LabKey Software™
PARTNERS IN SCIENCE

LabKey Software: Past Present & Future

Mark Igra

Partner, LabKey Software

marki@labkey.com

Presentation Topics

LabKey Software
PARTNERS IN SCIENCE

- Why Scientific Data Integration
- Evolution of LabKey
  - Software
  - Company
  - Common requirements for scientific data integration
- Where we are now
- Future Directions

Why Scientific Data Integration

LabKey Software
PARTNERS IN SCIENCE

- Data Volume
  - Hundreds of millions of results from thousands of high throughput assay runs
- Data Variety
  - Clinical, Demographic, Assay & Specimen Data
  - Scientific data annotation
- Collaboration
  - Investigators aren't all in the same place
  - Need secure data sharing
- Broad Range of Analyses
  - Queries, Reports, Domain Specific Tools

2003

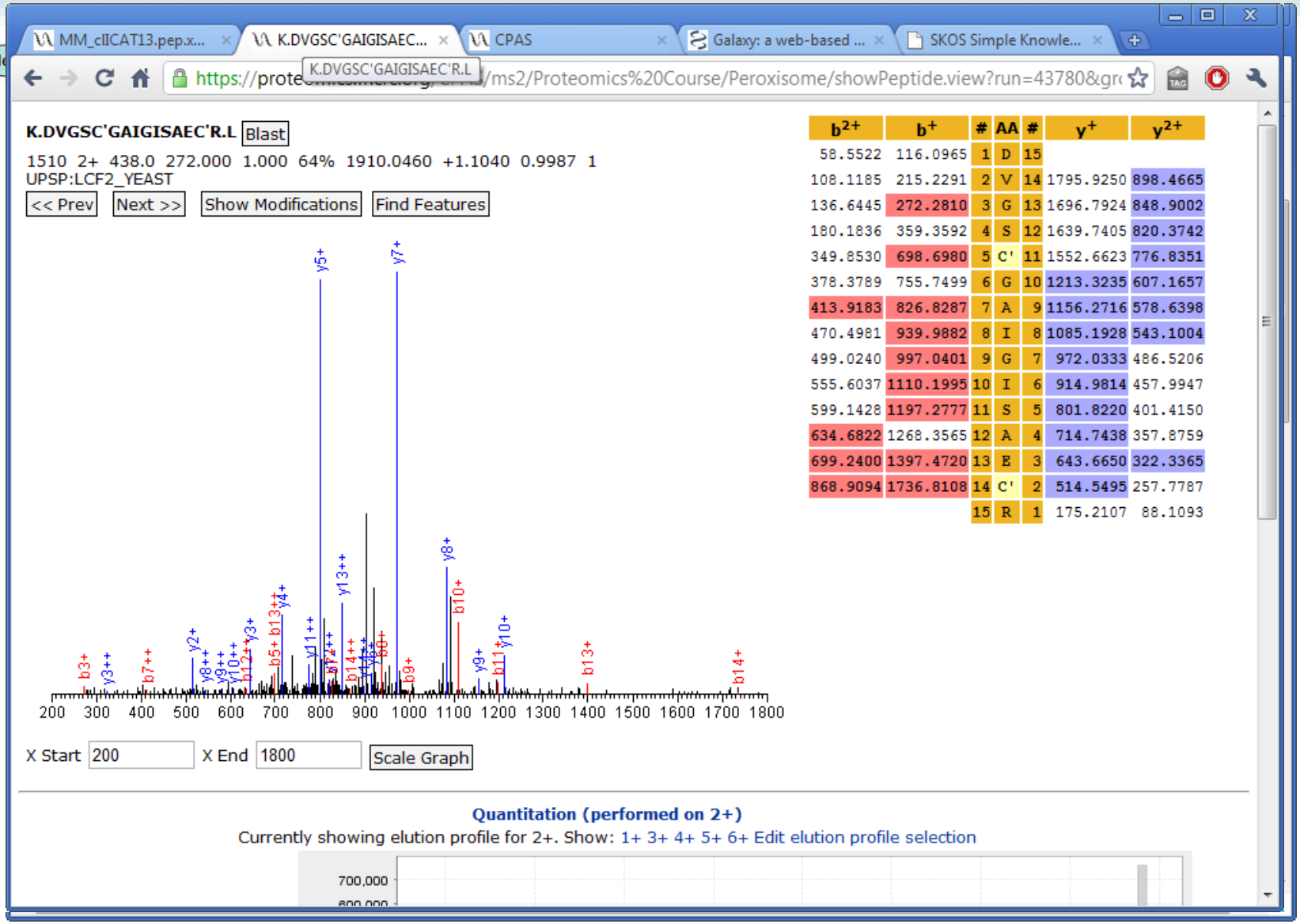
- Find Cancer Biomarkers using Proteomics
- Problems Faced
  - Number of reads from MS2-based proteomics assay were exploding & difficult to handle on existing tools
  - Analysis Pipeline on cluster was difficult to optimize

Proteomics.fhcrc.org

LabKey Software
PARTNERS IN SCIENCE

- Solution
  - The original LabKey Server (CPAS)
  - Pipeline to run MS2 analysis jobs via web browser and load results into Database
  - Web based analysis tools to view, combine & share results
  - Lots of Data
    - More than 90,000 MS2 Runs
    - More than 700,000,000 peptide identifications
- Rauch et al, Journal of Proteome Research, 5/2006

Proteomics Examples



b^{2+}	b^{+}	#	AA	#	y^{+}	y^{2+}
58.5522	116.0965	1	D	15		
108.1185	215.2291	2	V	14	1795.9250	898.4665
136.6445	272.2810	3	G	13	1696.7924	848.9002
180.1836	359.3592	4	S	12	1639.7405	820.3742
349.8530	698.6980	5	C'	11	1552.6623	776.8351
378.3789	755.7499	6	G	10	1213.3235	607.1657
413.9183	826.8287	7	A	9	1156.2716	578.6398
470.4981	939.9882	8	I	8	1085.1928	543.1004
499.0240	997.0401	9	G	7	972.0333	486.5206
555.6037	1110.1995	10	I	6	914.9814	457.9947
599.1428	1197.2777	11	S	5	801.8220	401.4150
634.6822	1268.3565	12	A	4	714.7438	357.8759
699.2400	1397.4720	13	E	3	643.6650	322.3365
868.9094	1736.8108	14	C'	2	514.5495	257.7787
		15	R	1	175.2107	88.1093

Software Requirement Summary

LabKey Software
PARTNERS IN SCIENCE

	Proteomics				
Data Size	700M Peptides 90K Runs				
Data Pipeline	✔				
Collaboration	✔				
Specimens	✔				
Security	✔				

2005 LabKey Software

- LabKey Software becomes independent company






2005 Flow Cytometry

- Measure intracellular cytokines in many samples
- Per-run quantity of data increasing
- Need consistent analysis within an experiment
- Cross-run analyses
- Each experiment type may have different statistics

Flow Cytometry

LabKey Software
PARTNERS IN SCIENCE

- Solution: LabKey Flow
  - High-throughput flow analysis engine
  - Loading of flow statistics
  - Adaptable data model for varying analysis types
  - Query tools for analyzing data
- Shulman et al, Cytometry A, Sep 8 2008

Flow Results

LabKey Software
PARTNERS IN SCIENCE

Flow Dashboard > labkey-analysis >

labkey-demo.xml analysis

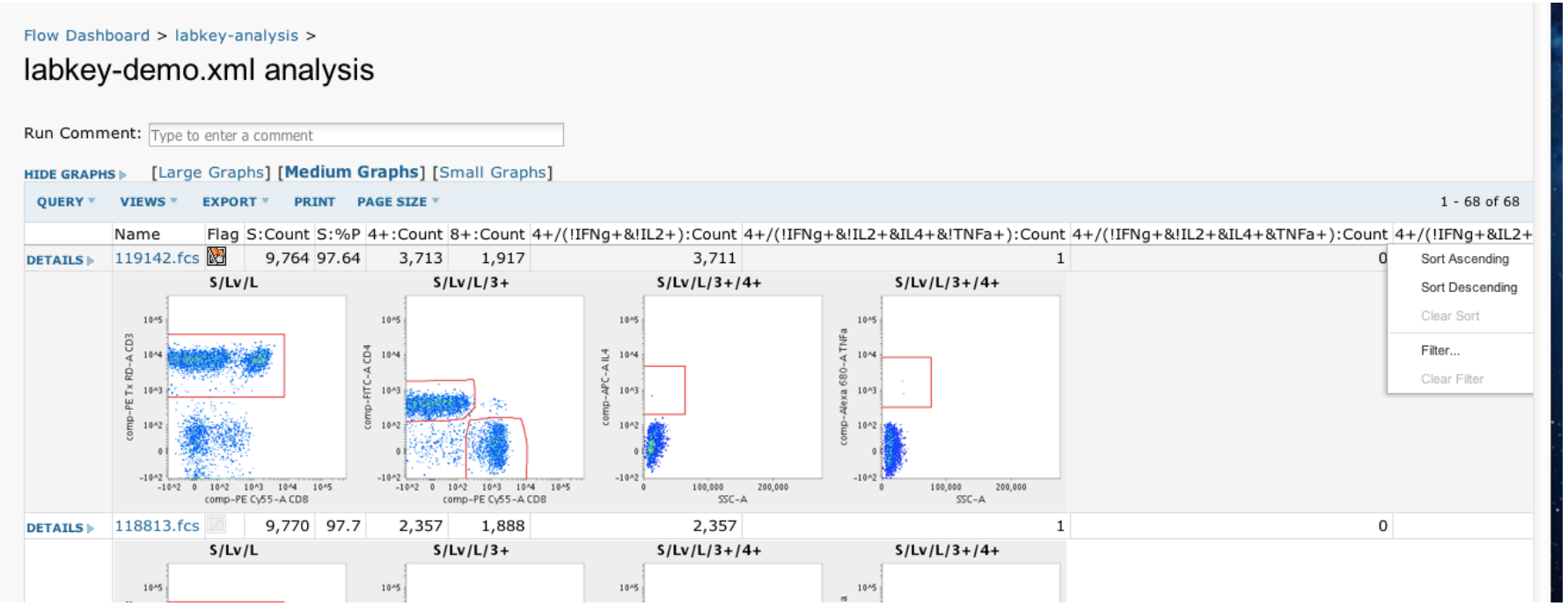
Run Comment: Type to enter a comment

SHOW GRAPHS

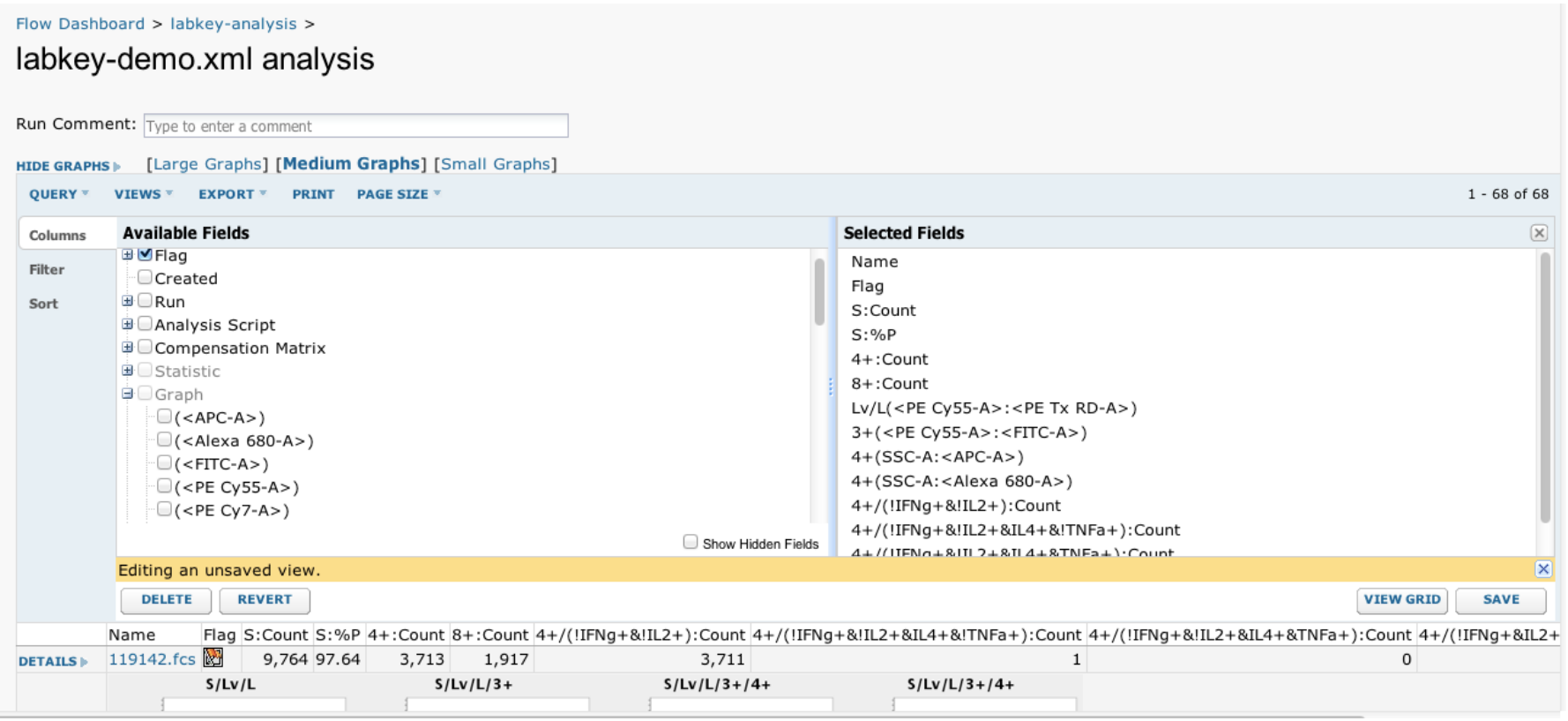
QUERY VIEWS EXPORT PRINT PAGE SIZE

	Name	Flag	S:Count	S:%P	4+:Count	8+:Count	4+/(!IFNg+&!IL2+):Count	4+/(!IFNg+&!IL2+&IL4+&!TNFa+):Count	4
DETAILS	119142.fcs		9,764	97.64	3,713	1,917	3,711	1	
DETAILS	118813.fcs		9,770	97.7	2,357	1,888	2,357	1	
DETAILS	118981.fcs		9,807	98.07	2,338	1,848	2,337	47	
DETAILS	118947.fcs		9,555	95.55	2,065	1,703	2,065	2	
DETAILS	119199.fcs		9,728	97.28	2,336	1,907	1,977	0	
DETAILS	119154.fcs		9,545	95.45	2,541	1,629	2,541	0	
DETAILS	118836.fcs		9,496	94.96	504	466	503	0	
DETAILS	119010.fcs		9,817	98.17	3,804	1,824	3,801	0	
DETAILS	118762.fcs		9,743	97.43	5	131	5	0	
DETAILS	118756.fcs		9,753	97.53	0	0	0	0	
DETAILS	119171.fcs		9,662	96.62	2,200	2,196	1,661	5	
DETAILS	118801.fcs		9,706	97.06	2,251	2,087	2,251	2	
DETAILS	118754.fcs		9,787	97.87	1	0	1	0	

View Graphs


Flow Dashboard > labkey-analysis >
labkey-demo.xml analysis
Run Comment: Type to enter a comment
HIDE GRAPHS [Large Graphs] [Medium Graphs] [Small Graphs]
QUERY VIEWS EXPORT PRINT PAGE SIZE
1 - 68 of 68
Name Flag S:Count S:%P 4+:Count 8+:Count 4+/(!IFNg+&!IL2+):Count 4+/(!IFNg+&!IL2+&IL4+&!TNFa+):Count 4+/(!IFNg+&!IL2+&IL4+&TNFa+):Count 4+/(!IFNg+&IL2+
DETAILS 119142.fcs 9,764 97.64 3,713 1,917 3,711 1 0
S/Lv/L
S/Lv/L/3+
S/Lv/L/3+/4+
S/Lv/L/3+/4+
comp-PE Tx RD-A CD3
comp-PE Cy55-A CD8
comp-FITC-A CD4
comp-PE Cy55-A CD8
comp-APC-A IL4
SSC-A
comp-Alexa 680-A TNFa
SSC-A
Sort Ascending
Sort Descending
Clear Sort
Filter...
Clear Filter
DETAILS 118813.fcs 9,770 97.7 2,357 1,888 2,357 1 0
S/Lv/L
S/Lv/L/3+
S/Lv/L/3+/4+
S/Lv/L/3+/4+

Customize View



Software Requirement Summary

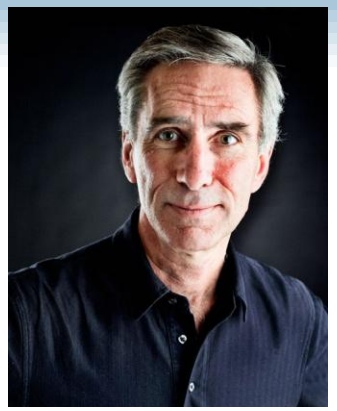
LabKey Software
PARTNERS IN SCIENCE

	Proteomics	Flow			
Data Size	700M Peptides 90K Runs	50M Statistics			
Data Pipeline	✔	✔			
Collaboration	✔				
Specimens	✔	✔			
Security	✔	✔			
Data Variety		✔			
Query		✔			

2006 Atlas – SCHARP Data Portal

LabKey Software
PARTNERS IN SCIENCE

- Combine many data types for HIV Vaccine studies
- Clinical Response Forms (CRF), Specimens, Many Assays
- Enable secure collaboration for scientists worldwide
- Allocate & distribute valuable specimens



CHAVI
CENTER FOR HIV/AIDS VACCINE IMMUNOLOGY

CAVD
The Collaboration for AIDS Vaccine Discovery

Atlas – SCHARP Data Portal

LabKey Software
PARTNERS IN SCIENCE

- Solution
  - Secure web portal for HIV Vaccine Enterprise Data
  - Used by several networks to share data
    - CHAVI, CAVD, HVTN, HPTN (3000 Users Worldwide)
  - Core software was written by LabKey
  - SCHARP runs Atlas
    - Defines available data and relationships
    - Manages security and permissions
    - Manages data loading
    - Builds custom modules
- Nelson et al, BMC Bioinformatics, March 2011
- Piehler et al, BMC Immunology, May 2011

Participant View

LabKey Software
PARTNERS IN SCIENCE

CHAVI 001 > Study Overview > Dataset: Binding Antibody, All Visits >

Participant - [redacted]

NEXT PARTICIPANT | SEARCH FOR '[redacted]'

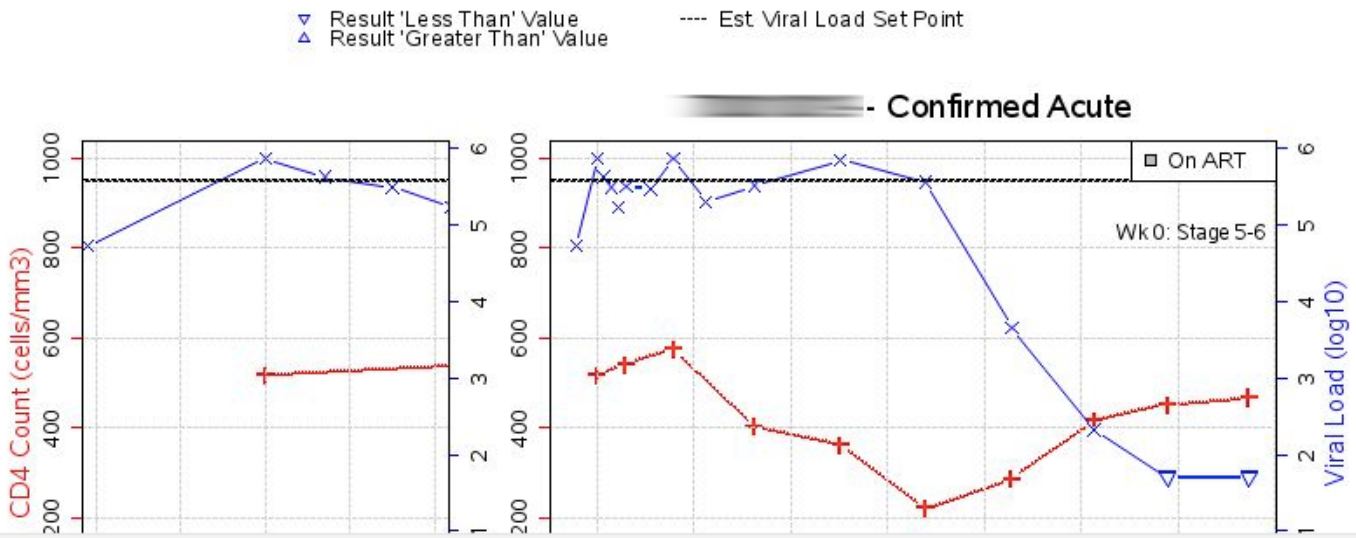
All Datasets | Specimen Timeline | Clinical Data | Requested Specimens | All Specimens | Binding Antibodies | Cytokines

Clinical Data

CRF Data

EXPORT TO EXCEL

Cohort:	Confirmed Acute
Age:	18
Gender:	Female
Race:	White
Tribe:	
Clade:	



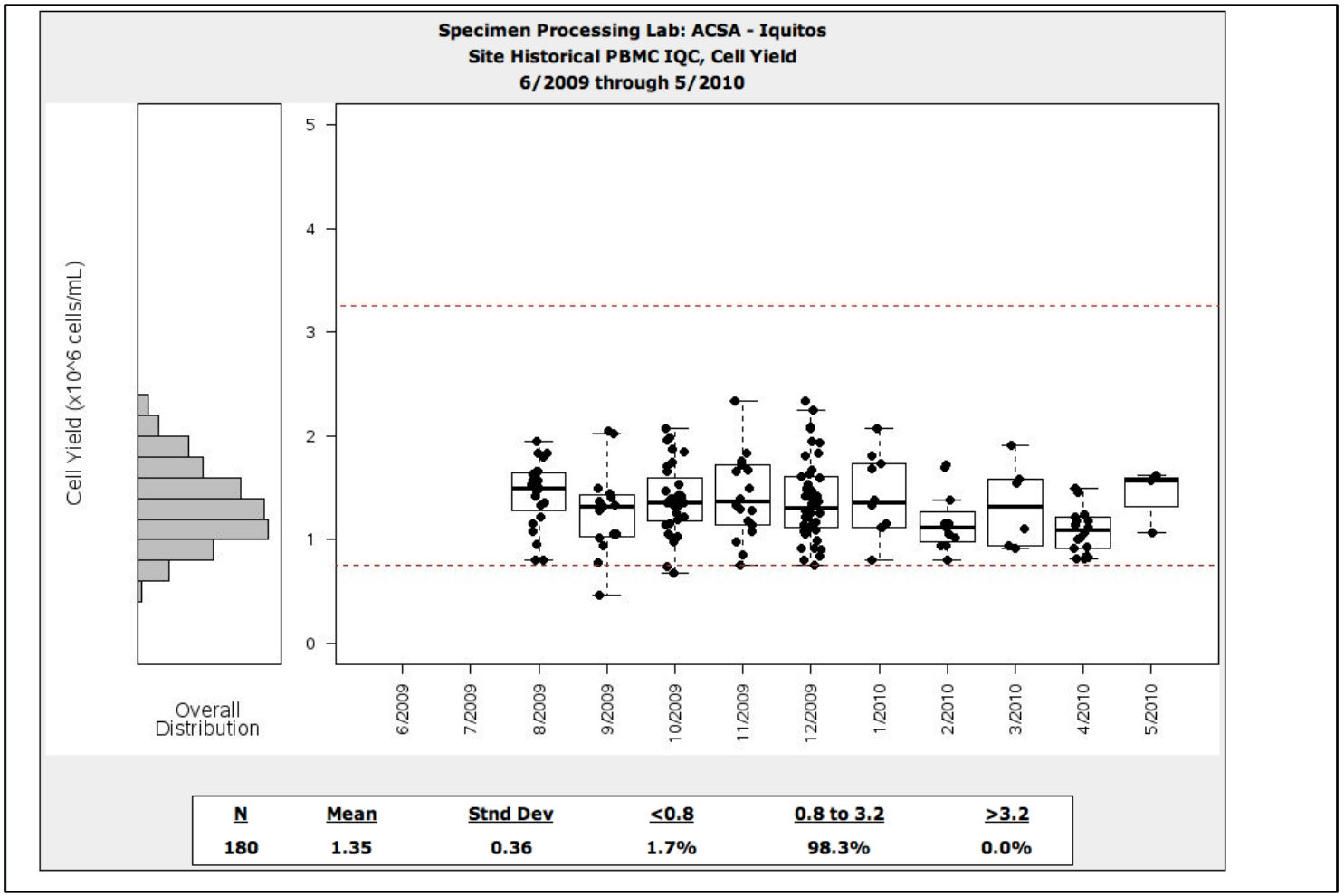
2007-2008 Custom Applications

LabKey Software
PARTNERS IN SCIENCE

- Dozens of small applications needed for different groups
  - Most used data already in LabKey System
  - But custom workflows, reports & analysis required
- SCHARP has detailed knowledge of the requirements
- LabKey provided an API and simple application building tools

HVTN IQC

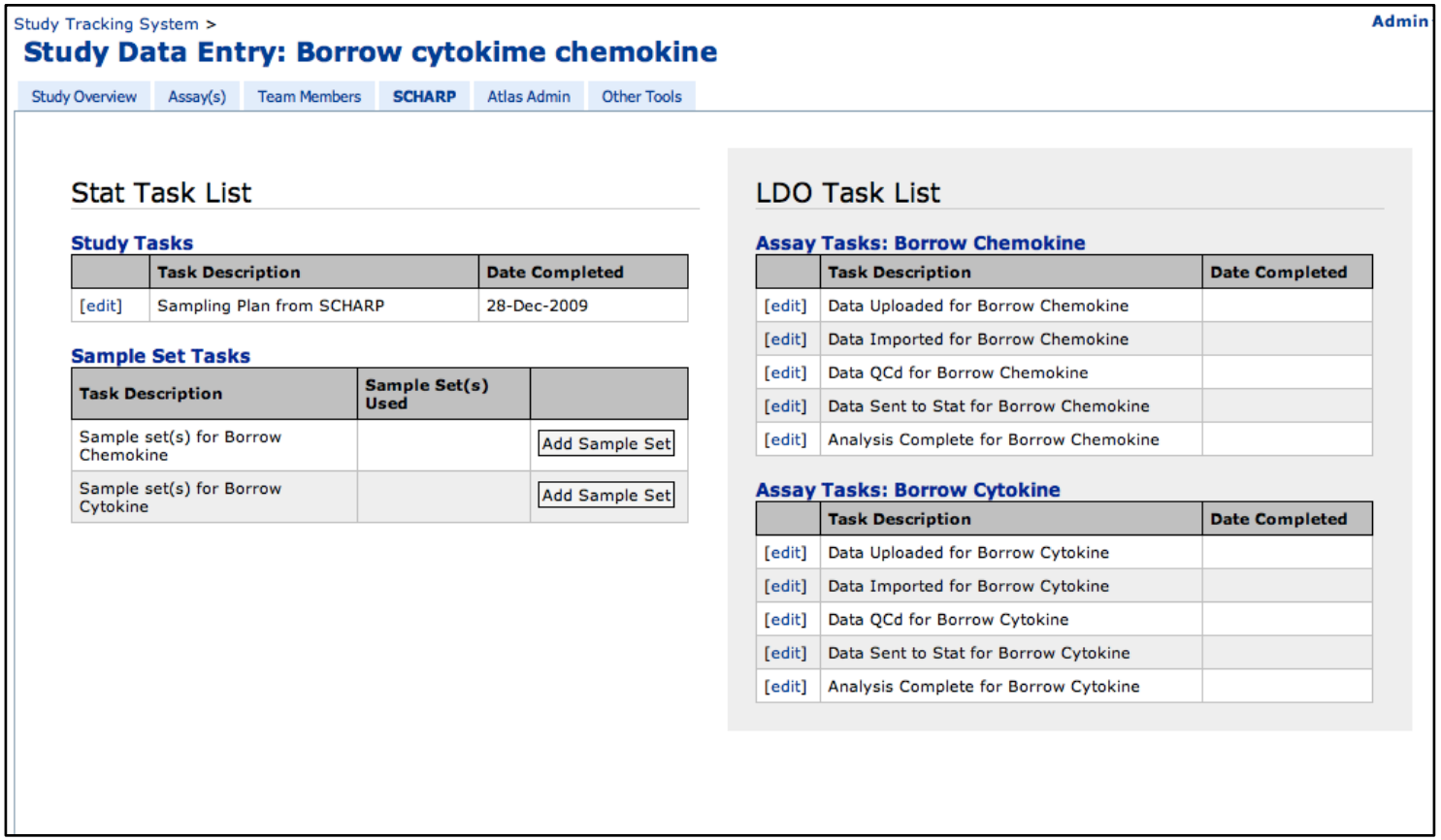
- Internal Quality Control for Sample Processing
  - Cell Yield & Processing Time



RV144 Followup Study Tracking System

LabKey Software
PARTNERS IN SCIENCE

- Tools for tracking information about studies



Software Requirement Summary

LabKey Software
PARTNERS IN SCIENCE

	Proteomics	Flow	Atlas (Studies)	Atlas (Labs)	
Data Size	700M Peptides 90K Runs	50M Statistics	20K Subjects 800K Specimens 1200 Datasets	30K Assay Runs	
Data Pipeline	✔	✔	✔	✔	
Collaboration	✔		✔	✔	
Specimens	✔	✔	✔	✔	
Security	✔	✔	✔	✔	
Data Variety		✔	✔	✔	
Query		✔	✔	✔	
Study Model			✔		
Custom Reports			✔	✔	
API			✔	✔	
Auditing			✔	✔	

2009-2010 Primate Electronic Health Record

LabKey Software
PARTNERS IN SCIENCE

- 30 years of daily data on thousands of animals
- Clinical staff & vets need health record
- Researchers need scientific data
- Colony Management is an ongoing problem

Primate Electronic Health Record

LabKey Software
PARTNERS IN SCIENCE

- EHR is a a specialized, ongoing “study”
- LabKey enhanced scalability of study solution
- LabKey enhanced API
- LabKey wrote tools to transfer data into new EHR within minutes of entry into old EHR
  - Now LabKey EHR is only one in use
- Wisconsin Primate Center built custom, forms, views & reports to enter & analyze the data
- Solution extending to Oregon National Primate Center

Primate EHR

LabKey Software
PARTNERS IN SCIENCE

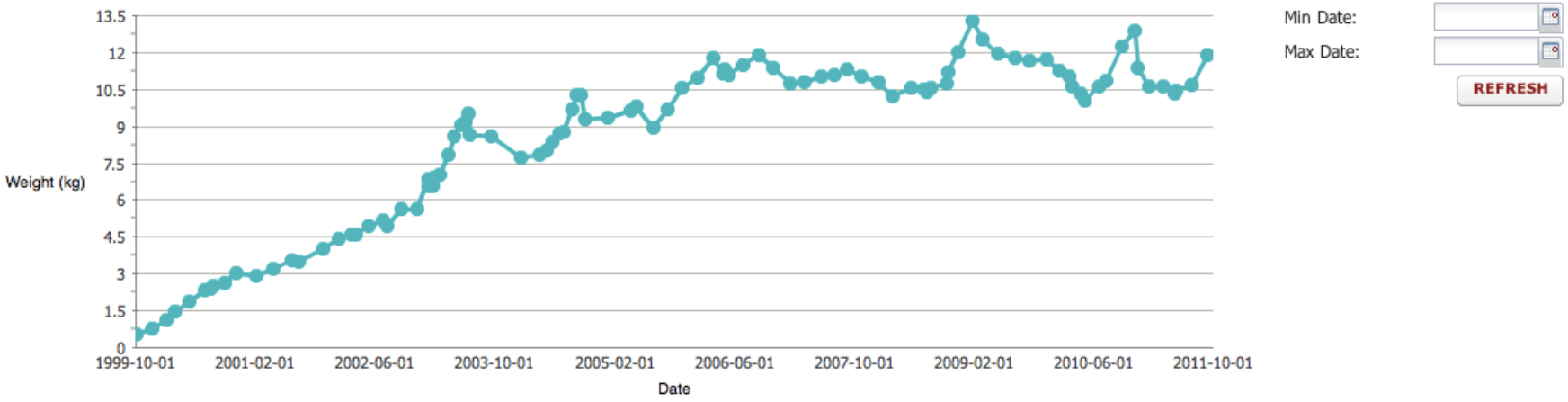
Problem List: [redacted]

VIEWS ▾ EXPORT ▾ PRINT PAGE SIZE ▾ MORE ACTIONS ▾ 1 - 12 of 12

		Id	Current Room	Current Cage	Problem Number	Date Observed ▾	Date Resolved	Category	Remark	Code
☐	DETAILS ▸	[redacted]			12	2011-10-04	2011-10-07	[redacted]	[redacted]	
☐	DETAILS ▸	[redacted]			11	2011-05-16	2011-05-20		[redacted]	
☐	DETAILS ▸	[redacted]			10	2010-11-22	2010-11-30		[redacted]	
☐	DETAILS ▸	[redacted]			9	2008-05-22	2008-05-23		[redacted]	
☐	DETAILS ▸	[redacted]			8	2007-09-05	2007-09-07		[redacted]	
☐	DETAILS ▸	[redacted]			7	2007-06-05	2007-06-08		[redacted]	
☐	DETAILS ▸	[redacted]			6	2005-01-12	2005-01-26		[redacted]	
☐	DETAILS ▸	[redacted]			5	2003-12-03	2003-12-03		[redacted]	
☐	DETAILS ▸	[redacted]			4	2003-03-17	2003-04-10		[redacted]	
☐	DETAILS ▸	[redacted]			3	2002-11-19	2002-11-19		[redacted]	
☐	DETAILS ▸	[redacted]			2	2002-07-22	2002-07-22		[redacted]	
☐	DETAILS ▸	[redacted]			1	2002-03-07	2002-03-19		[redacted]	

VIEWS ▾ EXPORT ▾ PRINT PAGE SIZE ▾ MORE ACTIONS ▾ 1 - 12 of 12

Weight: [redacted]

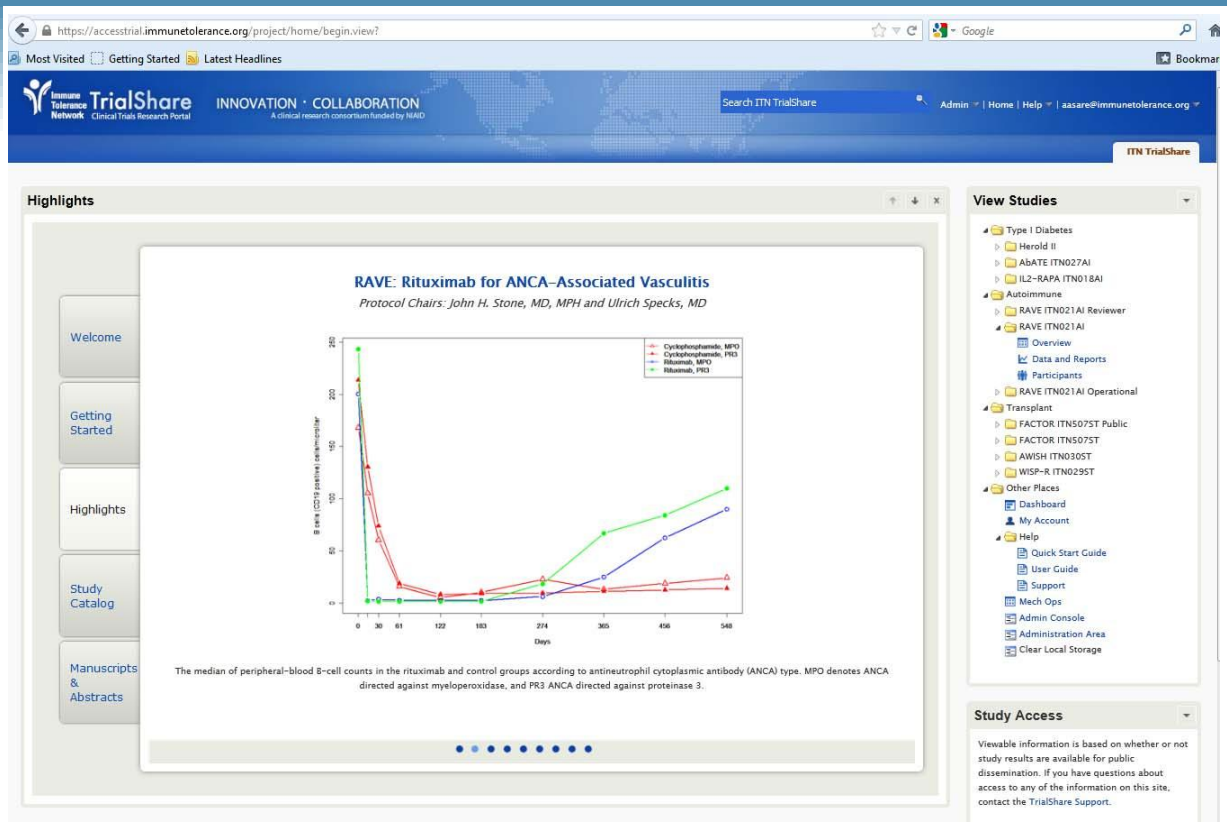


Software Requirement Summary

LabKey Software
PARTNERS IN SCIENCE

	Proteomics	Flow	Atlas (Studies)	Atlas (Labs)	Primate EHR
Data Size	700M Peptides 90K Runs	50M Statistics	20K Subjects 800K Specimens 1200 Datasets	30K Assay Runs	13K Animals 1.2M Drug Doses
Data Pipeline	✔	✔	✔	✔	
Collaboration	✔		✔	✔	✔
Specimens	✔	✔	✔	✔	
Security	✔	✔	✔	✔	✔
Data Variety		✔	✔	✔	✔
Query		✔	✔	✔	✔
Study Model			✔		✔
Custom Reports			✔	✔	✔
API			✔	✔	✔
Auditing			✔	✔	✔

2011 – ITN Polish, Analysis

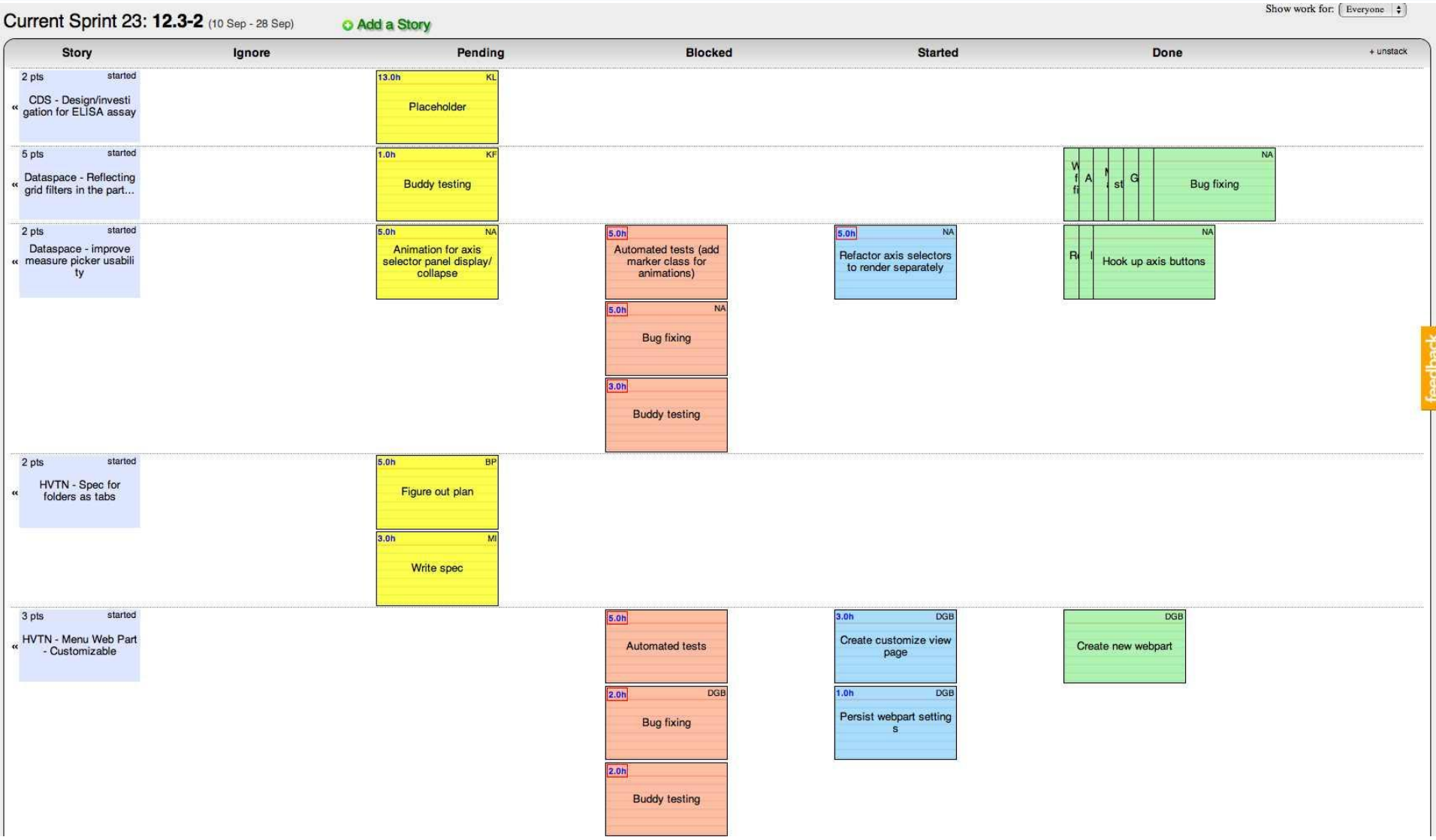


2011 – Refining our process

LabKey Software
PARTNERS IN SCIENCE

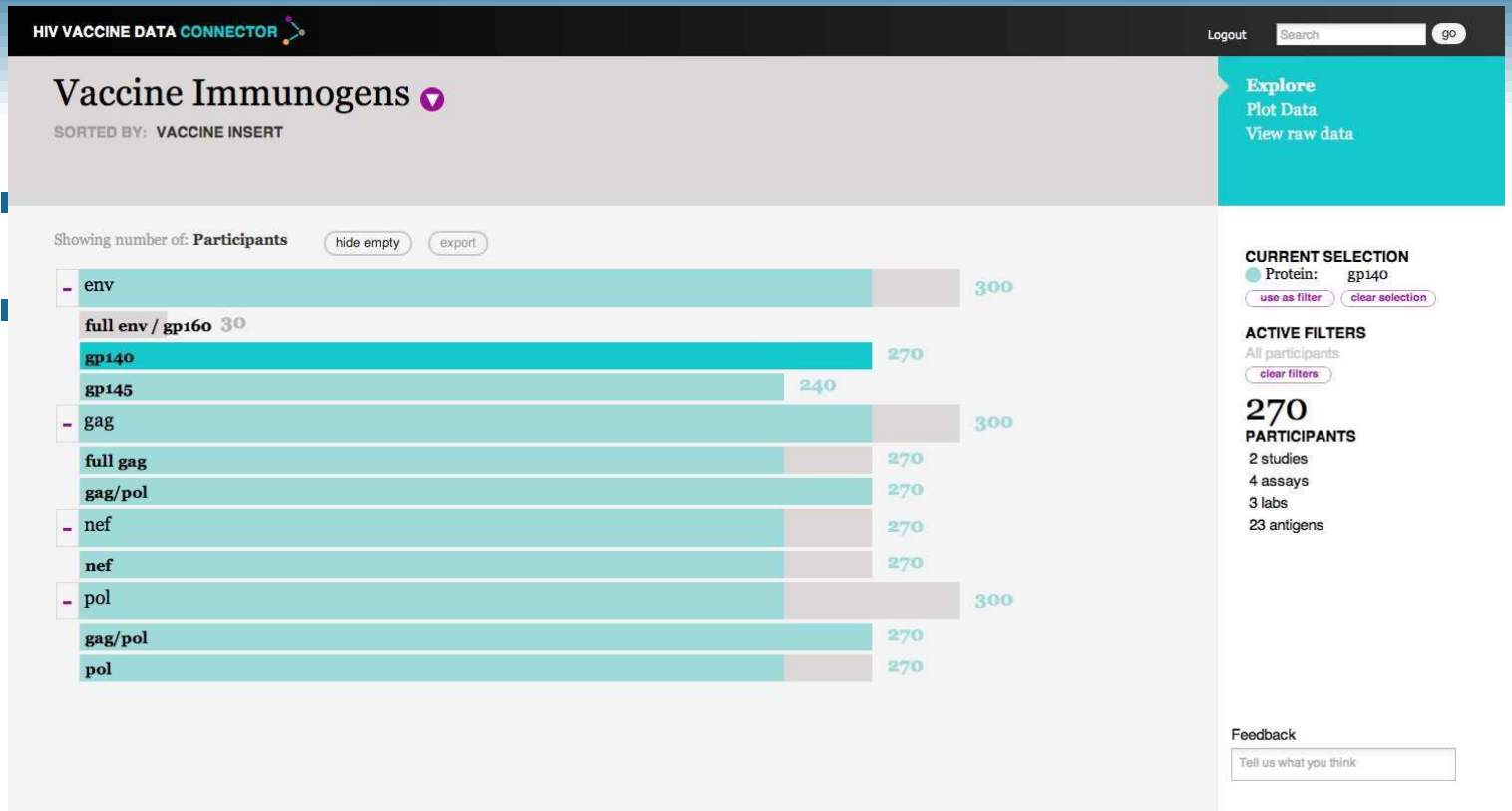
- Two teams
- Every month
  - Demo progress
  - Gather requirements & priorities
  - Write specs
  - Detailed cost estimates
    - Development, automated & manual testing, bug fixing
- Every day
  - Stand up meeting
  - Schedule progress
  - Quality metrics
  - Blocking Issues

What the process looks like



2012

LabKey Software
PARTNERS IN SCIENCE



Software Requirement Summary

LabKey Software
PARTNERS IN SCIENCE

	Proteomics	Flow	Atlas (Studies)	Atlas (Labs)	Primate EHR
Data Size	700M Peptides 90K Runs	50M Statistics	20K Subjects 800K Specimens 1200 Datasets	30K Assay Runs	13K Animals 1.2M Drug Doses
Data Pipeline	✔	✔	✔	✔	
Collaboration	✔		✔	✔	✔
Specimens	✔	✔	✔	✔	
Security	✔	✔	✔	✔	✔
Data Variety		✔	✔	✔	✔
Query	✔	✔	✔	✔	✔
Study Model			✔	✔	✔
Custom Reports	✔	✔	✔	✔	✔
API	✔	✔	✔	✔	✔
Auditing	✔	✔	✔	✔	✔

Key Requirements

LabKey Software
PARTNERS IN SCIENCE

- Data integration is the unifying theme
  - Samples to Assays
  - Assays to Subjects
  - Subjects to Studies
  - Translational Medicine involves integrating data from “Molecules to Populations” (Kuhn et al 2008)
- Data types to integrate are constantly evolving
  - Both file and “structured data types”
  - Can’t rebuild new systems for new data types
- Tools grow as fast as data types
- Scientists need to share data
  - Central lab doing work for distributed clients
  - Distributed group of labs contributing
  - Need to manage many users

Looking Forward

LabKey Software
PARTNERS IN SCIENCE

- Our collaborators guide our direction
  - But we have a few ideas of what's coming
- Networks/Consortia
  - Cross protocol data integration
  - Ancillary Studies
  - Deidentification
- Labs
  - Sequencing
  - Improved Throughput
  - Pipelines

Looking Forward

LabKey Software
PARTNERS IN SCIENCE

- Developers
  - Dependency & library management
  - Rserve – Build web tools with R back ends
  - Flexibility in display & navigation
- For Everyone
  - Integration with more systems (e.g. IMMPORT)
  - Visualization & Analysis
  - Terminology and Identity Management

Looking Forward – The Company

LabKey Software
PARTNERS IN SCIENCE

- Number & complexity of engagements has grown
  - We have evolved from 3 – 22 people in 9 years
  - Processes have evolved to scale
  - Can apply what we've learned to new engagements
- Open source model works
  - Inspires Trust
  - Fits the research market
- We're happy with our market, our growth & our business
- Thank You

Acknowledgements

- **Partners & Funders**
- Martin McIntosh (FHCRC)
  - NCI
  - Canary Foundation
- Steve Self (SCHARP)
  - CHAVI (NIH)
  - HVTN (NIH)
  - CAVD (Gates Foundation)
- David O'Connor (Wisconsin)
  - Primate EHR (ARRA)
  - Genotyping Tools (NIAID)
- Parag Mallick (USC)
- Michael Katze (UW)
- Immune Tolerance Network

- LabKey Development
  - Matthew Bellew
  - Adam Rauch
  - Britt Piehler
  - Josh Eckels
  - Kevin Krause
  - Brendan MacLean (now UW)
  - Nick Shulman (now UW)
  - Karl Lum
  - Nick Arnold
  - Cory Nathe
  - Ben Bimber (now ONPRC)
  - Alan Veniza
  - Dax Hawkins
  - Dave Bradlee
- Documentation
  - Elizabeth Nelson
  - Steve Hanson
- Test
  - Trey Chaddick
  - Elizabeth Van Nostrand
- Management & Operations
  - Britt Piehler
  - Kristin Fitzimmons
  - Peter Hussey
  - Ren Lis
  - Ben Hackett

Mark Igra
marki@labkey.com

LabKey Software™
PARTNERS IN SCIENCE

If you use LabKey Server for your research, please reference one of these publications about the platform:

General Use: Nelson EK, Piehler B, Eckels J, Rauch A, Bellew M, Hussey P, Ramsay S, Nathe C, Lum K, Krouse K, Stearns D, Connolly B, Skillman T, Igra M. LabKey Server: An open source platform for scientific data integration, analysis and collaboration. BMC Bioinformatics 2011 Mar 9; 12(1): 71.

Proteomics: Rauch A, Bellew M, Eng J, Fitzgibbon M, Holzman T, Hussey P, Igra M, Maclean B, Lin CW, Detter A, Fang R, Faca V, Gafken P, Zhang H, Whitaker J, States D, Hanash S, Paulovich A, McIntosh MW: Computational Proteomics Analysis System (CPAS): An Extensible, Open-Source Analytic System for Evaluating and Publishing Proteomic Data and High Throughput Biological Experiments. Journal of Proteome Research 2006, 5:112-121.

Flow Cytometry: Shulman N, Bellew M, Snelling G, Carter D, Huang Y, Li H, Self SG, McElrath MJ, De Rosa SC: Development of an automated analysis system for data from flow cytometric intracellular cytokine staining assays from clinical vaccine trials. Cytometry 2008, 73A:847-856.

LEARN APPLY SHARE
LabKey User Conference 2011
NOVEMBER 14-15, SEATTLE