FAST OPEN MODIFICATION SPECTRAL LIBRARY SEARCHING THROUGH APPROXIMATE NEAREST NEIGHBOR INDEXING

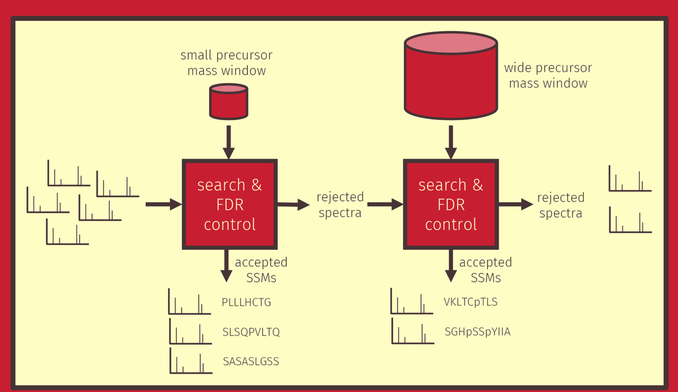
INTRODUCTION

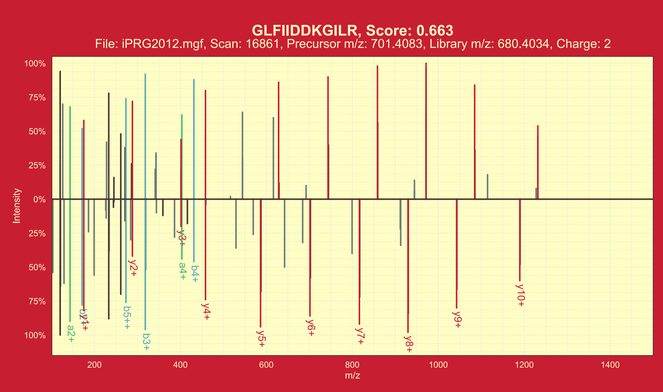
Open modification searching is a powerful strategy to identify modified spectra. By using a very wide precursor mass window modified query spectra are compared to their unmodified counterpart, implicitly considering any modification.
However, this results in a drastically **increased search space** and an excessive computational cost.

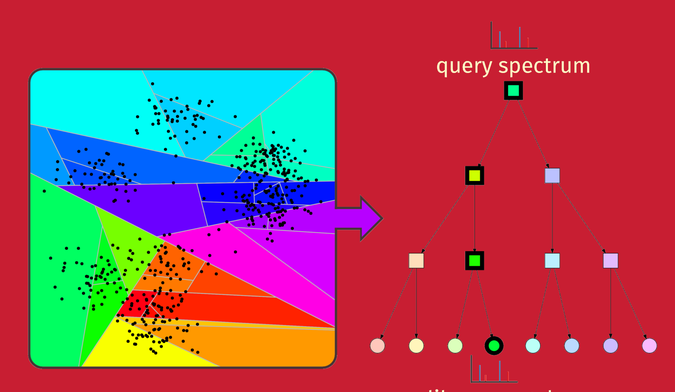
ANN-SOLO

The **ANN-SoLo** tool is optimized for fast and accurate open modification spectral library searching:

- The **cascade search strategy** maximally identifies both unmodified and modified spectra while strictly controlling the false discovery rate.
- The **shifted dot product** considers peaks that are shifted according to the precursor mass difference between two spectra to correctly identify modified spectra.
- **Approximate nearest neighbor indexing** is used to efficiently retrieve a limited number of library candidates and speed up open modification searching.





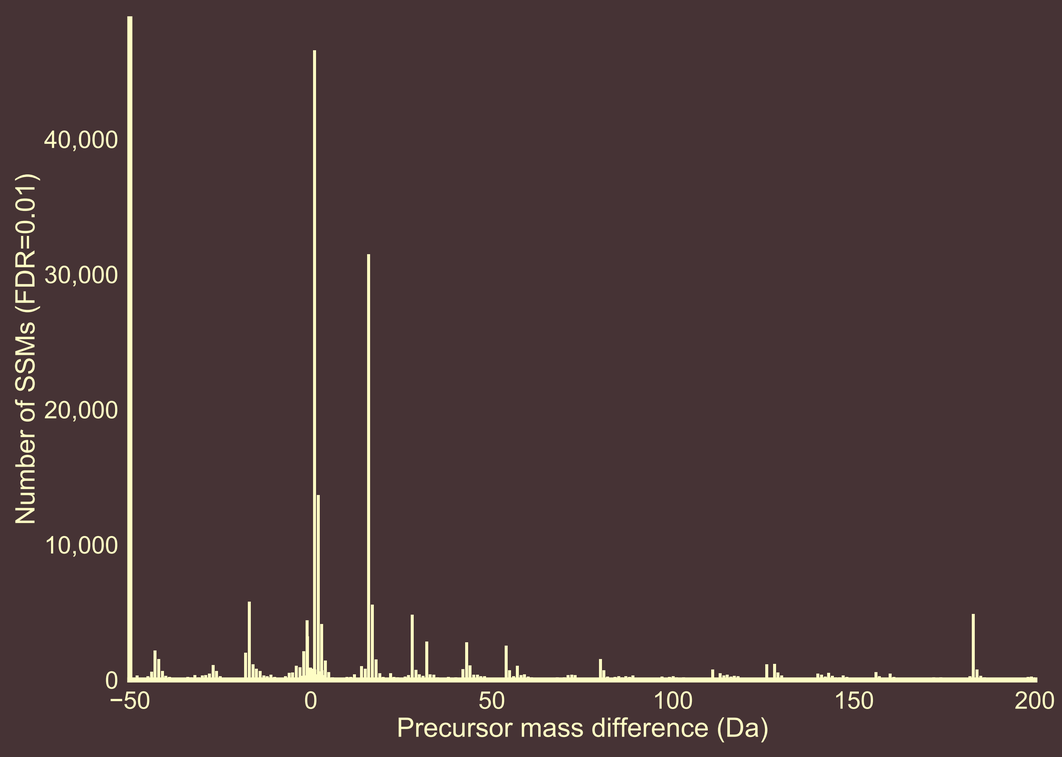


RESULTS

ANN-SoLo was evaluated against SpectraST and MSFragger:

- For open modification searching ANN-SoLo is an **order of magnitude faster** than SpectraST.
- ANN-SoLo identifies the **highest number of modified peptides**.

Search engine	Time (min)	# SSMs	# Peptides
Standard search			
MSFragger	0.7	344,998	104,672
SpectraST	5.2	369,079	102,077
ANN-SoLo	24.0	352,938	105,870
Open search			
MSFragger	34.7	526,027	126,364
SpectraST	1276.7	473,729	112,375
ANN-SoLo	108.5	647,469	153,605



Wout Bittremieux, Pieter Meysman, William Stafford Noble, Kris Laukens. *Journal of Proteome Research*, in press (2018).
10.1021/acs.jproteome.8b00359

https://github.com/bittremieux/ANN-SoLo

Wout Bittremieux[1,2*], Pieter Meysman[1], William Stafford Noble[2], Kris Laukens[1]

[1]University of Antwerp, [2]University of Washington
*wout@uw.edu