Skyline

Large Scale DIA
With Skyline

With

Brendan MacLean (Principal Developer, Skyline)

Agenda

- Welcome from the Skyline team!

- Large Scale DIA with Skyline

    - Introduction and overview with Brendan MacLean

    - Tutorial with Brendan MacLean

- Audience Q&A – submit questions to Google Form:

    https://skyline.ms/QA4Skyline.url

Webinar 2: Jump start DIA... November, 2014

Chromatography-based Quantification

- Hypothesis testing (Verification)

- SRM

- MS1 chromatogram extraction

- Targeted MS/MS (PRM)

- **Data independent acquisition (DIA/SWATH)**

| Acquisition	Targeted	Survey
More Selective	PRM	DIA
Less Selective	SRM	MS1

Exploratory or Confirmatory? HYPOTHESIS??

Prior Knowledge and Consistency

- Relative product ion abundance
  - Spectral libraries
- Retention time
  - iRT libraries

- Powerful enough to be used cross-lab / cross experiment
  - Measure, store, re-use
- More powerful run-to-run
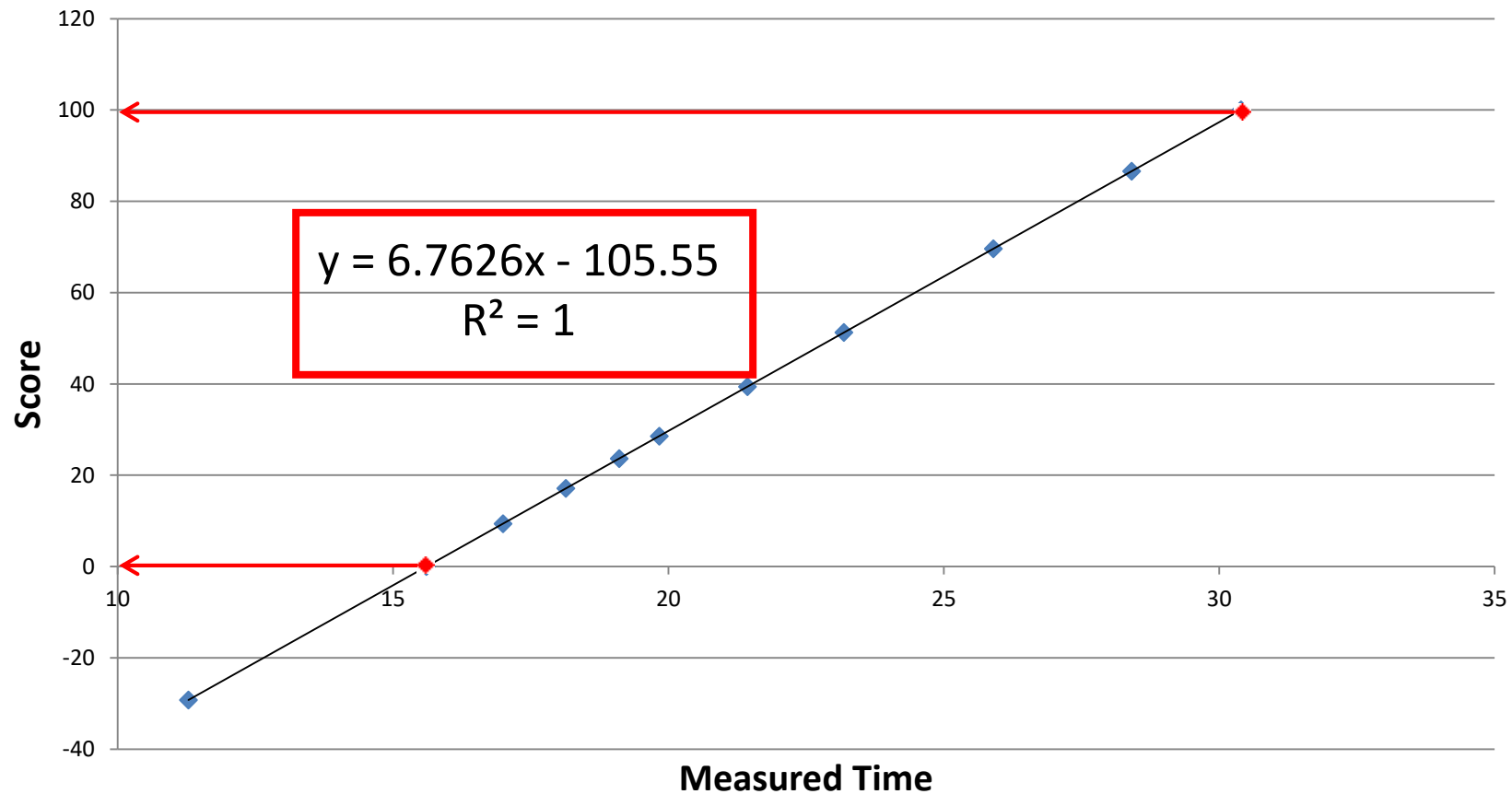
Relative Product Ion Abundance
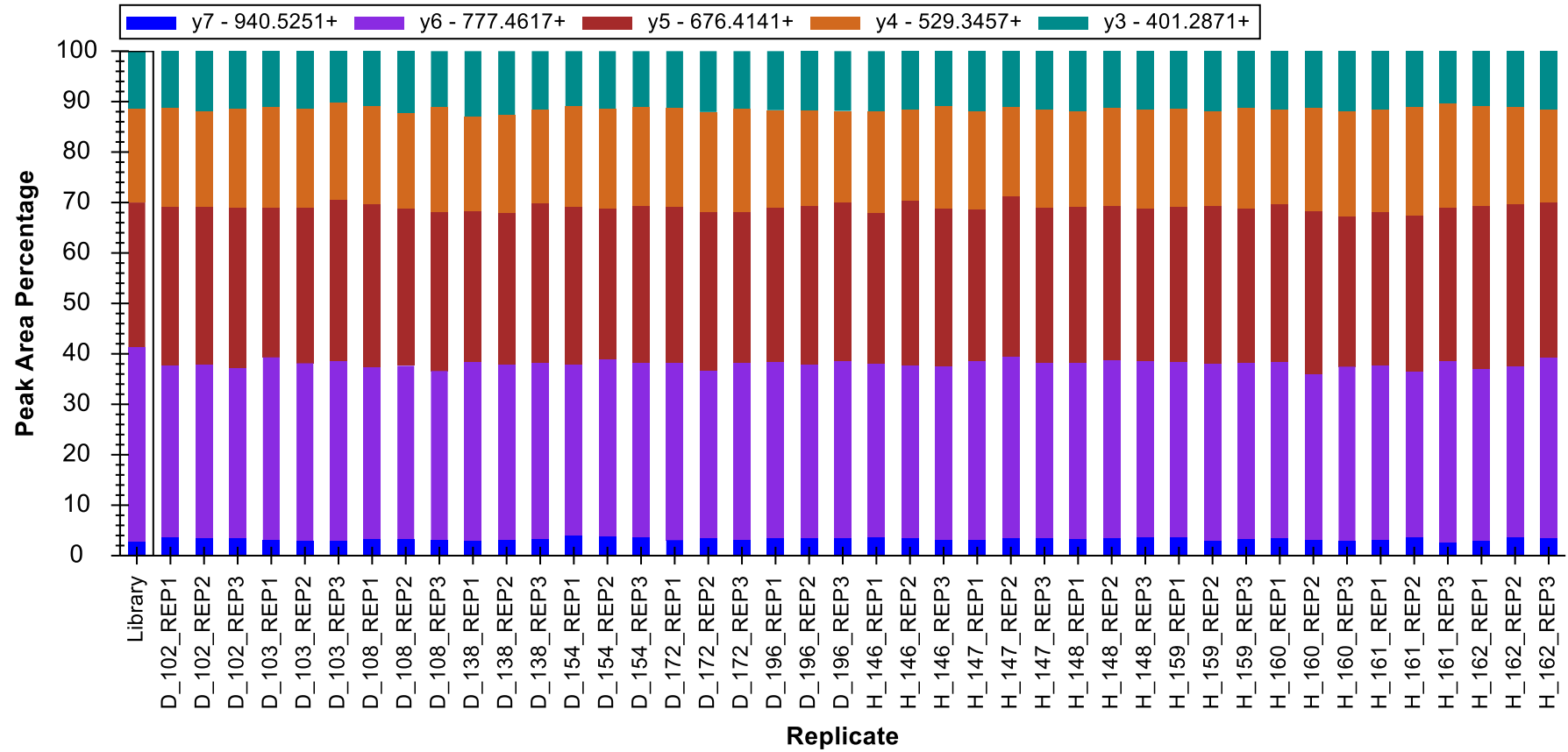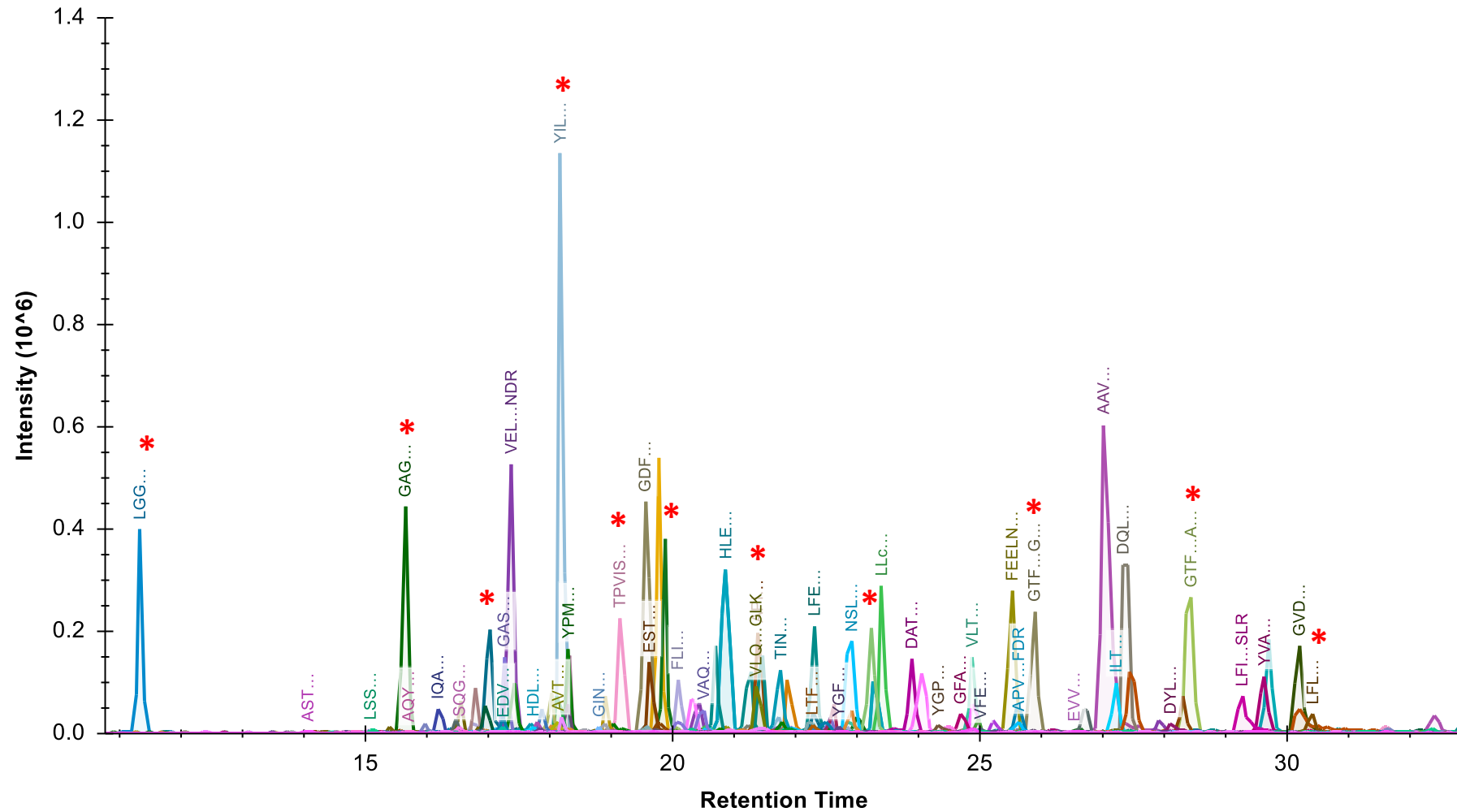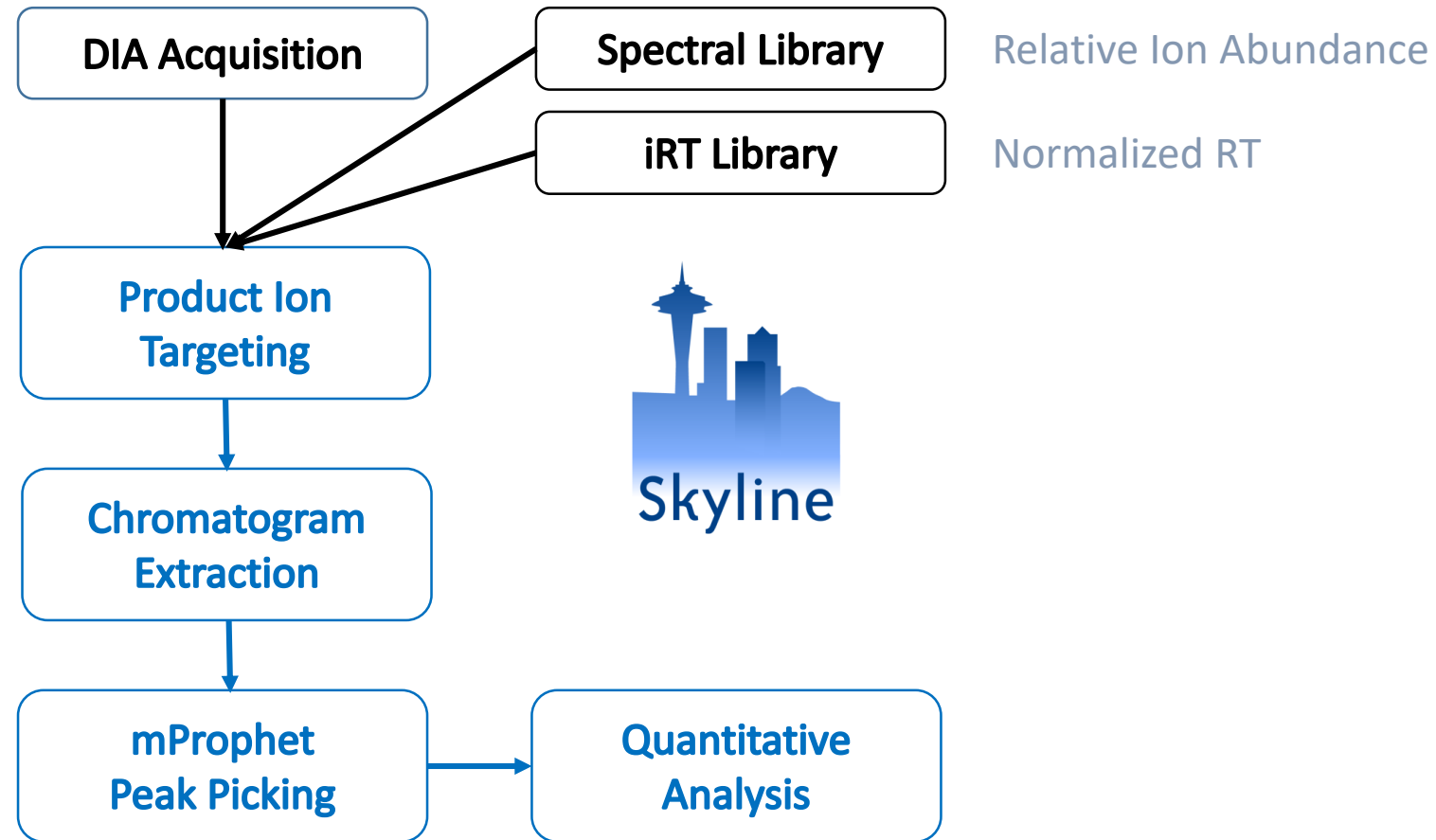
Webinar 7: iRT time prediction… May 2015

Escher, Reiter, et al. *Proteomics*, 2012
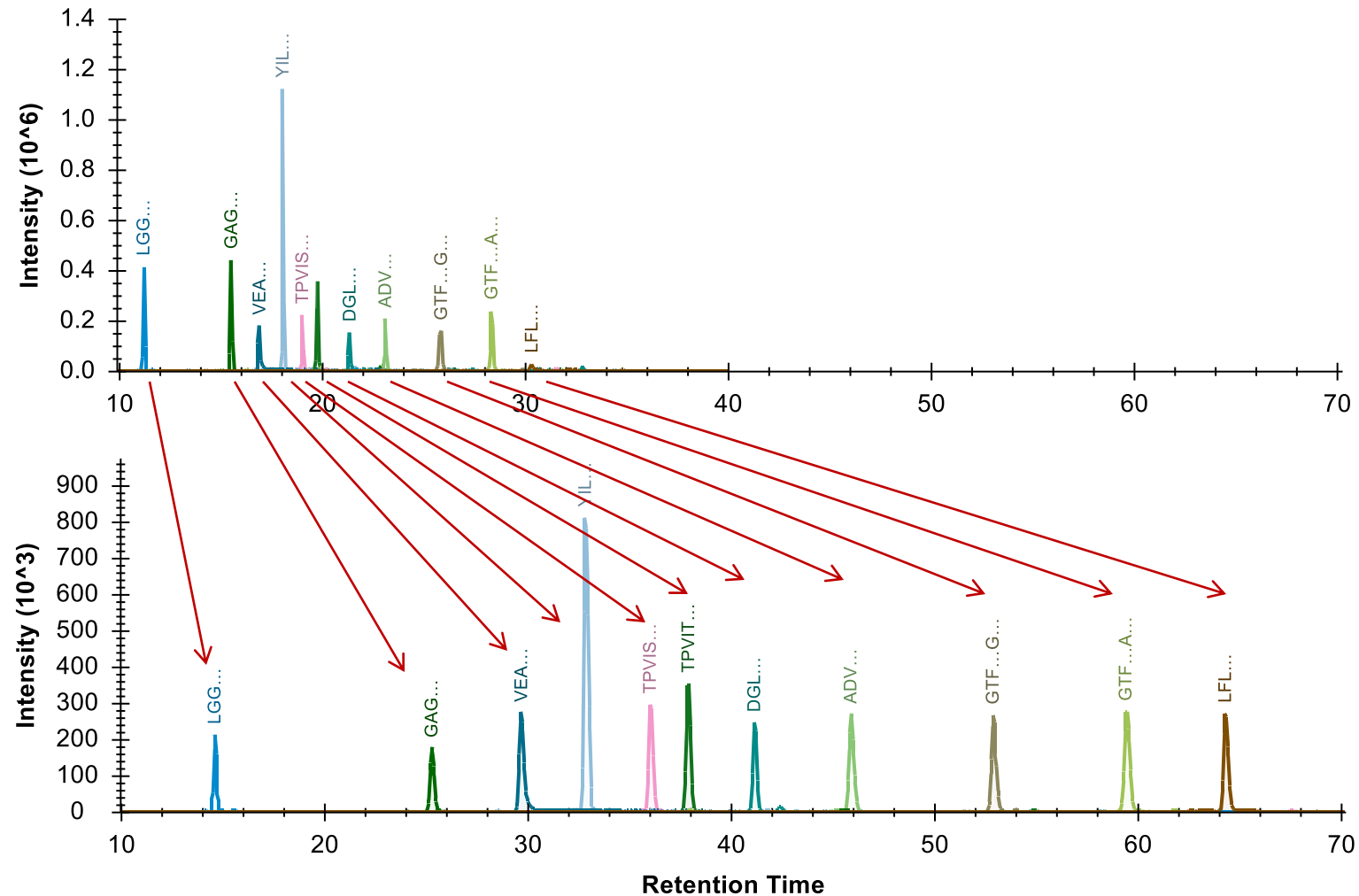
Defining an iRT Scale

- Points on a line (score = time * slope + intercept)

$$y = 6.7626x - 105.55$$
$$R^2 = 1$$

Score

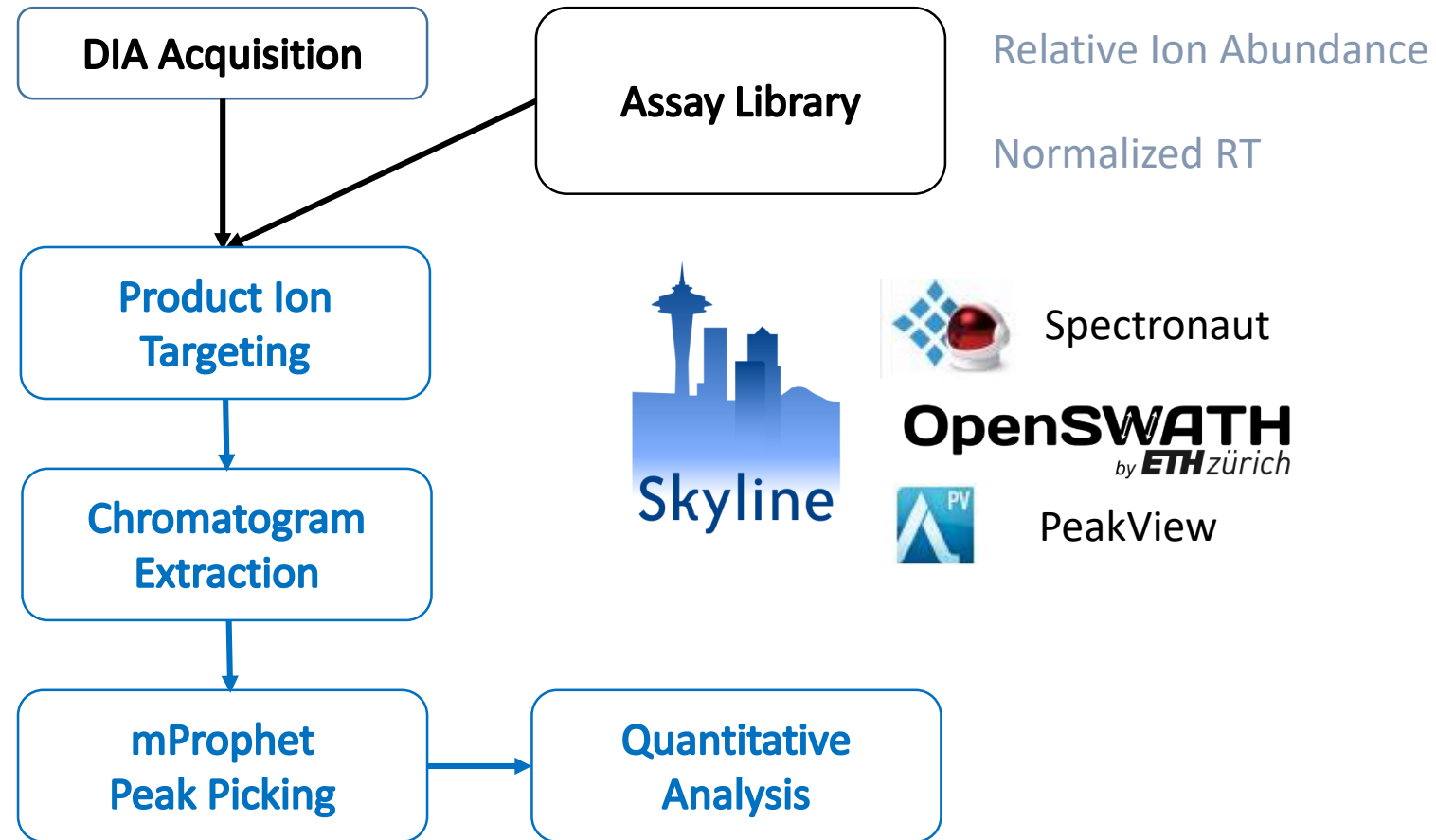Measured Time

Building an iRT Library

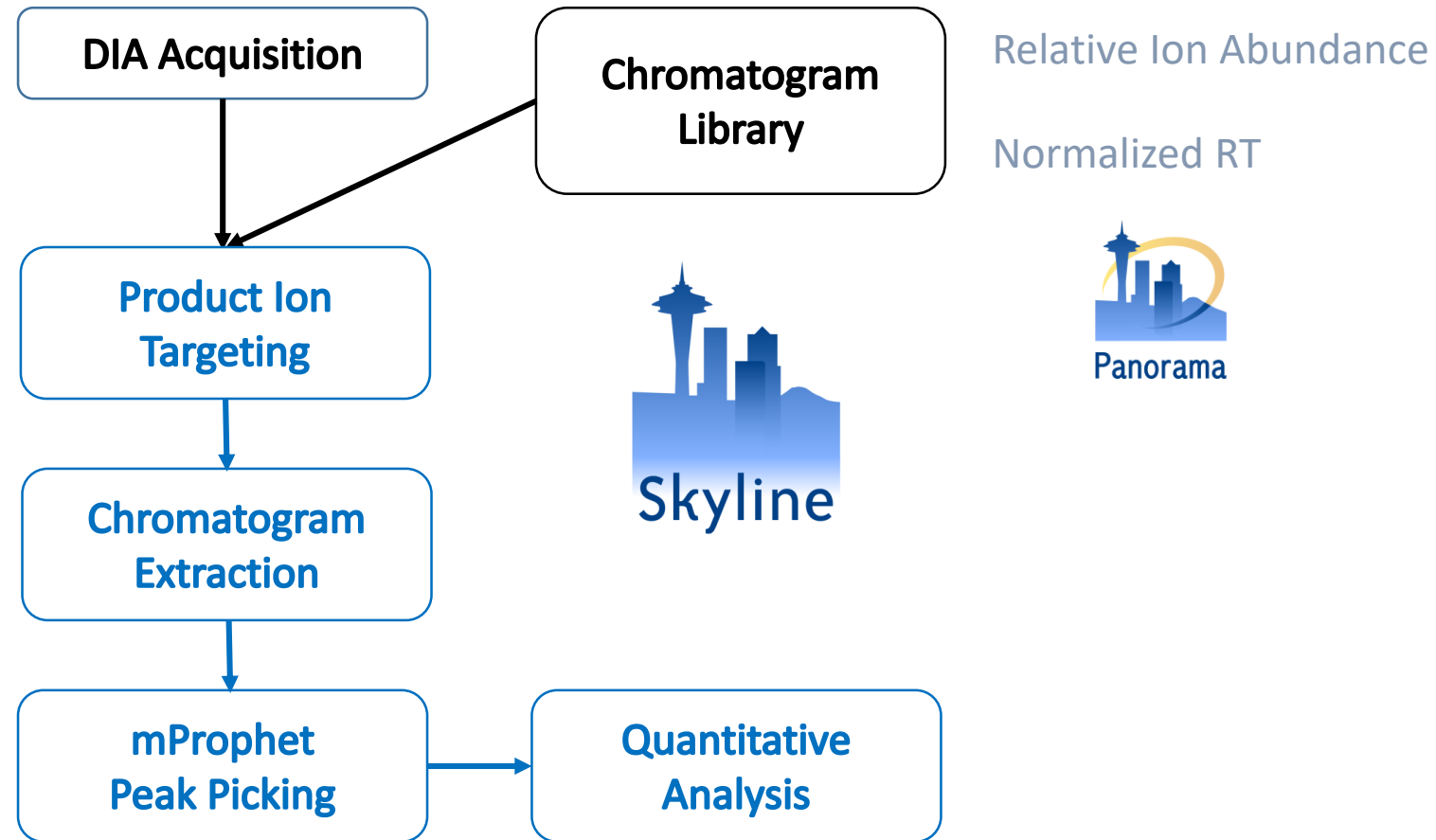# Prior Knowledge Workflow

Reiter, Rinner, et al.
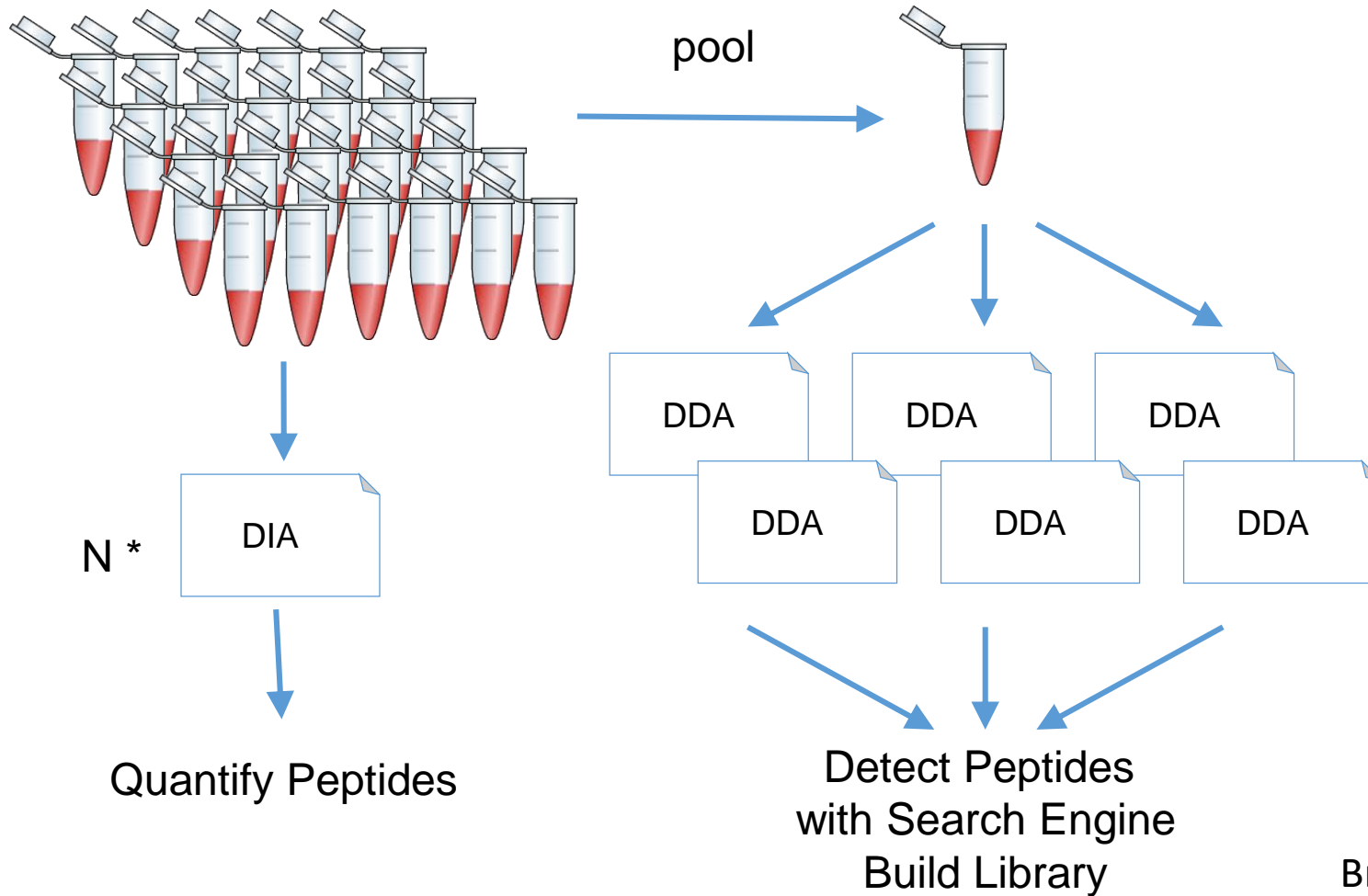*Nature Methods*, 2011

Prior Knowledge Workflow

Prior Knowledge Workflow

Two Phase Experiment

N * Quantitative Samples

1 * Detection Samples

pool

N * DIA

DDA   DDA   DDA

DDA   DDA   DDA

Quantify Peptides

Detect Peptides
with Search Engine
Build Library

Bruderer, et al. & Reiter, *MCP, 2015*

DIA/SWATH Software Comparison

- Driving lots of performance improvement
  - MaxQuant tutorial MS1 data imports 3x faster
  - Comparison pipeline tasks 20x to 200x faster
    - Single file import 5 hours on SSD with 3.1
    - 6 files import in 50 minutes on spinning drive with 3.5
    - Adjusting peaks took 2 hours with 3.1
    - Adjusting peaks takes 10 seconds with 3.5

- Looked bad for Skyline DIA quant
  - Turned out to be serious issues with settings

Navarro, et al. & Tenzer
A multicenter study, Nature Biotech. 2016

SkylineRunner Command-Line Interface

- Great for long batch runs
- Ready to be adapted to a compute cluster

Before

- Spectronaut 42,439 peptides identified
- OpenSWATH 40,387 peptides identified
- DIA Umpire 31,256 peptides identified
- PeakView 28,424 peptides identified
- Skyline 27,121 peptides identified

|		First Iteration				Second Iteration			
		Identifications	Relative quantifications	separation Y-H	separation E - H	Identifications	Relative quantifications	separation Y-H	separation E - H
DIAumpire	peptides	31256	24456	0.936	0.944				
	proteins (int)	4351	3952	0.952	0.975				
	proteins (TOP3)	3654	3046	0.952	0.957				
openSWATH	peptides	40387	35443	0.966	0.974	40366	35383	0.969	0.978
	proteins (TOP3)	4700	4308	0.967	0.993	4705	4304	0.975	0.984
PeakView	peptides	28424	28424	0.966	0.970	28393	21011	0.984	0.974
	proteins (int)	5749	5749	0.971	0.965				
	proteins (TOP3)	3926	3926	0.971	0.983	3842	3042	0.971	0.989
Skyline	peptides	27121	18505	0.956	0.876				
	proteins (TOP3)	3468	2633	0.946	0.901				
Spectronaut	peptides	42439	37119	0.966	0.969	42325	36292	0.968	0.984
	proteins (TOP3)	4810	4363	0.974	0.962	4801	4314	0.975	0.981

Finding the Problem
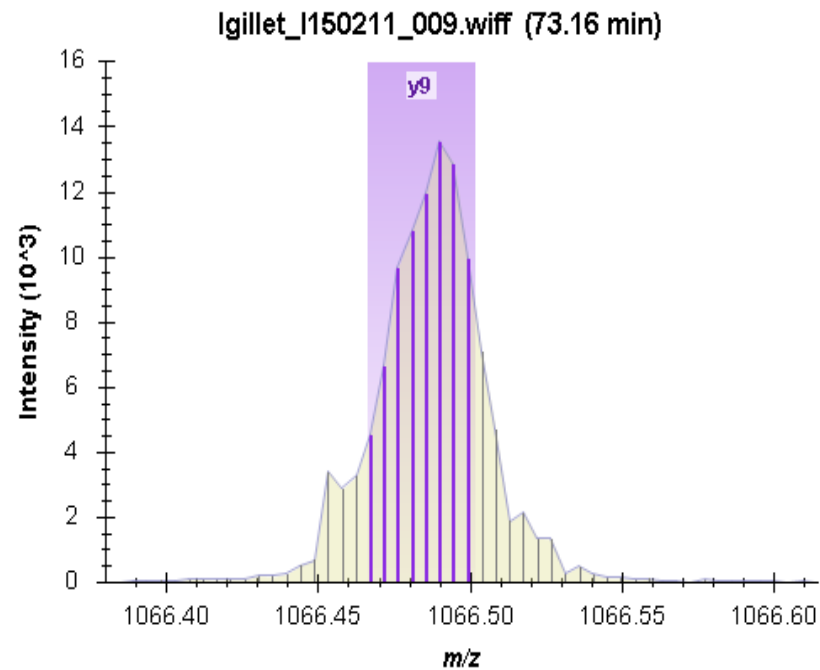
TPVIVTLK
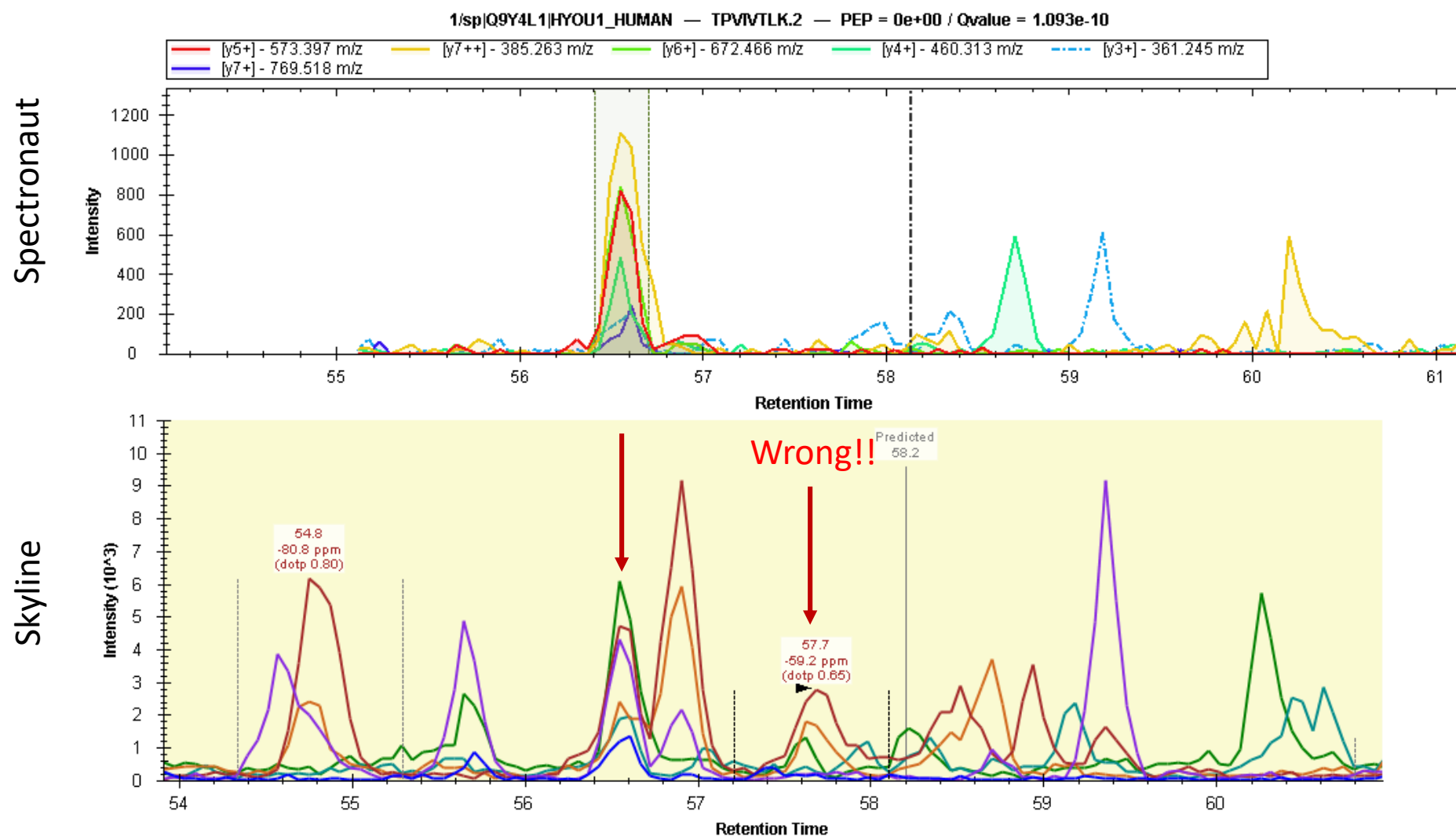
Extraction Width
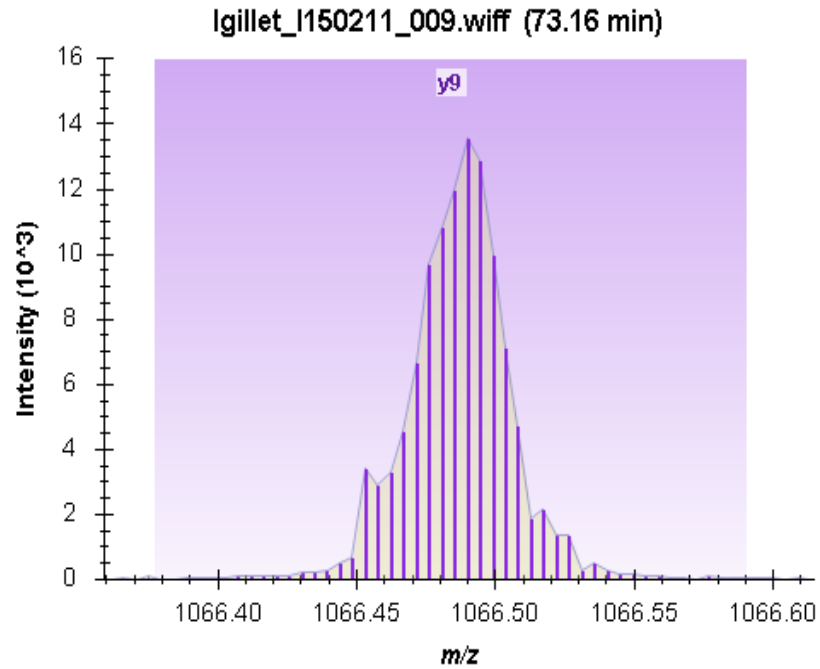
- 10,000 rp (default)   30,000 rp
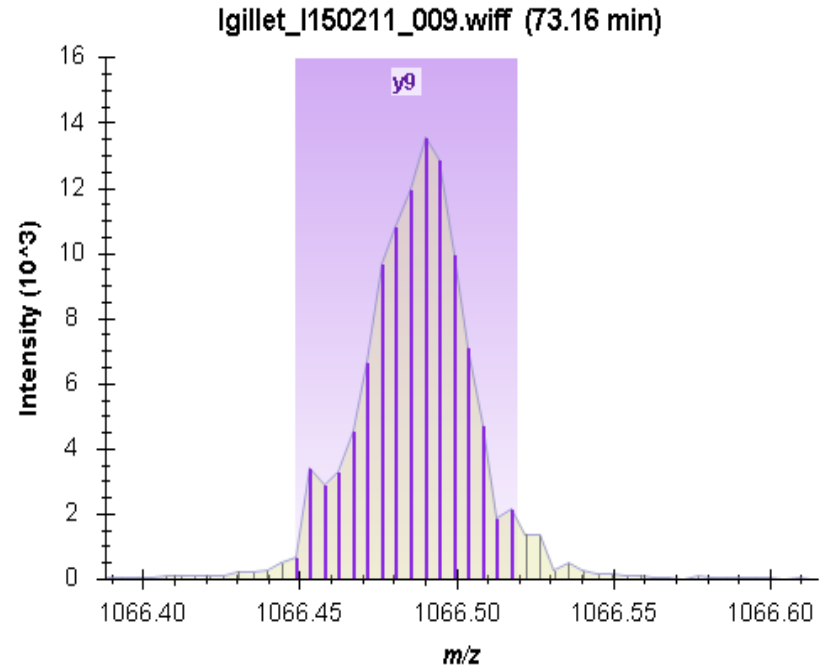
Extraction Width

- 60,000 rp

±10 ppm (centroided)
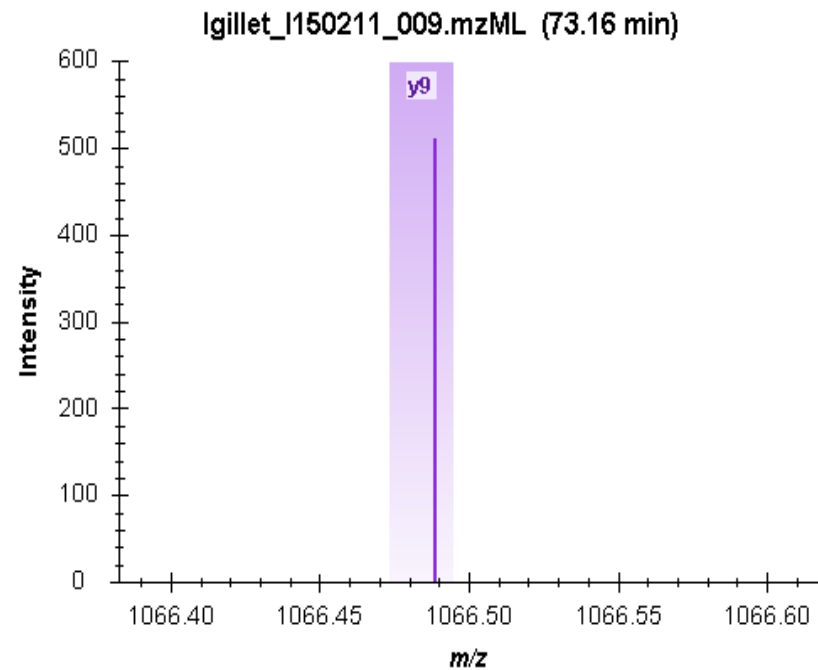
Extraction from Centroided Spectra

Extending SkylineRunner

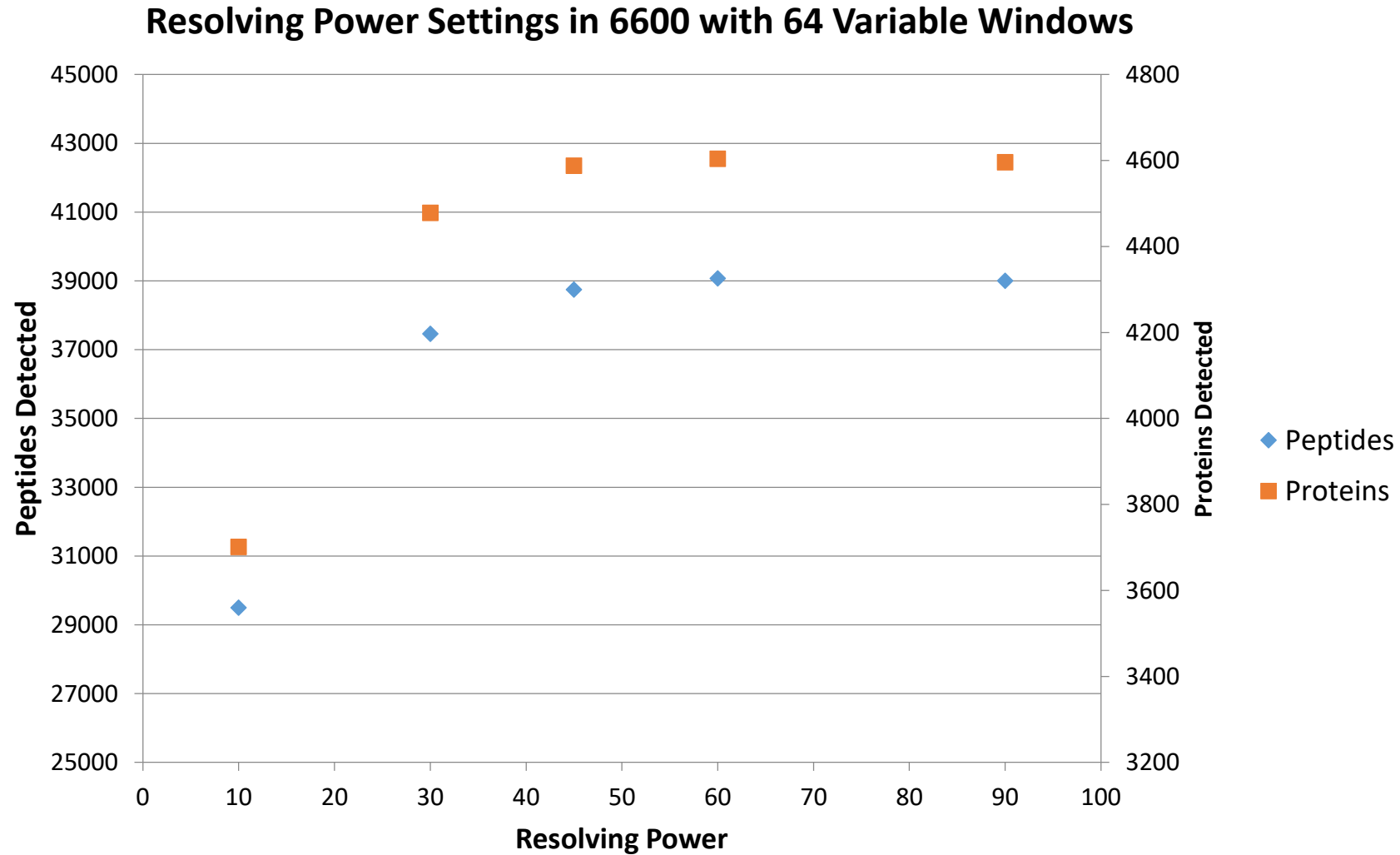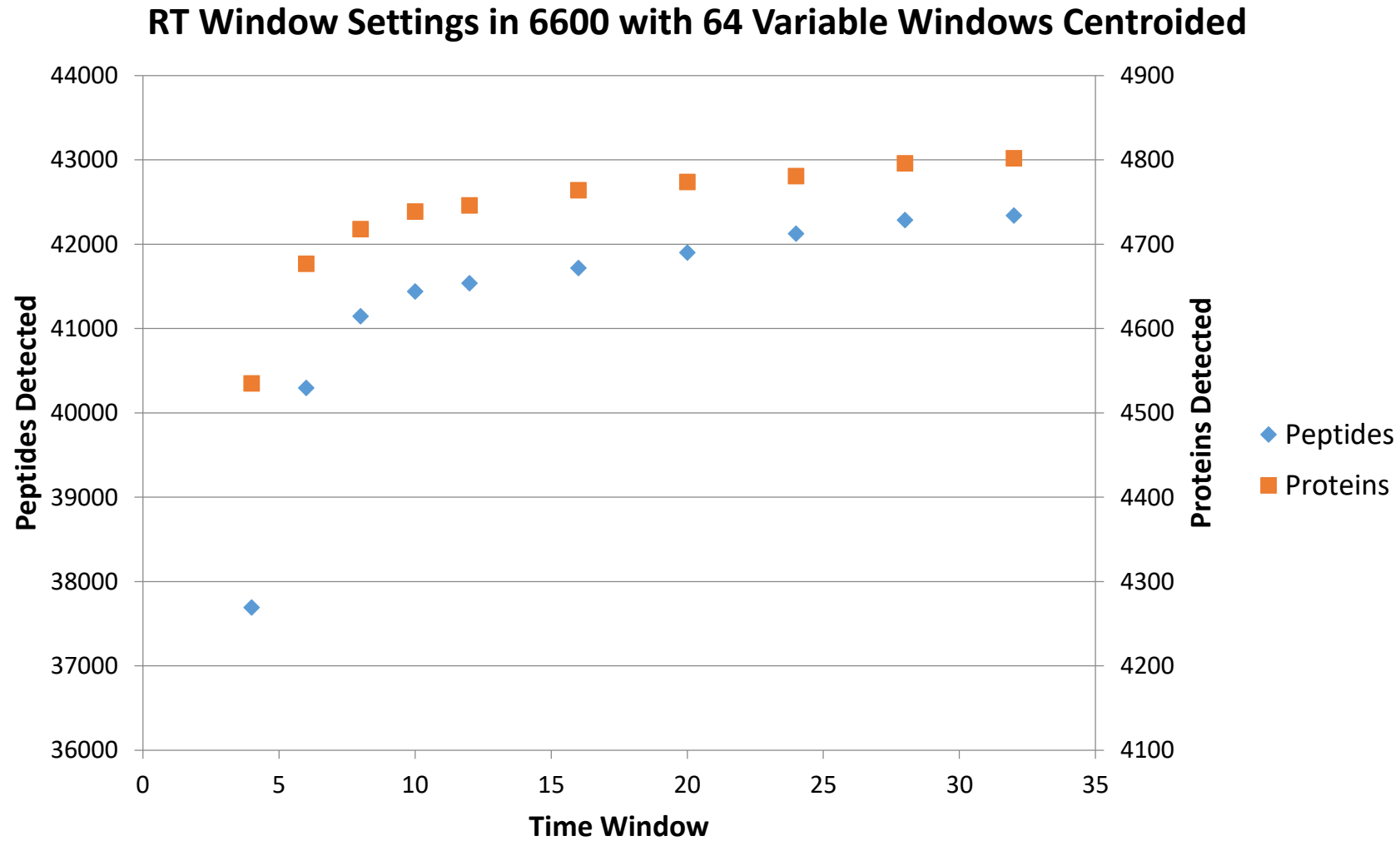|	
--reintegrate-model-name=<name>	The name of a scoring model to use for the reintegrate operation. The model can either be pre-defined (e.g. using the Edit > Refine > Reintegrate form) or created automatically during this operation by using --reintegrate-create-model.
--reintegrate-create-model	This option will cause a new model to be created, using the mProphet algorithm with all available scores for the results found in the document. (requires --reintegrate-model-name)
--reintegrate-annotate-scoring	Peaks will be annotated with q value and score annotations. (requires --reintegrate-model-name)
--reintegrate-overwrite-peaks	Existing manually integrated peaks will be overwritten with peaks chosen by the reintegration model. (requires --reintegrate-model-name)

Extending SkylineRunner

--full-scan-precursor-res=<resolving power>	Resolving power of the precursor mass analyzer.
--full-scan-precursor-res-mz=<m/z value>	The m/z value at which the precursor mass analyzer resolving power is specified. (applies only to orbitrap and ft_icr mass analyzers)
--full-scan-product_res=<resolving power>	Resolving power of the product mass analyzer.
--full-scan-precursor-res-mz=<m/z value>	The m/z value at which the product mass analyzer resolving power is specified. (applies only to orbitrap and ft_icr mass analyzers)
--full-scan-rt-filter-tolerance=<minutes>	The number of minutes on either side of the predicted time or MS/MS IDs, i.e. ± minutes. Defaults to.

Optimal Settings Analysis (RP)

Resolving Power Settings in 6600 with 64 Variable Windows

Optimal Settings Analysis (RT)

**RT Window Settings in 6600 with 64 Variable Windows Centroided**

After

- Skyline 42,517 peptides identified

- Spectronaut 42,325 peptides identified

- OpenSWATH 40,728 peptides identified

- DIA Umpire 36,249 peptides identified

- SWATH 2.0 35,489 peptides identified

			Iteration 1						Iteration 2			
		Median CV of human	Number of IDs	Valid quantification ratios	Overlap yeast−human (arctanh)	Overlap *E. coli*−human (arctanh)	Median CV of human	Number of IDs	Valid quantification ratios	Overlap yeast−human (arctanh)	Overlap *E. coli*−human (arctanh)
Peptides	OpenSWATH	6.5%	40,726	36,098	2.12	2.24	8.2%	40,728	35,944	2.10	2.26
	SWATH 2.0	6.6%	35,517	35,517	2.14	2.11	6.1%	35,489	26,303	2.49	2.52
	Skyline	7.4%	40,804	34,103	1.91	1.85	6.9%	42,517	37,977	2.13	2.14
	Spectronaut	5.9%	42,439	37,120	1.97	1.90	6.2%	42,325	36,292	2.11	2.26
	DIA-Umpire	13.2%	36,332	28,785	1.74	1.98	12.9%	36,249	25,677	1.82	2.18
Proteins	OpenSWATH	5.4%	4,632	4,343	2.30	2.56	6.4%	4,636	4,352	2.51	2.60
	SWATH 2.0	5.6%	4,323	4,323	2.37	2.36	6.1%	3,946	3,371	2.42	2.56
	SWATH 2.0 (built-in)	6.0%	6,178	6,178	2.23	2.03					
	Skyline	5.9%	4,518	4,140	2.03	2.15	5.5%	4,692	4,456	2.37	2.43
	Spectronaut	3.3%	4,692	4,346	2.13	2.18	3.3%	4,675	4,300	2.31	2.50
	DIA-Umpire	12.0%	3,795	3,379	2.12	2.30	12.3%	3,673	3,111	2.13	2.85
	DIA-Umpire (built-in)	13.1%	4,849	4,489	1.78	1.94					

Multi-File Parallel Loading
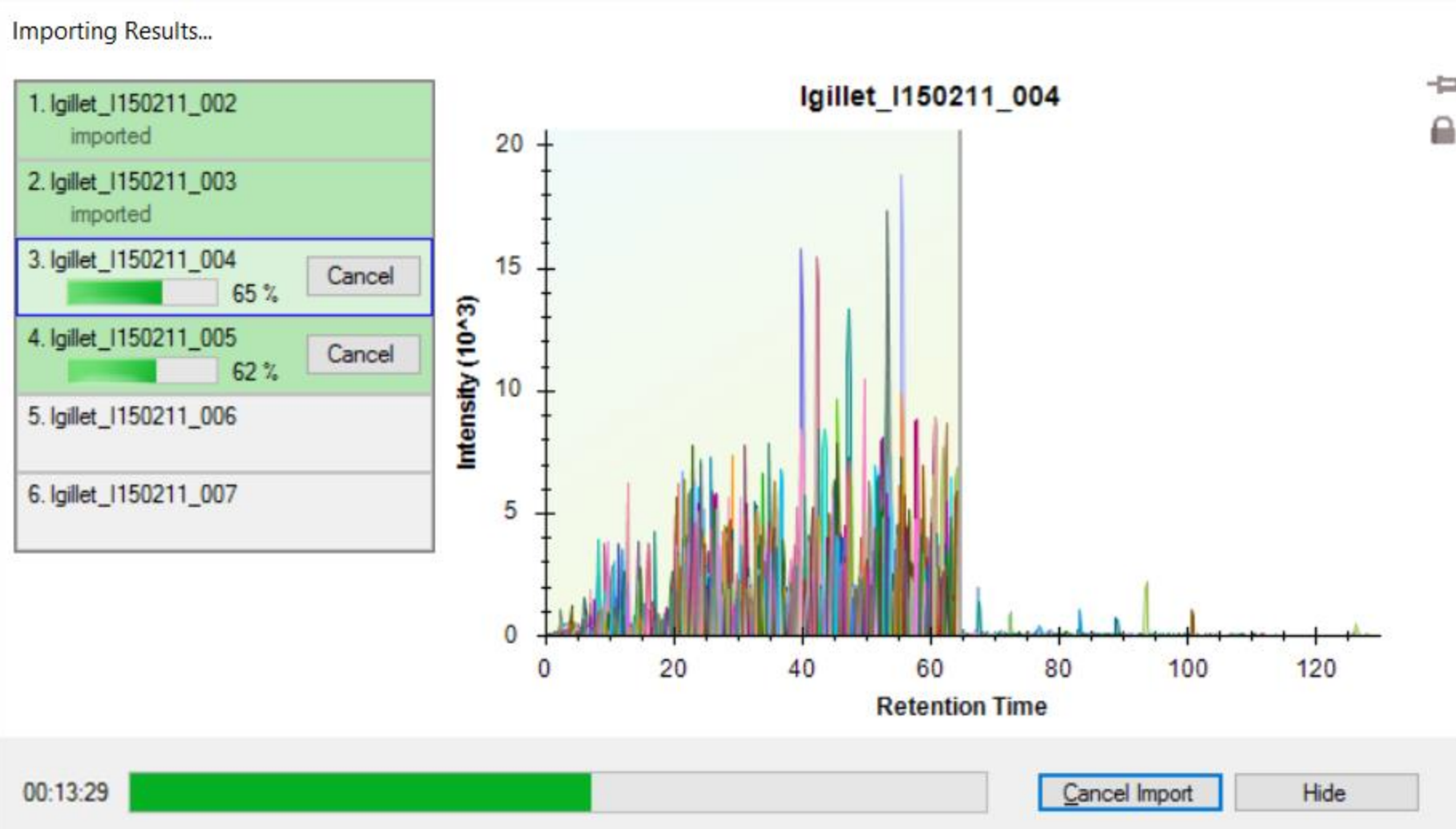
Multi-File Loading Performance

12 x WIFF – **24-core**, 178 GB, SSD

6 x WIFF – **4-core**, 12 GB, SSD
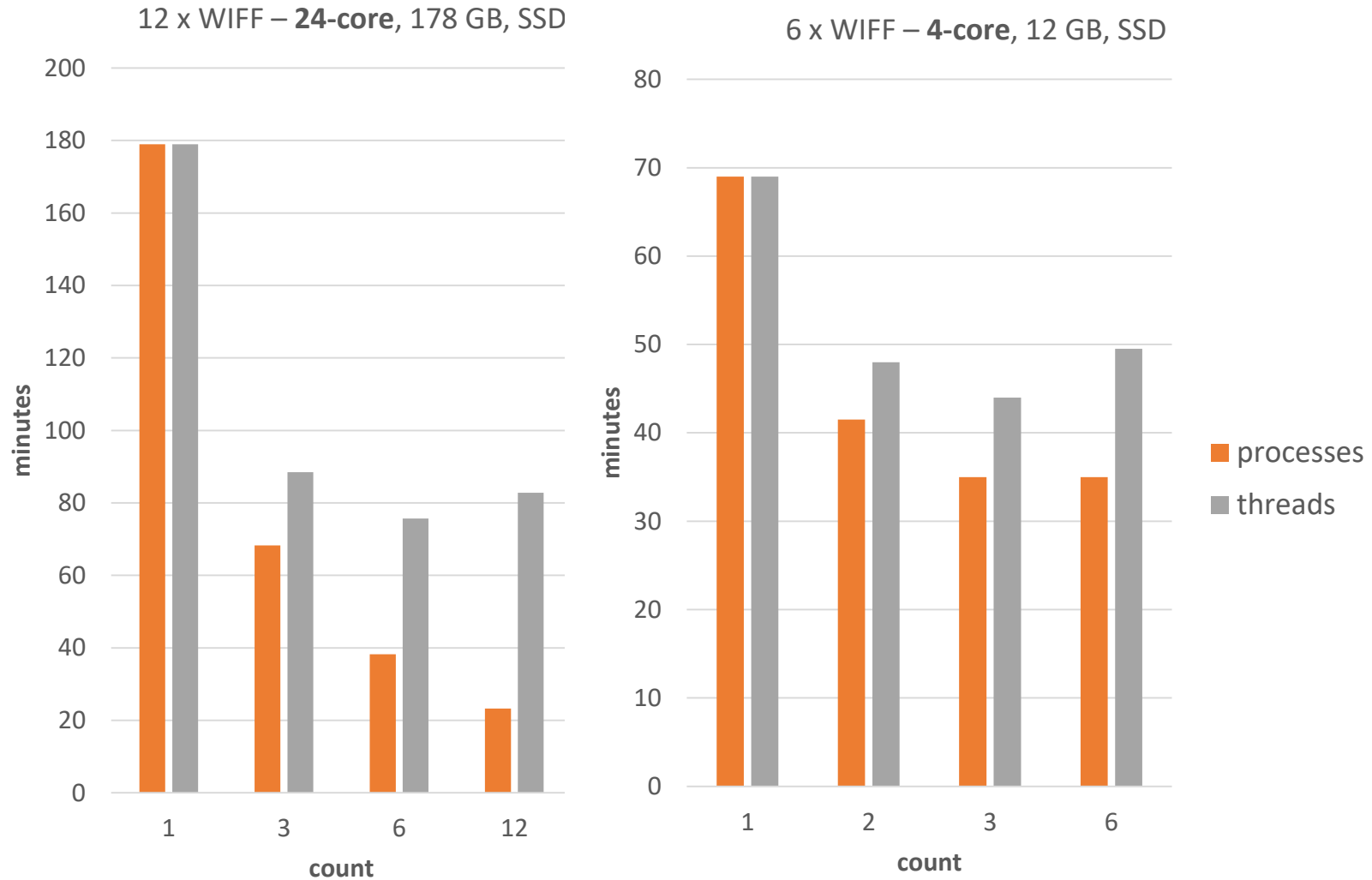
Analyzing Data From Navarro Paper

- Acquired by L. Gillet

Blib Document mProphet Model

Assay Library Document mProphet Model

Learn More

- Webinar #15: TBD (14 recorded webinars)

- Weeklong Courses 2017
  - Buck Institute, Feb 27 – Mar 3 - Full
  - Northeastern University, Boston – May 1-3 – Register now!
  - ETH, Zurich, June 26 – 30 – Register now!
  - University of Washington, Seattle – July 24-28

- Workshops and Conferences 2017
  - Skyline User Group Meeting at ASMS, Indianapolis – June 4

Listings updated in **Join Us** section of Skyline homepage:

https://skyline.ms/Skyline.url

Questions?

- Ask any questions at the following form:

  https://skyline.ms/QA4Skyline.url

- Take the post-webinar survey:

  https://skyline.ms/survey4webinar.url

Skyline Tutorial Webinar #14

This ends this Skyline Tutorial Webinar.

Please give us feedback on the webinar at the following survey:

https://skyline.ms/survey4webinar.url

A recording of today's meeting will be available shortly at the Skyline website.

We look forward to seeing you at a future Skyline Tutorial Webinar.