

OpenSpec Manual

Version.2025

Demo Data

The data was downloaded from previous publications¹. To facilitate the use and testing of the tool, the relevant files have been deposited to the ProteomeXchange Consortium (<http://proteomecentral.proteomexchange.org>) via the iProX^{2, 3} partner repository with the dataset identifier **PXD063366**. The link to the dataset in iProx: <https://www.iprox.cn/page/project.html?id=IPX0011827000>

- [1] Guzman UH, et al. (2024) Ultra-fast label-free quantification and comprehensive proteome coverage with narrow-window data-independent acquisition. Nat Biotechnol. 2024;42(12):1855-66.
- [2] Ma J, et al. (2019) iProX: an integrated proteome resource. Nucleic Acids Res. 47(D1), D1211-D1217.
- [3] Chen T, et al. (2021) iProX in 2021: connecting proteomics data sharing with big data. Nucleic Acids Res. 50(D1): D1522-D1527.

CONTENTS

- [Software Download](#)
- [OpenSpec Manual for Pseudo-MS/MS Spectra Generation](#)
- [OpenSpec Manual for Hybrid Spectral Library Construction](#)

Software Download

1 pFind Download

Downloads

pFind 3.1.6 for windows (64bit)

- released on Jan. 02, 2020 -

Please click [here](#) to download the user manual

For license application, please visit [i.pfind.org](#).

For technical support, please visit [forum.pfind.org](#).

For other issues, please contact [support@pfind.org](#).


Notice:

* The expiration date is set on **Jan. 10, 2023**.

**** 尊敬的用户您好，pFind 3.1.6此次更新并不影响此前的license使用。您在下载安装之后，可以直接导入此前申请的license文件，而不用再次申请。谢谢！**

**** License since version 3.1.3 is still valid for this version. After downloading and installing this version, you can just import the license applied previously instead of applying a new license. Thank you.**

* [.NET framework 4.5](#) or a higher version is required.

名称	修改日期	类型	大小
 pFind3.1_Setup_20200102.exe	2022/5/6 13:42	应用程序	33,222 KB

Login

<http://pfind.net/software/pFind/index.html#Downloads>

and download pFind3 at the bottom of the page.

2 DIA-NN Download

master 1 Branch 38 Tags


Go to file

Code

vdemichev Update README.md ✓ 70757b5 · 3 weeks ago 266 Commits

GUI	Add files via upload	last month
MiniDNN	Add files via upload	4 years ago
Presentations	Add files via upload	6 months ago
Setup	major update	5 years ago
Third-party	Add files via upload	last month
WiffReader	major update	5 years ago
cranium	update	6 years ago
diann	major update	5 years ago
eigen	update	5 years ago
mstoolkit	Delete Makefile	4 years ago
src	Delete diann.cpp	4 years ago
.gitignore	update	5 years ago
LICENSE.md	Update and rename LICENSE.txt to LICENSE.md	3 months ago
README.md	Update README.md	3 weeks ago
_config.yml	Set theme jekyll-theme-cayman	7 years ago
diann.sln	deep learning for spectra and RTs	6 years ago

README License



DIA-NN

About

DIA-NN - a universal automated software suite for DIA proteomics data analysis.

Readme

View license

Activity

337 stars

23 watching

62 forks

Report repository

Releases 38

DIA-NN 2.0 Latest on Jan 29

+ 37 releases

Packages

No packages published

Contributors 2

vdemichev Vadim Demichev

ewail Hao.Chen

Languages

C++ 64.1%

C 35.6%

C# 0.3%

Shell 0.0%

CMake 0.0%

Makefile 0.0%

Login <https://github.com/vdemichev/DiaNN>
and download **the latest version** at the top of
the page.

3 PepPre Download

Release

PepPre 1.3.0

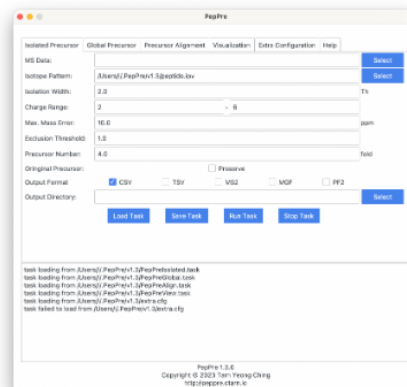
2023-12-29

[Linux](#) | [macOS \(pkg\)](#) | [macOS \(zip\)](#) | [Windows](#)

🦋 Our paper “[PepPre: Promote Peptide Identification Using Accurate and Comprehensive Precursors](#)” has been accepted by *Journal of Proteome Research*!

Note:

1. add experimental features for global peptide precursor analysis.
2. support .mes input, which is a binary format and is much faster than text-based formats.
3. able to split output into batches to support large data from instrument such as Astral.
4. support .pf2 output, which is commonly used by software from pFind Team.
5. improved UI.



Login <https://peppre.ctarn.io/>

and download **PepPre** at the top of the page.

PepPre 1.2.0

2023-5-7

[Linux](#) | [macOS \(pkg\)](#) | [macOS \(zip\)](#) | [Windows](#)

Note:

1. improved UI.

PepPre 1.1.1

2023-4-20

[Linux](#) | [macOS \(pkg\)](#) | [macOS \(zip\)](#) | [Windows](#)

Note:

1. change path for saving configuration.

PepPre 1.1.0

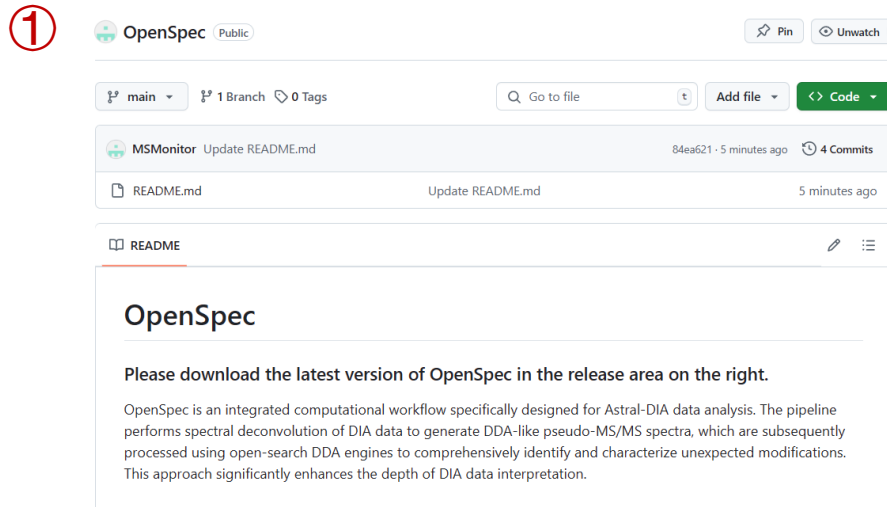
2023-4-19

[Linux](#) | [macOS](#) | [Windows](#)

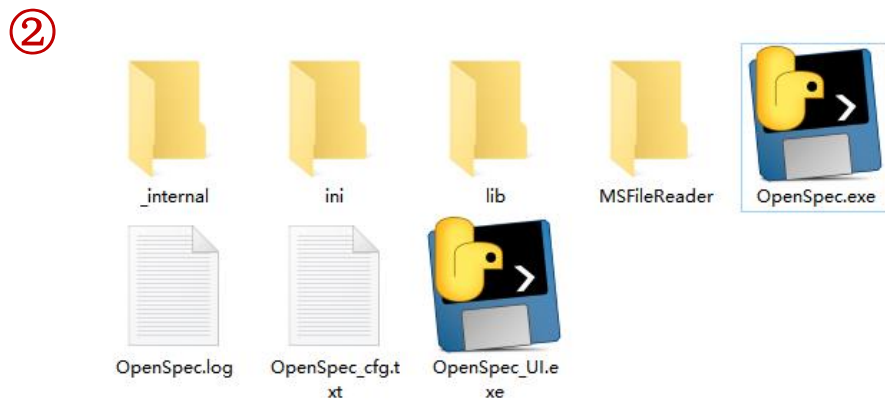
Note:

1. embed 'PepPepView' which can display deisotoped precursor ions and corresponding identification.
2. adjust arguments of CLI interface.

3 OpenSpec Download

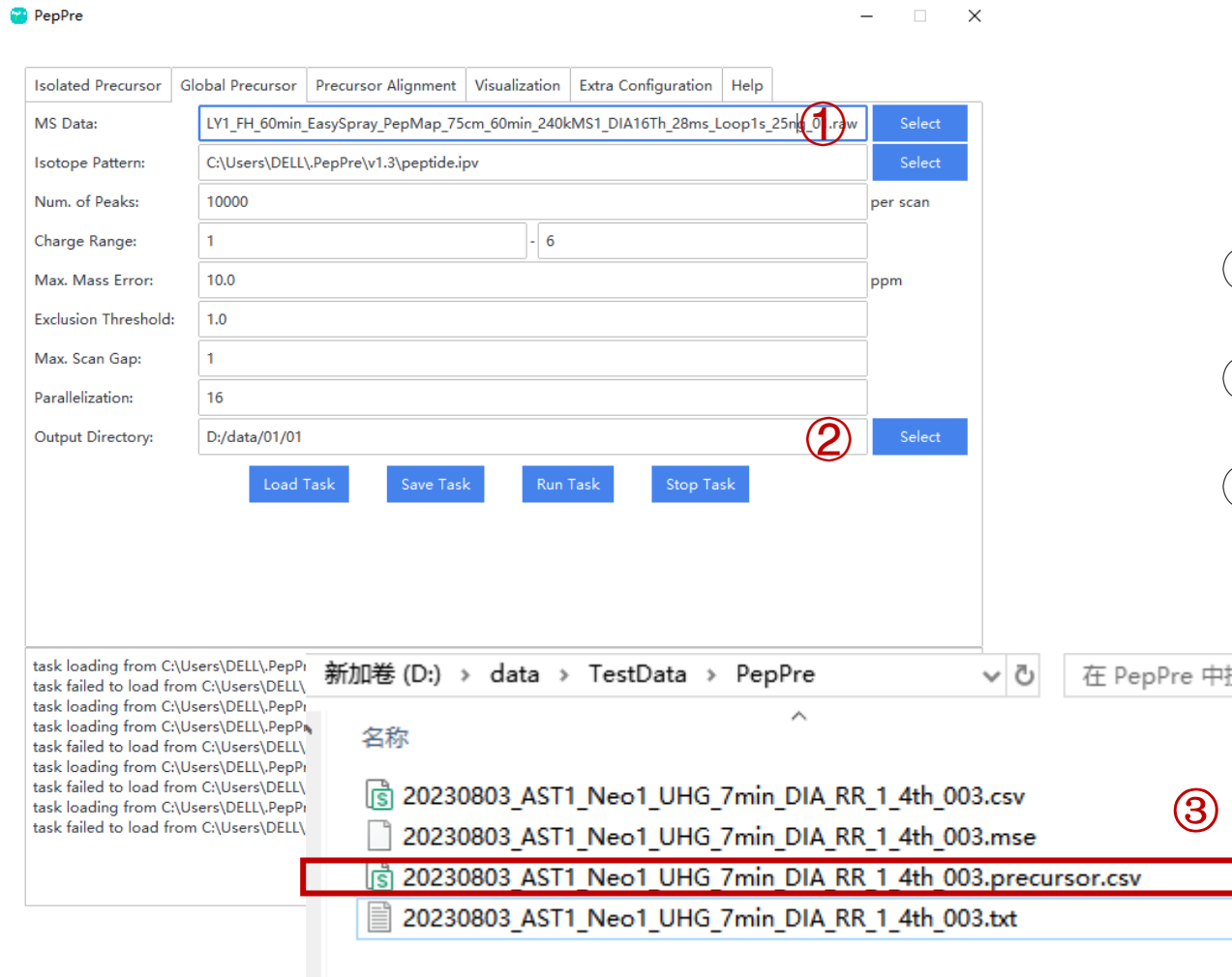


- ① Login <https://github.com/BUAA-LiuLab/OpenSpec.git> and download **the latest version** at the top of the page.
- ② Double click **OpenSpec_UI.exe to analyze.**



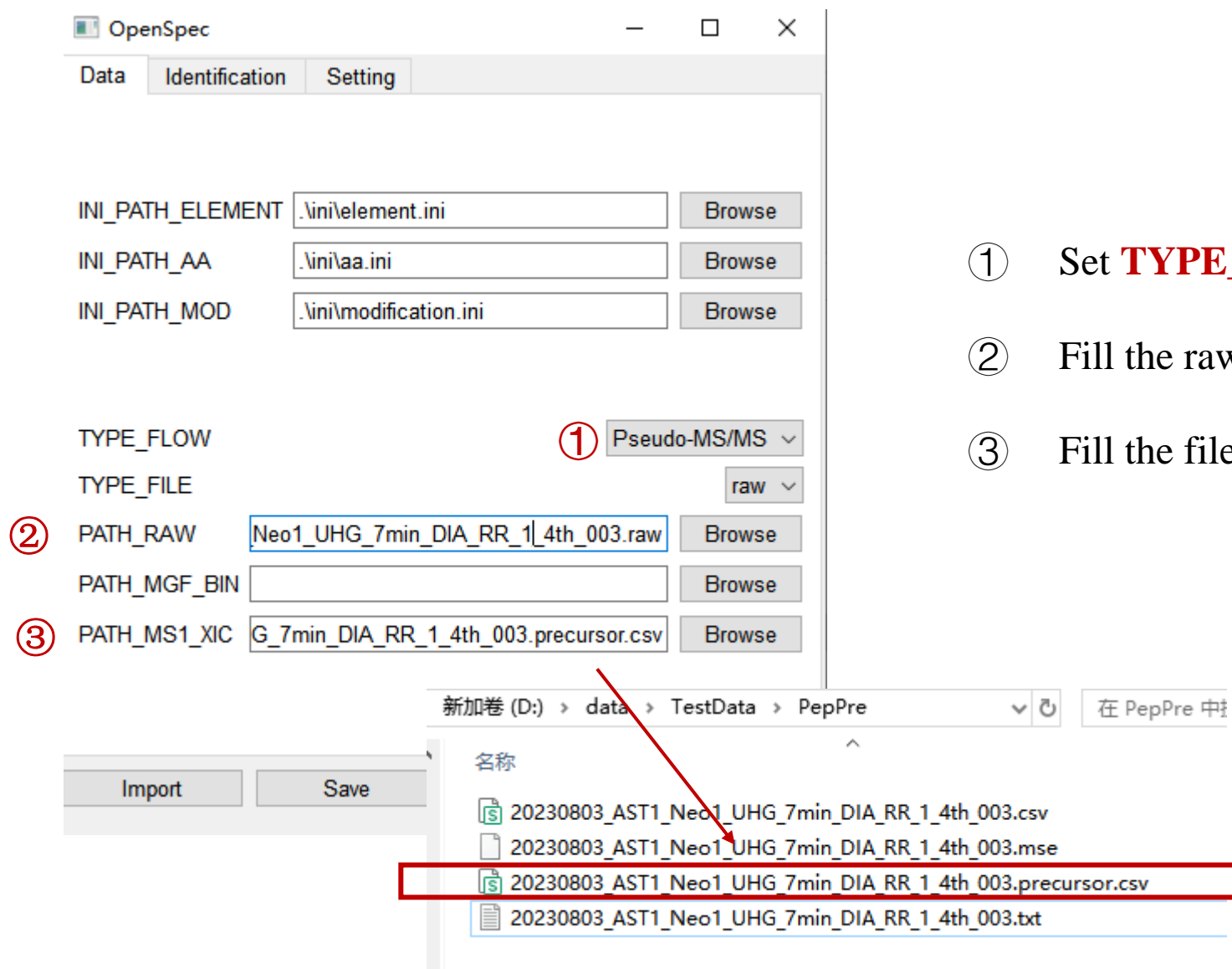
OpenSpec Manual for Pseudo-MS/MS Spectra Generation

1 Extract MS Features with PepPre



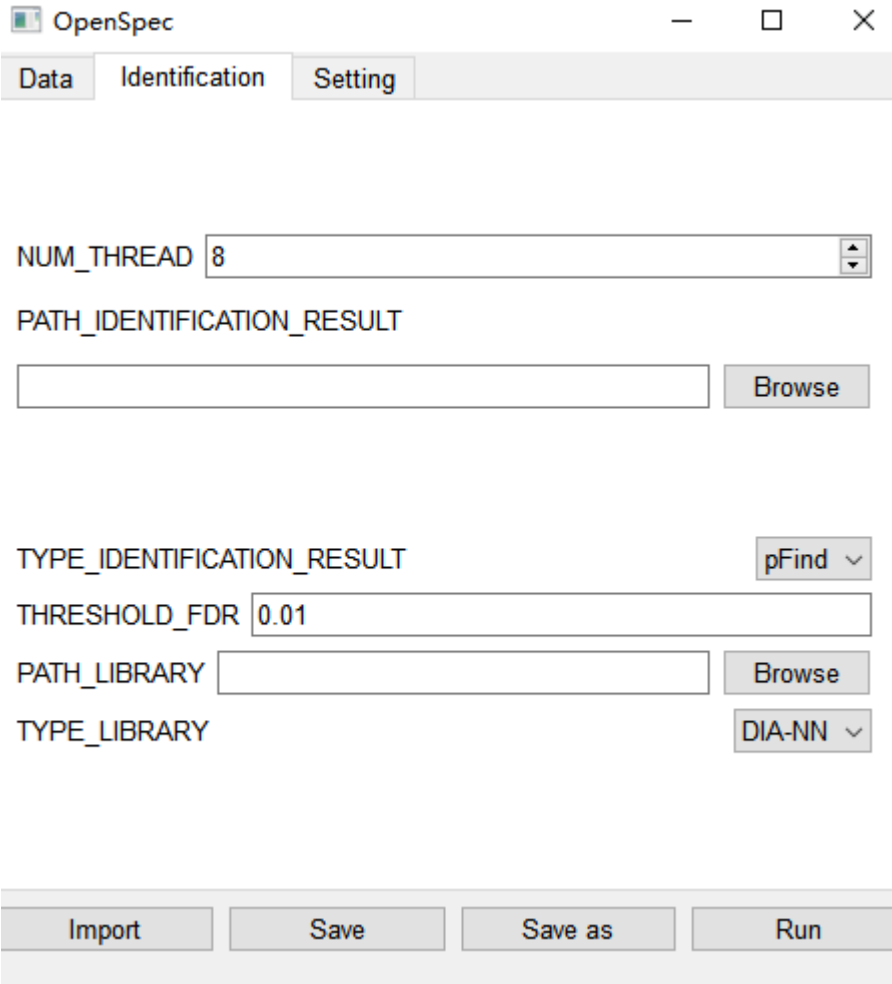
- ① Load raw file.
- ② Select output path.
- ③ The result file for subsequent analysis.

2 Pseudo-MS/MS Spectra Generation



- ① Set **TYPE_FLOW** as **Pseudo-MS/MS**.
- ② Fill the raw files into the **PATH_RAW**.
- ③ Fill the files into the **PATH_MS1_XIC**.

2 Pseudo-MS/MS Spectra Generation



The screenshot shows the 'OpenSpec' application window with the 'Identification' tab selected. The interface includes several input fields and buttons for configuring the identification process. A red circle with the number '1' highlights the 'NUM_THREAD' field, which is set to 8. Other fields include 'PATH_IDENTIFICATION_RESULT', 'TYPE_IDENTIFICATION_RESULT' (set to 'pFind'), 'THRESHOLD_FDR' (set to 0.01), 'PATH_LIBRARY', and 'TYPE_LIBRARY' (set to 'DIA-NN'). There are 'Browse' buttons for the 'PATH_IDENTIFICATION_RESULT' and 'PATH_LIBRARY' fields. At the bottom, there are buttons for 'Import', 'Save', 'Save as', and 'Run'.

OpenSpec

Data Identification Setting

① NUM_THREAD 8

PATH_IDENTIFICATION_RESULT

Browse

TYPE_IDENTIFICATION_RESULT pFind

THRESHOLD_FDR 0.01

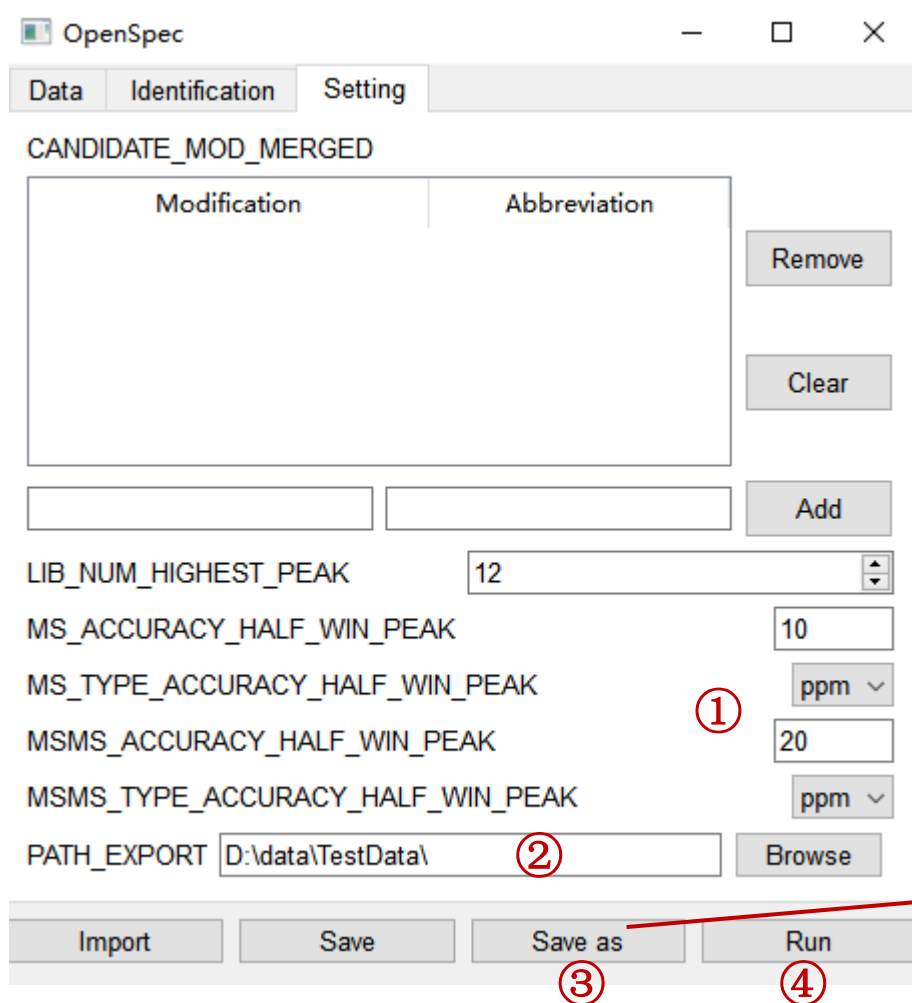
PATH_LIBRARY Browse

TYPE_LIBRARY DIA-NN

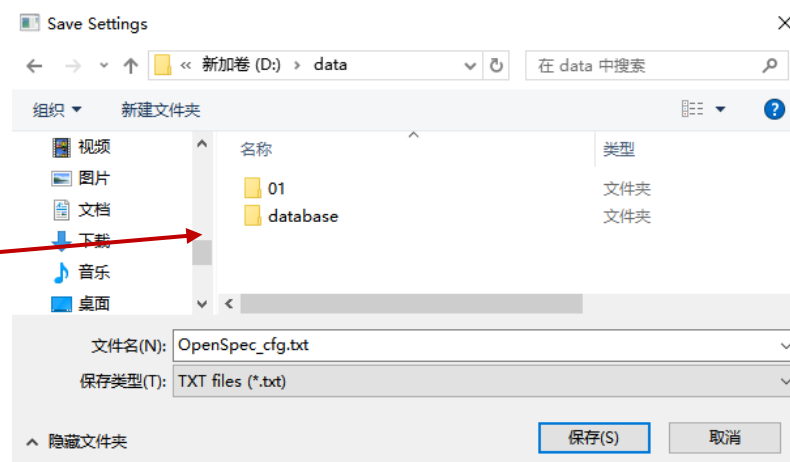
Import Save Save as Run

- ① Adjust **NUM_THREAD** according to your computer's performance; the default is 8.

2 Pseudo-MS/MS Spectra Generation



- ① Adjust the MS/MSMS accuracy based on your needs.
- ② Set the **PATH_EXPORT** for saving the results.
- ③ Click **Save as** button to save the config file.
- ④ Click **Run** button to start.





2 Pseudo-MS/MS Spectra Generation

```
[OpenSpec]    Checking the environment...
[OpenSpec]    Reading ini files...
[OpenSpec]    Extracting raw files...
[OpenSpec]    Extract RAW File: D:\data\TestData\20230803_AST1_Neo1_UHG_7min_DIA_RR_1_4th_003.raw
2025/4/27 20:37:07
[INFO] Checking File D:\data\TestData\20230803_AST1_Neo1_UHG_7min_DIA_RR_1_4th_003.raw
2025/4/27 20:37:07
[INFO] Reading scan titles from D:\data\TestData\20230803_AST1_Neo1_UHG_7min_DIA_RR_1_4th_003.raw
2025/4/27 20:37:13
[INFO] Writing all spectrum to D:\data\TestData\20230803_AST1_Neo1_UHG_7min_DIA_RR_1_4th_003.bin
2025/4/27 20:37:20
[INFO] Finish parsing file D:\data\TestData\20230803_AST1_Neo1_UHG_7min_DIA_RR_1_4th_003.raw

[OpenSpec]    Preprocessing...
0 / 660
0 / 81913
10000 / 81913
20000 / 81913
30000 / 81913
40000 / 81913
50000 / 81913
60000 / 81913
```

① The result files.

- ①
- | | | |
|---|--|------------|
|  | 20230803_AST1_Neo1_UHG_7min_DIA_RR_1_4th_003.mgf | MGF 文件 |
|  | 20230803_AST1_Neo1_UHG_7min_DIA_RR_1_4th_003.mgf_bin | MGF_BIN 文件 |

3 Open Search with pFind3

MS Data Identification Quantitation Summary

MS Data Format : MGF ①

MS Instrument : HCD-FTMS

Data File List

Files	Size
D:\data\TestData\20230803_AST1_Neo_UHG_7min_DIA_RR_1_	637.763MB

Add

Delete

Clear

MS Data Identification Quantitation Summary

Database Search

Database : HUMAN_swiss_prot ③

Enzyme : Trypsin KR_C Full-Specific Up to 3 missed cleavages

Precursor Tolerance \pm 20 ppm Fragment Tolerance \pm 20 ppm

☒ Open Search ④

Add Modification

Acetyl[AnyN-term]
Acetyl[K]

MS Data Identification Quantitation Summary

MS Data

Search

Filter

MS1 Quantitation

Save ⑤ Start Stop

- ① Set **MS Data Format** to **MGF**.
- ② Add **mgf file**.
- ③ Set the fasta database.
- ④ Set to **Open Search** is better.
- ⑤ Click **Save** and **Start** buttons, the progress information will be shown in Output interface.
- ⑥ The identification result files are used for hybrid spectral library construction.

名称

- 1.aa
- 1
- backup_pFind.cfg
- pFind.protein
- pFind.spectra
- pFind.summary
- pFind-Filtered.spectra

⑥

OpenSpec Manual for Hybrid Spectral Library Construction

1 First-Round Search with DIA-NN.

Input

Raw .d (DIA) Clear Convert to

D:\data\TestData\20230803_AST1_Neol_UHG_7min_DIA_RR_1_4th_003.raw

Output

☐ Reuse .quant file ☒ Quantities matrix ☐ XICs

Main output D:\data\TestData\DIANN-FirstR

Temp/.dia dir

☒ Generate spectral library

Output NN-FirstRound\topFive.parquet

Predicted library\data\TestData\DIANN-FirstR

☐ Generate Prosit input Threads 16

Precursor FDR (%) 1.0 Log Level 1

Additional options ☒ PDF

—var-mod UniMod:351,3.994915,W
—var-mod UniMod:425,31.989829,W
—var-mod UniMod:7,0.984016,N

FASTA

Add FASTA D:\data\TestData\uniprot-human_20422_20230615.fasta

Clear list

☐ Reannotate

☐ Contaminants

DIA-NN exe diann.exe

Precursor ion generation

☒ FASTA digest for library-free search / library

☒ Deep learning-based spectra, RTs and IMs prediction

Protease Trypsin/P Missed cleavages 1

Maximum number of variable modifications

☒ N-term M excision ☐ C carbamidomethylat

☒ Ox(M) ☒ Ac(N-term) ☐ Phospho ☐ K-GG

Peptide length range 6 - 50

Precursor charge range 1 - 6

Precursor m/z range 300 - 1800

Fragment ion m/z range 200 - 1800

Algorithm

Mass accuracy 0.0 ☒ MS/MS

MS1 accuracy 0.0 ☐ Unrelated runs

Scan window 0 ☒ Protein inference

Scoring Peptidoforms

Proteotypicity Genes

Machine learning NNs (cross-validated)

Quantification strategy QuantUMS (high precision)

Cross-run normalisation RT-dependent

Library generation IDs, RT & IM profiling

Speed and RAM usage Optimal results

Output files:

- report.gg_matrix.tsv
- report.log.txt
- report.manifest.txt
- report.parquet
- report.pg_matrix.tsv
- report.pr_matrix.tsv
- report.protein_description.tsv
- report.stats.tsv
- report.UniMod_1_sites_90.tsv
- report.UniMod_1_sites_99.tsv
- report.UniMod_7_sites_90.tsv
- report.UniMod_7_sites_99.tsv
- report.UniMod_35_sites_90.tsv
- report.UniMod_35_sites_99.tsv
- report.UniMod_351_sites_90.tsv
- report.UniMod_351_sites_99.tsv
- report.UniMod_425_sites_90.tsv
- report.UniMod_425_sites_99.tsv
- report.unique_genes_matrix.tsv
- report_runs.pdf
- report_trends.pdf
- topFive.parquet

Modifications:

Name	#Peptide	#Sites	Frequency
Oxidation[M]	1874 (11.68%)	1989	37.24%
Trp->Kynurenin[W]	400 (2.49%)	401	20.73%
Dioxidation[W]	390 (2.43%)	391	20.22%
Deamidated[N]	369 (2.30%)	378	4.39%
Acetyl[ProteinN-term]	196 (1.22%)	196	1.22%
Gln->pyro-Glu[AnyN-termQ]	147 (0.92%)	147	18.13%
DiDehydro[C]	29 (0.18%)	29	2.12%
Carbamidomethyl[C]	29 (0.18%)	29	2.12%
Trioxidation[C]	20 (0.12%)	20	1.46%

- ① Load raw file.
- ② Set the fasta database.
- ③ Set the **high-abundance modifications** according to open search results.
- ④ Select output path.
- ⑤ The identification result files are used for hybrid spectral library construction.

Note: since DIA-NN does not support certain symbols (e.g., ->), accession of UniMod are used uniformly instead.

2 Hybrid Spectral Library Construction

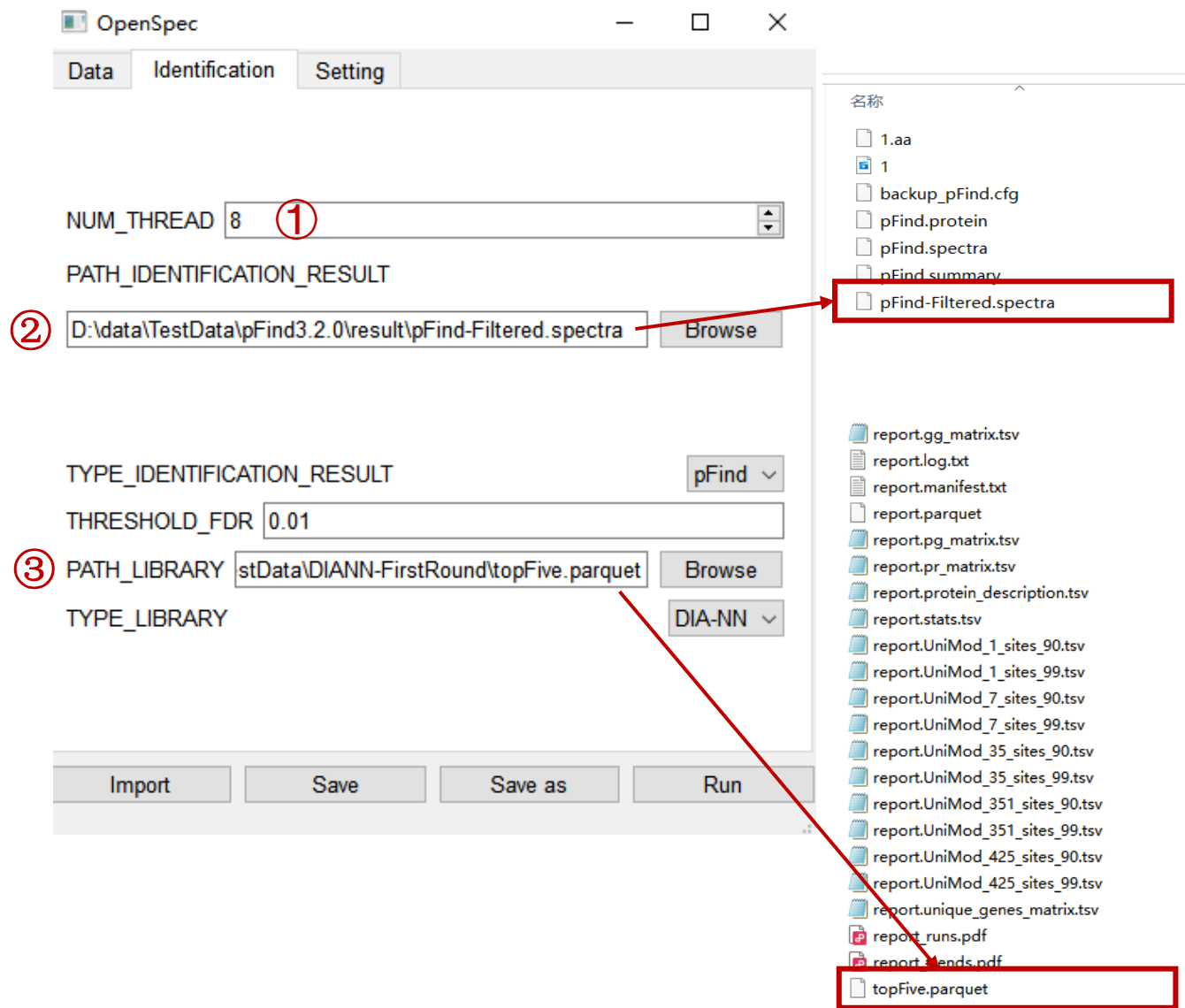
The screenshot shows the 'OpenSpec' application window with three tabs: 'Data', 'Identification', and 'Setting'. The 'Setting' tab is active. It contains several configuration fields and buttons:

- INI_PATH_ELEMENT**: Text box with '.\ini\element.ini' and a 'Browse' button.
- INI_PATH_AA**: Text box with '.\ini\aa.ini' and a 'Browse' button.
- INI_PATH_MOD**: Text box with '.\ini\modification.ini' and a 'Browse' button.
- TYPE_FLOW**: A dropdown menu with 'Hybrid-lib' selected, marked with a red circled '1'.
- TYPE_FILE**: A dropdown menu with 'raw' selected.
- PATH_RAW**: Text box with 'Neo1_UHG_7min_DIA_RR_1_4th_003.raw' and a 'Browse' button.
- PATH_MGF_BIN**: Text box with '1_UHG_7min_DIA_RR_1_4th_003.mgf_bin', marked with a red circled '2', and a 'Browse' button.
- PATH_MS1_XIC**: Empty text box with a 'Browse' button.

At the bottom of the window, there is a row of four buttons: 'Import', 'Save', 'Save as', and 'Run'.

- ① Set **TYPE_FLOW** as **Hybrid-lib**.
- ② Fill the files into the **PATH_MGF_BIN**.

2 Hybrid Spectral Library Construction



- ① Adjust **NUM_THREAD** according to your computer's performance; the default is 8.
- ② Put the identification result file **pFind-Filtered.spectra** from pFind here.
- ③ Put the **spectral library from first-round search with DIA-NN** here.

2 Hybrid Spectral Library Construction

pFind.summary - 记事本

文件(F) 编辑(E) 格式(O) 查看(V) 帮助(H)

Specific: 15098 94.13%
C-term Specific: 875 5.46%
N-term Specific: 67 0.42%
Non-Specific: 0 0.00%

Modifications:

Name	#Peptide	#Sites	Frequency
Oxidation[M]	1874 (11.68%)	1989	37.24%
Trp->Kynurenin[W]	400 (2.49%)	401	20.73%
Dioxidation[W]	390 (2.43%)	391	20.22%
Deamidated[N]	369 (2.30%)	378	4.39%
Acetyl[ProteinN-term]	196 (1.22%)	196	1.22%
Gln->pyro-Glu[AnyN-termQ]	147 (0.92%)	147	18.13%
DiDehydro[C]	29 (0.18%)	29	2.12%
Carbamidomethyl[C]	29 (0.18%)	29	2.12%
Trioxidation[C]	20 (0.12%)	20	1.46%

OpenSpec

①

CANDIDATE	MOD	MERGED
3	Gln->pyro-Glu[AnyN-termQ]	UniMod:28
4	Acetyl[ProteinN-term]	UniMod:1
5	Oxidation[M]	UniMod:35
6	Deamidated[N]	UniMod:7

②

PATH_EXPORT D:/data/

③

④

HybridLib.tsv

TSV 文件

Input: Raw .d (DIA) Clear Convert to

Output: Reuse .quant fill Quantities matrix XICs

Main output: D:/data/TestData/DIA-NN-FirstR

Temp/ dir:

Generate spectral library

Output: NW-FirstRound/topFive.parquet

Predicted library: D:/data/TestData/DIA-NN-FirstR

Generate Prosit input Threads: 16

Precursor FDR (%): 1.0 Log level:

Additional options: PDF

Run Finished Stop

Algorithm: Mass accuracy: 0.0 MS1 accuracy: 0.0 Scan window: 0 Scoring: Peptideforms Proteotypicity: Genes Machine learning: NNs (cross-validated) Quantification strat: QuantMS (high precisi) Cross-run normalisat: RT-dependent Library generation: IDs, RT & IM profiling Speed and RAM usage: Optimal results

- ① Set the high-abundance modifications. The **"Modification"** column contains the original modification names from the pFind, while the **"Abbreviation"** column contains the modified names, which should be consistent with the first-round search with DIA-NN.
- ② Set the **PATH_EXPORT** for saving the results.
- ③ Click **Save as** button to save the config file. Click **Run** button to start, the progress information will be shown in the command-line interface.
- ④ The result file.

3 Second-Round Search with DIA-NN.

Input

Raw .d (DIA) Clear Convert to

D:\data\TestData\20230803_AST1_Neol_UHG_7min_DIA_RR_1_4th_003.raw

①

⑤

HybridLib.tsv TSV 文件

②

③

Spectral D:\data\TestData\HybridLib.tsv

Add FASTA D:\data\TestData\uniprot-human_20422_20230615.fasta

Clear list

☐ Reannotate

☐ Contaminants

DIA-NN exe diann.exe

Precursor ion generation

☐ FASTA digest for library-free search / library

☐ Deep learning-based spectra, RTs and IMs prediction

Protease Trypsin/P Missed cleavages 1

Maximum number of variable modifications

☒ N-term M excision ☐ C carbamidomethylation

☒ Ox(M) ☒ Ac(N-term) ☐ Phospho ☐ K-GG

Peptide length range 6 - 50

Precursor charge range 1 - 6

Precursor m/z range 300 - 1800

Fragment ion m/z range 200 - 1800

Output

☐ Reuse .quant file ☒ Quantities matrix ☐ XICs

Main output .NN-SearchRound\report.parquet

Temp/.dia.dir

☒ Generate spectral library

Output SearchRound\HybridLib.parquet

Predicted library

☐ Generate Prosit input Threads 16

Precursor FDR (%1.0 Log level 1

Additional options ☒ PDF

④

Run Finished Stop

Algorithm

Mass accuracy 0.0 ☐ MBR

MS1 accuracy 0.0 ☐ Unrelated runs

Scan window 0 ☒ Protein inference

Scoring Peptidofoms

Proteotypicity Genes

Machine learning NNs (cross-validated)

Quantification strategy QuantUMS (high precision)

Cross-run normalisation RT-dependent

Library generation IDs, RT & IM profiling

Speed and RAM usage Optimal results

- ① Load raw file.
- ② Set the hybrid spectral library.
- ③ Set the fasta database.
- ④ Set the **high-abundance modifications**, which should be consistent with the first-round search with DIA-NN.
- ⑤ Select output path.
- ⑥ The final result.

- ⑥
- report.gg_matrix.tsv
 - report.log.txt
 - report.manifest.txt
 - report.parquet
 - report.pg_matrix.tsv
 - report.pr_matrix.tsv
 - report.protein_description.tsv
 - report.stats.tsv
 - report.UniMod_1_sites_90.tsv
 - report.UniMod_1_sites_99.tsv
 - report.UniMod_7_sites_90.tsv
 - report.UniMod_7_sites_99.tsv
 - report.UniMod_35_sites_90.tsv
 - report.UniMod_35_sites_99.tsv
 - report.UniMod_351_sites_90.tsv
 - report.UniMod_351_sites_99.tsv
 - report.UniMod_425_sites_90.tsv
 - report.UniMod_425_sites_99.tsv
 - report.unique_genes_matrix.tsv
 - report_runs.pdf
 - report_trends.pdf
 - topFive.parquet

Thanks!