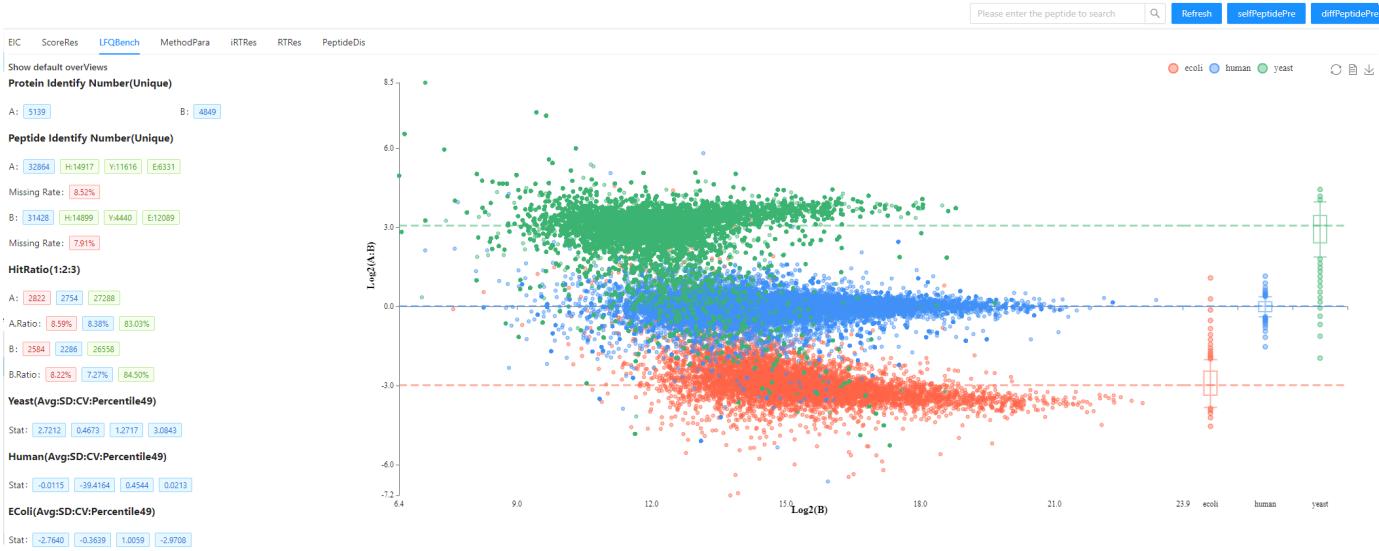
In the table. Each score has an arrow to indicate the score distribution. ↓ means less is better.

↑ means bigger is better.

3.5.4 LFQBench

LFQBench tag is a special tag. Which is only display when the project is HYE standard project.

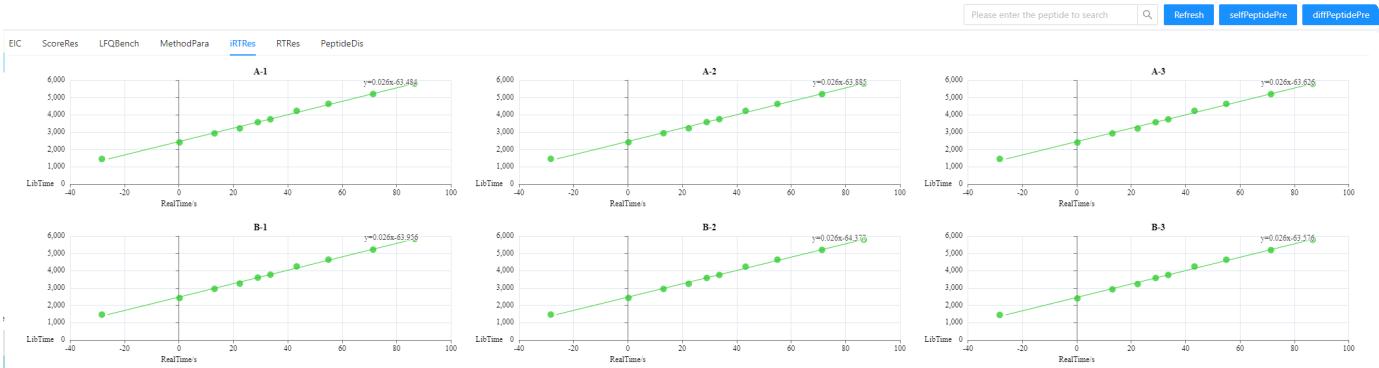
This function is used for some DIA analysis algorithm research.



HYE dataset is widely used to evaluate the quantitative and qualitative ability of DIA software. Different from the LFQBench code written by R language. We use a much friendly visualization to show the results. Each point in these charts can be clicked and will directly jump to the selected peptide's EIC chart.

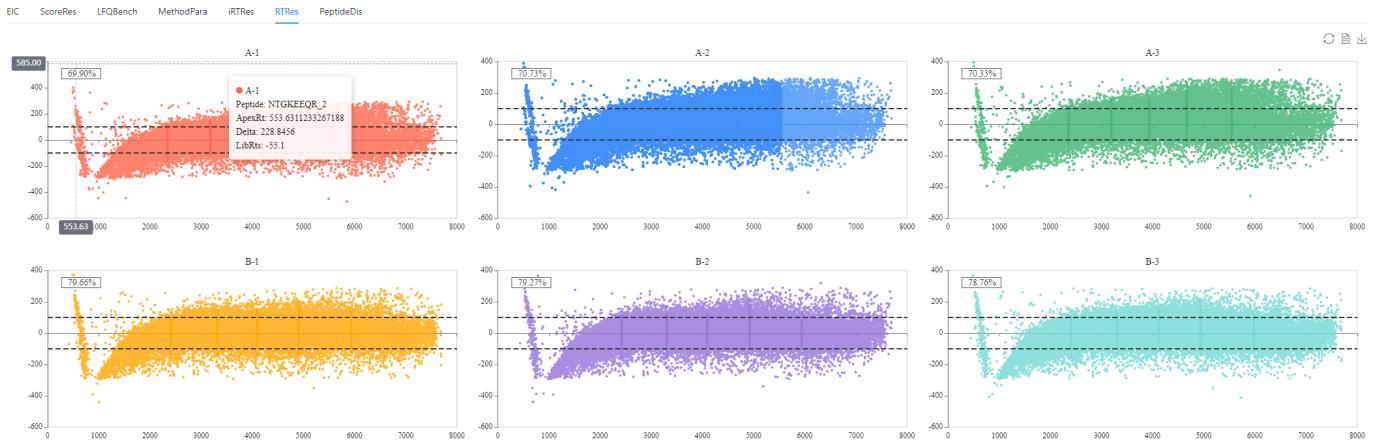
On the left panel is the statistic result for the dataset.

3.5.5 iRT Results



This function here is just as the same as that in the run list page.

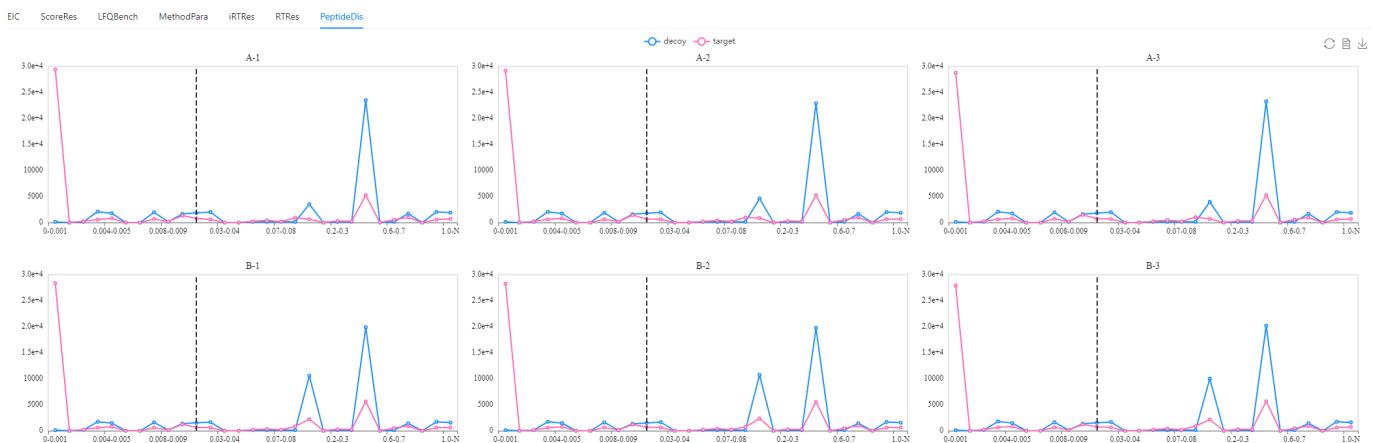
3.5.6 RT Distribution



RT distribution

This function is used to check if the real RT and predict RT is similar.

3.5.7 Target Decoy Distribution



Target Decoy Distribution for each overview

When using FDR. We need to pay attention to the distribution for target and decoy visually.

This function can show user the detail distribution for target and decoy.