Panorama
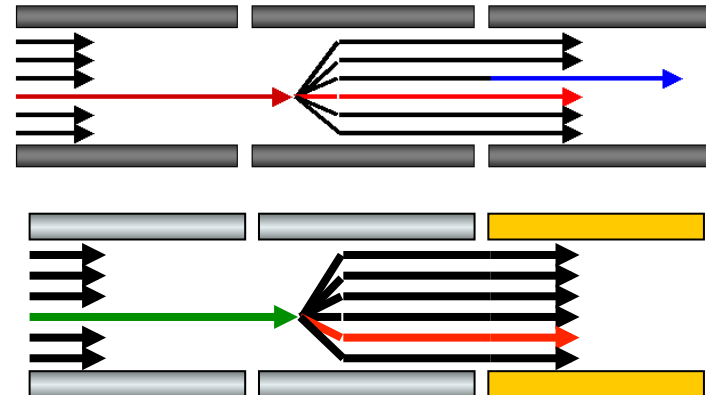
Targeted Proteomics Knowledge Base

Managing and Analyzing Large Data Sets

Webinar 3

Brendan MacLean
Vagisha Sharma
Josh Eckels

UNIVERSITY *of* WASHINGTON

An Ecosystem for Targeted Proteomics

- Raw data storage
- Chromatogram extraction
- Hosted on Amazon cloud

- Targeted method design
- Data processing and visualization

- Experiment repository
- Targeted library building
- Custom analysis and presentation
- Quality control (future)
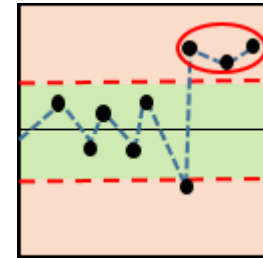
Ecosystem: External Tools

Msstats (Vitek)

grouped study statistics

QuaSAR (Carr)

response curve statistics

SProCoP (MacCoss)

system suitability

MS1 Probe (Gibson)

MS1 filtering statistics

Protter (Wollscheid)

transmembrane topology

Population Variation (Smith)

mutation frequency

Broudy, et. al, *Bioinformatics 2014*

Ecosystem: External Tools

Msstats (Vitek)

grouped study statistics

QuaSAR (Carr)

response curve statistics
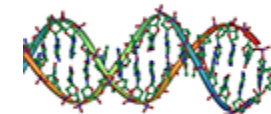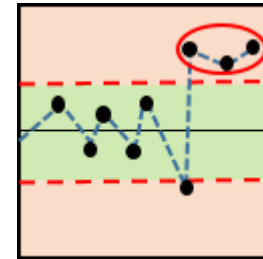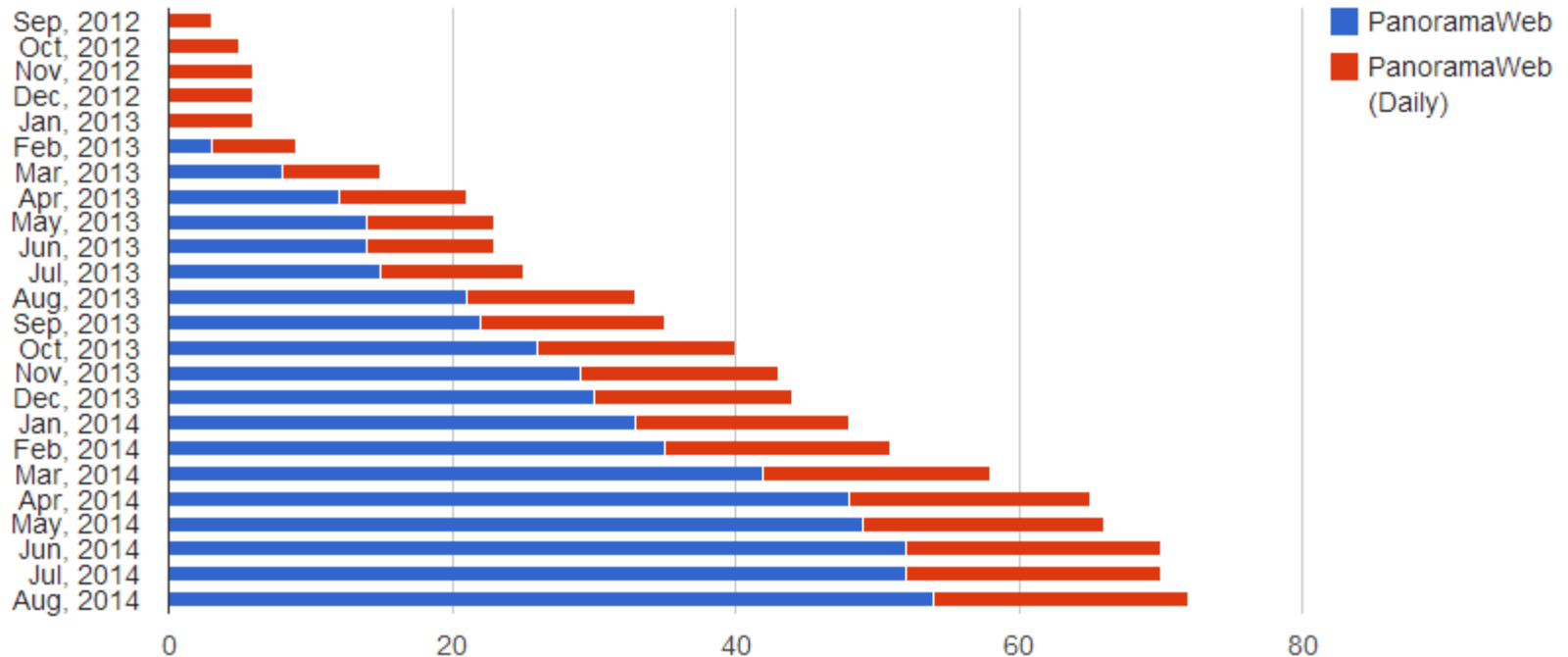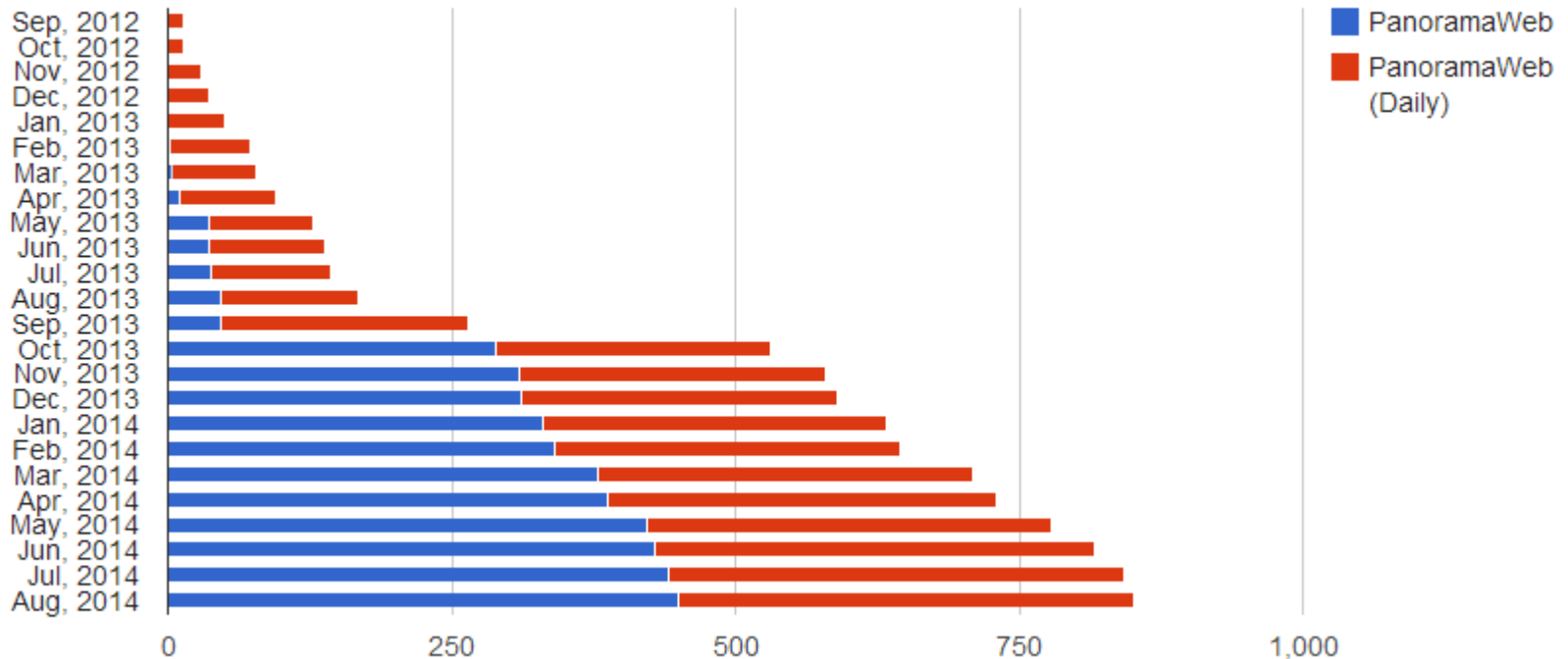
SProCoP (MacCoss)

system suitability

Panorama Nearly 2 Years Old

Projects on PanoramaWeb and PanoramaWeb (Daily)

Panorama Nearly 2 Years Old

Skyline documents on PanoramaWeb and PanoramaWeb (Daily)

PanoramaWeb.org

Funding

‣ LINCS through 2020

‣ Skyline R01 (NIGMS)

‣ Instrument vendor support

‣ Panorama Partners Program

  ‣ Seeking leading edge targeted proteomics infrastructure

  ‣ Local Panorama server installation

  ‣ Direct collaboration with Skyline/Panorama team

  ‣ Full support for 1 year

  ‣ Contact Josh Eckels (jeckels@labkey.com)

Two Prior Webinars

- An introduction to using PanoramaWeb (Feb 7, 2013)
  - Tutorial: Panorama Sharing Skyline Documents
- New features in Panorama v13.2 (Aug 8, 2013)
  - Tutorial: Panorama Chromatogram Libraries

- Managing and Analyzing Large Data Sets
  - Tutorial: Coming soon…

Panorama Paper Published

Panorama: A Targeted Proteomics Knowledge Base

Journal of Proteome Research, August 7, 2014

Vagisha Sharma , Josh Eckels , Greg K Taylor , Nicholas J Shulman , Andrew B. Stergachis , Shannon A Joyner , Ping Yan , Jeffrey R. Whiteaker , Goran N Halusa , Birgit Schilling , Bradford Wayne Gibson , Christopher M Colangelo , Amanda G. Paulovich , Steven A. Carr , Jacob D Jaffe , Michael J. MacCoss , and Brendan MacLean

Supporting Large Projects (2013-14)

- ## CPTAC Assay Portal
  - Assays.cancer.gov
  - CPTAC Assay Portal: a repository of targeted proteomic assays. *Nature Methods 2014*
  - http://tinyurl.com/Skyline-ugm-2014-whiteaker
- ## ABRF sPRG Study
  - https://panoramaweb.org/labkey/ABRF-sPRG-2013.url
  - http://tinyurl.com/Skyline-ugm-2014-colangelo
- ## LINCS
  - Lincscloud.org
  - https://daily.panoramaweb.org/labkey/LINCS.url
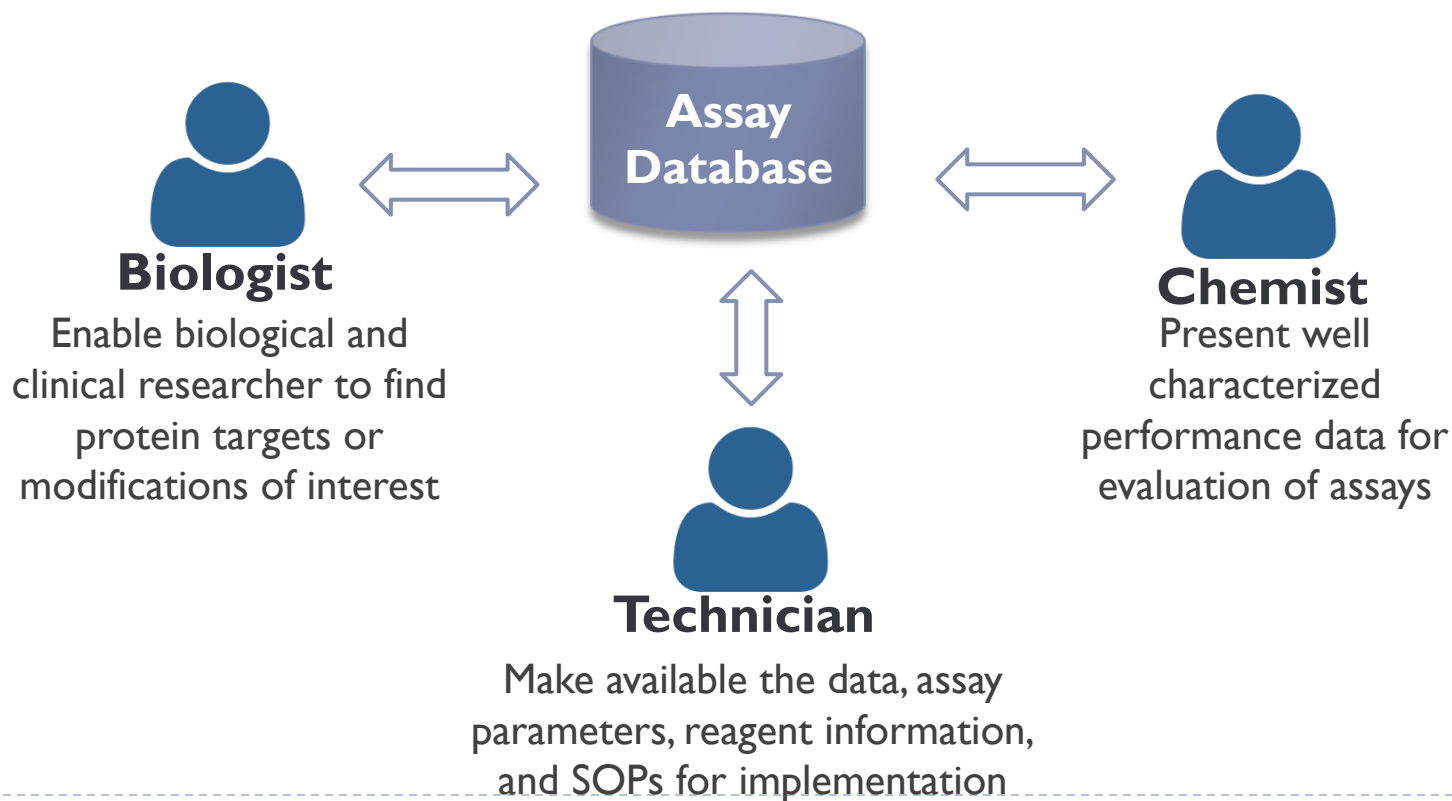
Panorama: beyond a data repository

▸ Multi-lab, collaborative projects on PanoramaWeb have allowed us to explore its potential beyond data management

  ▸ CPTAC Assay Portal

  ▸ ABRF sPRG 2013 study

  ▸ LINCS

▸ Used features already available in LabKey Server:

  ▸ Ability to write custom SQL queries and create specialized views on data tables

  ▸ Built-in user interface for analyzing data with the R

  ▸ Availability of client libraries to allow programmatic access to data

  ▸ JavaScript library includes APIs for building custom user interfaces

CPTAC Assay Portal

- **assays.cancer.gov**
  - CPTAC Assay Portal: a repository of targeted proteomic assays. *Nature Methods* (2014)
- Promote the development and dissemination of well characterized mass spectrometry-based proteomic assays
- Link a broad user community to a valuable assay resource

Biologist
Enable biological and clinical researcher to find protein targets or modifications of interest

Chemist
Present well characterized performance data for evaluation of assays

Technician
Make available the data, assay parameters, reagent information, and SOPs for implementation

Assay Database

CPTAC Assay Portal

- Integration with Panorama
  - Use PanoramaWeb as a data repository
  - Custom SQL queries provide input to R scripts developed by CPTAC for response curves and QC analysis
  - Analysis scripts stored and run on PanoramaWeb
  - Use the LabKey Client APIs to retrieve data and analysis results from Panorama for viewing in the web portal
  - Provide access to chromatogram libraries built in Panorama

ABRF sPRG (2013-2014 study)

▸ Focused on the ability of core facilities to determine relative quantitation of up to 1000 heavy/light peptide pairs in a single sample.

▸ Samples requested by 120 labs.

▸ Of the 90 participating labs, 47 returned data.

▸ Participants uploaded data to PanoramaWeb and the sPRG NIST server.

▸ Folders were created on PanoramaWeb for each participant.

  ▸ Participants were assigned a username and password for accessing their folder.

▸ Excel spreadsheets containing results from participants that did not use Skyline for data processing were added to PanoramaWeb by ABRF members.

ABRF sPRG (2013-2014 study)

- R scripts were developed by ABRF members to assess data quality.
- Analysis was done offline using the data downloaded from PanoramaWeb and Excel spreadsheets provided by participants.
- Summary results as well as results for each participant were uploaded to PanoramaWeb and added to wiki pages.
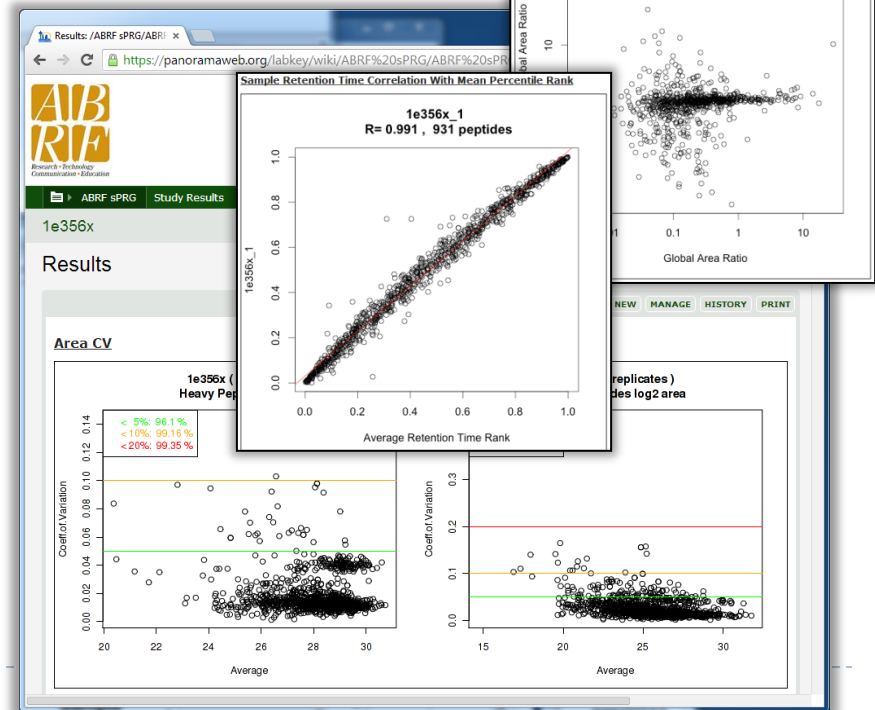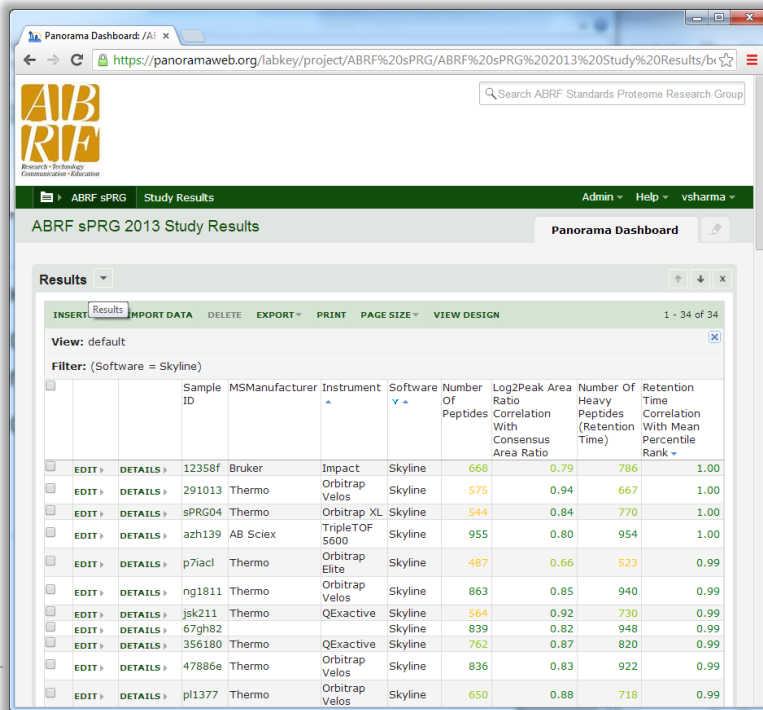
LINCS

‣ Library of Integrated Network-based Cellular Signatures

‣ PanoramaWeb hosts phosphoproteomic assays developed by the LINCS center at the Broad institute.

‣ P100 assay: detects and quantifies a representative set of ~100 phosphopeptide probes that are present in a wide range of cell types and have been demonstrated to be modulated via perturbations.

‣ More information on lincscloud.org

LINCS

▸ R scripts were written to generate GCT (Gene Cluster Text) files from annotated Skyline files uploaded to PanoramaWeb.

> ▸ GCT: tab-delimited file format that describes an expression dataset.

> ▸ Allows compatibility with the GENE-E visualization and analysis platform.

▸ Scripts for processing GCT data, developed at the Broad Institute, were also added to the LINCS project on PanoramaWeb.

▸ A custom user interface was developed to allow users access to the processed and unprocessed GCT files.

> ▸ GCT files can be viewed by launching the GENE-E viewer directly from Panorama.

LINCS

Demo

▸ Write a custom SQL query

    ▸ https://www.labkey.org/wiki/home/Documentation/page.view?name=customSQL

    ▸ Create input for CPTAC's R script for generating response curves

▸ Create an R view

    ▸ https://www.labkey.org/wiki/home/Documentation/page.view?name=rViews

    ▸ Run CPTAC's R script

▸ Build cutsom UI

    ▸ https://www.labkey.org/wiki/home/Documentation/page.view?name=javascriptTutorial

    ▸ Build a simple UI for generating the response curve for a selected protein and peptide

▸

Reporting – Full Spectrum of Capabilities

- **Simple and quick**
  - Built-in plotting/visualization tools
  - No development skills needed
  - Point-and-click

- **Extensible, scalable, and customizable**
  - Useful within the web server or externally
  - RServe
  - knitr

Built-In Visualization Tools

- ▸ Chart menu
  - ▸ Available as a button on most data grids
  - ▸ Point-and-click to create scatter and box plots
  - ▸ Accessible for all users
  - ▸ Rapidly expanding capabilities
- ▸ Conditional formatting
  - ▸ Highlight values that meet certain criteria
  - ▸ Bring attention to outliers
  - ▸ Configured by administrators
  - ▸ Similar to Excel

RServe

- Allows execution of R scripts on a remote server
- Transparent to R script developers
- Offloads work from the web server
- Avoids need to initialize R for each invocation
- Configured through LabKey Server's Admin Console
- Typically not required early on, but enables scaling
- http://rforge.net/Rserve

knitr

▸ Dynamic report generation with R

▸ Combines R script with HTML, Markdown, or other output languages

▸ Allows easy caching and reuse of previously generated results

▸ Enabled on a per-script basis in the LabKey Server R script editor

▸ http://yihui.name/knitr/

knitr Example Code

```
<!--begin.rcode
library(knitr)
opts_chunk$set(fig.width=5, fig.height=5)
end.rcode-->
```

```
<p>This is a minimal example which shows <strong>knitr</strong> working with HTML pages.</p>
```

```
<!--begin.rcode
## a simple calculator
1+1
## boring random numbers
set.seed(123)
rnorm(5)
end.rcode-->
```

```
<p>We can also produce plots (centered by the option <code>fig.align='center'</code>):</p>
```

```
<!--begin.rcode html-cars-scatter, message=FALSE, fig.align='center'
library(ggplot2)
plot(mpg~hp, mtcars)
qplot(hp, mpg, data=mtcars)+geom_smooth()
end.rcode-->
```

knitr Example Output