

Integration of Skyline into FragPipe for Streamlined Visualization

Skyline User Group Meeting
June 02, 2024

Fengchao Yu
Dan Polasky
University of Michigan
Ann Arbor, Michigan, United States

FragPipe

FragPipe (v22.0)

Config Workflow DIA Pseudo MS2 Database MSFragger Validation PTMs Glyco Quant (MS1) Quant (Isobaric) Spec Lib Quant (DIA) Skyline Run Downstream

Workflows

FragPipe supports multiple proteomic workflows. Select and load an option from the dropdown menu below to configure all the tools. Workflows can be customized, saved, and shared. [See the tutorial.](#)

Select a workflow: **Default** (dropdown menu) Load workflow or save current settings as workflow Save to built-in folder Save to custom folder Open built-in folder

Global settings

RAM (GB, 0=auto) (dropdown menu)

Input LC-MS Files

MS data type Regular MS IM-MS (ion mobility, timsTOF only)

Add files Add folder recursively Remove selected files Clear files Add recent F:\data\ccrcc_dda_phospho\mzml\01CPTAC_CCRCC_Phosphoproteome_JHU_20171106

Save as manifest Load manifest

Assign files to Experiments/Bioreplicates:

Set experiments Consecutive By parent directory By file name Custom Clear Set DDA Set DDA+ Set DIA

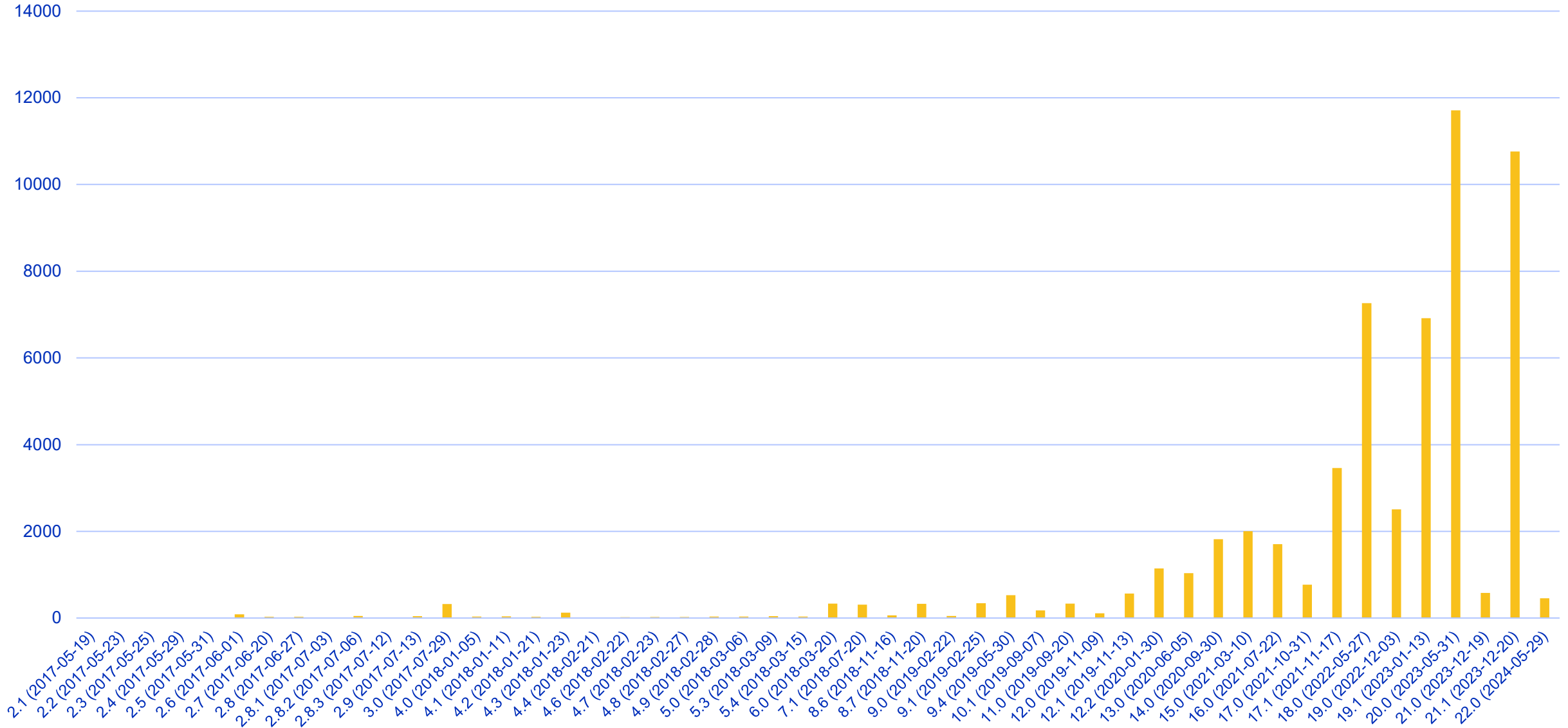
Set bioreplicates Consecutive Custom Clear Set DIA-Quant Set DIA-Lib

Total raw files: 39, Selected raw files: 0

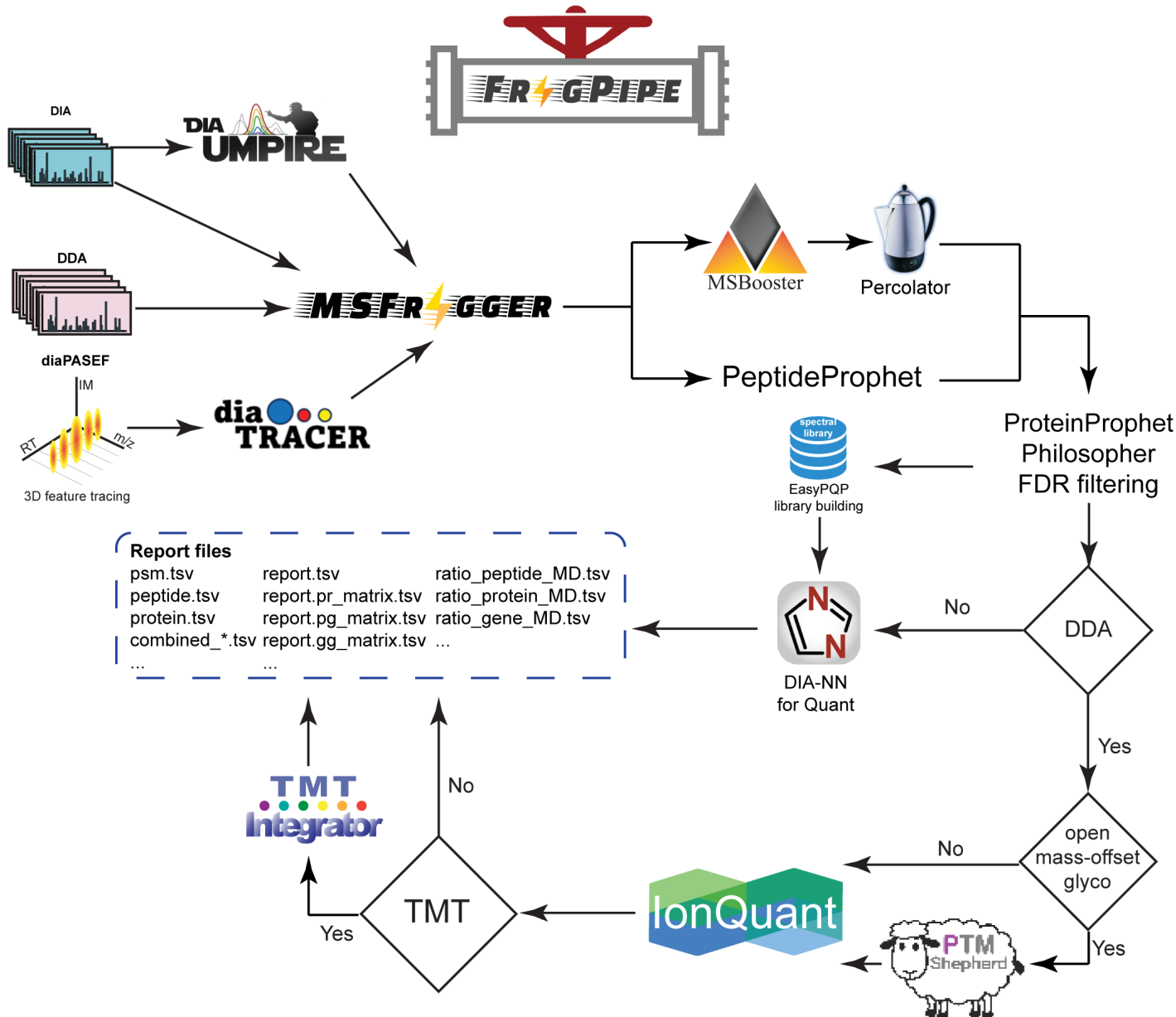
Path (can drag & drop from Explorer)	Experiment (can be empty, alp...	Bioreplicate (can be empty a...	Data type (DDA, DDA+, DIA,...
F:\data\ccrcc_dda_phospho\mzml\01CPTAC_CCRCC_Phosphoproteome_JHU_20171106\01CPTAC_CCRCC_P_JHU_20171106_LUMOS_f01.mzML	01CPTAC_CCRCC_Phosphoprote...		DDA
F:\data\ccrcc_dda_phospho\mzml\01CPTAC_CCRCC_Phosphoproteome_JHU_20171106\01CPTAC_CCRCC_P_JHU_20171106_LUMOS_f02.mzML	01CPTAC_CCRCC_Phosphoprote...		DDA
F:\data\ccrcc_dda_phospho\mzml\01CPTAC_CCRCC_Phosphoproteome_JHU_20171106\01CPTAC_CCRCC_P_JHU_20171106_LUMOS_f03.mzML	01CPTAC_CCRCC_Phosphoprote...		DDA
F:\data\ccrcc_dda_phospho\mzml\01CPTAC_CCRCC_Phosphoproteome_JHU_20171106\01CPTAC_CCRCC_P_JHU_20171106_LUMOS_f04.mzML	01CPTAC_CCRCC_Phosphoprote...		DDA
F:\data\ccrcc_dda_phospho\mzml\01CPTAC_CCRCC_Phosphoproteome_JHU_20171106\01CPTAC_CCRCC_P_JHU_20171106_LUMOS_f05.mzML	01CPTAC_CCRCC_Phosphoprote...		DDA
F:\data\ccrcc_dda_phospho\mzml\01CPTAC_CCRCC_Phosphoproteome_JHU_20171106\01CPTAC_CCRCC_P_JHU_20171106_LUMOS_f06.mzML	01CPTAC_CCRCC_Phosphoprote...		DDA
F:\data\ccrcc_dda_phospho\mzml\01CPTAC_CCRCC_Phosphoproteome_JHU_20171106\01CPTAC_CCRCC_P_JHU_20171106_LUMOS_f07.mzML	01CPTAC_CCRCC_Phosphoprote...		DDA
F:\data\ccrcc_dda_phospho\mzml\01CPTAC_CCRCC_Phosphoproteome_JHU_20171106\01CPTAC_CCRCC_P_JHU_20171106_LUMOS_f08.mzML	01CPTAC_CCRCC_Phosphoprote...		DDA
F:\data\ccrcc_dda_phospho\mzml\01CPTAC_CCRCC_Phosphoproteome_JHU_20171106\01CPTAC_CCRCC_P_JHU_20171106_LUMOS_f09.mzML	01CPTAC_CCRCC_Phosphoprote...		DDA
F:\data\ccrcc_dda_phospho\mzml\01CPTAC_CCRCC_Phosphoproteome_JHU_20171106\01CPTAC_CCRCC_P_JHU_20171106_LUMOS_f10.mzML	01CPTAC_CCRCC_Phosphoprote...		DDA
F:\data\ccrcc_dda_phospho\mzml\01CPTAC_CCRCC_Phosphoproteome_JHU_20171106\01CPTAC_CCRCC_P_JHU_20171106_LUMOS_f11.mzML	01CPTAC_CCRCC_Phosphoprote...		DDA
F:\data\ccrcc_dda_phospho\mzml\01CPTAC_CCRCC_Phosphoproteome_JHU_20171106\01CPTAC_CCRCC_P_JHU_20171106_LUMOS_f12.mzML	01CPTAC_CCRCC_Phosphoprote...		DDA
F:\data\ccrcc_dda_phospho\mzml\01CPTAC_CCRCC_Phosphoproteome_JHU_20171106\01CPTAC_CCRCC_P_JHU_20171106_LUMOS_fa.mzML	01CPTAC_CCRCC_Phosphoprote...		DDA
F:\data\ccrcc_dda_phospho\mzml\02CPTAC_CCRCC_Phosphoproteome_JHU_20171108\02CPTAC_CCRCC_P_JHU_20171108_LUMOS_f01.mzML	02CPTAC_CCRCC_Phosphoprote...		DDA
F:\data\ccrcc_dda_phospho\mzml\02CPTAC_CCRCC_Phosphoproteome_JHU_20171108\02CPTAC_CCRCC_P_JHU_20171108_LUMOS_f02.mzML	02CPTAC_CCRCC_Phosphoprote...		DDA
F:\data\ccrcc_dda_phospho\mzml\02CPTAC_CCRCC_Phosphoproteome_JHU_20171108\02CPTAC_CCRCC_P_JHU_20171108_LUMOS_f03.mzML	02CPTAC_CCRCC_Phosphoprote...		DDA
F:\data\ccrcc_dda_phospho\mzml\02CPTAC_CCRCC_Phosphoproteome_JHU_20171108\02CPTAC_CCRCC_P_JHU_20171108_LUMOS_f04.mzML	02CPTAC_CCRCC_Phosphoprote...		DDA

FragPipe

FragPipe downloads (56423 in total, 2024-05-31)



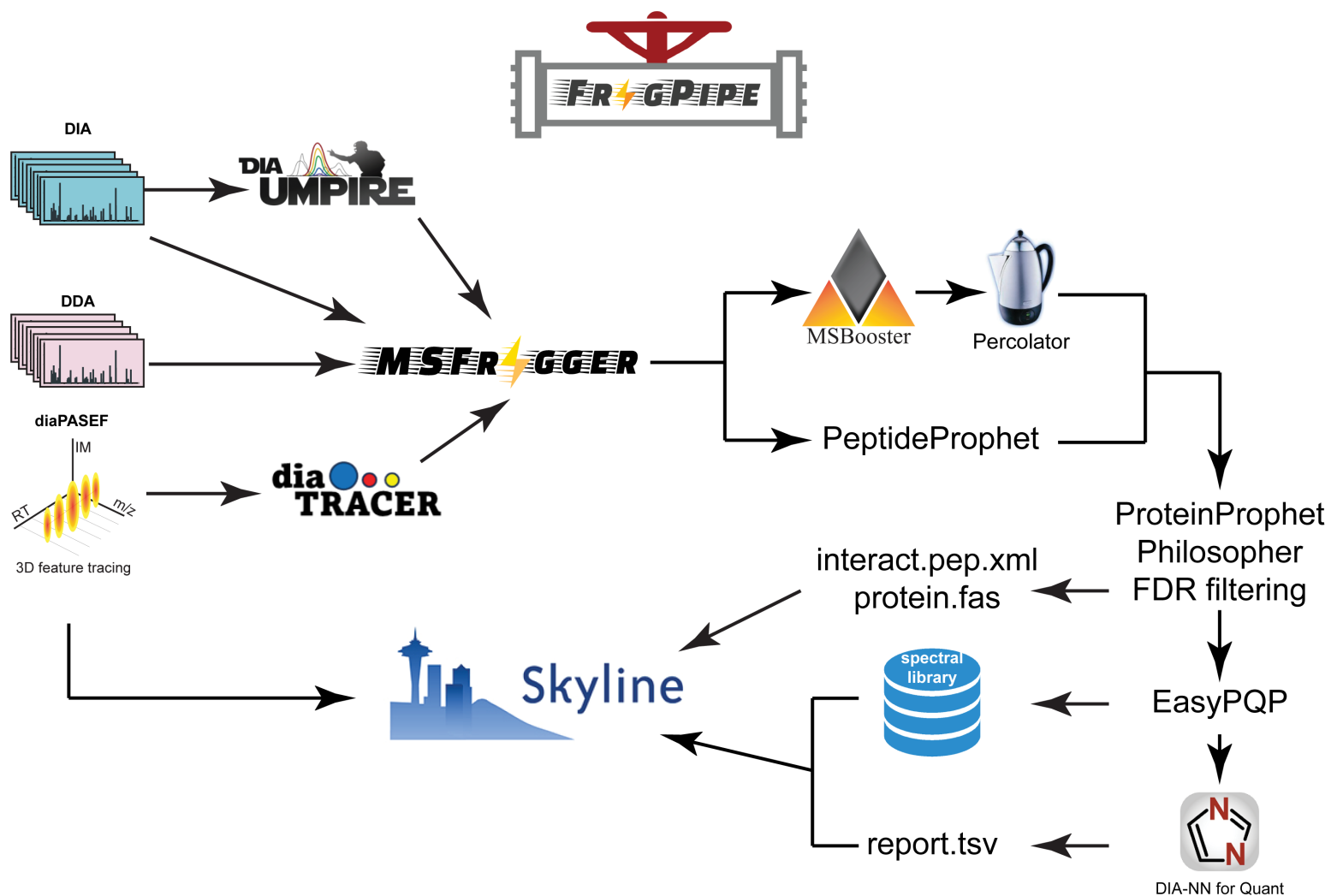
FragPipe



- DDA, DIA, ddaPASEF, and diaPASEF
- Conventional and open/mass-offset PTM searches
- Peptidomics
- Deep-learning prediction and rescoring
- Label-free quantification (LFQ-MBR and DIA)
- Isotopic- and isobaric-labeled quantifications
- Chemical proteomics (ABPP, PAL), FPOP
- Glycopeptide and labile PTM identification and quantification
- FragPipe has GUI and headless modes to run on desktops, clusters, or on the cloud

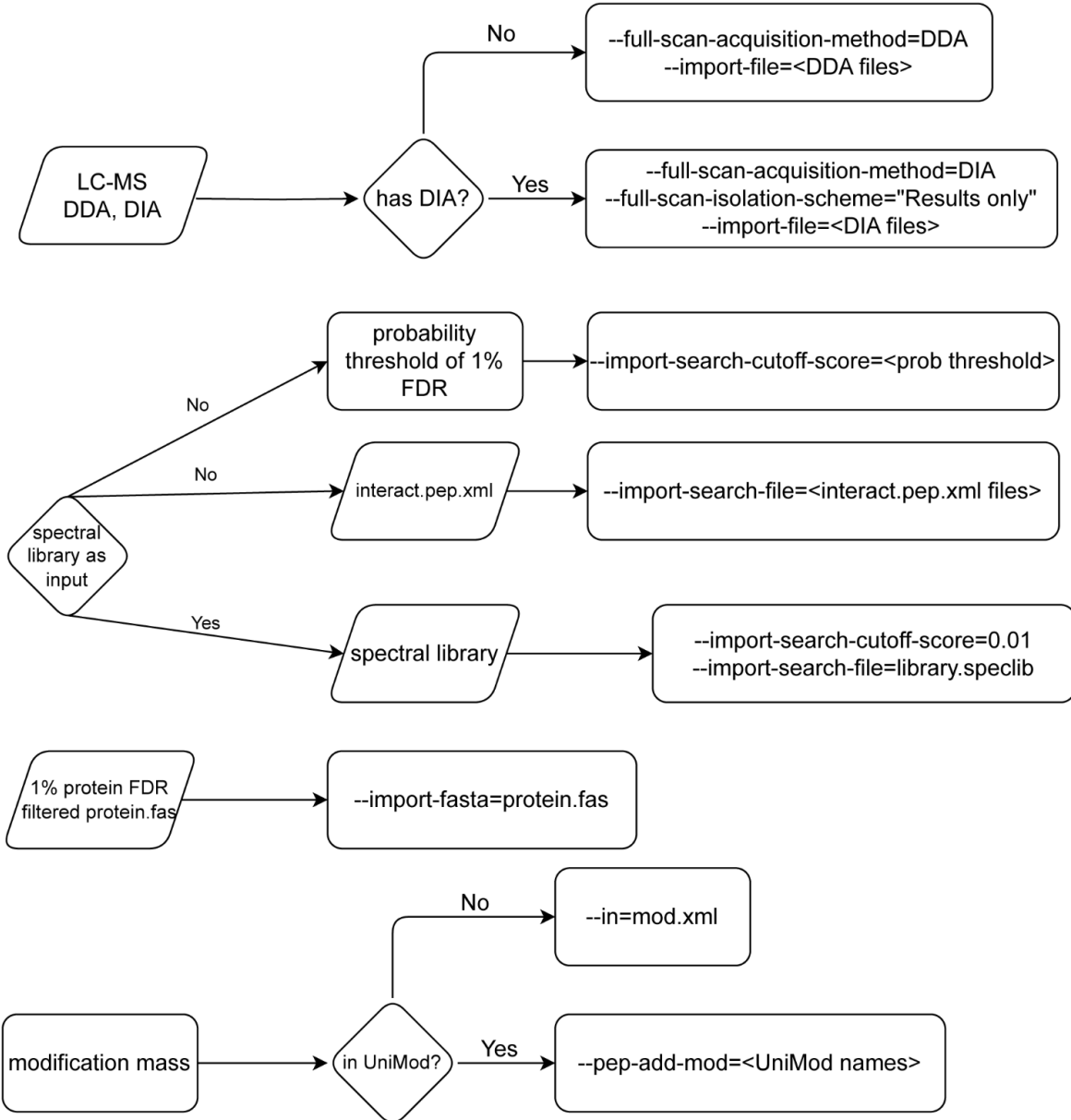
Skyline Integrated into FragPipe

ASMS poster: WP 428 Integration of Skyline and PDV into FragPipe for Advanced Visualization



- Skyline has been fully integrated into FragPipe
- It supports analysis including DDA, DIA, ddaPAEF, diaPASEF, N-Glyco, O-Glyco, open and mass-offset searches
- It runs seamlessly together with other modules in FragPipe
- It takes either pep.xml files or spectral library as result file input
- After FragPipe finishes, Skyline document is generated.

Skyline Integrated into FragPipe



ASMS poster: WP 428 Integration of Skyline and PDV into FragPipe for Advanced Visualization

- FragPipe extracts the parameters and converts them to arguments it passes to the Skyline command-line interface
- All steps are fully automated and processing is recorded to a log file
- Based on the configurations, FragPipe uses either interact.pep.xml or spectral library as Skyline input
- FragPipe either converts the modifications to UniMod names or writes non-UniMod ones to mod.xml
- All these steps happen under the hood

Skyline Integrated into FragPipe

- Load the workflow, specify the LC-MS files, and adjust the essential parameters
- Go to the “Skyline” tab and enable it
- Go to the “Run” tab, specify the output folder, and click “RUN”

The screenshot shows the 'Workflows' tab in FragPipe. It contains instructions on how to load and save workflows, and a table for assigning files to experiments and bioreplicates.

Workflows

FragPipe supports multiple proteomic workflows. Select and load an option from the dropdown menu below to configure all the tools. Workflows can be customized, saved, and shared. [See the tutorial.](#)

Select a workflow: Load workflow or save current settings as workflow Save to built-in folder Save to custom folder Open built-in folder

Complete workflow for DIA analysis using spectral library building and quantification using DIA-NN.

Spectral library can be built from DDA data and/or DIA data, with direct identification from DIA data using MSFragger-DIA. All data are processed together using MSFragger (in DDA or DIA mode, respectively) - MSBooster - Percolator - ProteinProphet (Philosopher) - EasyPQP. The library is filtered to 1% FDR at the protein and peptide levels. DIA files annotated as 'DIA-Quant' are used for quantification only (i.e. not used for spectral library building). If building the library from fractionated DDA data only, in EasyPQP tab change RT Calibration option to "cRT", or "IRT" if using organisms other than yeast or human (requires IRT peptides).

Supports DDA RAW/.d files, mzML, and MGF files. DIA files with staggered windows should be mzML files demultiplexed using Proteowizard. For quantification using DIA-NN, Thermo/Sciex DIA files should be in mzML format, and .d for Bruker's timsTOF. diaPASEF DIA data will only be used for quantification (i.e. automatically considered as 'DIA-Quant' data type).

Global settings

RAM (GB, 0=auto) Parallelism

Input LC-MS Files

MS data type Regular MS IM-MS (ion mobility, timsTOF only)

Add files Add folder recursively Remove selected files Clear files Add recent F:\data\ccrcc_dia_discovery_proteome_20\mzml

Save as manifest Load manifest

Assign files to Experiments/Bioreplicates:

Set experiments Consecutive By parent directory By file name Custom Clear Set DDA Set DDA+ Set DIA

Set bioreplicates Consecutive Custom Clear Set DIA-Quant Set DIA-Lib

Total raw files: 2, Selected raw files: 0

Path (can drag & drop from Explorer)	Experiment (can be empty, alphanumeric...)	Bioreplicate (can be empty and intege...	Data type (DDA, DDA+, DIA, DIA-Qua...
F:\data\ccrcc_dia_discovery_proteome_20\mzml\CPTAC_CCRCC_W_JHU_20190112_LUMOS_C3L-00004_NAT.mzML			DIA
F:\data\ccrcc_dia_discovery_proteome_20\mzml\CPTAC_CCRCC_W_JHU_20190112_LUMOS_C3L-00004_T.mzML			DIA

The screenshot shows the 'Skyline' tab in FragPipe. It features a 'Generate Skyline Document' button and options for Skyline running mode and special modifications mode.

Skyline

Generate Skyline Document

Skyline Skyline Daily Custom Browse

Skyline running mode Use speclib as input

Special Modifications Mode Default

Skyline Integrated into FragPipe

- Load the workflow, specify the LC-MS files, and adjust the essential parameters
- Go to the “Skyline” tab and enable it
- Go to the “Run” tab, specify the output folder, and click “RUN”

```
cmd: [START], Work dir: [G:\test_Skyline_Fragpipe\yufe_dia]
cmd: [CheckCentroid], Work dir: [G:\test_Skyline_Fragpipe\yufe_dia]
cmd: [WorkspaceCleanInit], Work dir: [G:\test_Skyline_Fragpipe\yufe_dia]
cmd: [MSFragger], Work dir: [G:\test_Skyline_Fragpipe\yufe_dia]
cmd: [MSBooster], Work dir: [G:\test_Skyline_Fragpipe\yufe_dia]
cmd: [Percolator], Work dir: [G:\test_Skyline_Fragpipe\yufe_dia]
cmd: [ProteinProphet], Work dir: [G:\test_Skyline_Fragpipe\yufe_dia]
cmd: [PhilosopherDbAnnotate], Work dir: [G:\test_Skyline_Fragpipe\yufe_dia]
cmd: [PhilosopherFilter], Work dir: [G:\test_Skyline_Fragpipe\yufe_dia]
cmd: [PhilosopherReport], Work dir: [G:\test_Skyline_Fragpipe\yufe_dia]
cmd: [WorkspaceClean], Work dir: [G:\test_Skyline_Fragpipe\yufe_dia]
cmd: [SpecLibGen], Work dir: [G:\test_Skyline_Fragpipe\yufe_dia]
cmd: [DIA-NN], Work dir: [G:\test_Skyline_Fragpipe\yufe_dia]
cmd: [Skyline], Work dir: [G:\test_Skyline_Fragpipe\yufe_dia]

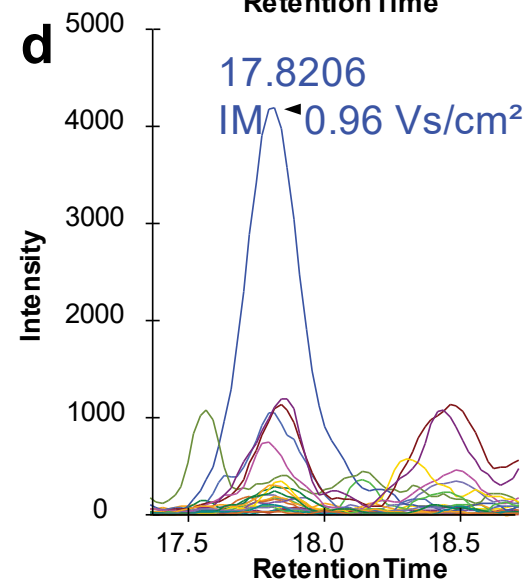
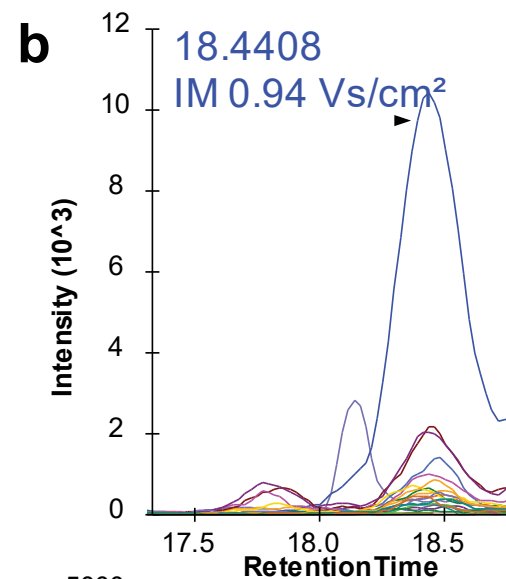
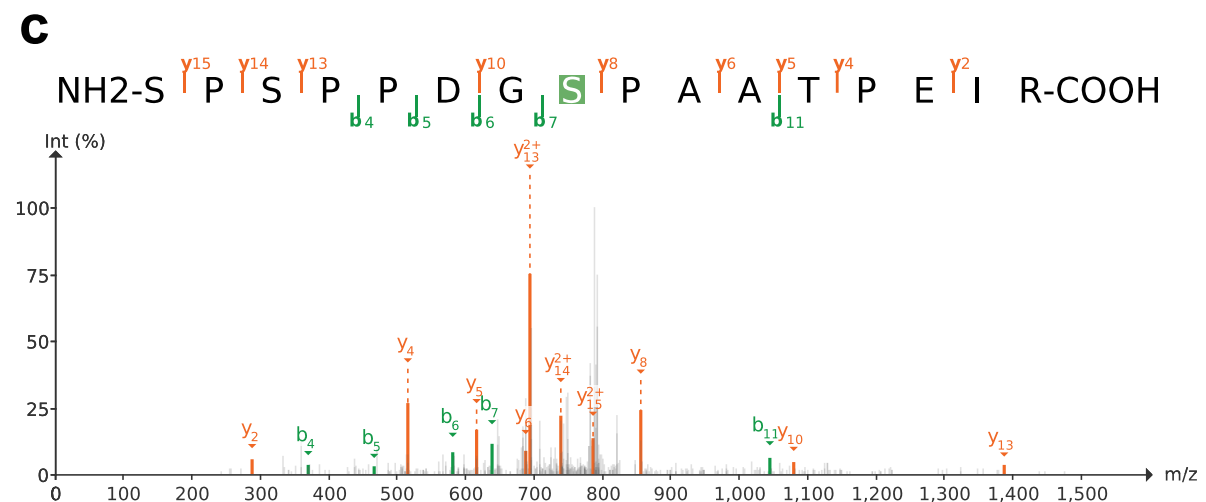
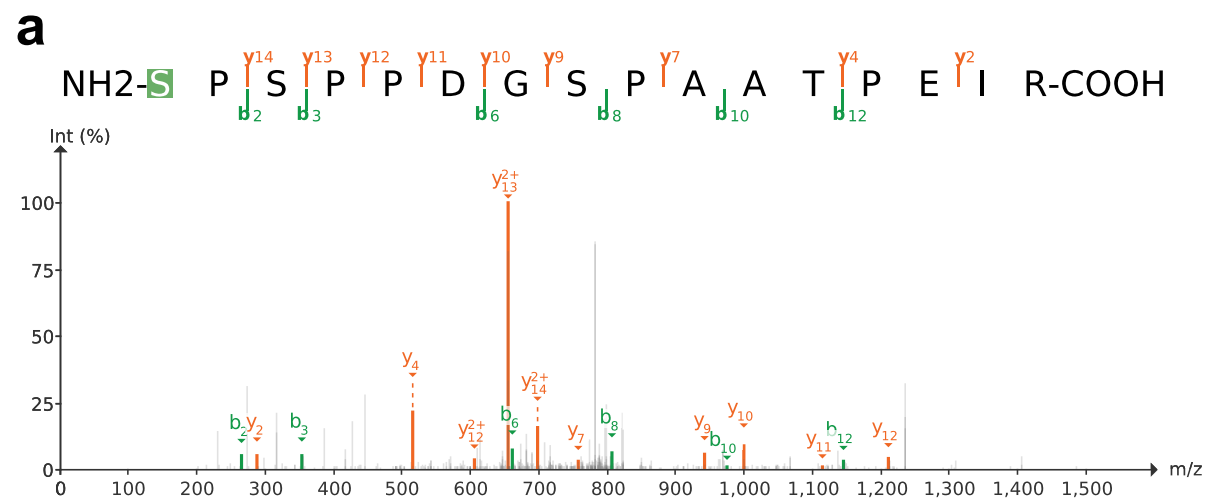
-----Sample of G:\dev\2021-03-23-decoys-reviewed-contam-UP000005640.fas-----
>BiologySIIRT-RLT_wr_fusion
>rev_sp|P06023|DHX16_HUMAN Pre-mRNA-splicing factor ATP-dependent RNA helicase DHX16 OS=Homo sapiens OX=9606 GN=DHX16 PE=1 SV=2
>rev_sp|P05053|GARG1_HUMAN c antigen 4 OS=Homo sapiens OX=9606 GN=GARG1 PE=5 SV=1
>rev_sp|P09093|AT1A2_HUMAN Sodium/potassium-transporting ATPase subunit alpha-2 OS=Homo sapiens OX=9606 GN=AT1A2 PE=1 SV=1
```

```
[2024/05/31 16:02:10] [1] 50% [2] 50%
[2024/05/31 16:02:02] [1] 52% [2] 53%
[2024/05/31 16:02:04] [1] 55% [2] 55%
[2024/05/31 16:02:07] [1] 57% [2] 59%
[2024/05/31 16:02:09] [1] 61% [2] 61%
[2024/05/31 16:02:12] [1] 64% [2] 64%
[2024/05/31 16:02:14] [1] 66% [2] 67%
[2024/05/31 16:02:16] [1] 69% [2] 70%
[2024/05/31 16:02:18] [1] 72% [2] 73%
[2024/05/31 16:02:21] [1] 76% [2] 74%
[2024/05/31 16:02:24] [1] 79% [2] 79%
[2024/05/31 16:02:27] [1] 83% [2] 82%
[2024/05/31 16:02:29] [1] 86% [2] 86%
[2024/05/31 16:02:31] [1] 90% [2] 88%
[2024/05/31 16:02:33] [1] 92% [2] 92%
[2024/05/31 16:02:38] [1] 95%
[2024/05/31 16:02:38]
[2024/05/31 16:02:41] Joining file G:\test_Skyline_Fragpipe\yufe_dia\skyline-output\CPTAC_OCCRC_W_JHU_20190112_LUMOS_C3L-00004_NAT_2117407036.mzML.skyd
[2024/05/31 16:02:43] 59% - updating peak statistics
[2024/05/31 16:02:45] 100%
[2024/05/31 16:02:45] saving file...
[2024/05/31 16:02:47] fragpipe.sky
[2024/05/31 16:02:49] 62%
[2024/05/31 16:02:51] 96%
[2024/05/31 16:02:51] 100%
[2024/05/31 16:02:51] File fragpipe.sky saved.
DONE! The Skyline files locate in G:\test_Skyline_Fragpipe\yufe_dia
Process 'Skyline' finished, exit code: 0

Please cite:
(Any searches) MSFragger: ultrafast and comprehensive peptide identification in mass spectrometry-based proteomics. Nat Methods 14:513 (2017)
(Any searches) Fast deisotoping algorithm and its implementation in the MSFragger search engine. J. Proteome Res. 20:498 (2021)
(timsTOF ddaPASEF) Fast quantitative analysis of timsTOF PASEF data with MSFragger and IonQuant. Mol Cell Proteomics 19:1575 (2020)
(Open search) Identification of modified peptides using localization-aware open search. Nat Commun. 11:4065 (2020)
(Glyco search) Fast and comprehensive N- and O-glycoproteomics analysis with MSFragger-Glyco. Nat Methods 17:1125 (2020)
(Labile search) MSFragger-Labile: A Flexible Method to Improve Labile PTM Analysis in Proteomics. Mol Cell Proteomics 22:100538 (2023)
(MSBooster) MSBooster: improving peptide identification rates using deep learning-based features. Nat Commun. 14:4539 (2023)
(PSM validation with Percolator) Semi-supervised learning for peptide identification from shotgun proteomics datasets. Nat Methods 4:923 (2007)
(Protein inference with ProteinProphet) A statistical model for identifying proteins by tandem mass spectrometry. Anal. Chem. 75:4646 (2003)
(FDR filtering and reporting) Philosopher: a versatile toolkit for shotgun proteomics data analysis. Nat Methods 17:869 (2020)
(Spectral library generation and DIA analysis) Analysis of DIA proteomics data using MSFragger-DIA and FragPipe computational platform. Nat Commun. 14:4154 (2023)
(DIA quantification with DIA-NN) dia-PASEF data analysis using FragPipe and DIA-NN for deep proteomics of low sample amounts. Nat Commun. 13:3944 (2022)
(Visualization with FragPipe-PDV) PDV: an integrative proteomics data viewer. Bioinformatics. 35(7):1249 (2019)
(Visualization with Skyline) Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. Bioinformatics. 26(7):966 (2010)

-----ALL JOBS DONE IN 22.0 MINUTES-----
```


Skyline Integrated into FragPipe



- Phospho-enriched diaPASEF data
- Pseudo-MS/MS was generated by diaTracer
- The XIC was extracted from the original diaPASEF data by Skyline

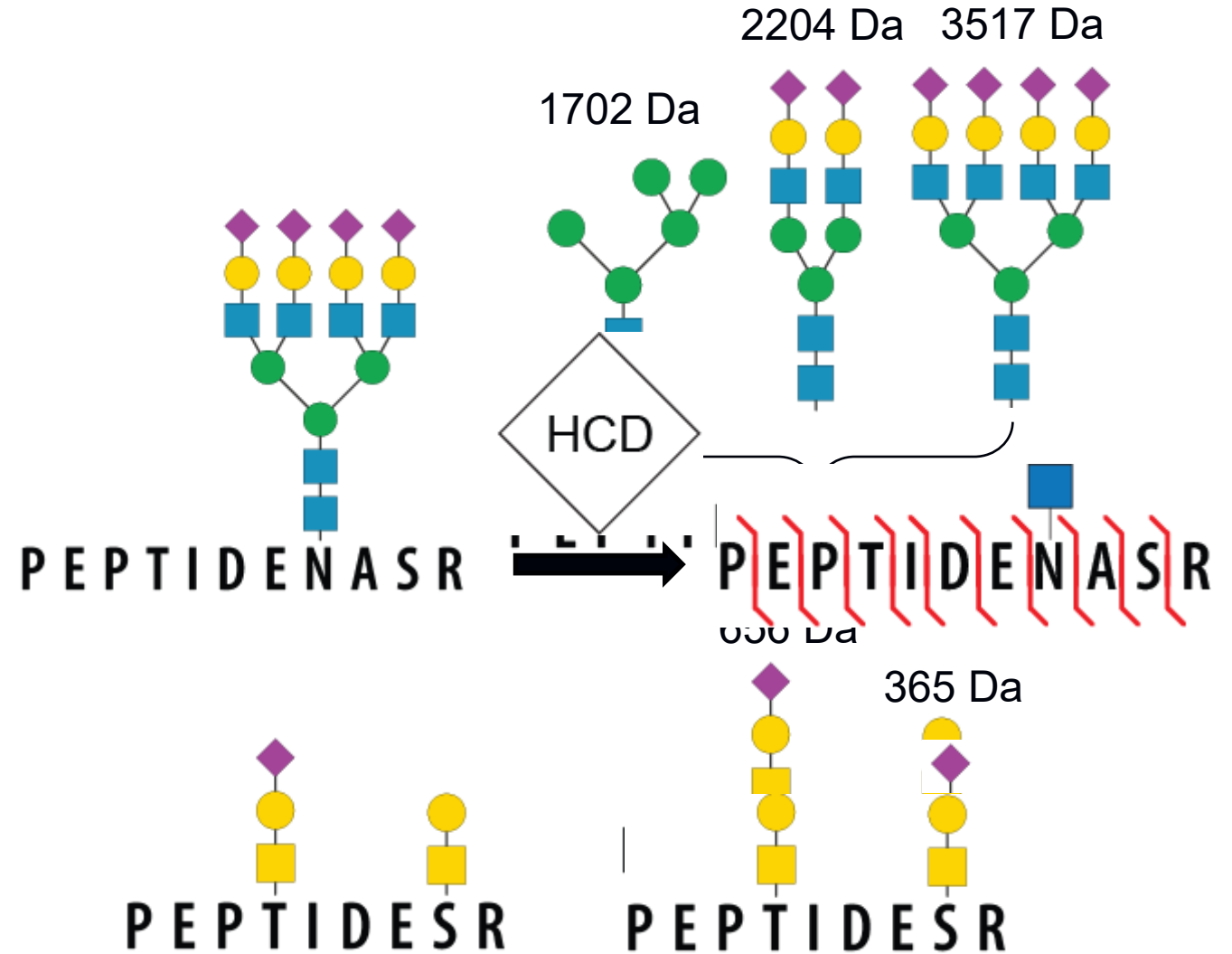
Skyline with FragPipe Example: Glycoproteomics

Dan Polasky



Glycoproteomics provides many challenges

- Many modifications
 - Can search for 100s or 1000s of glycans
 - Many glycans are not in Unimod
- Glycans are labile in tandem MS
 - Fragment ions require special handling
- Many similar glycoforms to untangle
 - Positional isomers
 - Isobaric compositions
 - Structural isomers



Adapting FragPipe and Skyline for Glycopeptides

- Modifications not in Unimod

- Write custom mod definitions file (mod.xml)
- Automatically generated from search settings
- Non-glycan mods still matched to Unimod

- Glycan fragmentation

- Encode custom neutral losses for each mod
- Override Unimod definitions

- Loading many modifications

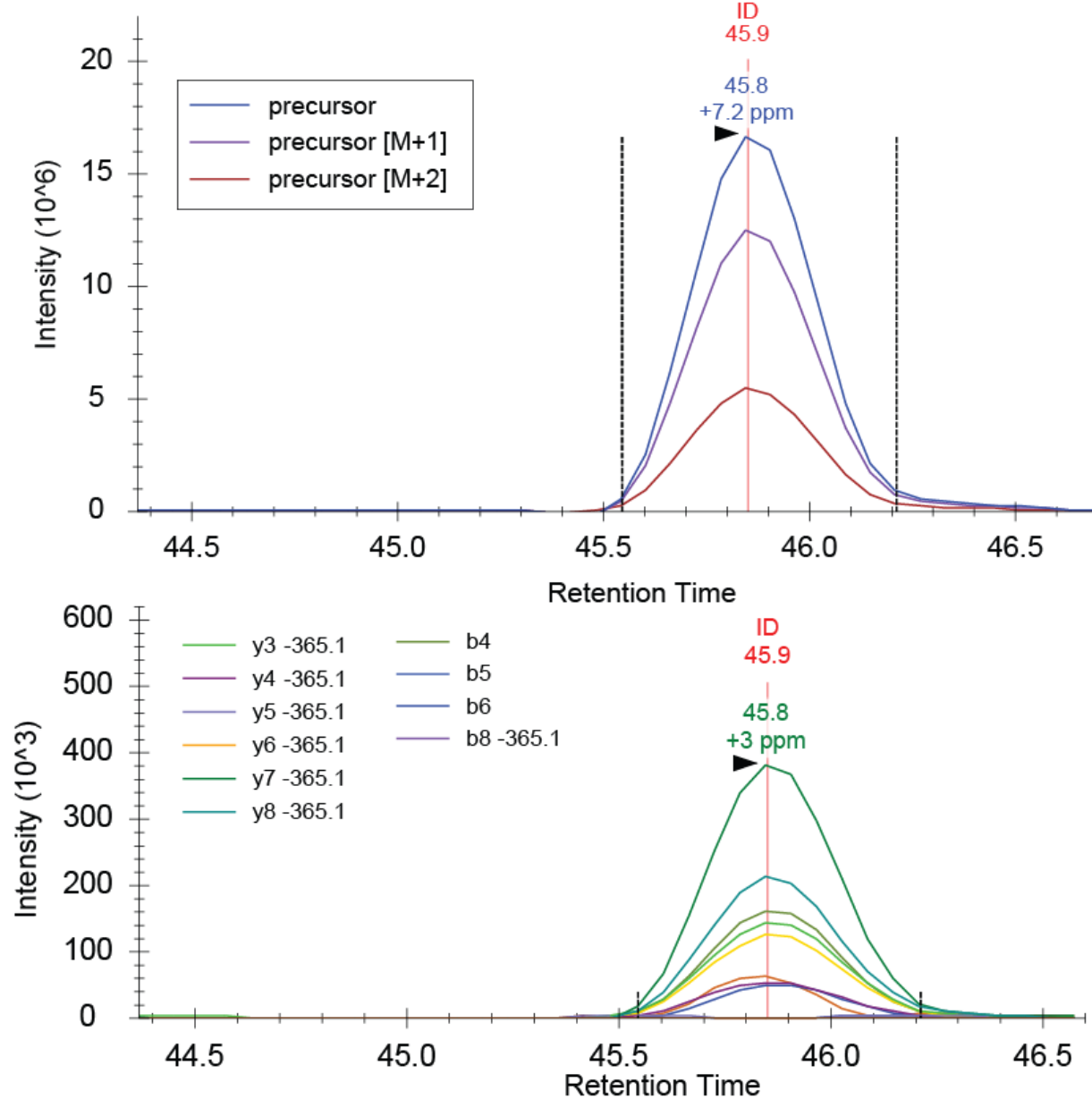
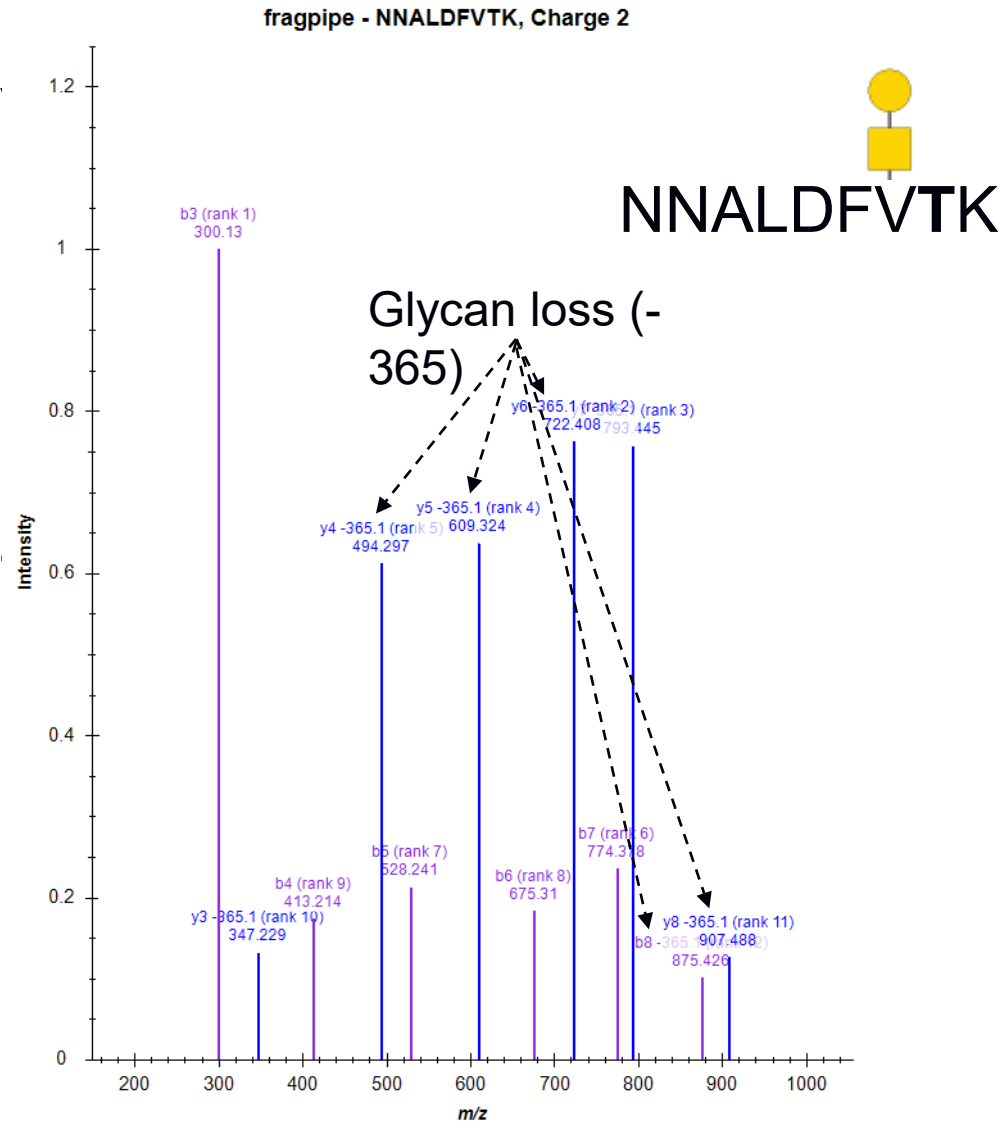
- Development ongoing...
- Currently limit 1 glycan per peptide by default
 - User can change once the Skyline document is loaded

```
<?xml version="1.0" encoding="utf-8"?>
<srm_settings format_version="23.1" software_version="Skyline (64-bit) 23.1.0.455">
  <settings_summary name="Extra Mods">
    <peptide_settings>
      <peptide_modifications>
        <static_modifications>
          <static_modification name="HexNAc(1)" aminoacid="N" variable="true" formula="C8H13N1O5">
          </static_modification>
          <static_modification name="HexNAc(2)Hex(1)" aminoacid="N" variable="true" formula="C22H36N2O15">
          <potential_loss massdiff_monoisotopic="365.13217" massdiff_average="365.13217" formula="C14H23N1O10" />
          </static_modification>
          <static_modification name="HexNAc(2)Hex(3)" aminoacid="N" variable="true" formula="C34H56N2O25">
          <potential_loss massdiff_monoisotopic="689.2378" massdiff_average="689.2378" formula="C26H43N1O20" />
          </static_modification>
        </static_modifications>
      </peptide_modifications>
    </peptide_settings>
  </settings_summary>
</srm_settings>
```

mod.xml

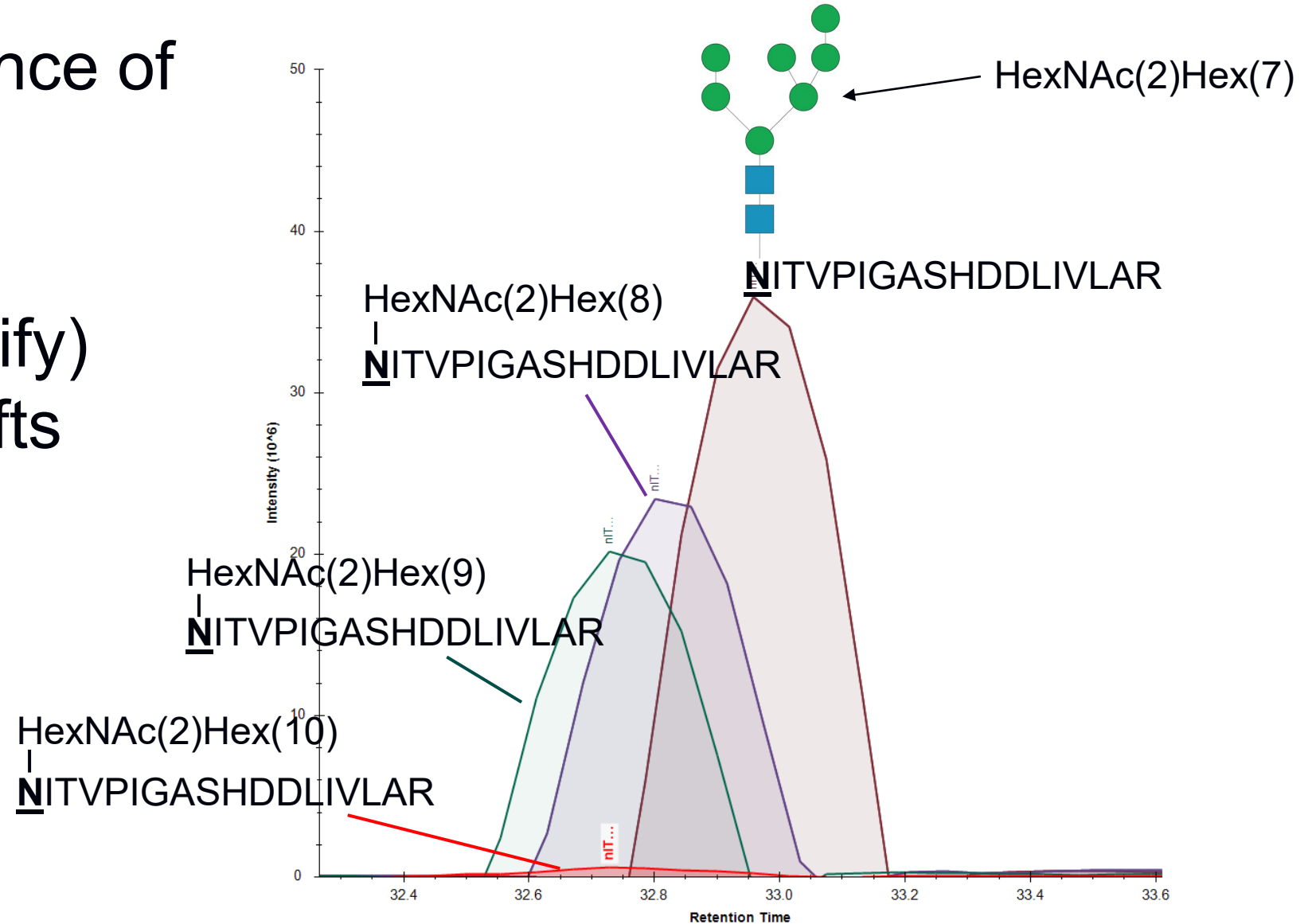
(etc...)

Glycopeptides in Skyline



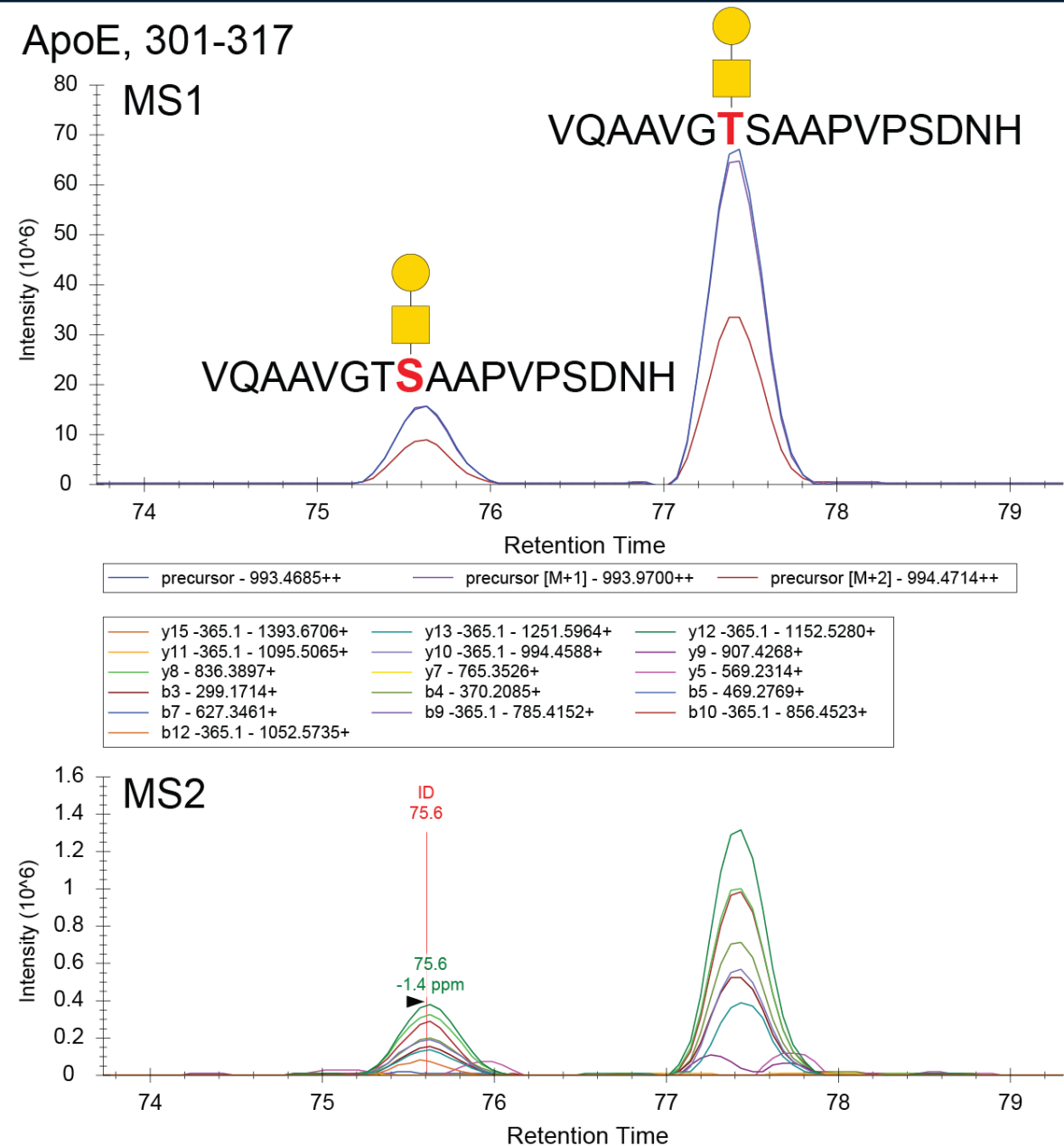
Quantifying Glycoforms

- Visualize abundance of each glycoform
- Observe (and verify) retention time shifts



Positional Isomers

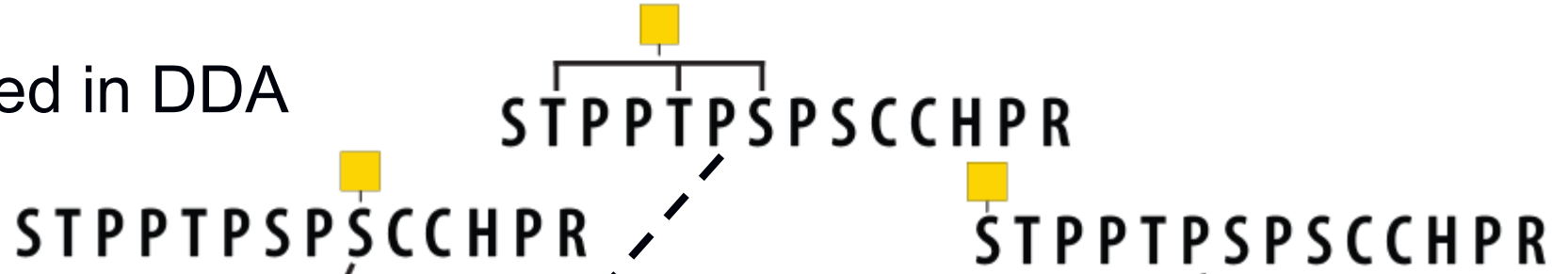
- XIC traces in Skyline are very helpful for tracking different glycoforms



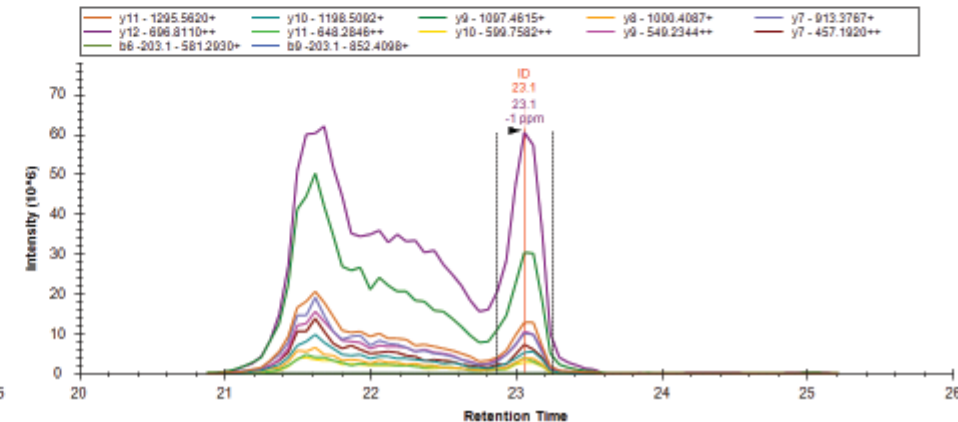
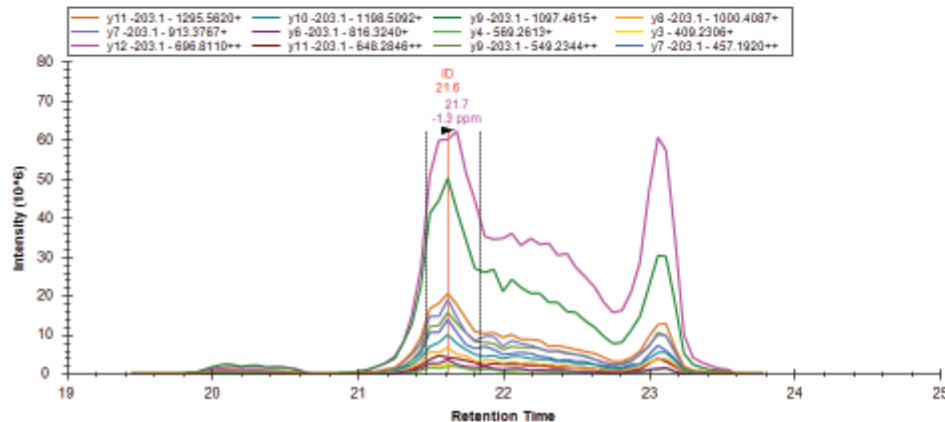
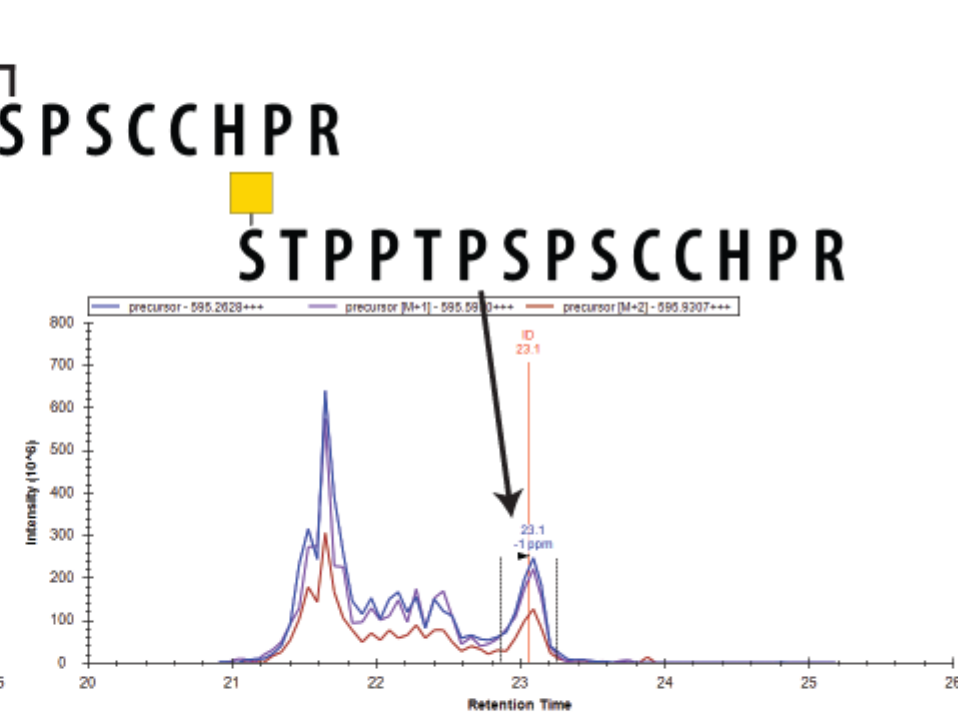
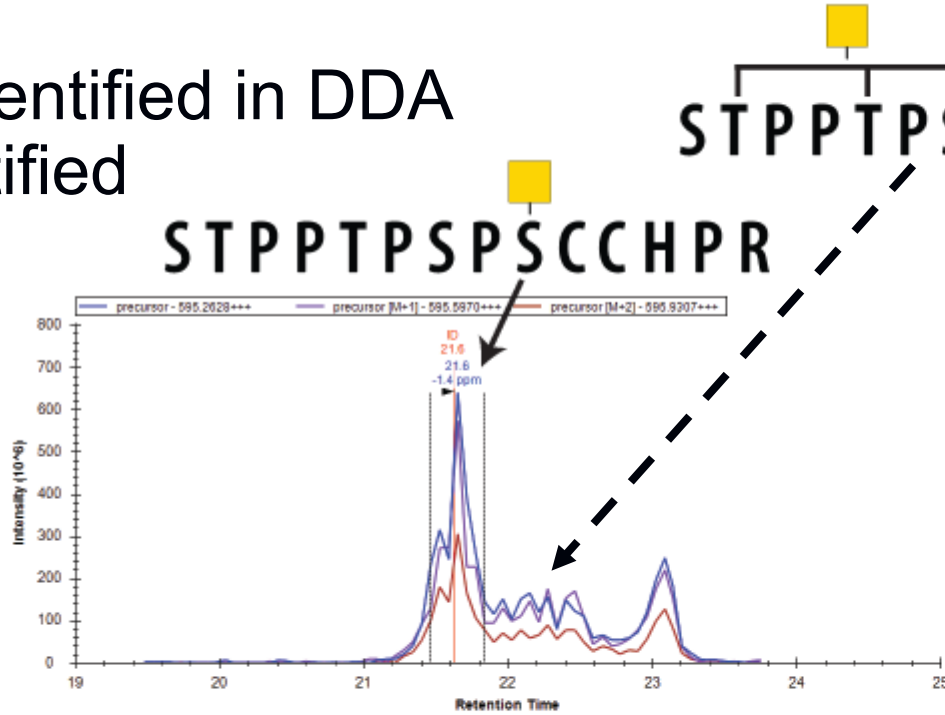
Visualization helps diagnose complex issues

- Example:

- 2 glycosites identified in DDA library & quantified



- Unresolved additional site(s) observed in XIC



Summary and Ongoing Work

- FragPipe 22.0 release can now generate Skyline documents automatically
- Several areas of continuing development
 - Improve speed with many modifications
 - Improve support for DDA glycoproteomics data
 - Support non-unimod modifications from .speclib format

Acknowledgements

Nesvizhskii lab

Skyline Team

- Brendan MacLean
- Matthew Chambers



Collaborators:

Lukas Käll (Percolator)

Vadim Demichev (DIA-NN)

Bo Wen, Bing Zhang (PDV)

George Rosenberger (EasyPQP)

David Shteynberg (PTM-Prophet)

Michael Shortreed (University of Wisconsin-Madison)

Ryan Kelly (BYU)

Keri Backus Lab (UCLA)

Ralser Lab (Francis Crick)

Stephan Hacker Lab (Leiden U)

Brendan MacLean (University of Washington)

Michael Shortreed (University of Wisconsin-Madison)

Ying Zhu, Chris Rose, Meena Choi, Dennis Wolan (Genentech)



Thank you

Q & A