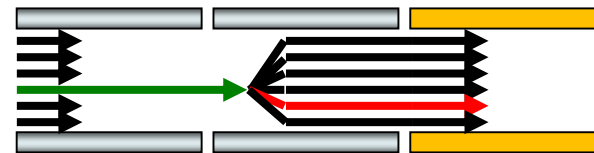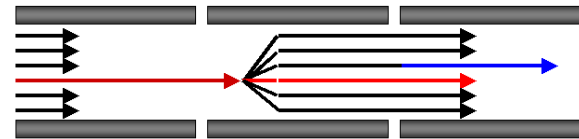# Skyline

Targeted Proteomics Environment

Building a Software Ecosystem for Targeted Proteomics

Brendan MacLean
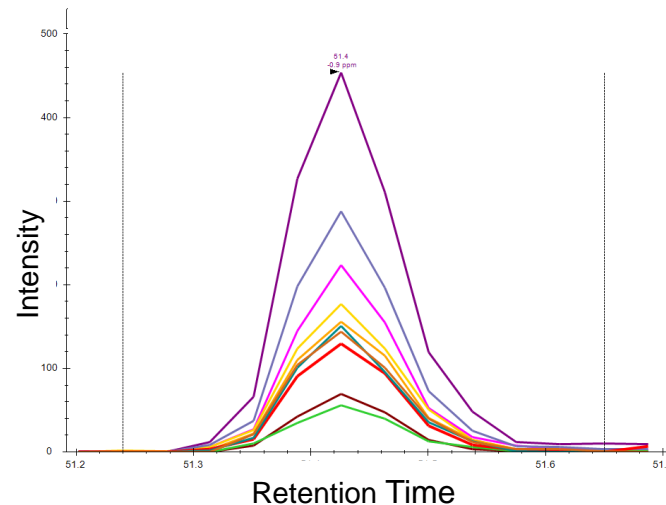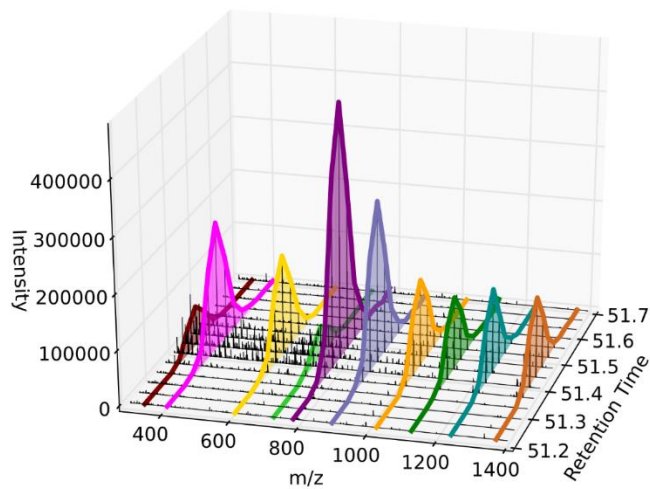MacCoss Lab

UNIVERSITY *of* WASHINGTON
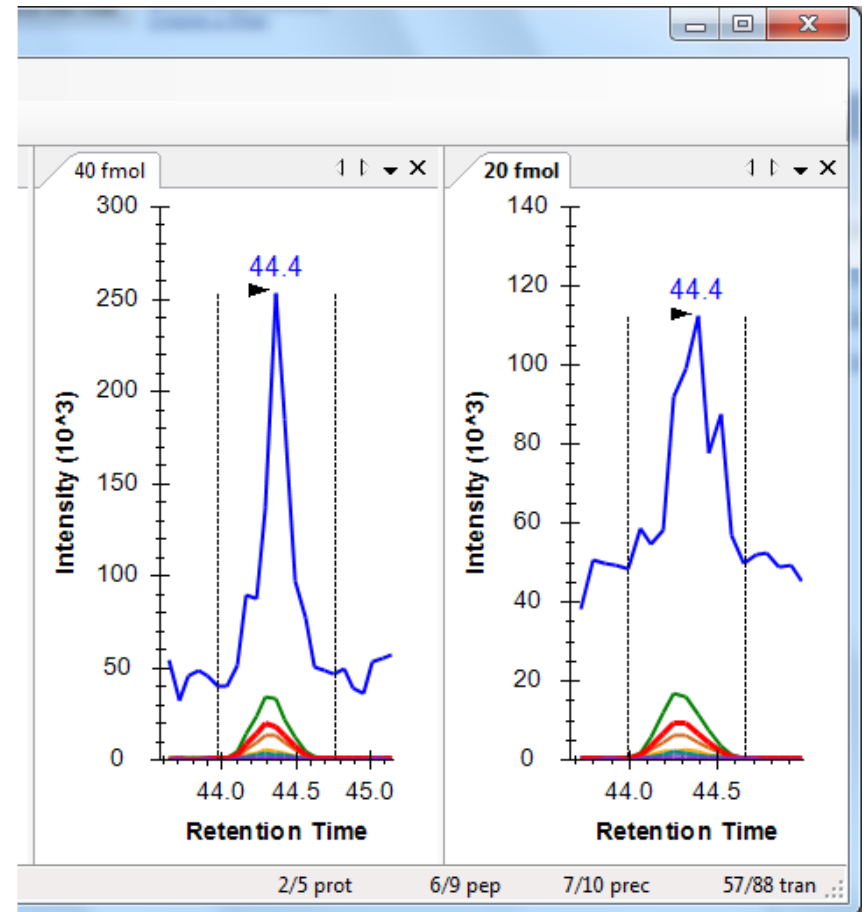
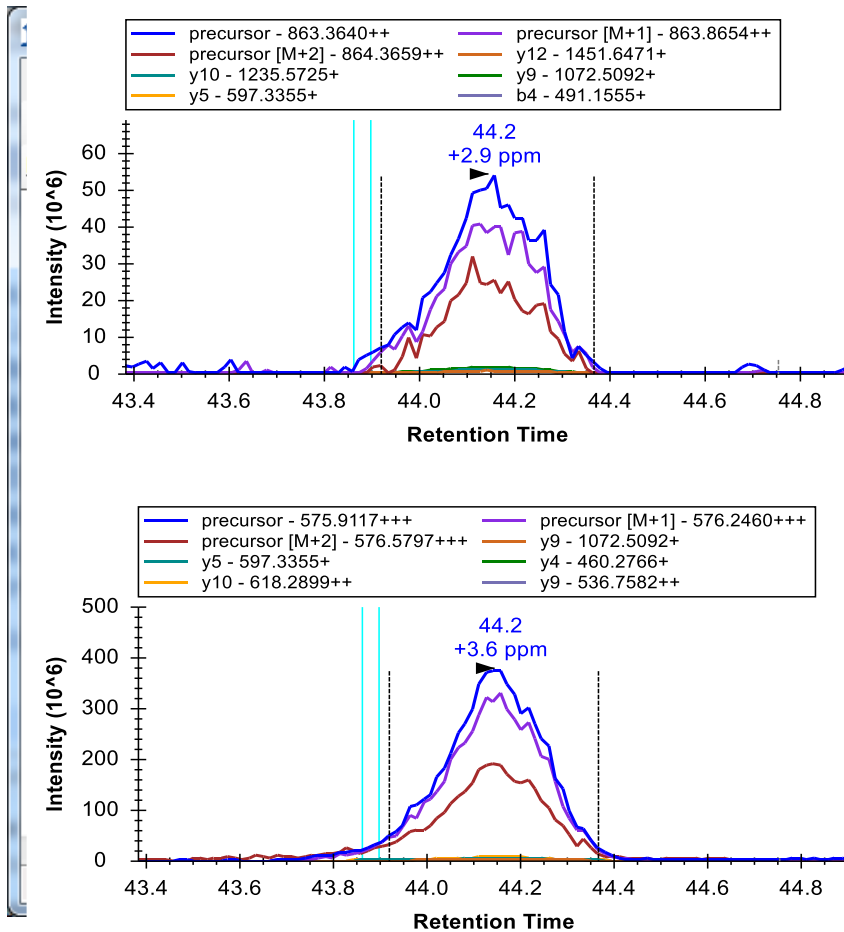# Chromatography-based Quantification

- **SRM – Selected ion chromatograms**
- PRM – Extracted ion chromatograms
- DIA – Extracted ion chromatograms
- DDA – Extracted ion chromatograms from MS1-only

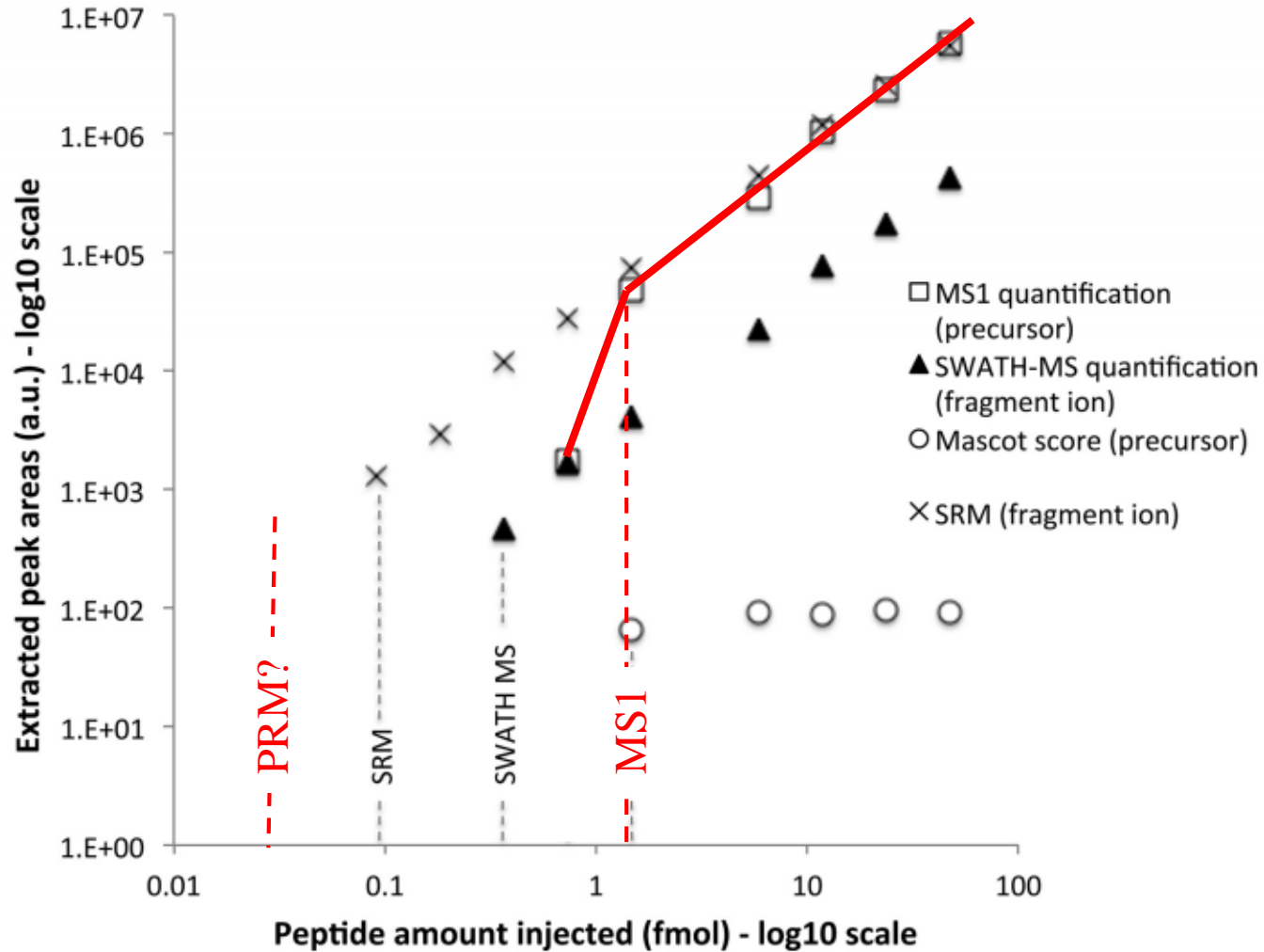| Acquisition	Targeted	Survey
More Selective	PRM	DIA
Less Selective	SRM	DDA

When Selectivity Becomes Sensitivity

Decreasing noise increases signal-to-noise

Dynamic Range

Skyline Project Overview

- 7th year in development
- 6 developers
- over 400,000 lines of code
- 60,000 lines of code for testing
- over 5000 registered users
- over 7000 instances started each week
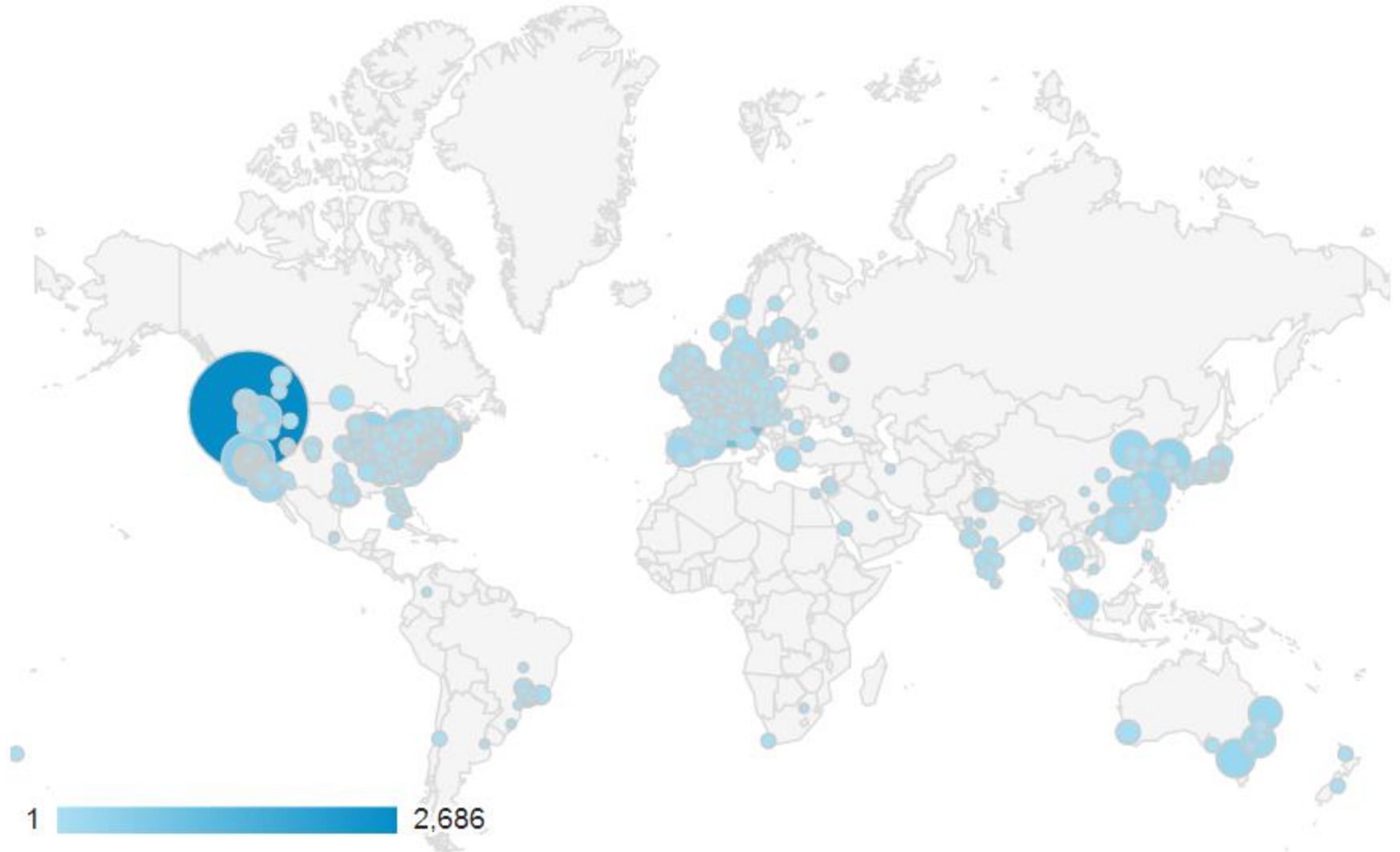- funded by NIH and 6 instrument vendors
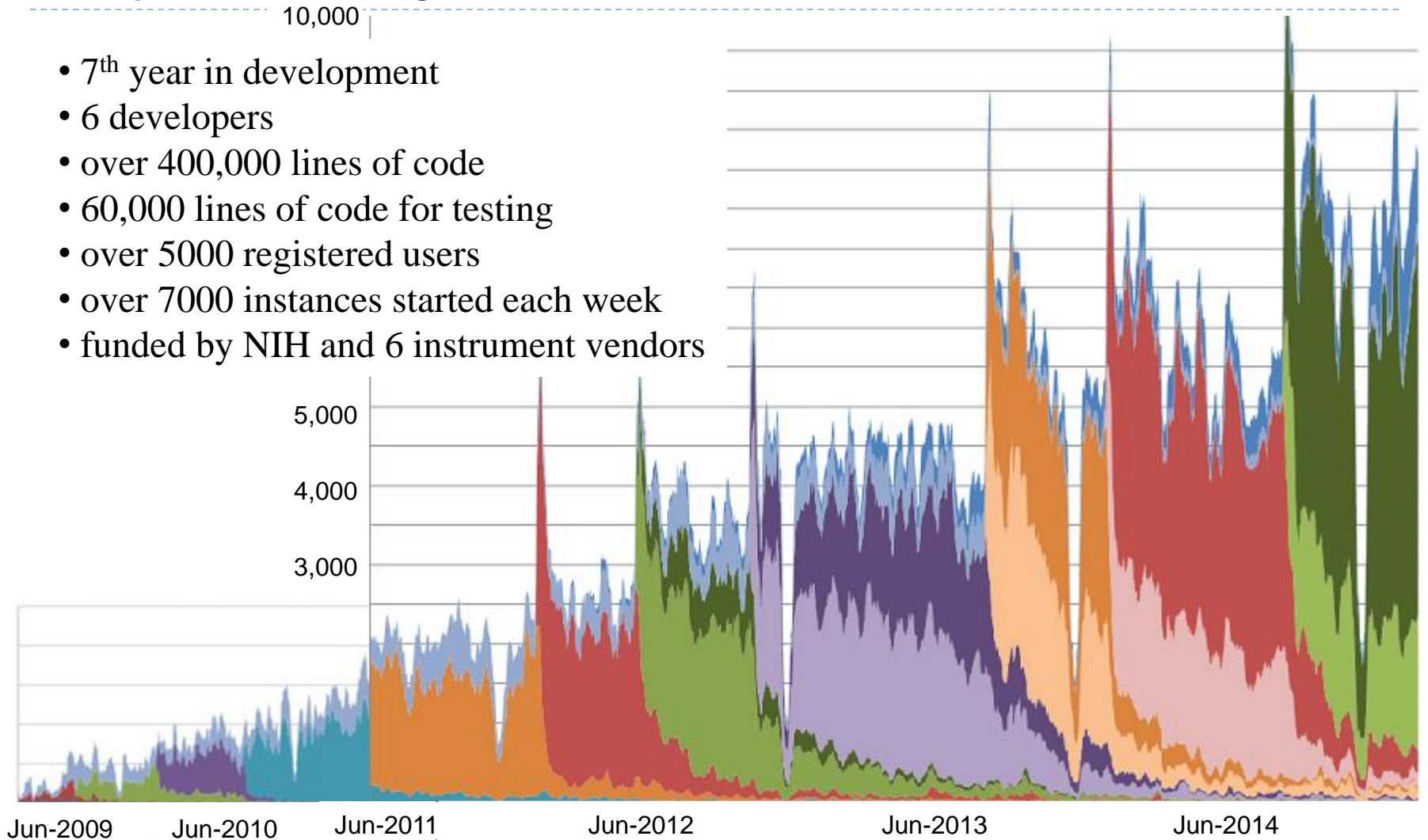
Multiple Instrument Vendors

Skyline Web Site Visits (past 6 months)

1 ▭▭▭▭▭ 2,686

"Method of the year 2012" Nature Methods

Targeted proteomics

Analysis of a preselected group of proteins delivers more precise, quantitative, sensitive data to more biologists. Vivien Marx reports.

Although the number and identity of protein-coding genes in humans and many other organisms are known to a certain level of approximation, the numbers of proteins produced by each of these genes remains a mystery. Further complicating matters, given the many possible splice forms and post-translational modifications, the potential number of proteins is "staggering," says Arizona State University researcher Josh LaBaer, who is also president-elect of the US Human Proteome Organization. A protein is also dynamic. "It's phosphorylated this minute; it's not phosphorylated the next minute." he says. This is fascinating science.

"I personally can't wait until we stop hearing about someone describing how big of a list of proteins, peptides or phosphopeptides they detected," says one researcher critical of discovery proteomics who did not wish to be identified. Proteomics has been doing "my list is bigger than your list" for far too long. "It is way more important to measure the one right protein than 10,000 wrong ones."

Scientists wanting to follow well-founded hunches about dozens or hundreds of proteins seek a focused, reproducible, quantitative view of a small subset of the whole proteome in their
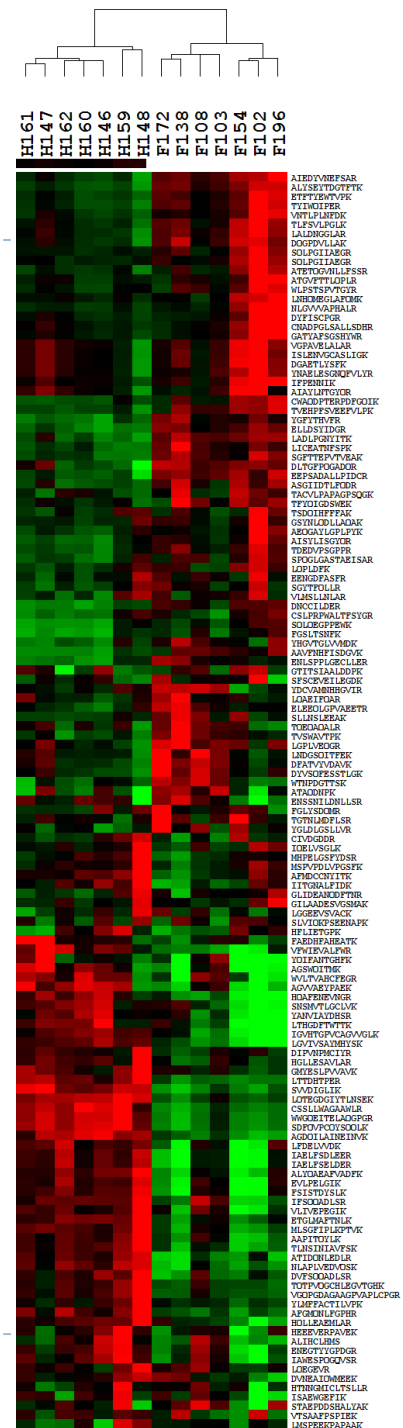
Why Go Targeted?

- Dynamic range
- Precision
- Cost effective verification & validation
- Repeatability

- Multiple hypothesis testing & error rates
  - Reiter, MCP 2009
  - Ezkurdia, JPR 2014

- DDA runs identify 4,500 peptides at 1% FDR
 - What is the error rate for quantification?

On Error Rates

‣ DDA runs identify 4,500 peptides with 1% FDR

‣ Each is assigned a quantitative value of 10 runs

‣ 100 seem to be changing by at least 2-fold

‣ 90 are already explained by the literature

‣ 10 seem novel and exciting

‣ 1,000 labs have the same experience
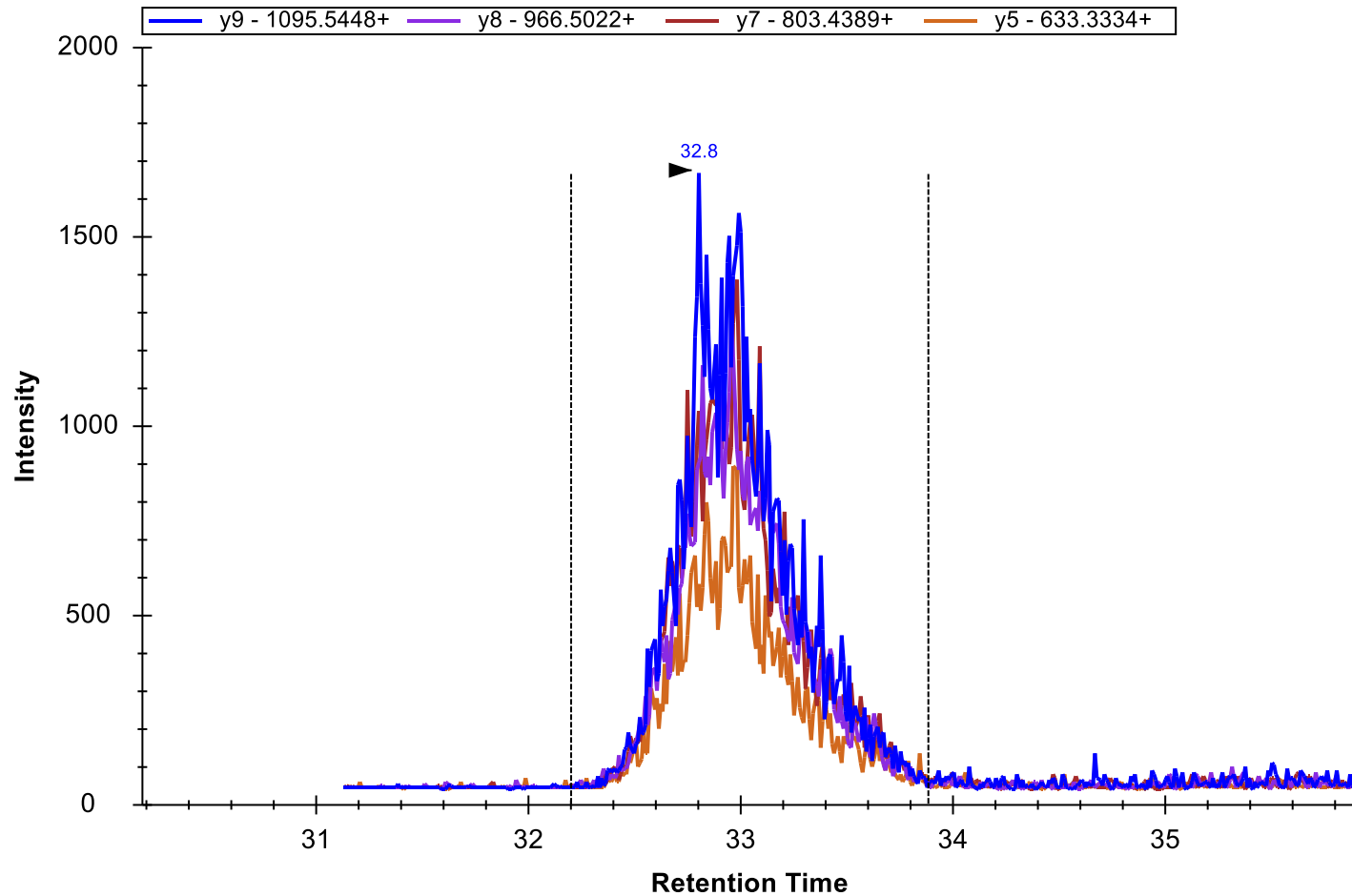
‣ What is the error rate among the 10,000 novel peptides?

Depends on Chromatography
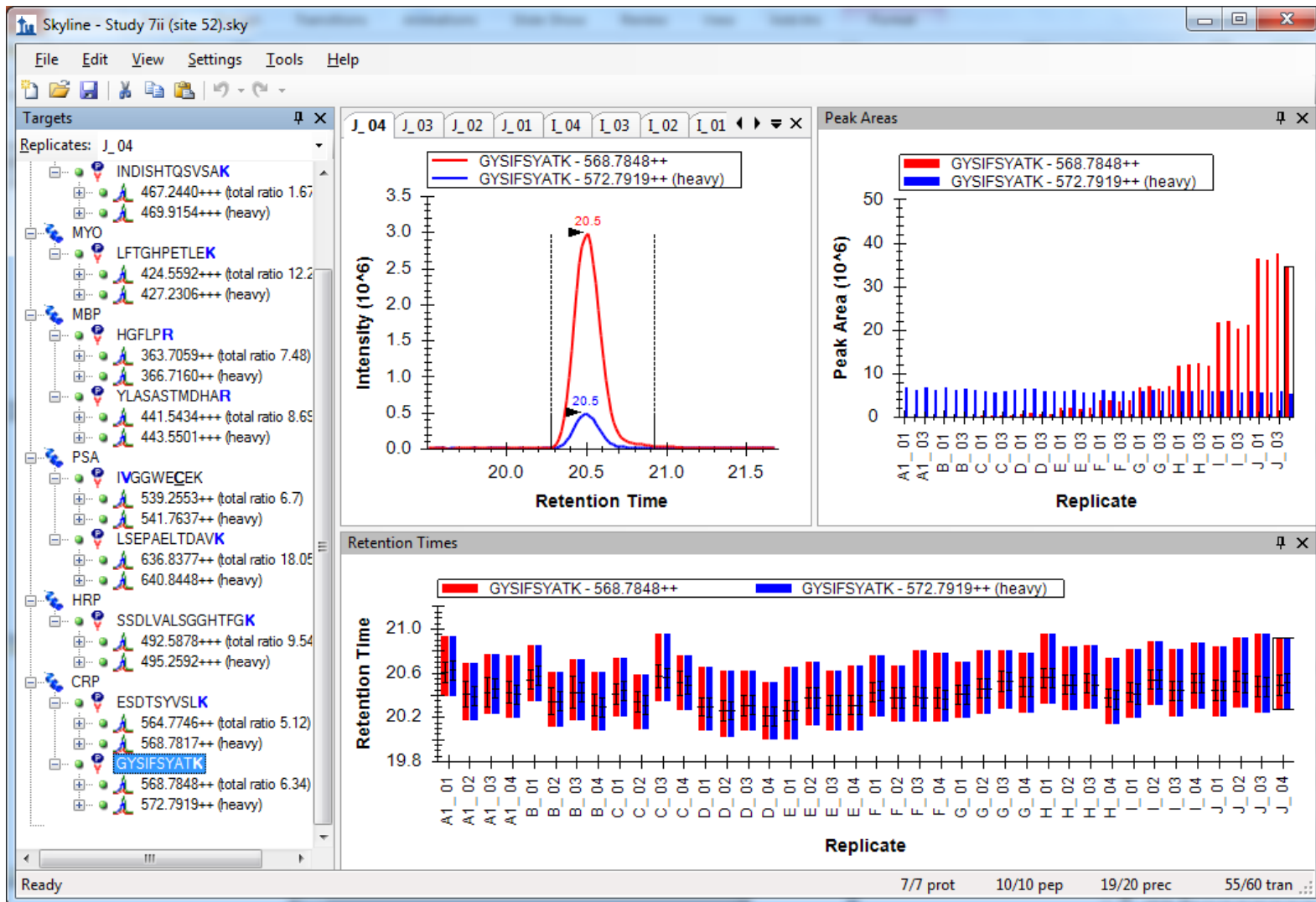
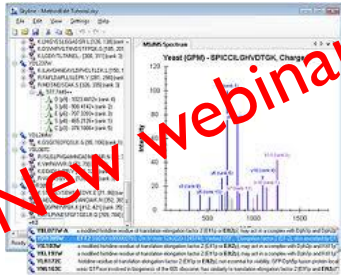Very good chromatography!
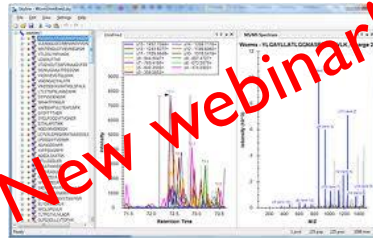
Depends on Chromatography

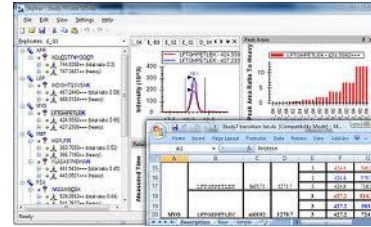Poor chromatography

12 Tutorials (5 translated)

Method Editing

New webinar!

Method Refinement

New webinar!

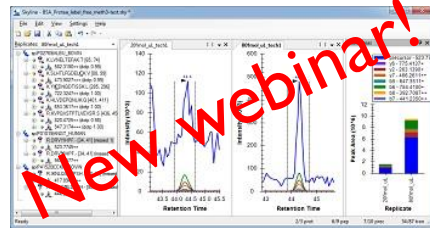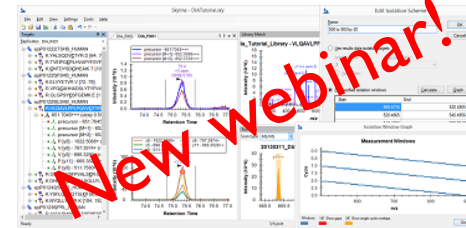**Existing Experiments**

Absolute Quantification

MS1 Filtering

New webinar!

Targeted MS/MS (PRM)
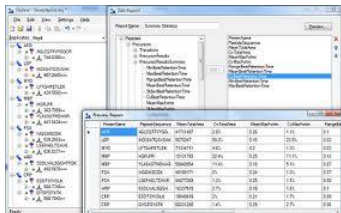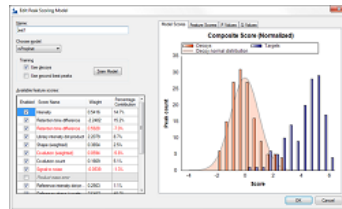

New webinar!

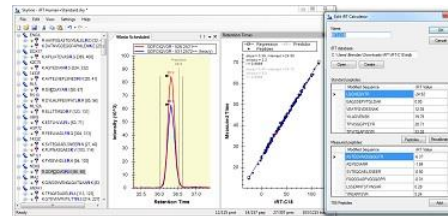**DIA with DDA**

New webinar!

Library Explorer
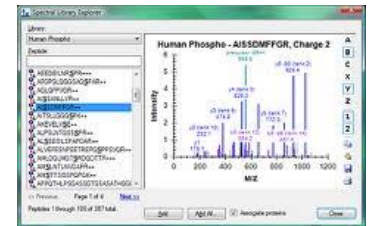
Custom Reports

Advanced Peak Picking

iRT

CE Optimization
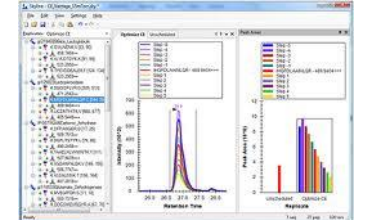
Growing…

Chinese Skyline (2014)

Learn More

▸ Webinar #6: Effective Data Processing & Interrogation

  ▸ Tuesday, April 14.

▸ Weeklong Course at PRBB, Barcelona 2014 – videos

  ▸ http://tinyurl.com/skyline-embo-2014

▸ **Workshop at US HUPO, Tempe, March 15**

▸ Workshop at ASMS, St. Louis, May 30-31

▸ Workshop in Rio de Janiero, August 31-September 2

▸ Weeklong Course at PRBB, Barcelona

  ▸ November 15-20

▸ More in 2016…

▸

6 External Tools
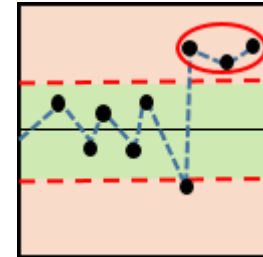
Msstats (Vitek)

grouped study statistics

QuaSAR (Carr)

response curve statistics

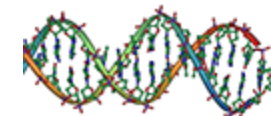### SProCoP (Bereman)

system suitability

MS1 Probe (Gibson)

MS1 filtering statistics

Protter (Wollscheid)

transmembrane topology

Population Variation


mutation frequency

Broudy, et. al, *Bioinformatics 2014*

Aggregating and Publishing

▸ Publish fully annotated Skyline documents

▸ Build chromatogram libraries

▸ Aggregate lab QC data

▸ Free hosted version (http://panoramaweb.org)

    ▸ >90 separate projects so far (CPTAC, LINCS & ABRF sPRG)

    ▸ >1000 data sets uploaded

    ▸ User controlled security

▸ Locally installable server application

    ▸ Genentech, J&J, Merck and Roche

▸ Free and open source (Apache 2.0)
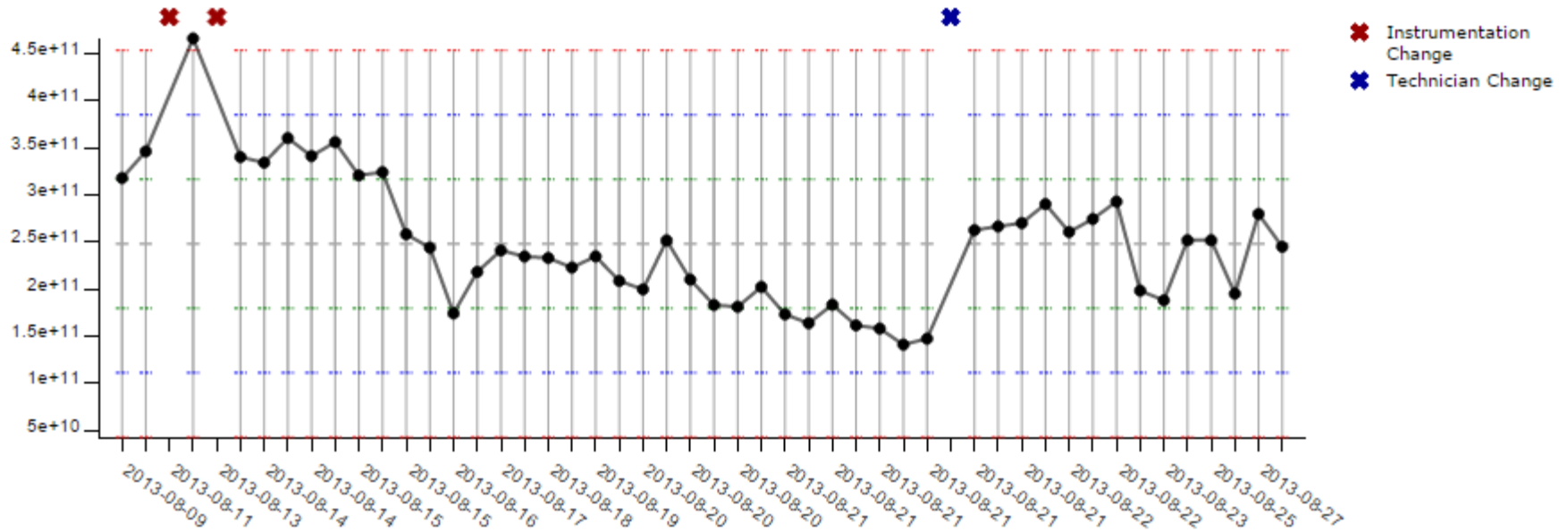
Skyline – Panorama Workflow
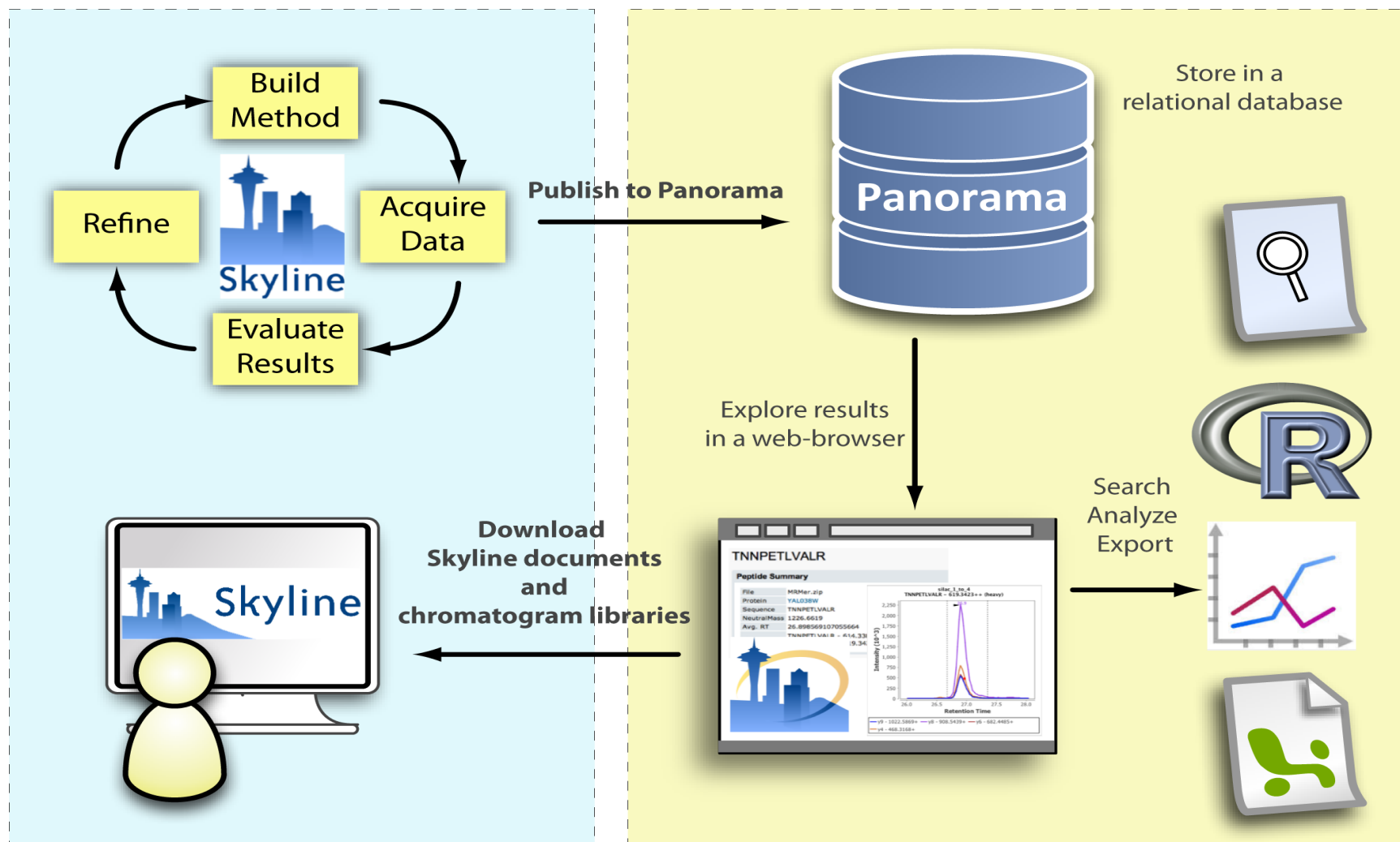
Panorama QC Folder

Chorus For Mass Spec File Storage

- Google Docs-like interface
- Lab-centered security model
- Cloud storage for raw data
  - 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 16 months

Usage Statistics

- 864 Users
- 13.77 Data volume (TB)
- 47,973 Data files
- 63 Public projects
- 187 Public Experiments

Chorus

>1 TB Downloaded per Month

Remote Extraction of Chromatograms

Skyline Requests Extracted Ion Chromatograms

CHORUS Generates Chromatograms and Returns Skyline Cache