Data-Independent Acquisition Mass Spectrometry-based proteomics and Software Tools: a Glimpse in 2020

Fangfei Zhang, Weigang Ge, Guan Ruan, Xue Cai, Tiannan Guo

WILEY

Top Cited Article 2020-2021

Congratulations to:

Fangfei Zhang

whose paper has been recognized as a **top cited** paper in:

PROTEOMICS

Data-Independent Acquisition Mass Spectrometry-Based Proteomics and Software Tools: A Glimpse in 2020

WILEY

Top Downloaded Article

Congratulations

Fangfei Zhang

whose paper was one of the most downloaded* during its first 12 months of publication in:

PROTEOMICS

Data-Independent Acquisition Mass Spectrometry-Based Proteomics and Software Tools: A Glimpse in 2020

*Among work published between 1 January 2019 – 31 December 2020.

Outline

- DIA-MS data acquisition methods

- DIA software tools for DIA data analysis

- User recommendation for DIA-MS and DIA tools

- Future directions

The emergence of DIA-MS

- LC-MS/MS for holistic understanding of complex biological system

- DIA technique can acquire MS2 spectra for all precursor ions within isolation windows to obtain complete MS/MS data for full mass ranges without any prior knowledge about precursors

- Major challenge in DIA-based methods are computational deconvolution of the multiplexed fragmented spectra thus acquired.

Implementations of DIA-MS

Shotgun CID Purvine et al., 2003	**AIF** Geiger et al., 2010	**WiSIM** Koopmans et al., 2018
DIA Venable et al., 2004	**SWATH** Gillet et al., 2012	**BoxCar DIA** Meier et al., 2018
MS^E Plumb et al., 2006	**MSX** Egertson et al., 2013	**ScanningSWATH** Messner et al., 2019
PAcIFIC Panchaud et al., 2009	**SONAR** Moseley et al., 2018	**diaPASEF** Meier et al., 2019
		PulseDIA Cai et al, 2020

Legend:
- ◆ SCIEX Triple TOF 5600/6600
- ◆ SCIEX Triple TOF 6600+
- ▲ Thermo Orbitrap (QE-HF, QE-HFX, Lumos, Exploris 480)
- ▲ Thermo LTQ
- ★ Waters QTOF
- ● Bruker timsTOF Pro

Software tools for interpreting DIA data

- Spectral Library-based approaches

  - Tools for searching and building libraries

  - Tools for predicting library assays

  - Library-based scoring peptide identification

- Spectral Library-based approaches

Spectral Library Building Tools

Experimental Spectra Searching				
Comet Eng et al., 2013	X!Tandem Craig et al., 2003	Mascot Brosch et al., 2009	Sequest Eng et al., 1994	OMMSA Brewis et al., 2010
MS GF+ Kim et al., 2014	ProteinPilot Shilov et al., 2007	pFind Chi et al., 2018	MSFragger Kong et al., 2017	MaxQuant Cox et al., 2008

Library Building		
SpectraST Lam et al., 2007	Skyline MacLean et al., 2010	Pulsar Spectronaut

Library Predicting		
Deep Mass Tiwary et al., 2019	Prosit Gessulat et al., 2019	pDeep Zhou et al., 2017

DIA Analysis Tools

Library-based Analysis				
OpenSWATH Roest et al., 2014 ◆ ▲ ●	**Spectronaut** Bruderer et al., 2015 ◆ ▲ ●	**EncyclopeDIA** Searle et al., 2018 ◆ ▲	**Skyline** MacLean et al., 2010 ◆ ▲	**PeakView** Gillet et al., 2012 ◆ ◆
Spector Peckner et al., 2018 ◆ ▲	**SWATHProphet** Keller et al., 2015 ◆ ◆	**PIQED** Meyer et al., 2017 ▲	**DIA-NN** Demichev et al., 2018 ◆ ▲ ◆ ●	

Library-free Analysis				
directDIA Spectronaut ◆ ▲ ●	**Peaks** Tran et al., 2019 ▲	**DIA-Umpire** Tsou et al., 2015 ◆ ▲	**Group-DIA** Li et al., 2015 ◆ ▲	**PECAN** Ting et al., 2017 ▲

Statistical Validation		
mProphet Reiter et al., 2011	**pyProphet** Teleman et al., 2015	**Percolator** Kall et al., 2007

◆ SCIEX Triple TOF 5600/6600   ▲ Thermo Orbitrap (QE-HF, QE-HFX, Lumos, Exploris 480)

◆ SCIEX Triple TOF 6600+   ▲ Thermo LTQ   ★ Waters QTOF   ● Bruker timsTOF Pro

The schematic workflows of selected DIA analysis tools
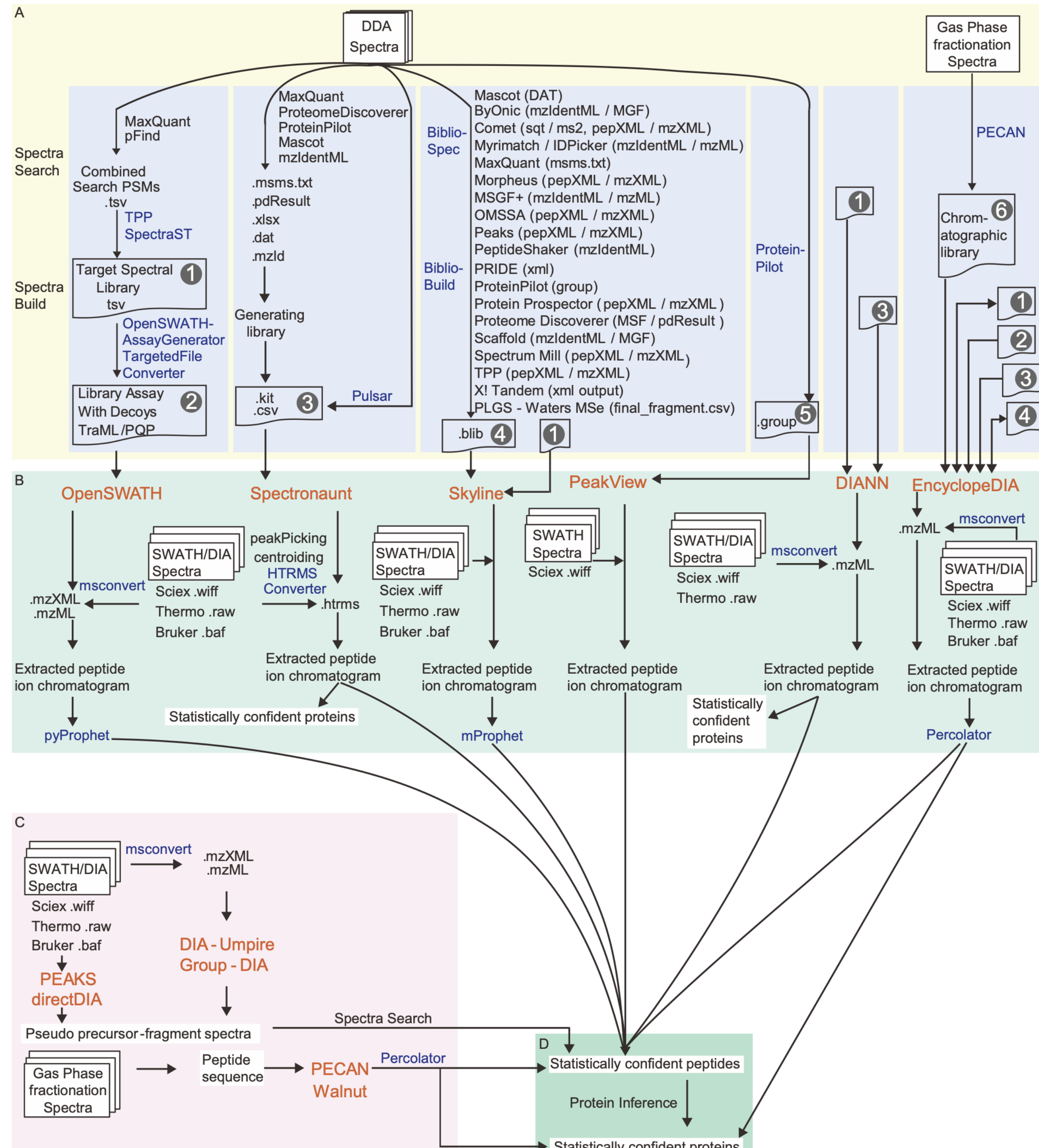
A user perspective for choosing DIA-MS and tools

DIA-MS is not the method of choice if the main goal of a study is to identify maximal number of proteins lower than 20

DIA-MS is mostly applied in single-shot analysis. Therefore, it requires only 0.1-1 microgram peptides for a proteomic analysis.

We recommend building a comprehensive sample specific library to maximize the proteome depth.

DIA-NN is a recently published tool which is packaged into a GUI as well as a command-line interface with fast running speed and deep proteomic coverage.

Outlook

Multiple issues remain to be solved:

- Inability to identify low-abundance and non-canonical or modified peptides

- Precise estimation of peptide and protein FDR

- Technical missing values in large data sets

- The integration of search results from different search engines

- Library free methods to detect novel proteoforms