

Searching plexDIA data with DIA-NN

Please see <https://github.com/vdemichev/DiaNN>
for more information about DIA-NN

Last updated December 8th, 2021

1. Download and install DIA-NN v1.8.1 beta7

DIA-NN v1.8.1 beta7 is available for download on the plexDIA website. A link to the download can be found in the “Download data” section with the hyperlinked “DIA-NN”.

<https://plexdia.slavovlab.net>

Or, a direct link is here:

https://drive.google.com/drive/u/1/folders/1p538GxhbZ7CllodVKVc0zFjRm_OmoQo1

2. Download plexDIA data

Data can be found on MassIVE (MSV000088302)

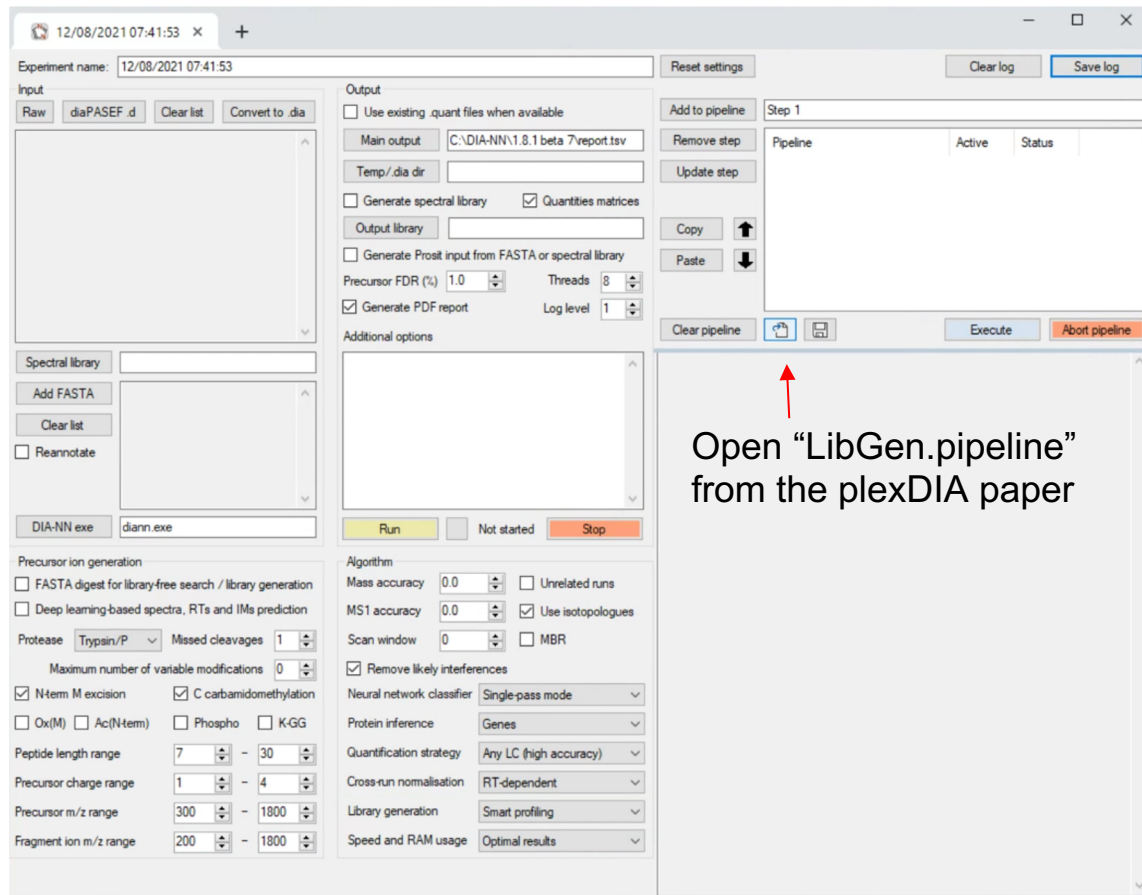
1. FASTA
2. Raw files
3. Spectral libraries (optional; will show how to generate these from FASTA next)
4. .pipeline files (this makes it easier to set up searches)

3. In-silico spectral library generation

In DIA-NN, it is possible to save search settings as a pipeline to make reproducibility easier.

Before, we search the raw data, we can generate the spectral libraries from the human, yeast, ecoli FASTA we provide at MSV000088302.

1. Open/load “LibGen.pipeline” which is available at MSV000088302

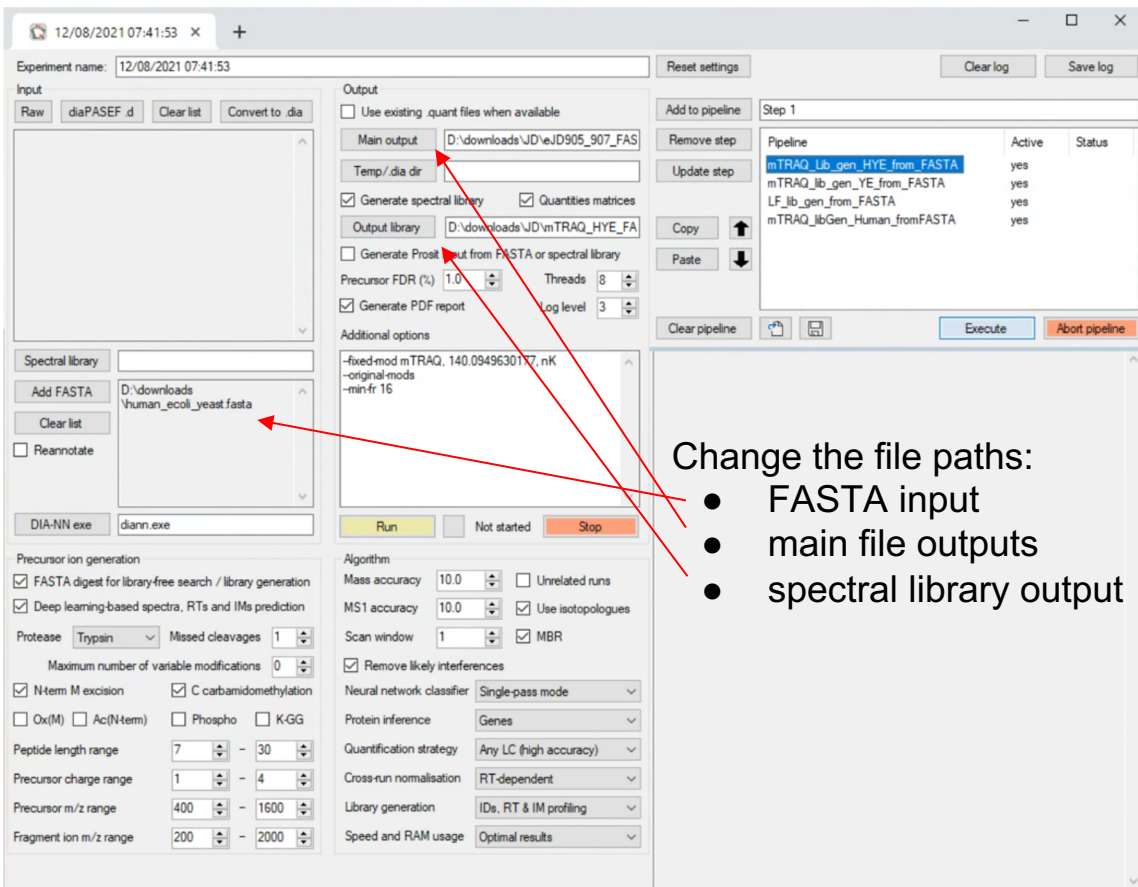


3. In-silico spectral library generation

In DIA-NN, it is possible to save search settings as a pipeline to make reproducibility easier.

Before, we search the raw data, we can generate the spectral libraries from the human, yeast, ecoli FASTA we provide at MSV000088302.

1. Open/load “LibGen.pipeline” which is available at MSV000088302
2. Change the input and output file paths



3. In-silico spectral library generation

In DIA-NN, it is possible to save search settings as a pipeline to make reproducibility easier.

Before, we search the raw data, we can generate the spectral libraries from the human, yeast, ecoli FASTA we provide at MSV000088302.

1. Open/load “LibGen.pipeline” which is available at MSV000088302
2. Change the input and output files paths
3. Check the commands

Experiment name: 12/08/2021 07:41:53

Input: Raw | diaPASEF.d | Clear list | Convert to .dia

Output: ☐ Use existing quant files when available

Main output: D:\downloads\JD\mTRAQ_907_FAS

Temp/.dia dir:

☒ Generate spectral library ☒ Quantities matrices

Output library: D:\downloads\JD\mTRAQ_HYE_FA

☐ Generate Proxit input from FASTA or spectral library

Precursor FDR (%): 1.0 Threads: 8

☒ Generate PDF report Log level: 3

Additional options: --fixed-mod mTRAQ, 140.0949630177, nK
--original-mods
--min-fr 16

Run | Not started | Stop

Precursor ion generation: ☒ FASTA digest for library-free search / library generation
☒ Deep learning-based spectra, RTs and IMs prediction

Protease: Trypsin Missed cleavages: 1

Maximum number of variable modifications: 0

☒ N-term M excision ☒ C carbamidomethylation

☐ Ox(M) ☐ Ac(N-term) ☐ Phospho ☐ K-GG

Peptide length range: 7 - 30

Precursor charge range: 1 - 4

Precursor m/z range: 400 - 1600

Fragment ion m/z range: 200 - 2000

Algorithm: Mass accuracy: 10.0 ☐ Unrelated runs
MS1 accuracy: 10.0 ☒ Use isotopologues
Scan window: 1 ☒ MBR
☒ Remove likely interferences

Neural network classifier: Single-pass mode

Protein inference: Genes

Quantification strategy: Any LC (high accuracy)

Cross-run normalisation: RT-dependent

Library generation: IDs, RT & IM profiling

Speed and RAM usage: Optimal results

Reset settings | Clear log | Save log

Add to pipeline | Remove step | Update step | Copy | Paste

Pipeline table:

Pipeline	Active	Status
mTRAQ_Lib_gen_HYE_from_FASTA	yes	
mTRAQ_lib_gen_YE_from_FASTA	yes	
LF_lib_gen_from_FASTA	yes	
mTRAQ_libGen_Human_fromFASTA	yes	

Clear pipeline | Execute | Abort pipeline

These commands will add the delta0 mTRAQ mass to N-terminal peptides and lysine residues. Downstream, we will use this mass tag to add masses for the other mTRAQ tags as channels for the search.

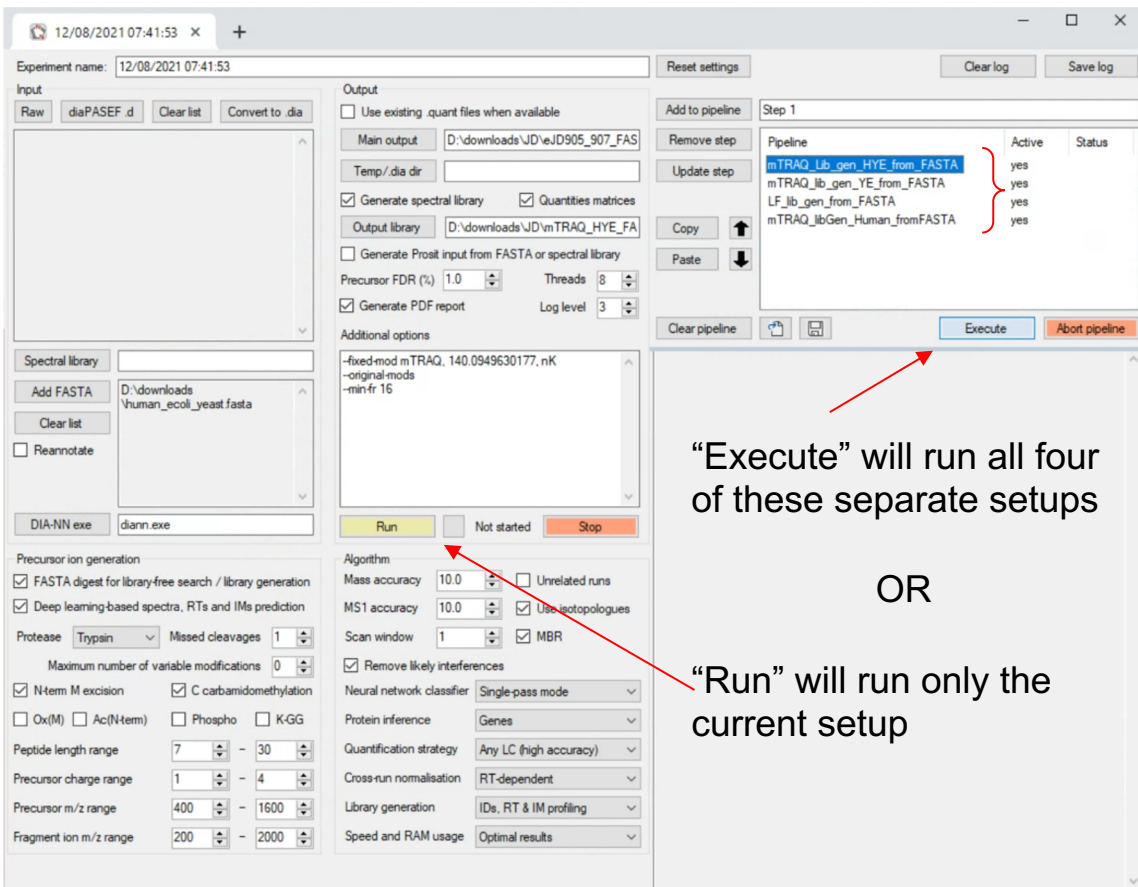
The --min-fr 16 is optional. It will predict 16 fragments for each precursor.

3. In-silico spectral library generation

In DIA-NN, it is possible to save search settings as a pipeline to make reproducibility easier.

Before, we search the raw data, we can generate the spectral libraries from the human, yeast, ecoli FASTA we provide at MSV000088302.

1. Open/load “LibGen.pipeline” which is available at MSV000088302
2. Change the input and output files paths
3. Check the commands
4. Click run to run the current setup, or click execute to run the entire pipeline.



3. In-silico spectral library generation

In DIA-NN, it is possible to save search settings as a pipeline to make reproducibility easier.

Before, we search the raw data, we can generate the spectral libraries from the human, yeast, ecoli FASTA we provide at MSV000088302.

1. Open/load “LibGen.pipeline” which is available at MSV000088302
2. Change the input and output files paths
3. Check the commands
4. Click run to run the current setup, or click execute to run the entire pipeline.

The screenshot displays the DIA-NN software interface. The 'Experiment name' is '12/08/2021 07:41:53'. The 'Input' section shows 'Raw' data. The 'Output' section shows 'Main output' as 'D:\downloads\JD\veJD905_907_FAS'. The 'Spectral library' section shows 'Add FASTA' with the path 'D:\downloads\human_ecoli_yeast.fasta'. The 'Precursor ion generation' section has checkboxes for 'FASTA digest for library-free search / library generation' and 'Deep learning-based spectra, RTs and IMs prediction'. The 'Algorithm' section shows 'Mass accuracy' and 'MS1 accuracy' both set to 10.0. The 'Neural network classifier' is set to 'Single-pass mode'. The 'Protein inference' is set to 'Genes'. The 'Quantification strategy' is set to 'Any LC (high accuracy)'. The 'Cross-run normalisation' is set to 'RT-dependent'. The 'Library generation' is set to 'IDs, RT & IM profiling'. The 'Speed and RAM usage' is set to 'Optimal results'. The 'Run' button is highlighted in yellow. The 'Pipeline' table on the right shows the status of the pipeline steps.

Pipeline	Active	Status
mTRAQ_Lib_gen_HYE_from_FASTA	yes	
mTRAQ_lib_gen_YE_from_FASTA	yes	
LF_lib_gen_from_FASTA	yes	
mTRAQ_LibGen_Human_fromFASTA	yes	

When the library generation is finished running, you should find a .predicted.speclib file in the “Output library” path

→ This is the predicted spectral library that can be used for future searches.

3. In-silico spectral library generation

In DIA-NN, it is possible to save search settings as a pipeline to make reproducibility easier.

Before, we search the raw data, we can generate the spectral libraries from the human, yeast, ecoli FASTA we provide at [MSV000088302](https://www.ebi.ac.uk/MSV000088302).

1. Open\load “LibGen.pipeline” which is available at
2. Change the input and output files paths
3. Check the commands
4. Click run to run the current setup, or click execute to run the entire pipeline.

Note:

The predicted.speclib file can be converted to .tsv by loading the predicted spectral library, deleting any existing commands, and unchecking the two boxes in the “Precursor Generation” tab, then clicking “Run”.

4. Search raw plexDIA data

1. Open/load “Searches.pipeline”

12/08/2021 07:41:53

Experiment name: 12/08/2021 07:41:53

Input

Raw diaPASEF.d Clear list Convert to .dia

D:\downloads\JD\wJD804.raw
D:\downloads\JD\wJD803.raw
D:\downloads\JD\wJD815.raw
D:\downloads\JD\wJD805.raw
D:\downloads\JD\wJD808.raw
D:\downloads\JD\wJD807.raw
D:\downloads\JD\wJD806.raw

Spectral library D:\downloads\JD\mTRAQ_HYE_FA

Add FASTA D:\downloads\Human_ecoli_yeast.fasta

Clear list

☐ Reannotate

DIA-NN exe diann.exe

Precursor ion generation

☐ FASTA digest for library-free search / library generation

☐ Deep learning-based spectra, RTs and IMs prediction

Protease Trypsin Missed cleavages 1

Maximum number of variable modifications 0

☒ N-term M excision ☒ C carbamidomethylation

☐ Ox(M) ☐ Ac(N-term) ☐ Phospho ☐ K-GG

Peptide length range 7 - 30

Precursor charge range 1 - 4

Precursor m/z range 400 - 1600

Fragment ion m/z range 200 - 2000

Output

☐ Use existing .quant files when available

Main output D:\downloads\JD\w1.8b7_10102021

Temp./dia dir

☐ Generate spectral library ☒ Quantities matrices

Output library

☐ Generate Proxit input from FASTA or spectral library

Precursor FDR (%) 1.0 Threads 8

☒ Generate PDF report Log level 3

Additional options

-fixed-mod mTRAQ_140.0949630177_nK
-channels mTRAQ_0.0.0.4:4.0070994.8:
8.0141988132
-peak-translation
-original-mods
-report-lib-info
-ms1-isotope-quant
-ms1-subtract 2

Run Not started Stop

Algorithm

Mass accuracy 10.0 ☐ Unrelated runs

MS1 accuracy 5.0 ☒ Use isotopologues

Scan window 1 ☒ MBR

☒ Remove likely interferences

Neural network classifier Single-pass mode

Protein inference Genes

Quantification strategy Peak height

Cross-run normalisation RT-dependent

Library generation IDs, RT & IM profiling

Speed and RAM usage Optimal results

Reset settings Clear log Save log

Add to pipeline Step 1

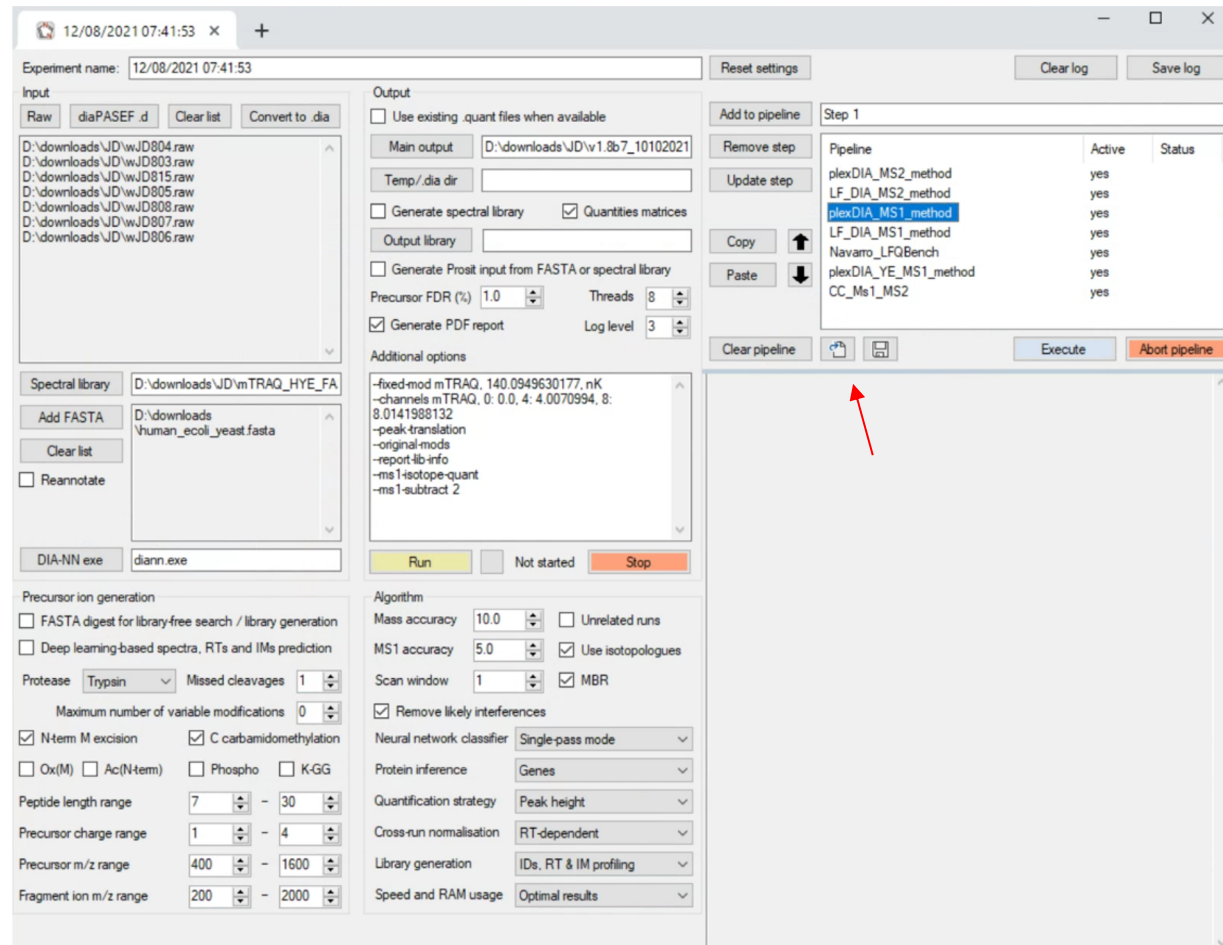
Remove step

Update step

Copy Paste

Clear pipeline Execute Abort pipeline

Pipeline	Active	Status
plexDIA_MS2_method	yes	
LF_DIA_MS2_method	yes	
plexDIA_MS1_method	yes	
LF_DIA_MS1_method	yes	
Navarro_LFQBench	yes	
plexDIA_YE_MS1_method	yes	
CC_Ms1_MS2	yes	



4. Search raw plexDIA data

1. Open/load “Searches.pipeline”
1. Select an appropriate search-setup from the pipeline

The screenshot displays the plexDIA software interface with the following sections:

- Experiment name:** 12/08/2021 07:41:53
- Input:** Raw, diaPASEF.d, Clear list, Convert to .dia. A list of raw files is shown, including D:\downloads\JD\wJD804.raw through D:\downloads\JD\wJD806.raw.
- Output:** Use existing .quant files when available. Main output: D:\downloads\JD\w1.8b7_10102021. Temp/.dia dir. Generate spectral library (checked), Quantities matrices (checked), Output library. Generate Proxit input from FASTA or spectral library. Precursor FDR (%): 1.0, Threads: 8, Log level: 3. Generate PDF report (checked).
- Additional options:** -fixed-mod mTRAQ, 140.0949630177, nK -channels mTRAQ, 0.0, 4: 4.0070994, 8: 8.0141988132 -peak-translation -original-mods -report-lib-info -ms1-isotope-quant -ms1-subtract 2
- Precursor ion generation:** FASTA digest for library-free search / library generation (unchecked), Deep learning-based spectra, RTs and IMs prediction (unchecked), Protease: Trypsin, Missed cleavages: 1, Maximum number of variable modifications: 0, N-term M excision (checked), C carbamidomethylation (checked), Ox(M) (unchecked), Ac(N-term) (unchecked), Phospho (unchecked), K-GG (unchecked), Peptide length range: 7 - 30, Precursor charge range: 1 - 4, Precursor m/z range: 400 - 1600, Fragment ion m/z range: 200 - 2000.
- Algorithm:** Mass accuracy: 10.0, MS1 accuracy: 5.0, Scan window: 1, Remove likely interferences (checked), Neural network classifier: Single-pass mode, Protein inference: Genes, Quantification strategy: Peak height, Cross-run normalisation: RT-dependent, Library generation: IDs, RT & IM profiling, Speed and RAM usage: Optimal results.
- Pipeline:** A table showing the pipeline steps and their status. A red bracket highlights the first five steps.

Pipeline	Active	Status
plexDIA_MS2_method	yes	
LF_DIA_MS2_method	yes	
plexDIA_MS1_method	yes	
LF_DIA_MS1_method	yes	
Navarro_LFQBench	yes	
plexDIA_YE_MS1_method	yes	
CC_Ms1_MS2	yes	

Buttons: Run, Not started, Stop, Execute, Abort pipeline.

4. Search raw plexDIA data

1. Open/load “Searches.pipeline”
1. Select an appropriate search-setup from the pipeline
1. Change the file paths

The screenshot shows the DIA-NN software interface. Red arrows point from the list on the left to specific fields in the software:

- Arrow 1 points to the **Raw** input field, which currently shows a list of raw files.
- Arrow 2 points to the **Spectral library** field, which currently shows a list of spectral libraries.
- Arrow 3 points to the **FASTA** input field, which currently shows a list of FASTA files.
- Arrow 4 points to the **Main output** field, which currently shows a path to a directory.

Other visible fields and settings include:

- Experiment name:** 12/08/2021 07:41:53
- Output:** Use existing .quant files when available, Main output: D:\downloads\JD\v1.8b7_10102021
- Generate PDF report:** Checked
- Precursor FDR (%):** 1.0
- Threads:** 8
- Log level:** 3
- Additional options:** -fixed-mod mTRAQ, 140.0949630177, nK; -channels mTRAQ, 0.0-0.4:4.0070994, 8:8.0141988132; -peak-translation; -original-mods; -report-lib-info; -ms1-isotope-quant; -ms1-subtract-2
- Run:** Run, Not started, Stop
- Algorithm:** Mass accuracy: 10.0, MS1 accuracy: 5.0, Scan window: 1, Remove likely interferences: Checked, Neural network classifier: Single-pass mode, Protein inference: Genes, Quantification strategy: Peak height, Cross-run normalisation: RT-dependent, Library generation: IDs, RT & IM profiling, Speed and RAM usage: Optimal results
- Pipeline table:**

Pipeline	Active	Status
plexDIA_MS2_method	yes	
LF_DIA_MS2_method	yes	
plexDIA_MS1_method	yes	
LF_DIA_MS1_method	yes	
Navarro_LFQBench	yes	
plexDIA_YE_MS1_method	yes	
CC_Ms1_MS2	yes	

Change the file paths:

- raw data input
- spectral library input
- FASTA input
- Search results output

4. Search raw plexDIA data

1. Open/load “Searches.pipeline”
1. Select an appropriate search-setup from the pipeline
1. Change the file paths
1. Modify the commands and settings (optional)

The screenshot displays the plexDIA software interface with various search parameters and pipeline steps. Red boxes highlight specific areas of interest:

- Additional options:** A text area containing search parameters: `--fixed-mod mTRAQ_140.0949630177.mK`, `--channels mTRAQ_0.0.0,4:4.0070994,8:8.0141988132`, `--peak-translation`, `--original-mods`, `--report-lib-info`, `--ms1-isotope-quant`, and `--ms1-subtract 2`.
- Algorithm:** A section with dropdown menus for Mass accuracy (10.0), MS1 accuracy (5.0), Scan window (1), and checkboxes for Unrelated runs, Use isotopologues, MBR, and Remove likely interferences.
- Neural network classifier:** A dropdown menu set to "Single-pass mode".
- Pipeline:** A table showing the sequence of steps in the search pipeline.

Pipeline	Active	Status
plexDIA_MS2_method	yes	
LF_DIA_MS2_method	yes	
plexDIA_MS1_method	yes	
LF_DIA_MS1_method	yes	
Navarro_LFQBench	yes	
plexDIA_YE_MS1_method	yes	
CC_Ms1_MS2	yes	

For general plexDIA experiments, it may be worthwhile to set scan window to 0 (automatic). However, for long duty cycles, a small scan window may be beneficial.

“--channels ...” specifies the added mass for each channel from the fixed-mod mTRAQ ($\Delta 0$, $\Delta 4$, $\Delta 8$)

“--ms1-isotope-quant” integrates MS1 signal at M+1

“--ms1-subtract 2” subtracts isotopic envelope carryover between mTRAQ channels

4. Search raw plexDIA data

1. Open/load “Searches.pipeline”
1. Select an appropriate search-setup from the pipeline
1. Change the file paths
1. Modify the commands and settings (optional)

The screenshot displays the plexDIA software interface with various search parameters and pipeline steps. Red boxes highlight specific areas:

- Additional options:** A text area containing search parameters: `-fixed-mod mTRAQ, 140.0949630177, nK`, `-channels mTRAQ, 0.0, 4.4.0070994, 8:`, `8.0141988132`, `-peak-translation`, `-original-mods`, `-report-ib-info`, `-ms1-isotope-quant`, and `-ms1-subtract 2`.
- Algorithm:** A section with dropdowns for Mass accuracy (10.0), MS1 accuracy (5.0), Scan window (1), and checkboxes for Unrelated runs, Use isotopologues, MBR, and Remove likely interferences.
- Neural network classifier:** A dropdown menu set to "Single-pass mode".
- Pipeline:** A table showing the sequence of steps in the search pipeline.

Pipeline	Active	Status
plexDIA_MS2_method	yes	
LF_DIA_MS2_method	yes	
plexDIA_MS1_method	yes	
LF_DIA_MS1_method	yes	
Navarro_LFQBench	yes	
plexDIA_YE_MS1_method	yes	
CC_Ms1_MS2	yes	

Note:

The boxes in “Precursor ion generation” are unchecked because we’ve already generated the library.

If we did not have the spectral library pre-made, we could check these two boxes and then it would predict the spectral library and subsequently use it to search these runs.

4. Search raw plexDIA data

1. Open/load “Searches.pipeline”
1. Select an appropriate search-setup from the pipeline
1. Change the file paths
1. Modify the commands and settings (optional)
1. Click run to run the current setup, or click execute to run the entire pipeline.

The screenshot displays the plexDIA software interface with various configuration panels. The 'Input' panel shows a list of raw data files. The 'Output' panel includes options for using existing quant files, generating a spectral library, and setting precursor FDR and threads. The 'Additional options' panel contains a list of search parameters. The 'Protease' panel shows settings for trypsin digestion. The 'Precursor ion generation' panel includes options for FASTA digest, deep learning-based spectra, and maximum number of variable modifications. The 'Algorithm' panel shows mass accuracy, MS1 accuracy, scan window, and removal of likely interferences. The 'Neural network classifier' panel shows single-pass mode. The 'Protein inference' panel shows genes. The 'Quantification strategy' panel shows peak height. The 'Cross-run normalisation' panel shows RT-dependent. The 'Library generation' panel shows IDs, RT & IM profiling. The 'Speed and RAM usage' panel shows optimal results.

On the right side, the 'Pipeline' panel shows a list of steps. A red bracket groups the steps: plexDIA_MS2_method, LF_DIA_MS2_method, plexDIA_MS1_method, LF_DIA_MS1_method, Navarro_LFQBench, plexDIA_YS1_method, and CC_Ms1_MS2. The 'Execute' button is highlighted with a red arrow. Below the 'Execute' button, the text "Execute" will run every step/setup in pipeline is shown. The word "OR" is displayed in the center. Below "OR", the text "Run" will run only the current setup is shown, with a red arrow pointing to the 'Run' button.

Execute

OR

Run

5. DIA-NN search outputs

- When the runs finish searching, the output folder should hold most of the following data
- The main .tsv file will have all the required information for analysis.
- For more information about the column outputs, please refer to:

<https://github.com/vdemichev/DiaNN#main-output-reference>

Note: “First-pass” refers to run-specific results (before MBR).

Name	Date modified	Type	Size
Report.auto.pipeline	10/11/2021 7:13 AM	PIPELINE File	2 KB
Report.gg_matrix.tsv	10/11/2021 7:13 AM	TSV File	387 KB
Report.log.txt	10/11/2021 7:13 AM	TXT File	337 KB
Report.pdf	10/11/2021 7:14 AM	Microsoft Edge PDF ...	106 KB
Report.pg_matrix.tsv	10/11/2021 7:13 AM	TSV File	857 KB
Report.pr_matrix.tsv	10/11/2021 7:12 AM	TSV File	41,732 KB
Report.pr_matrix_channels.tsv	10/11/2021 7:12 AM	TSV File	19,778 KB
Report.pr_matrix_channels_ms1.tsv	10/11/2021 7:12 AM	TSV File	20,972 KB
Report.pr_matrix_channels_ms1_extracted.tsv	10/11/2021 7:13 AM	TSV File	30,302 KB
Report.pr_matrix_channels_ms1_translated.tsv	10/11/2021 7:13 AM	TSV File	20,482 KB
Report.pr_matrix_channels_translated.tsv	10/11/2021 7:12 AM	TSV File	19,599 KB
Report.stats.tsv	10/11/2021 7:13 AM	TSV File	2 KB
Report.tsv	10/11/2021 7:12 AM	TSV File	1,273,146 KB
Report.unique_genes_matrix.tsv	10/11/2021 7:13 AM	TSV File	345 KB
Report-first-pass.gg_matrix.tsv	10/11/2021 7:03 AM	TSV File	357 KB
Report-first-pass.pg_matrix.tsv	10/11/2021 7:03 AM	TSV File	830 KB
Report-first-pass.pr_matrix.tsv	10/11/2021 7:03 AM	TSV File	39,297 KB
Report-first-pass.pr_matrix_channels.tsv	10/11/2021 7:03 AM	TSV File	18,756 KB
Report-first-pass.pr_matrix_channels_ms1.tsv	10/11/2021 7:03 AM	TSV File	19,805 KB
Report-first-pass.pr_matrix_channels_ms1_extra...	10/11/2021 7:03 AM	TSV File	28,134 KB
Report-first-pass.pr_matrix_channels_ms1_transl...	10/11/2021 7:03 AM	TSV File	19,428 KB
Report-first-pass.pr_matrix_channels_translated....	10/11/2021 7:03 AM	TSV File	18,538 KB
Report-first-pass.stats.tsv	10/11/2021 7:03 AM	TSV File	2 KB
Report-first-pass.tsv	10/11/2021 7:03 AM	TSV File	1,105,781 KB
Report-first-pass.unique_genes_matrix.tsv	10/11/2021 7:03 AM	TSV File	317 KB

Questions? Please reach out to:

Derks J, Leduc A, Huffman RG, Specht H, Ralser M, Demichev V, Slavov N.
(2021) **Increasing the throughput of sensitive proteomics by plexDIA**
bioRxiv 2021.11.03.467007; doi: <https://doi.org/10.1101/2021.11.03.467007>
<https://plexdia.slavovlab.net>

Jason Derks (derks.j{at}northeastern.edu)

Vadim Demichev (vadim.demichiev{at}crick.ac.uk)

Nikolai Slavov (nslavov{at}alum.mit.edu)