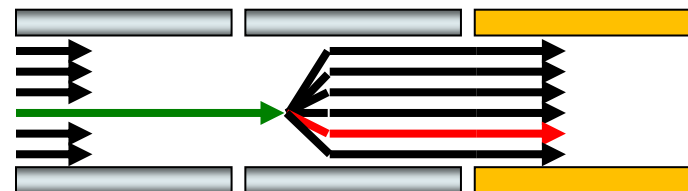# Skyline

Targeted Proteomics Environment

Rapid Processing of Large Scale Quantitative Proteomics Projects: Integration of Skyline with the CHORUS Cloud

Brendan MacLean; Andrey Bondarenko; Nick Shulman; Oleksii Tymchenko; Christine Wu; Nathan Yates; Michael J. Maccoss

UNIVERSITY *of* WASHINGTON

Chromatography-based Quantification

▸ Freely-available, and vendor neutral

▸ SRM

▸ MS1 chromatogram extraction

▸ Targeted MS/MS (PRM)

▸ **DIA / SWATH**

Acquisition	Targeted	Survey
More Selective	Targeted-MS/MS	DIA
Less Selective	SRM	MS1

DIA Chromatogram Extraction

- How many chromatograms to make DIA interesting?

- 10,000 peptides? (50,000+ transitions)
- Whole proteome? (500,000+ transitions)

- Hypothesis driven inquiries?

- "Kind of defeats the purpose of SWATH"?

DIA Fit-for-Purpose

▸ Discovery Proteomics
  ▸ Systematic acquisition without missing data
  ▸ Petpide searching tools less mature than DDA
  ▸ Chimeric spectra hard to search

▸ Targeted Proteomics
  ▸ High level of multiplexing without scheduled acquisition
  ▸ Ability to test new hypotheses after acquisition
  ▸ Gain selectivity over MS1
  ▸ Lose selectivity from SRM and PRM

DIA versus SRM Multiplexing

‣ SRM

  ‣ 100 transitions unscheduled

  ‣ 20-30 peptides label-free

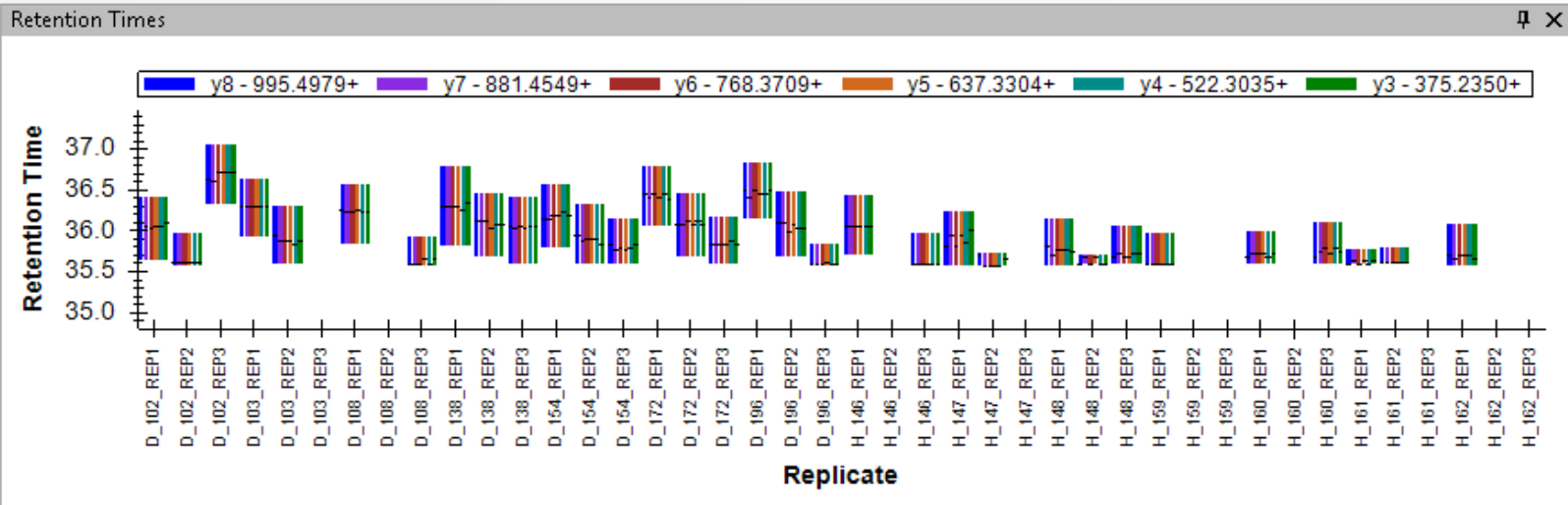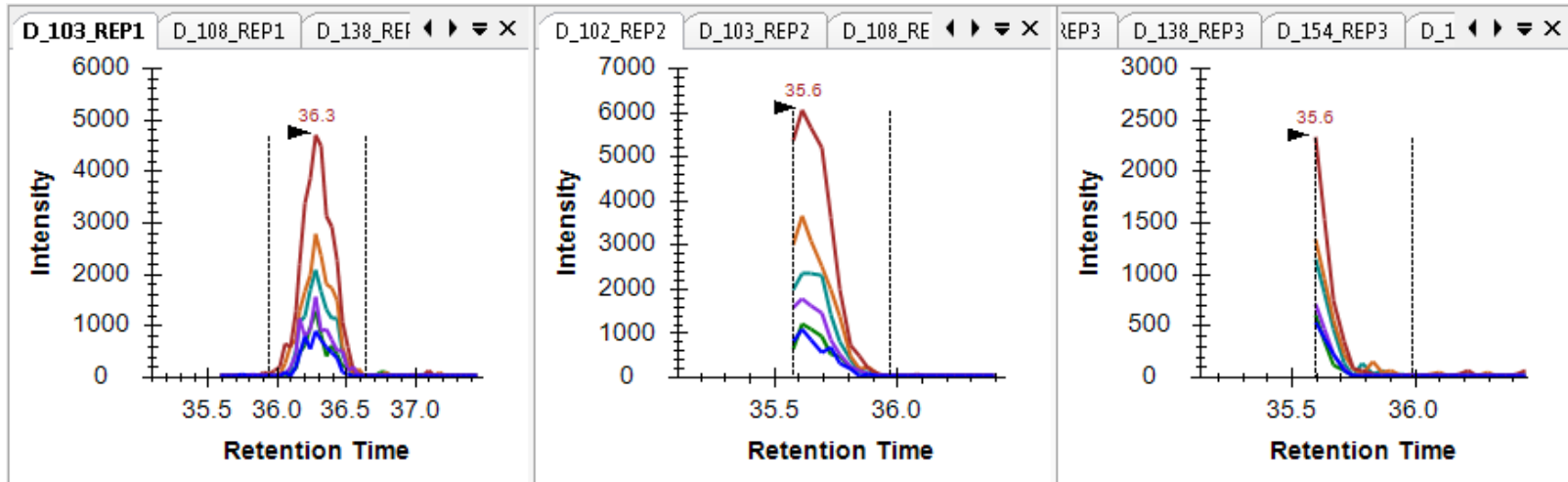 ‣ 10-15 peptides with labeled pairs

‣ DIA

  ‣ Unlimited

‣ Problems with scheduling

  ‣ Shifts in chromatography can compromise measurement

  ‣ Add setup time and complexity

  ‣ More susceptible to human error

Truncated and Missing Peaks

TGTNLMDFLSR

DIA versus SRM Files and Performance

- ## SRM
  - Size: 5 to 20 MB
  - Import time: seconds to a few minutes
- ## DIA
  - Size: 200 MB to 4000 MB (with IMS 2000 to 8000 MB)
  - Import time: 30 seconds to tens of minutes

DIA versus SRM 50 Runs

- SRM
    - Size: 0.5 GB
    - Import time: 10 minutes
- DIA
    - Size: 100 GB
    - Import time: 6 hours

- **100x Storage and Performance Impact**

Chorus For Mass Spec File Storage

- Google Docs-like interface
- Lab-centered security model
- Raw data file storage
  - Upload as acquired
  - Translated into distributed data structure
  - Massively parallel cloud data access
- Fast chromatogram extraction
- Fast single spectrum access
- Scalable

In Beta Release for 12 months

Usage Statistics
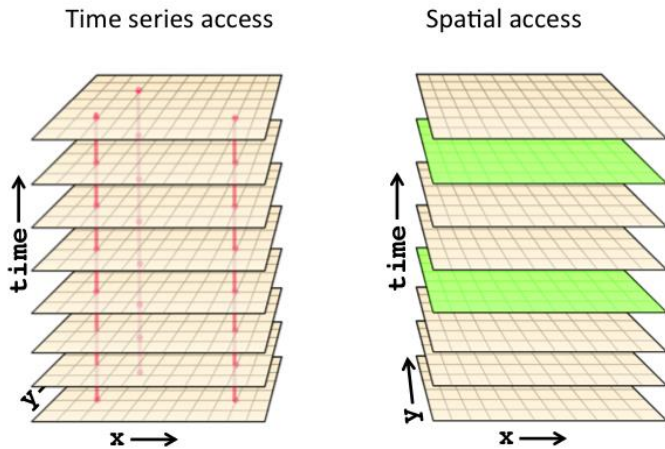
- 552 Users
- 7.09 Data volume (TB)
- 26,689 Data files
- 40 Public projects
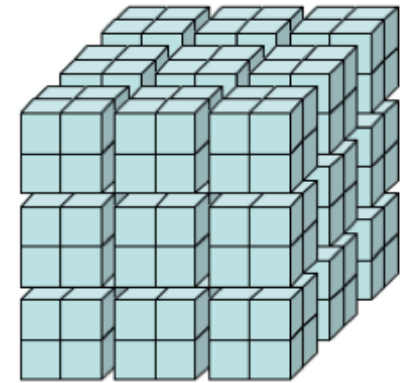- 111 Public Experiments

>1 TB Downloaded per Month

Using a Distributed Data Structure

Traditional Data file storage

Chorus Data Storage

Time series access   Spatial access

time
y
x

time
y
x

chunked

- Fast to get a spectrum
- Slow to get a chromatogram
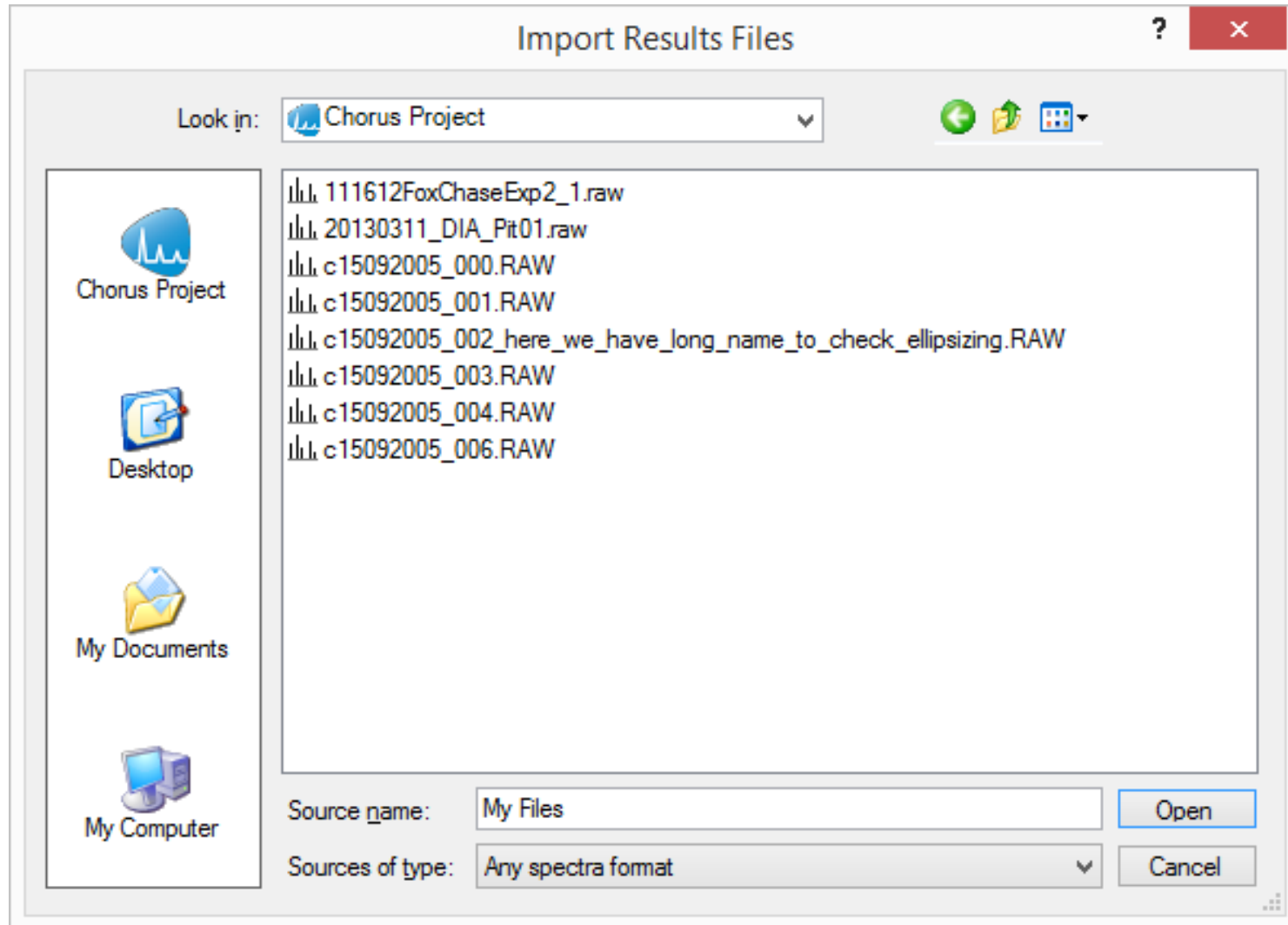
- Random access to the file
- Many processes can be used to extract many chromatograms/spectra using MapReduce

Skyline Direct Access to Chorus

Performance Tests Systems

▸ Desktop

- ▸ CPU: i7@3.5 GHz (7.8)
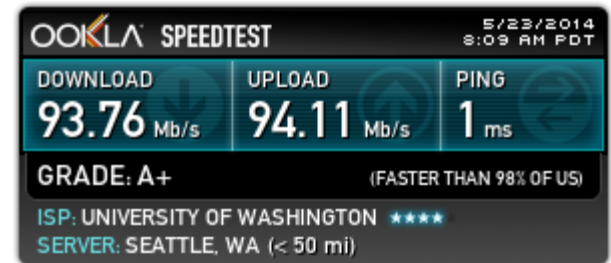- ▸ RAM: 16 GB (7.8)
- ▸ Drive: SSD (7.9)

▸ Laptop

- ▸ CPU: i7@1.8 GHz (6.9)
- ▸ RAM: 8 GB (7.6)
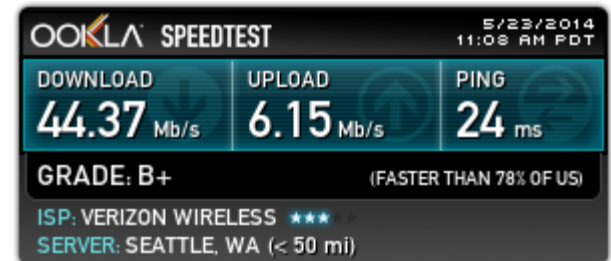- ▸ Drive: SSD (8.65)

Performance Tests Networks

▶ **University of Washington**

  ▶ Download: 93.76 Mb/s

  ▶ Upload: 94.11 Mb/s

▶ **Verizon**

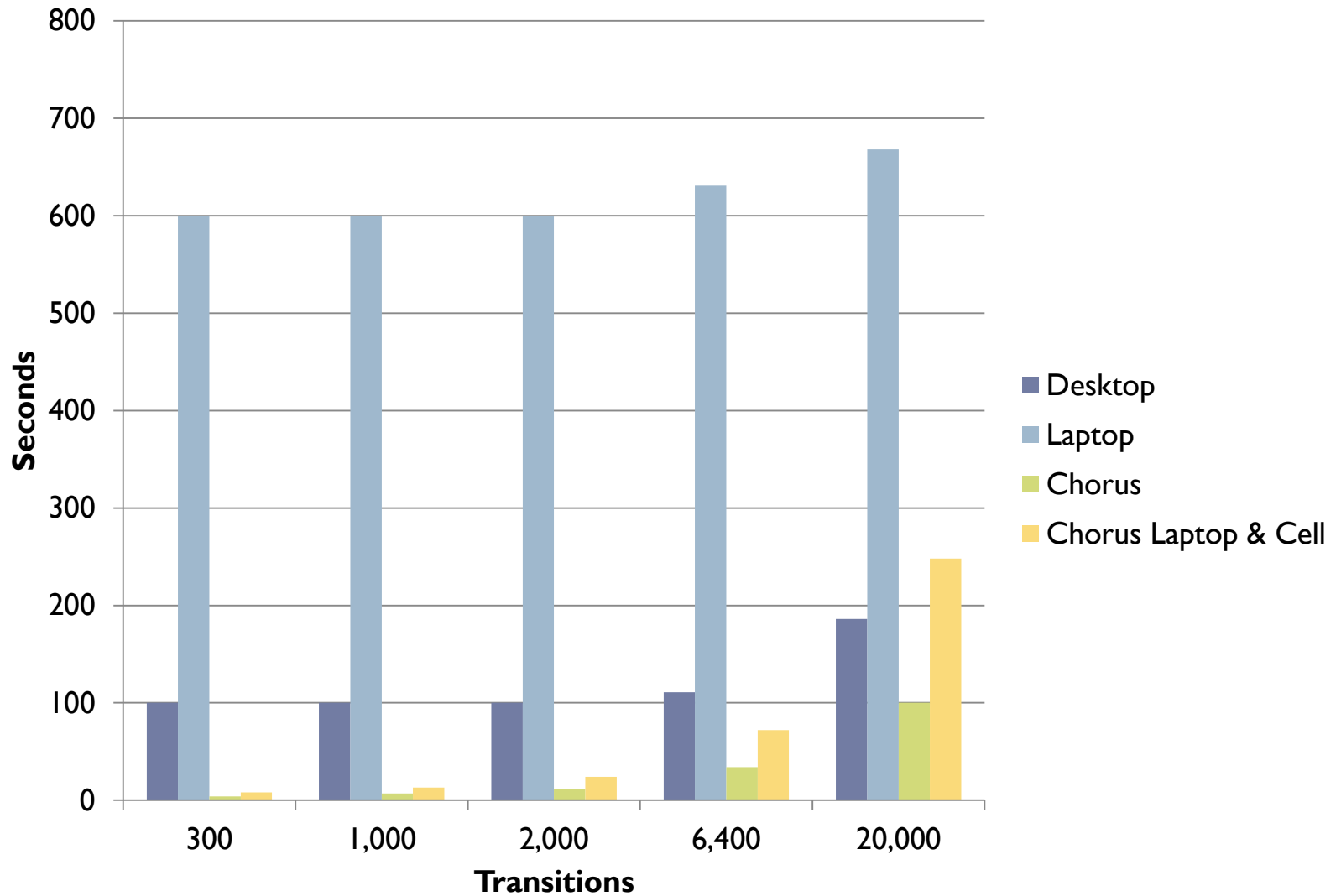 ▶ Download: 44.37 Mb/s

  ▶ Upload: 6.15 Mb/s
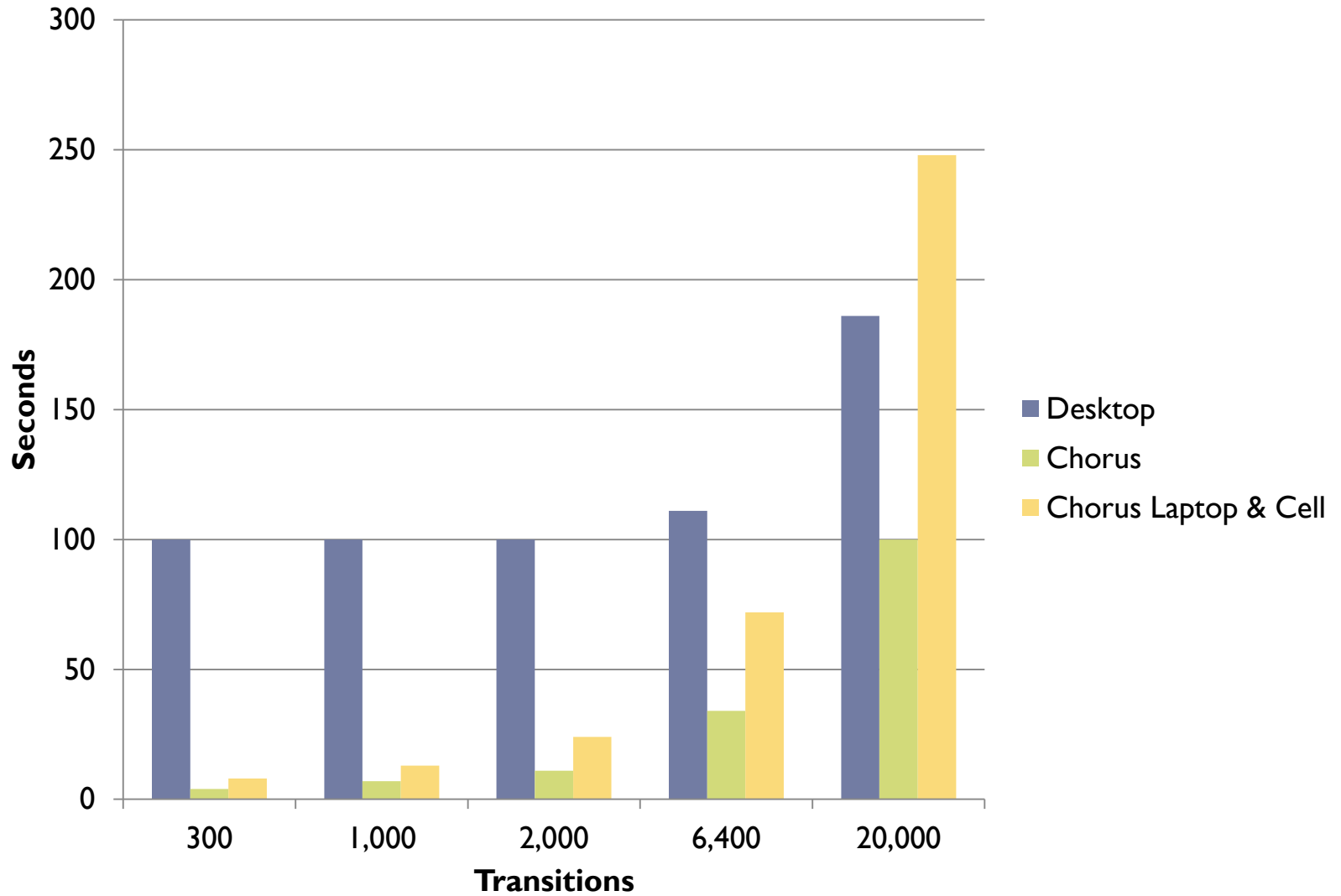

▶ **Baltimore Hilton**
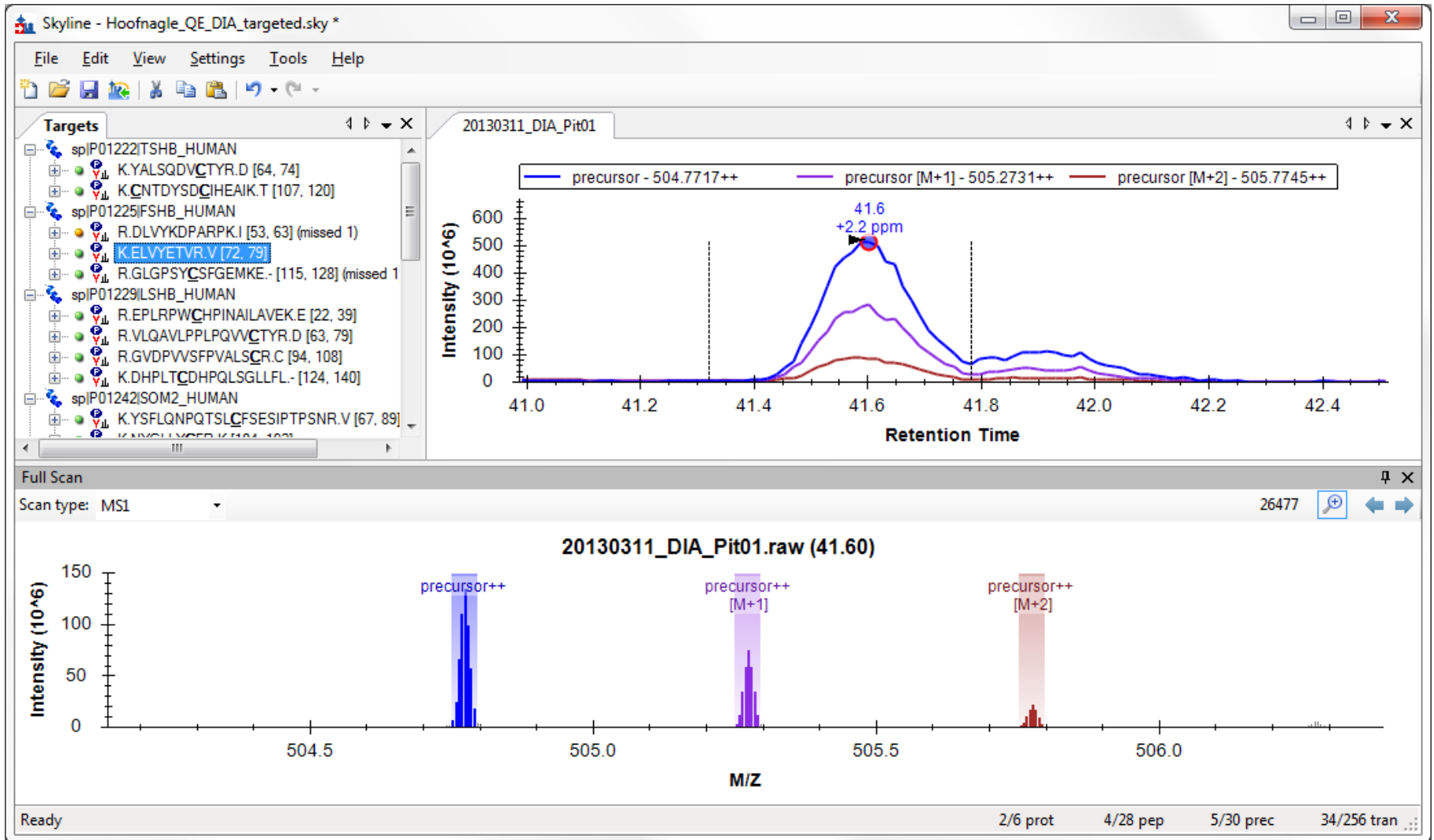
  ▶ Download: 11.76 Mb/s
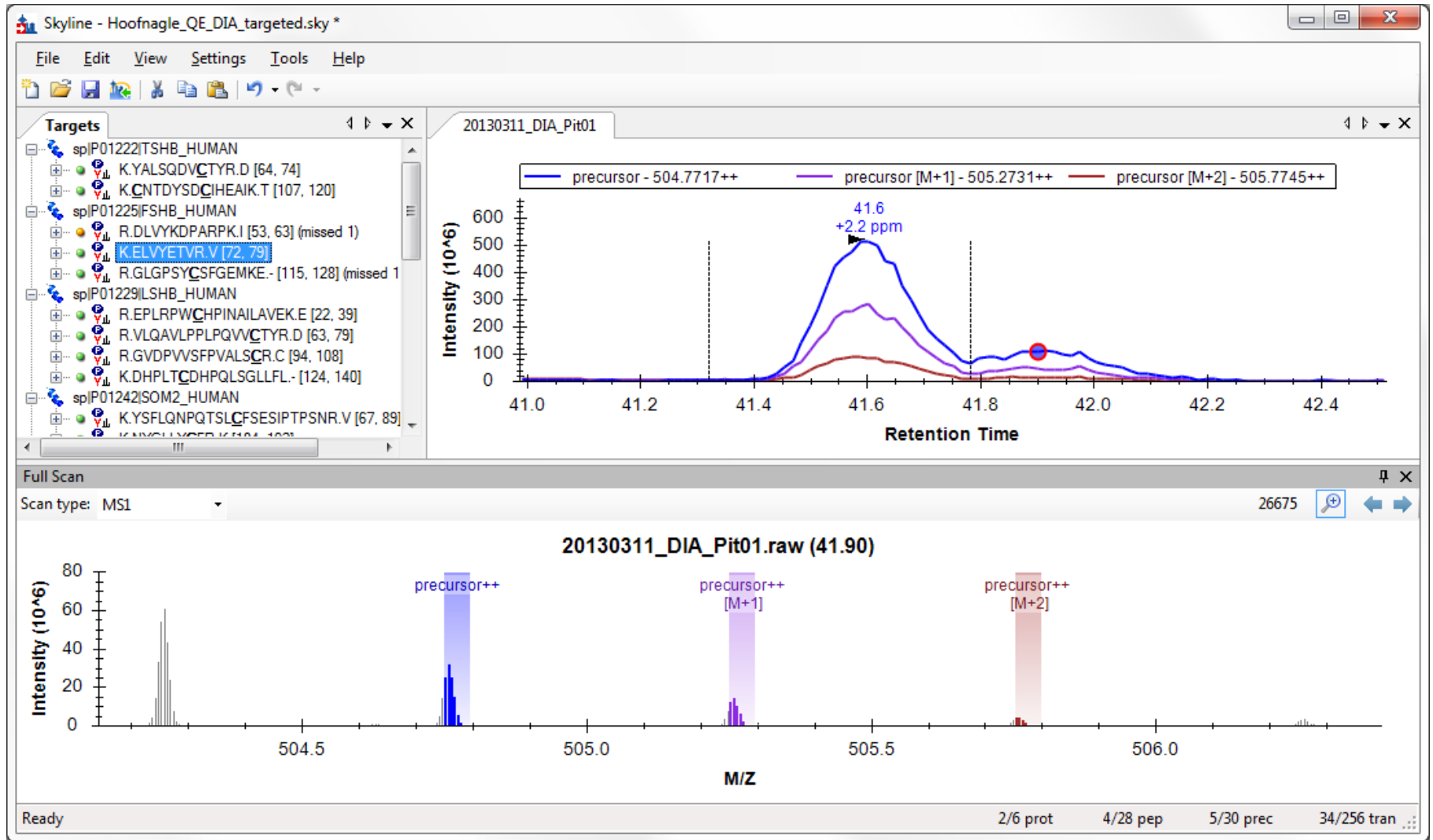
  ▶ Upload: 6.34 Mb/s
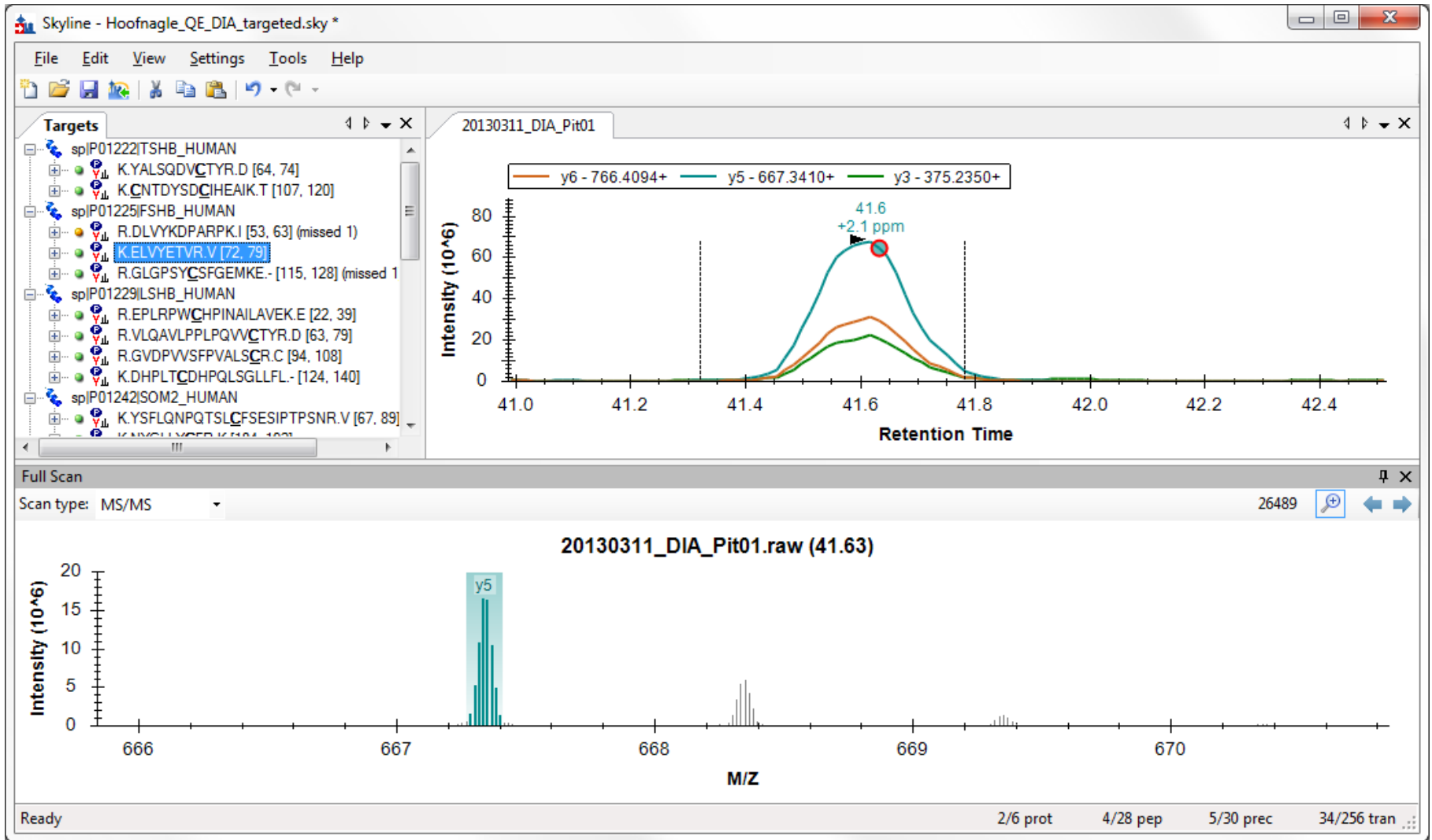
Data Import Performance
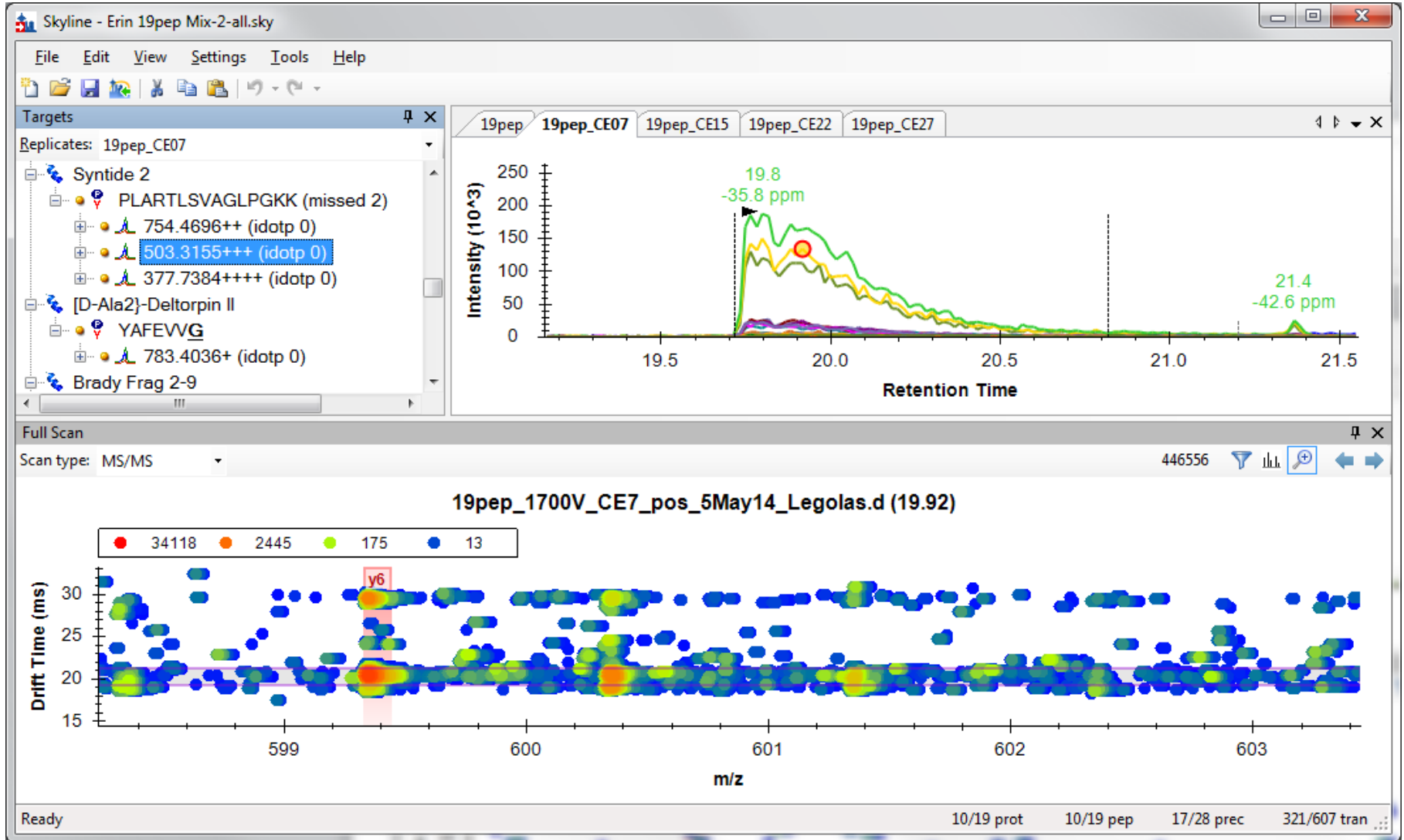
Data Import Performance

Full Scan Spectrum View

Exploring MS1 Scans

Exploring MS/MS Scans

Exploring Ion Mobility MS/MS Scans

Exploring Ion Mobility MS/MS Scans

19pep_1700V_CE7_pos_5May14_Legolas.d (19.92)
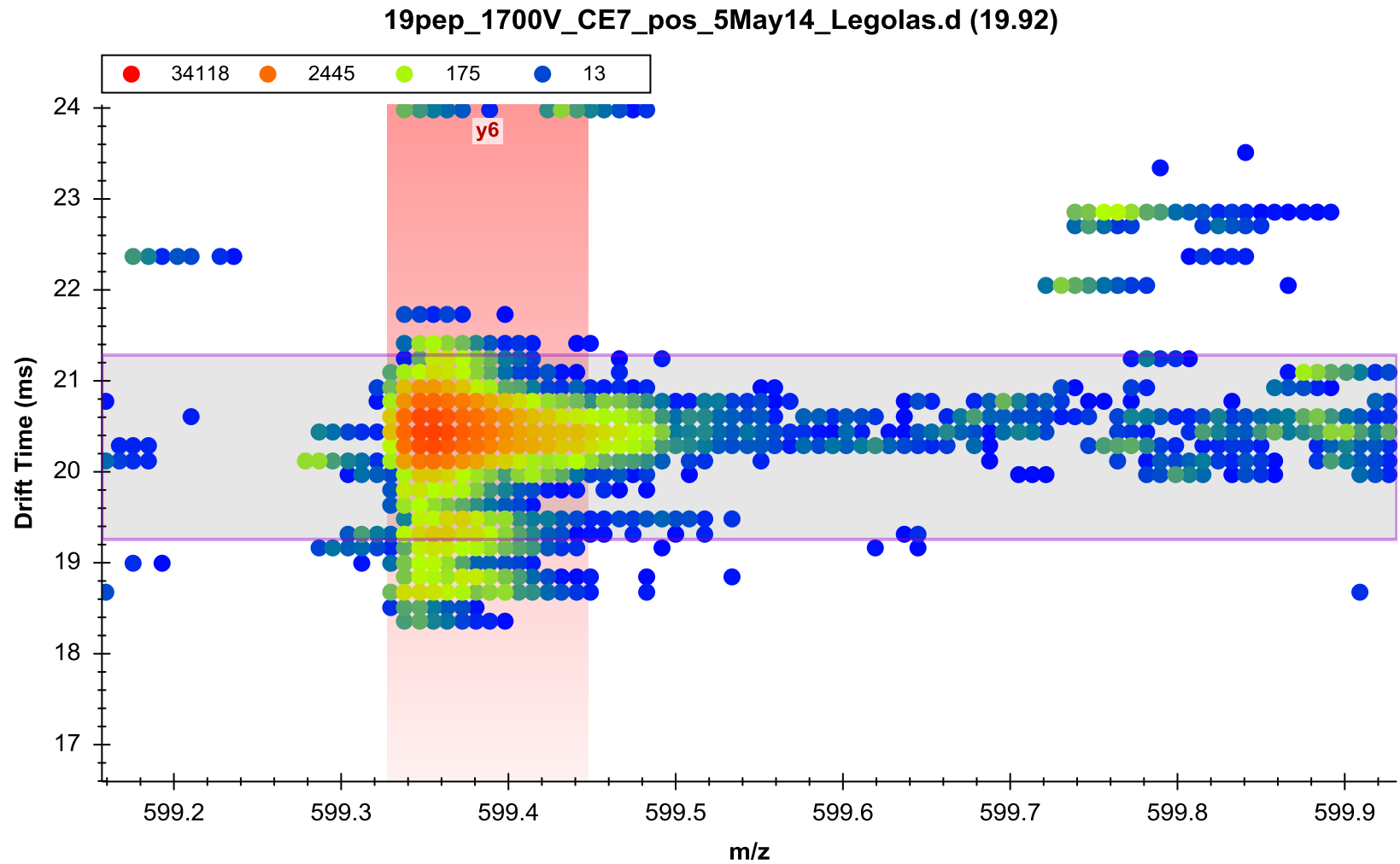
Imagine

▸ Files automatically posted to Chorus

▸ Fast chromatogram extraction to Skyline

▸ Sharable Skyline documents

▸ Reprocess data on a laptop without download

▸ Spectrum access from anywhere

▸ More processing and viewing options on Chorus

▸ Processed Skyline documents on Panorama

▸ Integrated systems: Chorus, Skyline and Panorama

Skyline Team

- Nick Shulman

- Don Marsh

- Brian Pratt

- Dario Amodei

- Vagisha Sharma

- Kaipo Tamura

- Yuval Boss
- Jarrett Egertson
- Max Horrowitz-Gelb
- Danny Broudy
- Trevor Killeen

Collaborators:

- **U. of Wa.**
  - Michael Bereman
  - Jim Bolinger
  - Jimmy Eng
  - Andrew Stergachis
  - Sonia Ting
- **Broad Institute**
  - Jake Jaffe
  - Steve Carr
  - Hasmik Keshishian
  - D. R. Mani

- **Buck Institute**
  - Birgit Schilling
  - Matthew Rardin
  - Brad Gibson
- **Duke**
  - Will Thompson
  - Arthur Moseley
- **IMSB**
  - Rudolph Aebersold
  - Christina Ludwig
  - Olga Schubert
  - Hannes Röst
  - George Rosenburger
  - Lucia Espona Pernas

- **PNNL**
  - Sam Payne
  - Sangtae Kim
- **Purdue**
  - Meena Choi
  - Olga Vitek
- **Stanford**
  - Dario Amodei
  - Parag Mallick
- **Vanderbilt**
  - Matthew Chambers
  - Daniel Liebler
  - David Tabb

Instrument Vendor Collaborators

- **Agilent Technologies**
  - Christine Miller
  - Joe Roark
  - Juli Salcedo
  - Shripad Torvi
- **Bruker**
  - Carsten Baessmann
  - Marius Kallhardt
  - Stephanie Kaspar
  - Pierre-Olivier Schmit
- **AB Sciex**
  - David Cox
  - Christie Hunter
  - Brent Lefebvre
  - Steve Tate

- **Shimadzu**
  - Alan Baynes
  - Junko Iida
  - Neil Loftus
  - Kiriko Matsuo
- **Thermo-Scientific**
  - Sue Abbatiello
  - Markus Kellmann
  - Andreas Kuehn
  - Vlad Zabrouskov
- **Waters**
  - James Langridge
  - Roy Martin
  - Kieran Neeson
  - Keith Richards