EncyclopeDIA and OpenSWATH in DIA data analysis

Lina,Lu

Department of Biology, Box 118, 221 00, Lund University, Sweden

Abstract

This project aims to compare the performance of two software: EncyclopeDIA and OpenSWATH in peptides detection and protein quantification of Data Independent Acquisition (DIA) Mass Spectrometry (MS) data. Pooled human plasma sample was prepared using S-trap before collecting data using Data independent acquisition (DIA) mass spectrometer. The collected data is analyzed using EncyclopeDIA workflow and OpenSWATH workflow respectively. In order to ensure the consistency of other variables, the two software uses the same chromatogram library, which is generated by the EncyclopeDIA workflow. The result shows that EncyclopeDIA doesn't detect as much peptides as OpenSWATH, but it has higher repeatability and accuracy. Moreover, OpenSWATH cannot generate a chromatogram library, it relies on the chromatogram library generated from EncyclopeDIA or other platforms such as DIA-Umpire, which makes the analysis of DIA data using OpenSWATH more complicated. Therefore, EncyclopeDIA workflow instead of the OpenSWATH workflow is written into the Sankemake workflow. This Snakemake workflow realizes the automation of DIA data analysis and ensures the high efficiency and high accuracy of peptide detection and protein quantification.

Introduction

In traditional data-dependent acquisition (DDA), a proteomic sample is digested into peptides, ionized and analyzed by mass spectrometry. Peptide signals that rise above the noise in a full-scan mass spectrum are selected for fragmentation, producing tandem (MS/MS) mass spectra that can be matched to spectra in a database [1]. Although extremely powerful, the mass spectrometer randomly samples peptides for fragmentation and is biased to pick those with the strongest signal [1]. Thus, it remains a challenge to reproducibly quantify especially low-abundance peptides.

DIA MS changed how proteomic data are generated. In DIA mode, the instrument fragments all precursors generated from a sample that are within a predetermined mass-to-charge ratio (m/z) and retention-time range [2]. The analysis is repeated as the mass spectrometer marches up the full m/z range, which results in accurate peptide quantification without being limited to profiling predefined peptides of interest [3].

However, DIA Data analysis is generally challenging as the resulting fragment ion spectra are highly multiplexed, multiple peptides in an m/z window are fragmented together in DIA, the

resulting MS/MS spectra are very complex and require deconvolution [1]. Therefore, a robust data analysis tools are required for DIA data analysis.

Many software used to analyze DIA data are published, such as EncyclopeDIA[4], OpenSWATH[6]. However, it is not clear which software has better performance in detecting peptides and has greater potential to facilitate future analysis of new DIA data. DIA data is usually very large. Analysis of DIA data includes data format conversion, library generation and quantitative report generation. It requires the participation of multiple software such as MSconvert, EncyclopeDIA or OpenSWATH. These two reasons make the step-by-step analysis very time-consuming, so a workflow that can obtain high-quality quantitative reports from raw data is necessary. This project aims to evaluate which of EncyclopeDIA and OpenSWATH is better in peptide quantification. The excellent one is written as a Snakemake workflow to facilitate future DIA data analysis.

The same data is conducted EncyclopeDIA workflow and OpenSWATH workflow respectively. Result shows that OpenSWATH detects more peptides than EncyclopeDIA, but has lower accuracy and lower repeatability within runs. However, EncyclopeDIA not only performs better (higher accuracy and repeatability) on quantitative analysis, but also can generate library which can be used in other platforms. Since OpenSWATH can't generate library, analyzing DIA data using OpenSWATH is not complete automatic, so EncyclopeDIA is written into snakemake workflow for future DIA data analysis.

This report also introduces several other software and some important concepts in DIA data analysis.

Methods

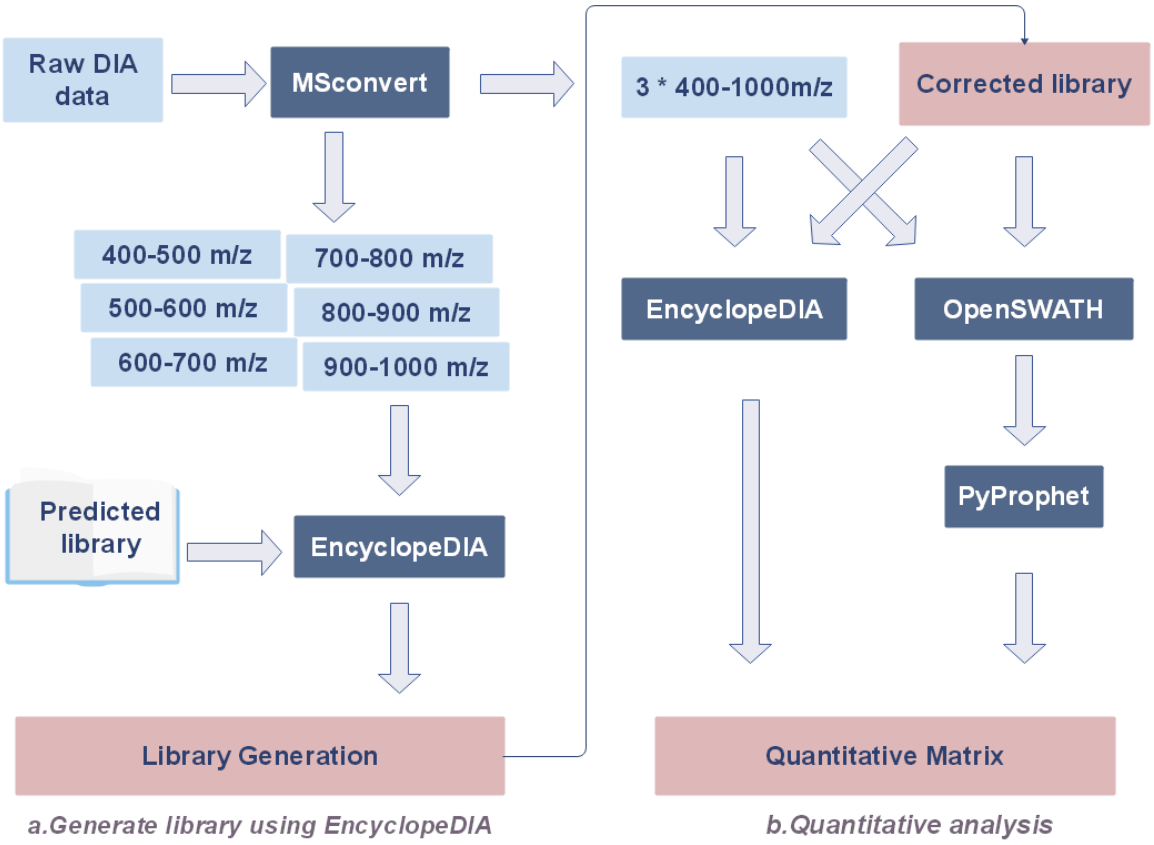

Figure 1. Workflow of EncyclopeDIA and OpenSWATH in DIA data analysis.

3 full window raw DIA data and 6 narrow window raw DIA data are collected. Figure 1. a.Generate library using EncyclopeDIA: raw data should be converted to mzML [4] format first so that EncyclopeDIA can deconvolves these narrow window data and make 6 narrow window runs have a precursor separation of 2 m/z, and then detects peptide anchors from these runs so that chromatographic data of each peptide such as retention time, peak shape , the fragment ion intensity can be stored in the chromatogram library. Figure 1. b.Quantitative analysis: this library is used to OpenSWATH workflow and EncyclopeDIA quantitative workflow. Either for OpenSWATH or EncyclopwDIA, they search iron data of 3 full window runs, match them with library and output quantitative peptides report which were filtered to a 1% False discovery rate (FDR) on peptide-level.

Data collection

Preparation of human plasma sample was performed by using standard protocol from ： "https://www.protifi.com/s-trap/". Peptides were eluted from the filter and then dried in the speed-vac. Right after peptides were C18 cleaned by using a C18 micro spin column (the

nest group, USA) and eluted in 50% ACN/5% FA. Cleaned peptides were dried again in speedvac. Peptides were diluted in 0.1% FA prior to peptides quantification by NanoDrop. A total amount of 400 ng was injected in the Q-Exactive HF-X Hybrid Quadrupole-Orbitrap DIA Mass Spectrometer.

In order to use EncyclopeDIA workflow ([7] Searle, B. C 2020) to generate library, in addition to setting a wide window of 400-1000 m/z, 6 narrow windows: 400-500 m/z, 500-600 m/z, 600-700 m/z, 700-800 m/z, 800-900 m/z, 900-1000 m/z are set.

Different format of DIA data

The first step of DIA data analysis is to convert the original data into "mzML" format. Raw DIA data is directly comes from mass spectrometers. Each of the vendors of these mass spectrometers uses a different proprietary binary output file format, which has hindered data sharing and the development of open-source software for downstream analysis [4].

The "mzML" format is based on an open XML format to encode the output files of the mass spectrometer, which can be archived and shared. The DIA analysis software are also use the DIA data in this format as input instead of raw data. Therefore, format conversion is a necessary step.

File conversion can be done by using the ProteoWizard[6] tool MSConvert. MSConvert is a command line tool for converting between various file formats, it run in a docker container.

Library generation

Not all software requires a library, Walnut can be used for searching DIA files without a library. However, library is recommended for maximum sensitivity. So, before a targeted, peptide-centric DIA analysis can be performed, a spectral library containing peptide-query parameters needs to be generated. Such a spectral library can be created either from DDA runs of the same or related samples, from the DIA runs directly.

Spectral libraries are built from fragment ion spectra that are assigned with high confidence to a peptide sequence (peptide spectrum matches, PSMs). To establish this match, fragment Ion are subjected to sequence database searching. At this stage, it is important that the protein sequence database (typically in FASTA format) contains the sequences of the retention time reference peptides to allow for retention time normalization at a later step. To control the FDR of the PSMs, the protein sequence database also needs to contain a decoy entry for every protein. Even though protein sequence reversal is the most commonly used method to generate decoy peptides, decoys most precisely reflecting target peptides are generated by pseudo-reversal of target peptide sequences (in a protein sequence the trypsin peptide sequences are reversed but keeping the last amino acid intact (K or R)) [7].
According to Brian C. Searle's workflow ([7] Searle, B. C., Swearingen, K. E., Barnes, C. A., Schmidt, T., Gessulat, S., Küster, B., & Wilhelm, M. (2020) Generating high quality libraries for DIA MS with empirically corrected peptide predictions. Nature Communications, 11(1).

https://doi.org/10.1038/s41467-020-15346-1.), I used EncyclopeDIA to correct libraries from peptide predictions. Predicted library is generated by a recently developed deep neural network, Prosit14, it generates a predicted spectrum library of fragmentation patterns and retention times for every +2H and +3H tryptic peptide in a FASTA database, with up to one missed cleavage. Fragmentation prediction in Prosit adjusts based on normalized collision energy (NCE), and the NCE parameter is tuned for each peptide charge state to account for DIA-specific fragmentation [7].

Predicted library "uniprot*human*25apr2019.fasta.z2*nce33.dlib" and protein sequence "*uniprot*human_25apr2019.fasta" are downloaded directly from "https://www.proteomicsdb.org/prosit/libraries/[15]" website. The optimal parameters such as the number of tolerated trypsin termini, missed cleavages, precursor mass tolerance and variable modifications is set to default.

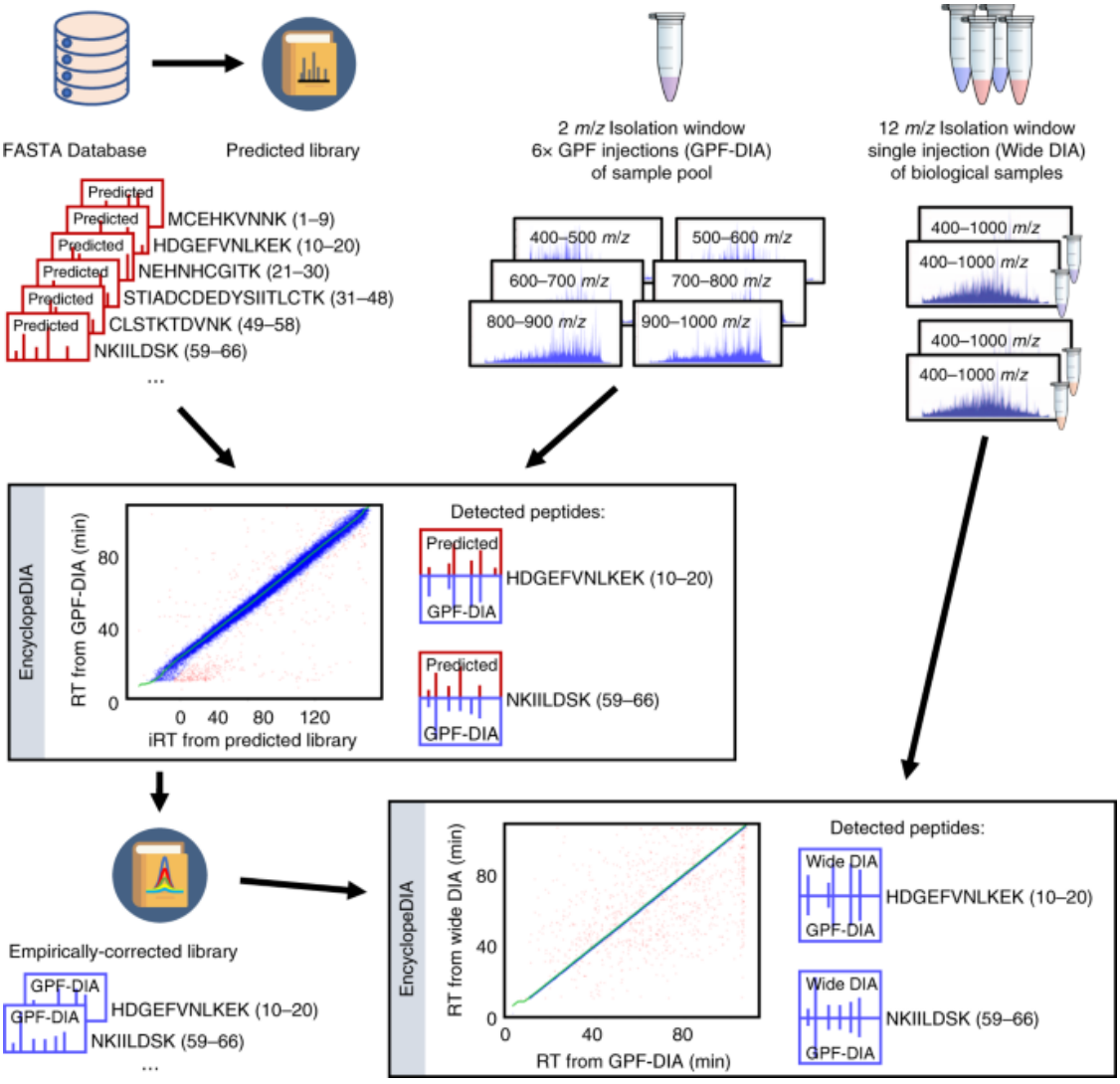

Figure 2. Workflow for generating empirically corrected libraries[7].

Fragmentation patterns and indexed retention times (iRTs) are generated with Prosit for all possible tryptic peptides in a FASTA database, and these predictions are compiled into a predicted spectrum library. EncyclopeDIA can search Gas-phase fractionation [8] (GPF)-DIA acquisitions of a sample pool with that library, and peptides detection results are compiled into an experiment-specific, empirically corrected library. This new library contains fragmentation patterns and retention times extracted from the GPF-DIA data for only the detected peptides (blue empirical spectra). Since GPF-DIA and single-injection GPF-DIA have the same instrumentation and on-column matrix, retention times and fragmentation patterns in the empirically corrected library are more closely aligned than the original predictions [7].

Use the corrected library across platforms

OpenSWATH uses corrected library generated from EncyclopeDIA, but this library file has to be converted to OpenSWATH usable format. This step is done by Encyclopedia GUI (encyclopedia-0.9.5-executable) at windows system since it's Unachievable with encyclopedia command line.

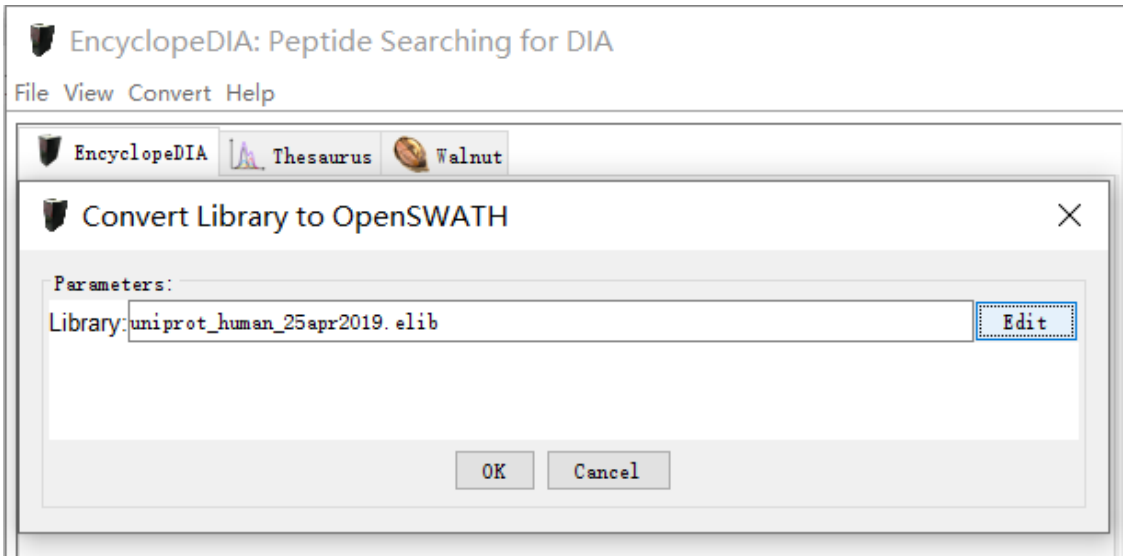

Figure 3. Converting library to OpenSWATH-usable library by using EncyclopeDIA GUI.

EncyclopeDIA

EncyclopeDIA workflow contain chromatogram library generation and searching spectrum and chromatogram libraries and output quantitative reports.

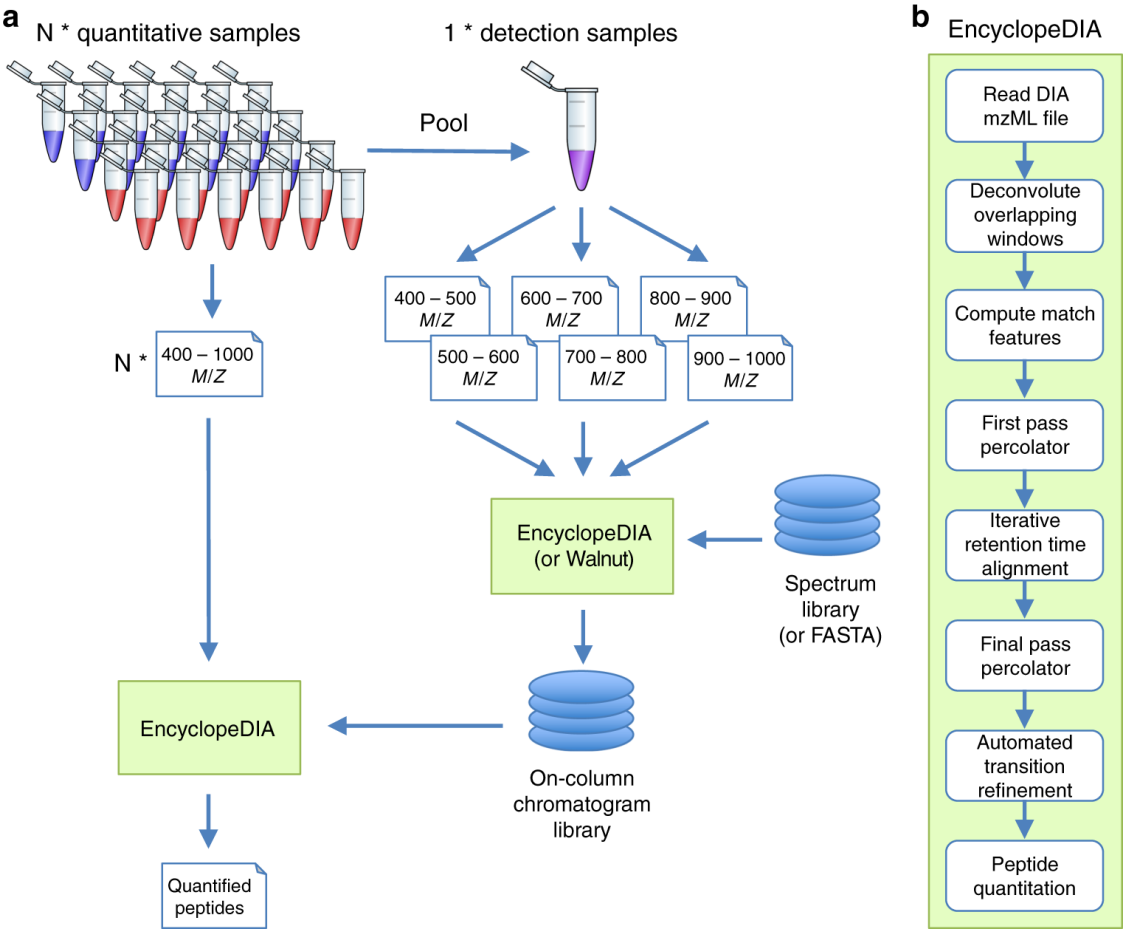

Figure 4. EncyclopeDIA workflow [9].

a. The chromatogram library generation workflow. Briefly, in addition to collecting wide-window DIA experiments on each quantitative replicate, a pool containing peptides from every condition is measured using several staggered narrow-window DIA experiments by EncyclopeDIA. After deconvolution, these narrow-window experiments have 2 m/z precursor isolation, which is analogous to targeted parallel reaction monitoring (PRM) experiments, except effectively targeting every peptide between 400 and 1000 m/z. We detect peptide anchors from these experiments using either EncyclopeDIA (searching a DDA spectrum library) or PECAN/Walnut (using a protein database) and chromatographic data about each peptide is stored in a chromatogram library with retention times, peak shape, fragment ion intensities, and known interferences tuned specifically for the LC/MS/MS setup. EncyclopeDIA then uses these precise coordinates for m/z, time, and intensity to detect peptides in the quantitative samples [9]. **b.** The EncyclopeDIA algorithmic workflow for searching spectrum and chromatogram libraries. After reading and uncurling DIA raw files, EncyclopeDIA calculates several retention time independent feature scores for each peptide that are amalgamated and FDR corrected with Percolator. Using high confidence peptide detections, EncyclopeDIA retention time aligns detections to the library, determines the retention time accuracy, and reconsiders outliers. After a second FDR correction with

Percolator, EncyclopeDIA autonomously picks fragment ion transitions that fit each non-parametrically calculated peak shape and quantifies peptides using these ions [9].

EncyclopeDIA version used for this project is “encyclopedia-0.9.5”, the "-libexport" function can be used to generate spectral library file and global quantitative results. With the "-a false" flag, the "-libexport" function can be used to generate a chromatogram library. This library file can be used for quantitative analysis of OpenSWATH; With the "-a true" flag, this command performs retention time alignment (match-between-runs) and global FDR estimation. After run the "-libexport" function with “-a true” flag, all of the results are filtered to a 1% FDR with Percolator at the peptide and protein level [16].

OpenSWATH

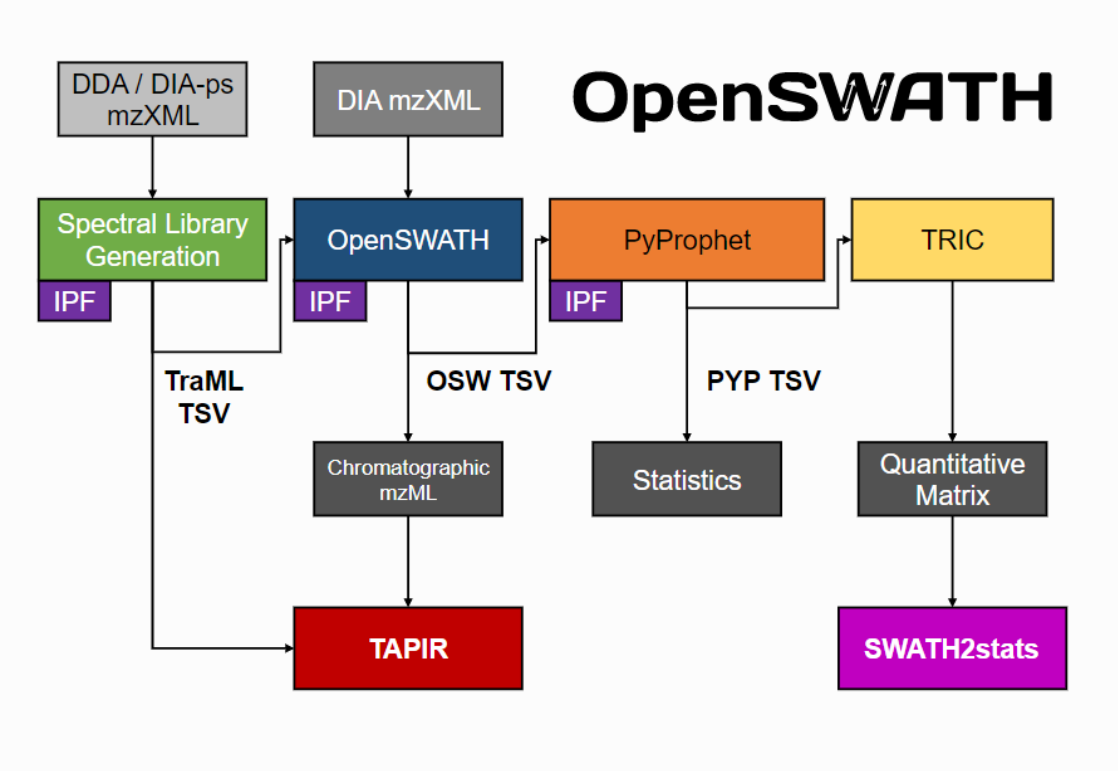

Figure 5. OpenSWATH workflow [10].

OpenSWATH allows targeted analysis of DIA data in an automated, high-throughput fashion. The OpenSWATH algorithm can be summarized in the following five steps:
1) Data conversion
OpenSWATH takes as input the acquired SWATH-MS data and an assay library. These are first converted to suitable open file formats (mzML and TraML). The assay library contains precursor- and fragment-ion m/z values (transitions) as well as relative fragment-ion intensities and normalized peptide retention times. Decoy assays are appended to the target assay library for later classification and error rate estimation [10].

2) Retention-time alignment
Each run is aligned against a previously determined normalized retention-time space using reference peptides whose mappings to the normalized space are known (for example, spiked-in peptides). Outliner detection is subsequently applied to remove wrongly assigned reference peptides and to evaluate the quality of the alignment [10].
3) Chromatogram extraction
Using the m/z and retention-time information from the assay library, the workflow extracts an ion chromatogram from the corresponding MS/MS map, producing integrated fragmentation counts versus retention-time data. The extraction function (Top-hat or Bartlett) and m/z window-width can be specified to account for the instrument-specific MS/MS resolution [10].
4) Peak-group scoring
The core algorithm identifies 'peak groups' (that is, the position where multiple fragment traces are replaced) and scores them using multiple, orthogonal scores. These scores are based on the elution profiles of the fragment ions, the correspondence of the peak group with the expected retention time and fragment-ion intensity from the assay library, as well as the properties of the full MS/MS spectrum at the chromatographic peak apex [10].
5) Statistical analysis
The separation between true and false signal is achieved using a set of decoy assays that were scored exactly the same way as the target assays. The false-discovery rate (FDR = false positives/ (true positives + false positives)) can then be estimated, for example by the PyProphet algorithm. If multiple runs are present, a peak-group alignment can be performed to annotate signals that could not be confidently assigned using data from a single run alone [10].

FDR estimation

All searches are performed using the target/decoy strategy [11]. As previously described [9], EncyclopeDIA generates decoy peptide sequences by keeping the first and last amino acids in place, but reversing the remaining in between sequence. Decoy spectra are generated by moving all fragment ions corresponding to amino acids to the mass appropriate for the new decoy sequence. Each decoy peptide retains the same retention time as the corresponding target peptide. Retention time is only used as a feature (not a filter), so every peptide (decoy or target) can be assigned at any retention time. This approach is designed to give decoys a chance to produce higher scores and better model truly incorrect peptides. EncyclopeDIA search results were filtered to a 1% peptide-level using Percolator 3.1 [12][13]. Proteins are then parsimoniously allocated to protein groups and filtered to a 1% protein-level FDR [7]. OpenSWATH search results were filtered to a 1% peptide-level using pyProphet, if peptide and protein inference in the global context was conducted.

Snakemake workflow

DIA data analysis methods are mainly divided into two types: library-based (peptide-centered) and non-library-based (spectrum-centered). Generally, library-based methods tend to be more sensitive and can produce accurate quantitative results, especially if the assay library is prepared from the same sample. Two software used in this project are both library-based and can achieve high-quality DIA data analysis. However, although EncyclopeDIA can generate both a library and quantitative reports, because the DIA data is very large, step-by-step operation is very time-consuming, for OpenSWATH, it cannot generate a library, which makes the analysis process more complicated and time-consuming. To this end, a reusable Snakemake workflow is written (Supplementary Data 1), this workflow allows users to get library for new DIA data first, then this library file is used to output quantitative results by EncyclopeDIA. With this Snakemake workflow, users can get EncyclopeDIA quantitative reports from raw DIA data with one command line, which simplifies the DIA data analysis process and realizes automated analysis.

This workflow is very simple to use. In order to execute Snakemake workflow, the installation of java, python, docker, EncyclopeDIA and snakemake is required. And the directory of the files to generate the library and to get quantitative results should be pass to the workflow as a variable value via the command line.

Snakefile can be download from https://github.com/Lina0125/DIA_workflows/blob/master/snakefile, EncyclopeDIA can be downloaded from https://bitbucket.org/searleb/encyclopedia/downloads/encyclopedia-0.9.5-executable.jar and should be moved into “bin” directory. Besides, predicted library and protein fasta file should be downloaded from https://www.proteomicsdb.org/prosit/libraries/[17], put 2 files in “ref” directory.

Set the directory as follows:

```
.
├── bin
|   └── encyclopedia-0.9.5-executable.jar
├── ref
|   ├── uniprot_human_25apr2019.fasta
|   └── uniprot_human_25apr2019.fasta.z2_nce33.dlib
└── snakefile
```

Figure 6. Setting of snakemake work directory.

Then execute snakemake on the command line:

```
$Library_dir='your path of files to generate the library' Quantitation_dir='your path of files to generate the library' snakemake
```

Result

	01 full window run	02 full window run	03 full window run
EncyclopeDIA	5412 peptides	5300 peptides	3604 peptides
OpenSWATH	4758 peptides	4748 peptides	849 peptides

Table 1. Total number of peptides detected in 3 runs by 2 software.

Peptides report either from EncyclopeDIA or OpenSWATH contains the results from three separate DIA runs of the same sample. There are 7878 peptides in library in total. In the two workflows, a total of 7463 unique peptides were detected. All peptides are from human.

Comparing the two software, EncyclopeDIA detected more peptides than OpenSWATH in same run, but OpenSWATH identified more unique peptides in 3 runs than EncyclopeDIA. Among the three independent runs, the number of peptides detected in the third run was significantly less, the relevant experimenters proved that the third sample does have quality problems. it was not injected properly. Probably a lack of material in the vial or sample was not picked up. Thus, the file size is small and the chromatogram very poor. So, the third run is excluded in comparison because of quality problem.

In order to compare the number of overlapping peptides between the two software and 2 runs more intuitively, UpSet is used to generate visualized figure.

Comparison of EncyclopeDIA and OpenSWATH peptides report

Repeatability within runs

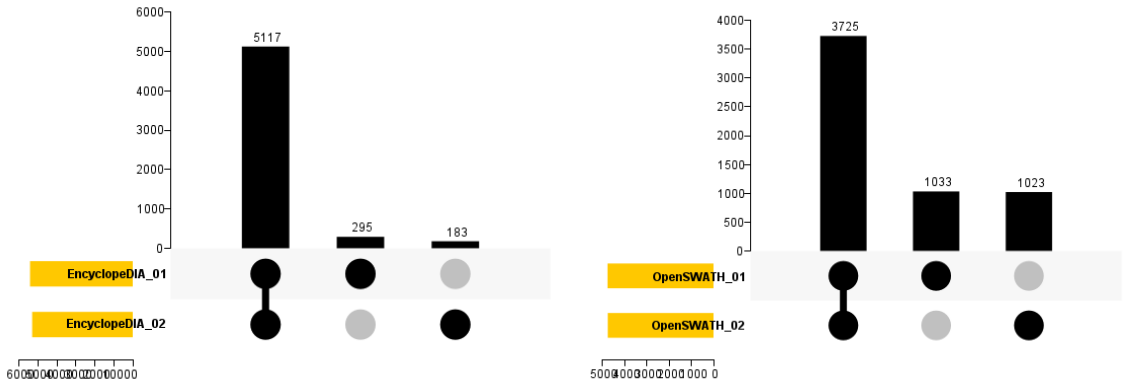

Figure 7. Overlapping peptides detected by EncyclopeDIA or OpenSWATH in 2 runs.

As we can see from *Figure 7*, 5117 out of 5595 overlapping peptides are detected by EncyclopeDIA in 2 runs with 91.45% repetition rate. 3752 out of 5781 overlapping peptides are detected by OpenSWATH in 2 runs, with 64.90% repetition rate. In total, OpenSWATH detected more peptides, while EncyclopeDIA ensures higher repetition rate.

OpenSWATH runs independently in each run, which may result in inconsistent quantification due to changes in peak boundaries, so that different peaks are selected in each run [14], which explains lower repeatability within runs.

Accuracy

Figure 8. Peptides detected by EncyclopeDIA or OpenSWATH in library.

Figure 8 shows that 5289 out of 5595 peptides detected by EncyclopeDIA are in library, the ratio is 94.5%. 4930 out of 5781 peptides detected by OpenSWATH are in library (85.27%), which indicates that OpenSWATH is not as accurate as EncyclopeDIA.

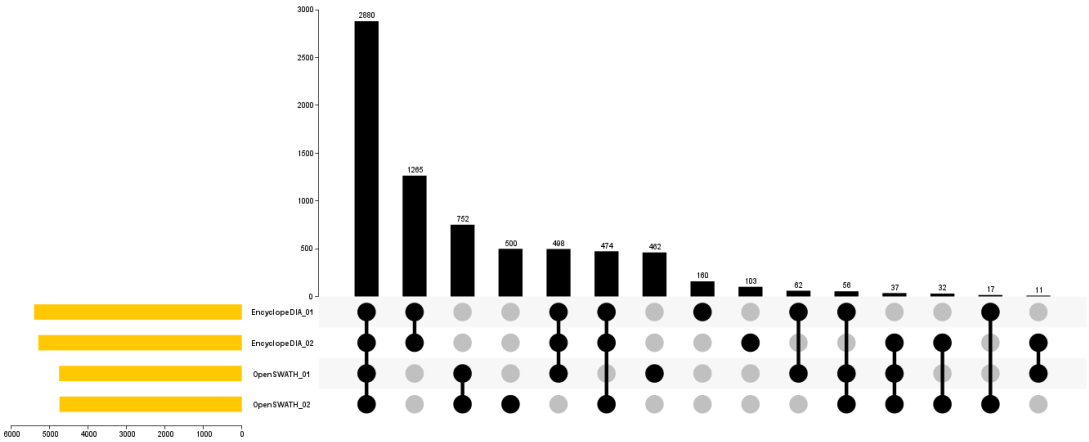

Figure 9. Overlapping peptides detected by EncyclopeDIA and OpenSWATH in 2 runs.

Figure 9 shows that 2880 out of 7309 overlapping peptides were detected by EncyclopeDIA and OpenSWATH. The second column shows that there are 1265 peptides detected by EncyclopeDIA but not detected by OpenSWATH. On the contrary, there are only 752 peptides detected by OpenSWATH but not detected by EncyclopeDIA, which indicates that EncyclopeDIA may has higher accuracy.

Discussion

As mentioned previously, there are two DIA data analysis methods: Library-Based Approaches and Spectral Library-Free Approaches (spectral-based) [5]. Although the library is not necessary, using the same sample to generate a library will greatly improve the peptide detection and protein quantification. accuracy. Both OpenSWATH and EncyclopeDIA are based on libraries.

OpenSWATH workflow requires several different commands and the process is not completely automatic. OpenSWATH executes the following steps in order: Reading of the raw input file (provided as mzML, mzXML or sqMass) and RT normalization transition list; Computing the retention time transformation using RT normalization peptides; Reading of the transition list; Extracting the specified transitions; Scoring the peak groups in the extracted ion chromatograms (XIC); Reporting the peak groups and the chromatograms. Then PyProphet conducts semi-supervised learning and error-rate estimation and exports quantitative reports.

On the contrary, EncyclopeDIA workflow only needs one command and all algorithms will be executed. Briefly, EncyclopeDIA workflow starts with reading raw MS/MS data in mzML files into an SQLite database designed for querying fragment spectra across precursor isolation windows. If fragment spectra are collected using overlapping windows, they are deconvoluted on the fly during file reading. Libraries are read as DLIB (DDA-based spectrum libraries) or ELIB (DIA-based chromatogram libraries). EncyclopeDIA determines the

highest scoring retention time point corresponding to each library spectrum (as well as a paired reverse sequence decoy) using a scoring system modeled after the X!Tandem HyperScore17. Fifteen auxiliary match features (not based on retention time) are calculated at this time point. These features are aggregated and submitted to Percolator 3.118, a semi-supervised SVM algorithm for interpreting target/decoy peptide detections, for a first pass validation. EncyclopeDIA generates a retention time model from peptides detected at 1% FDR using a non-parametric kernel density estimation algorithm that follows the density mode across time. The retention time-curated feature sets are submitted to Percolator for final pass validation at 1% peptide FDR [9].

We can infer that EncyclopeDIA is more accurate than OpenSWATH may because it generates retention time alignment mixture model, any target or decoy peptide in the feature set that does not match the retention time model is reconsidered up to five times until find a highest scoring retention time point that matches the model in EncyclopeDIA workflow [9], and it takes long time. OpenSWATH computes the retention time transformation using RT normalization peptides and extract chromatogram without generating any model, so it doesn't take long to run all the commands and get quantitative results.

In summary, the advantage of EncyclopeDIA is that it is more accurate and the peptides detected within runs have higher repeatability, the disadvantage is the long running time caused by the complexity of the algorithm. However, OpenSWATH has a shorter running time due to its simpler algorithm, but at the expense of accuracy.

Acknowledgement

I thank Fredrik Levander for his support and help to me throughout this project.

References

[1] Doerr, A. (2014). DIA mass spectrometry. In Nature Methods. https://doi.org/10.1038/nmeth.3234.

[2] Röst, H. L., Rosenberger, G., Navarro, P., Gillet, L., Miladinoviä, S. M., Schubert, O. T., … Aebersold, R. (2014). OpenSWATH enables automated, targeted analysis of data-independent acquisition MS data. Nature Biotechnology. https://doi.org/10.1038/nbt.2841.

[3] Gillet, L. C., Navarro, P., Tate, S., Röst, H., Selevsek, N., Reiter, L., … Aebersold, R. (2012). Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: A new concept for consistent and accurate proteome analysis. Molecular and Cellular Proteomics, 11(6). https://doi.org/10.1074/mcp.O111.016717.

[4] Deutsch, E. W. (2010). Mass spectrometer output file format mzML. Methods in Molecular Biology (Clifton, N.J.), 604. https://doi.org/10.1007/978-1-60761-444-9_22.

[5] Zhang, F., Ge, W., Ruan, G., Cai, X., & Guo, T. (2020). Data-Independent Acquisition Mass Spectrometry-Based Proteomics and Software Tools: A Glimpse in 2020. PROTEOMICS, 20. https://doi.org/10.1002/pmic.201900276.

[6] Kessner, D., Chambers, M., Burke, R., Agus, D., & Mallick, P. (2008). ProteoWizard: Open source software for rapid proteomics tools development. Bioinformatics, 24(21). https://doi.org/10.1093/bioinformatics/btn323.

[7] Searle, B. C., Swearingen, K. E., Barnes, C. A., Schmidt, T., Gessulat, S., Küster, B., & Wilhelm, M. (2020). Generating high quality libraries for DIA MS with empirically corrected peptide predictions. Nature Communications, 11(1). https://doi.org/10.1038/s41467-020-15346-1.

[8] Panchaud, A., Scherl, A., Shaffer, S. A., Von Haller, P. D., Kulasekara, H. D., Miller, S. I., & Goodlett, D. R. (2009). Precursor acquisition independent from ion count: How to dive deeper into the proteomics ocean. Analytical Chemistry, 81(15). https://doi.org/10.1021/ac900888s.

[9] Searle, B. C., Pino, L. K., Egertson, J. D., Ting, Y. S., Lawrence, R. T., MacLean, B. X., … MacCoss, M. J. (2018). Chromatogram libraries improve peptide detection and quantification by data independent acquisition mass spectrometry. Nature Communications, 9(1). https://doi.org/10.1038/s41467-018-07454-w.

[10] Röst, H. L., Rosenberger, G., Navarro, P., Gillet, L., Miladinoviä, S. M., Schubert, O. T., … Aebersold, R. (2014). OpenSWATH enables automated, targeted analysis of data-independent acquisition MS data. Nature Biotechnology. https://doi.org/10.1038/nbt.2841.

[11] Elias, J. E., & Gygi, S. P. (2007). Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. Nature Methods, 4(3). https://doi.org/10.1038/nmeth1019.

[12] Käll, L., Canterbury, J. D., Weston, J., Noble, W. S., & MacCoss, M. J. (2007). Semi-supervised learning for peptide identification from shotgun proteomics datasets. Nature Methods, 4(11). https://doi.org/10.1038/nmeth1113.

[13] The, M., MacCoss, M. J., Noble, W. S., & Käll, L. (2016). Fast and Accurate Protein False Discovery Rates on Large-Scale Proteomics Data Sets with Percolator 3.0. Journal of the American Society for Mass Spectrometry, 27(11). https://doi.org/10.1007/s13361-016-1460-7.

[14] Gupta, S., & Röst, H. (2020). Automated Workflow For Peptide-level Quantitation From DIA/ SWATH-MS Data. BioRxiv. https://doi.org/10.1101/2020.01.21.914788.

[15] https://www.proteomicsdb.org/prosit/libraries/ Access time: June-November,2020

[16] https://bitbucket.org/searleb/encyclopedia/wiki/Home Access time: June-November,2020

Electronic supplementary material

Supplementary Data 1