Labkey User's Meeting

Richard Green
greener@uw.edu

November 14th 2011

Outline

- Background
- Pepsi challenge
- Labkey in a production system
- Katze lab integration of labkey server: some real world examples
- Customizing labkey tools
- R/Bioconductor tools
- Future directions

Why We Use LabKey Server

- A development platform to archive and mine data (proper data dissemination)
- To answer biological questions
- We don’t always know the questions we will have: We often revisit and repurpose data

Why We Use LabKey Server

- We no longer work in silos

Having a platform where you and your collaborators and exchange data/analysis is necessary

Extend the Platform

- To submit data in the public repositories and integrate information out of the repositories (GEO for microarray and PRIDE for proteomics) * Focused in grants
- Web analytics is necessary
- Meta data will become the next challenge
- Data integration (protein, arrays, NGS, more)
- Data sets will only grow larger and more complex (NGS)

Pepsi Challenge

- SBEAMS and CPAS
- Demo installations in Fall 2005

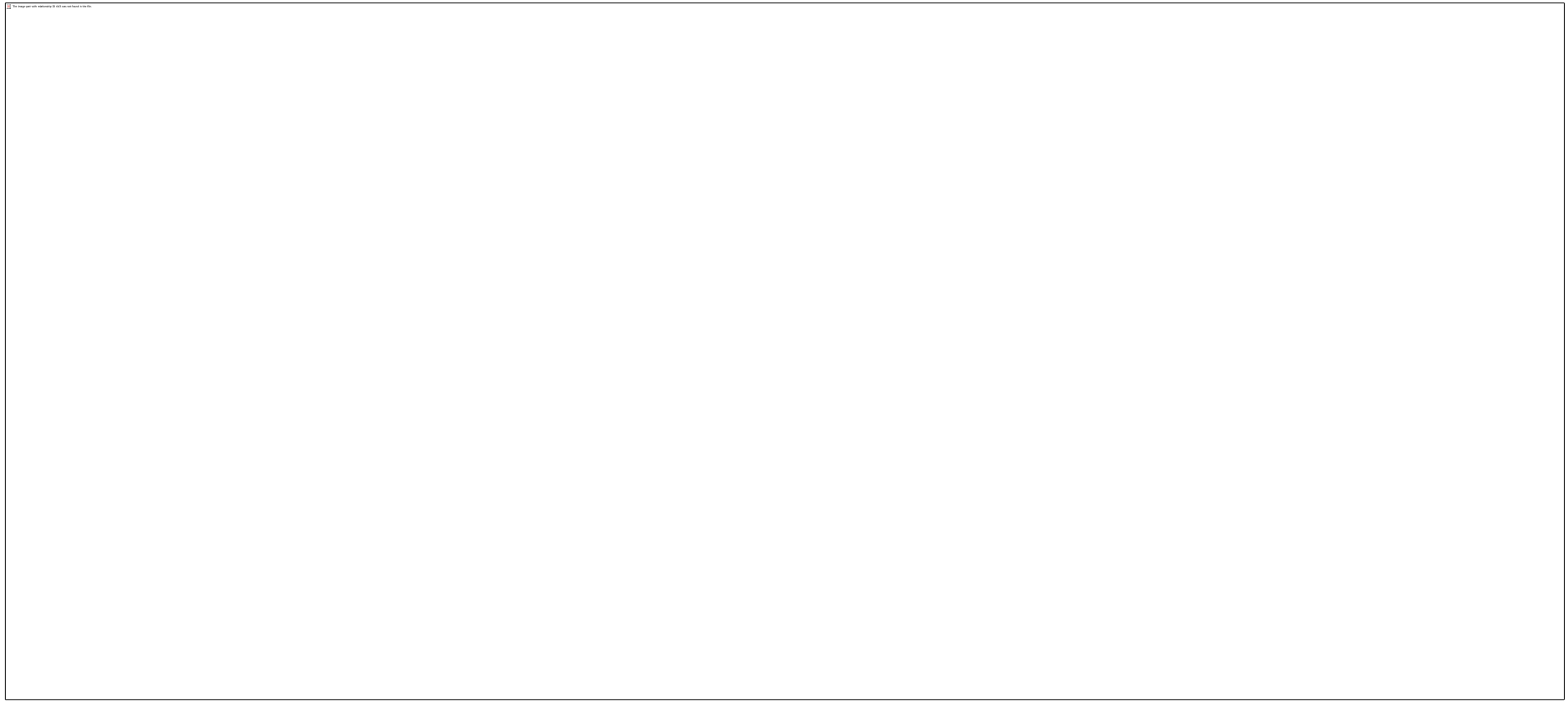
Canary Grant 2007

- Expanded to operating in a production environment
- Integration with high performance cluster
- Customized proteomics pipelines
- Proteomics: development in analyzing data sets
- Beginning of the enterprise pipeline
- Plan ahead for data size

Katze Lab

- 3 installations of labkey: Katze informatics, Systems Virology, RIG-I adjuvant contract
- LIMS-like system: Tracking experimental design, custom wikis for results, procedures, tutorials
- Recording multiple data types: Proteomics, Genomics, NGS: meta data

Manuscript Tracker

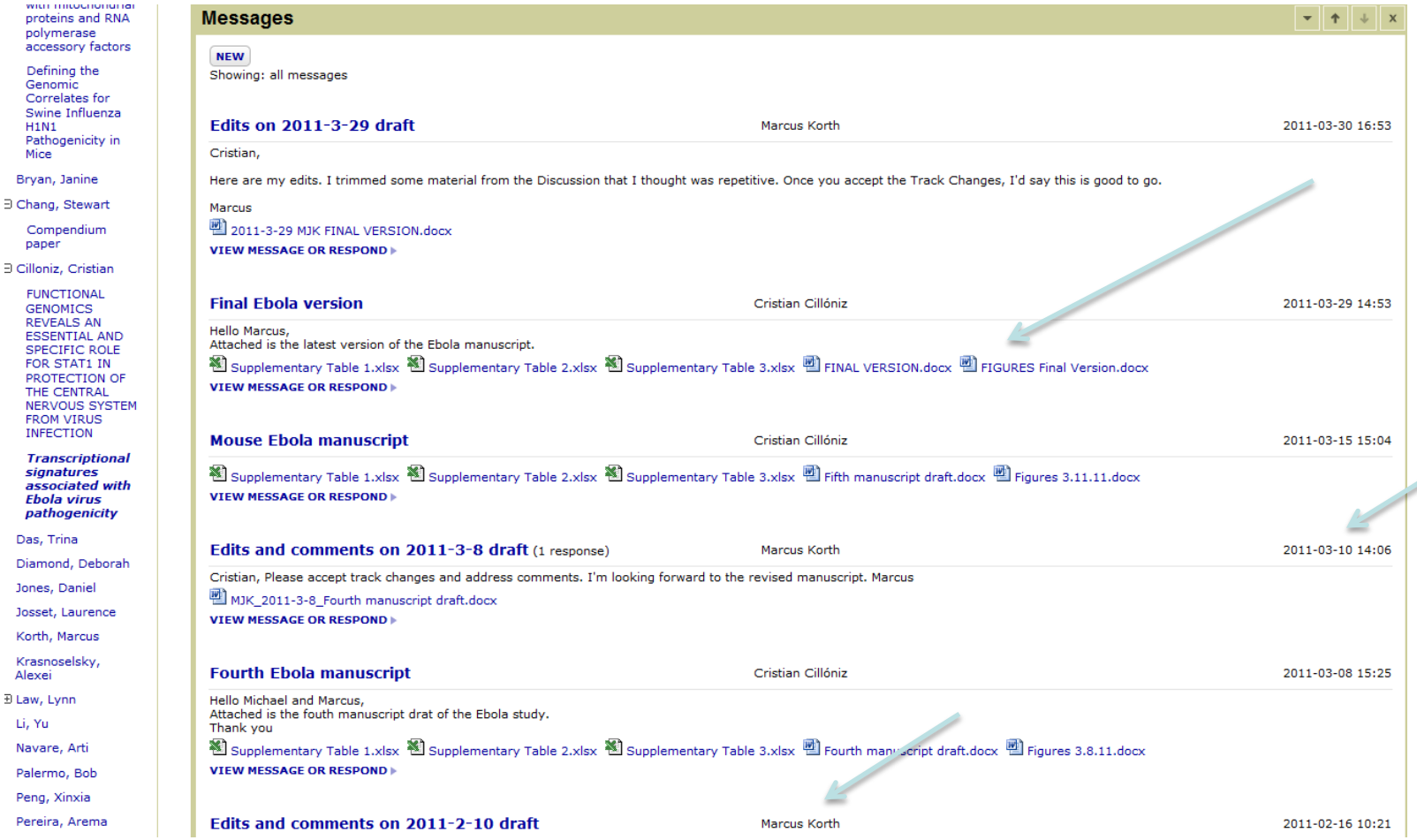


Uniform way to track all manuscripts generated by Katze Lab

1) Tracker queue shows manuscripts by status

2) Discussion board for revisions of manuscripts and discussions

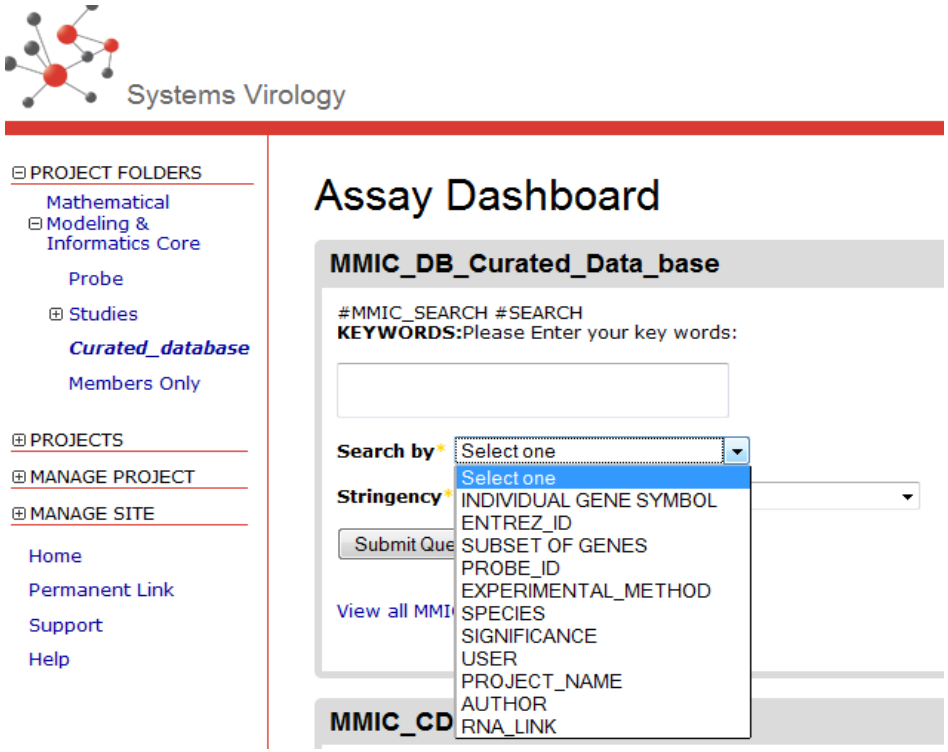
Manuscript Tracker



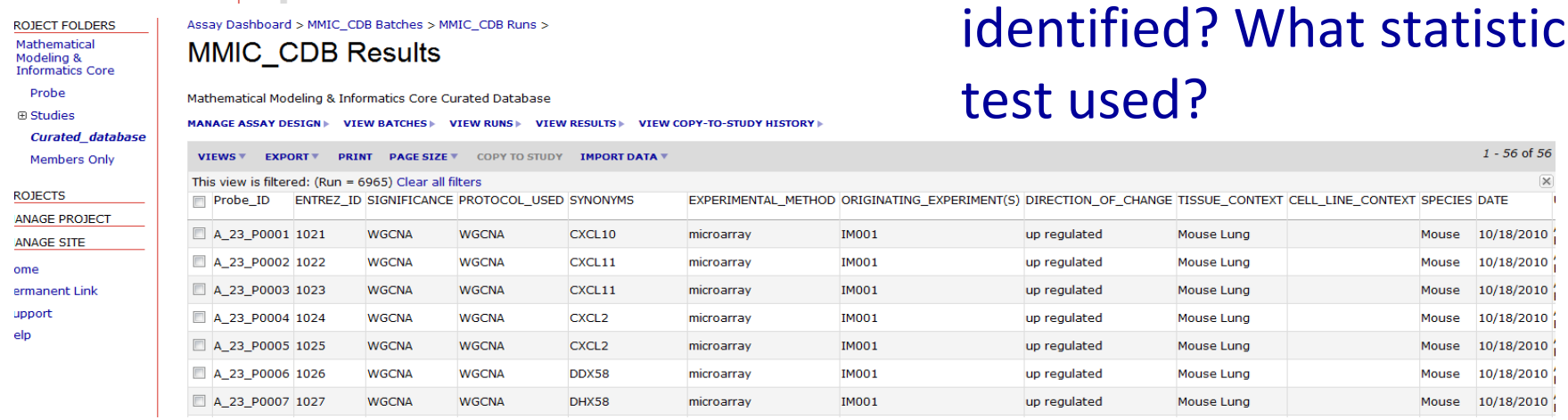
- Screen shot dialogue
- Changes to figures, tables, edits and comments dated and identified by user

Systems Virology LabKey Development

SV / MMIC Curated Database



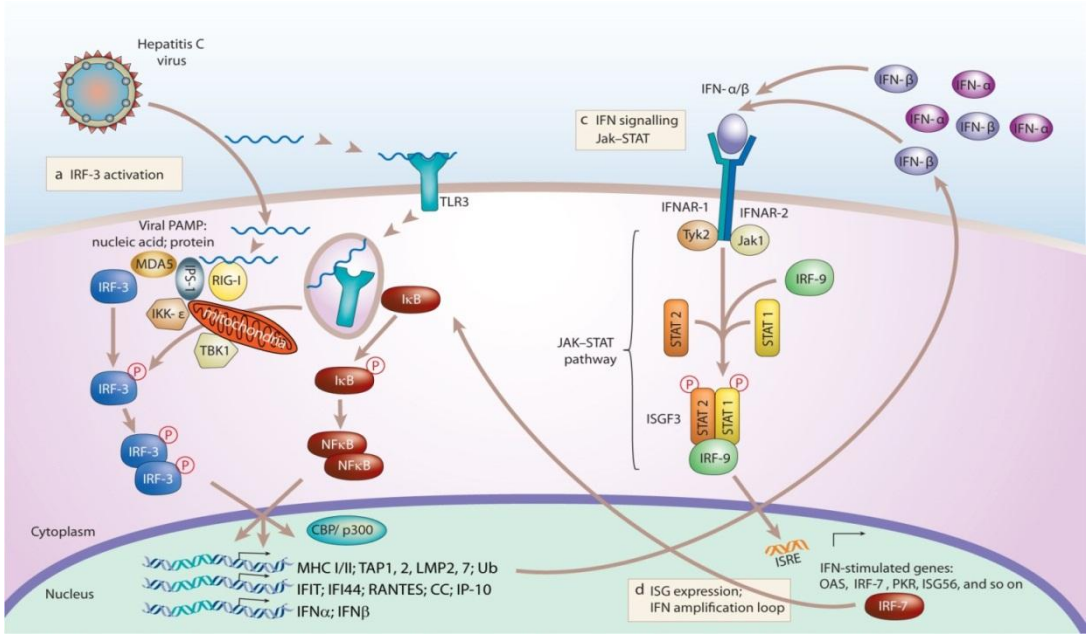
- Allows SV MMIC members to query manually curated datasets
- Members upload datasets by text, Excel, or manually
- Stores results into database
- Why are these genes interesting? How were they identified? What statistical test used?



Probe_ID	ENTREZ_ID	SIGNIFICANCE	PROTOCOL_USED	SYNONYMS	EXPERIMENTAL_METHOD	ORIGINATING_EXPERIMENT(S)	DIRECTION_OF_CHANGE	TISSUE_CONTEXT	CELL_LINE_CONTEXT	SPECIES	DATE
A_23_P0001	1021	WGCNA	WGCNA	CXCL10	microarray	IM001	up regulated	Mouse Lung		Mouse	10/18/2010
A_23_P0002	1022	WGCNA	WGCNA	CXCL11	microarray	IM001	up regulated	Mouse Lung		Mouse	10/18/2010
A_23_P0003	1023	WGCNA	WGCNA	CXCL11	microarray	IM001	up regulated	Mouse Lung		Mouse	10/18/2010
A_23_P0004	1024	WGCNA	WGCNA	CXCL2	microarray	IM001	up regulated	Mouse Lung		Mouse	10/18/2010
A_23_P0005	1025	WGCNA	WGCNA	CXCL2	microarray	IM001	up regulated	Mouse Lung		Mouse	10/18/2010
A_23_P0006	1026	WGCNA	WGCNA	DDX58	microarray	IM001	up regulated	Mouse Lung		Mouse	10/18/2010
A_23_P0007	1027	WGCNA	WGCNA	DHX58	microarray	IM001	up regulated	Mouse Lung		Mouse	10/18/2010

RIG-I Contract

- Find new ways to stimulate, enhance innate antiviral immunity against pathogenic RNA viruses.
- Screen candidate adjuvants for biological activity



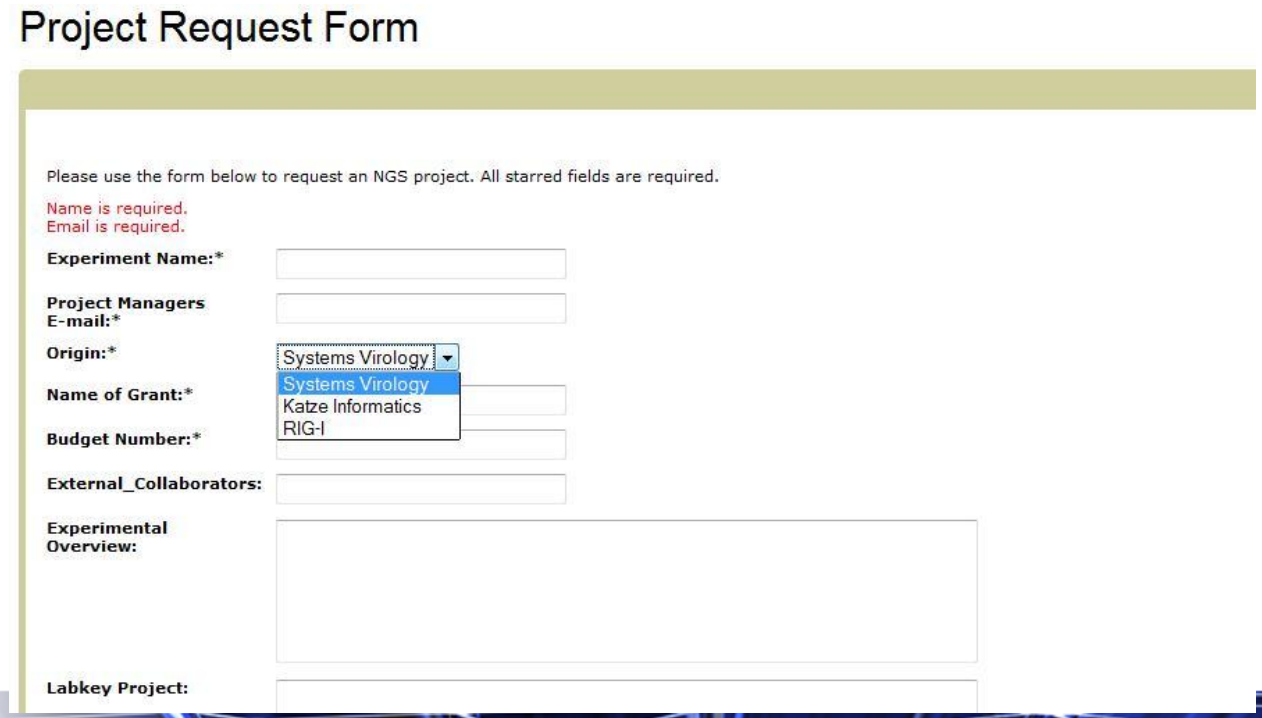
Stores assay results:

- Flow data
- Luminex
- Microarray raw data

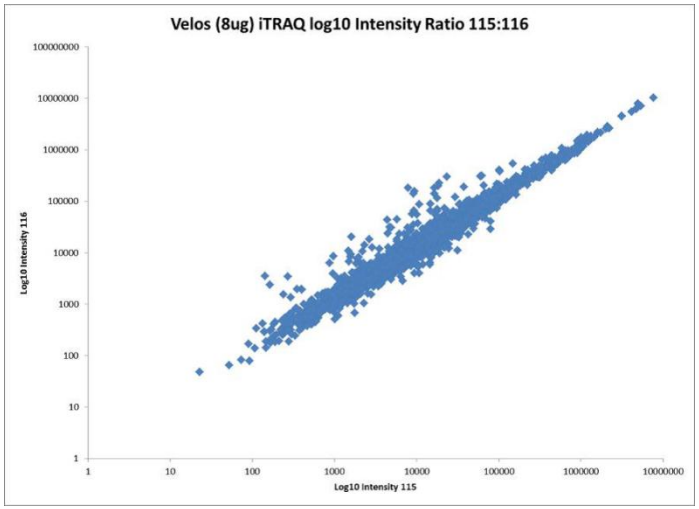
Customizing LabKey Tools

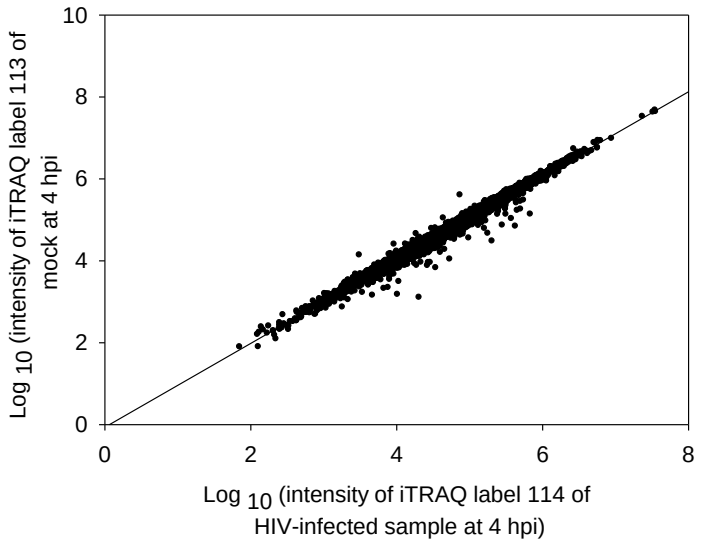
Custom Assays: UI

- User interfaces
- Content stored into the data base
- Metrics to be generated for future Grants or Contracts



Integration of Libra in 11.3

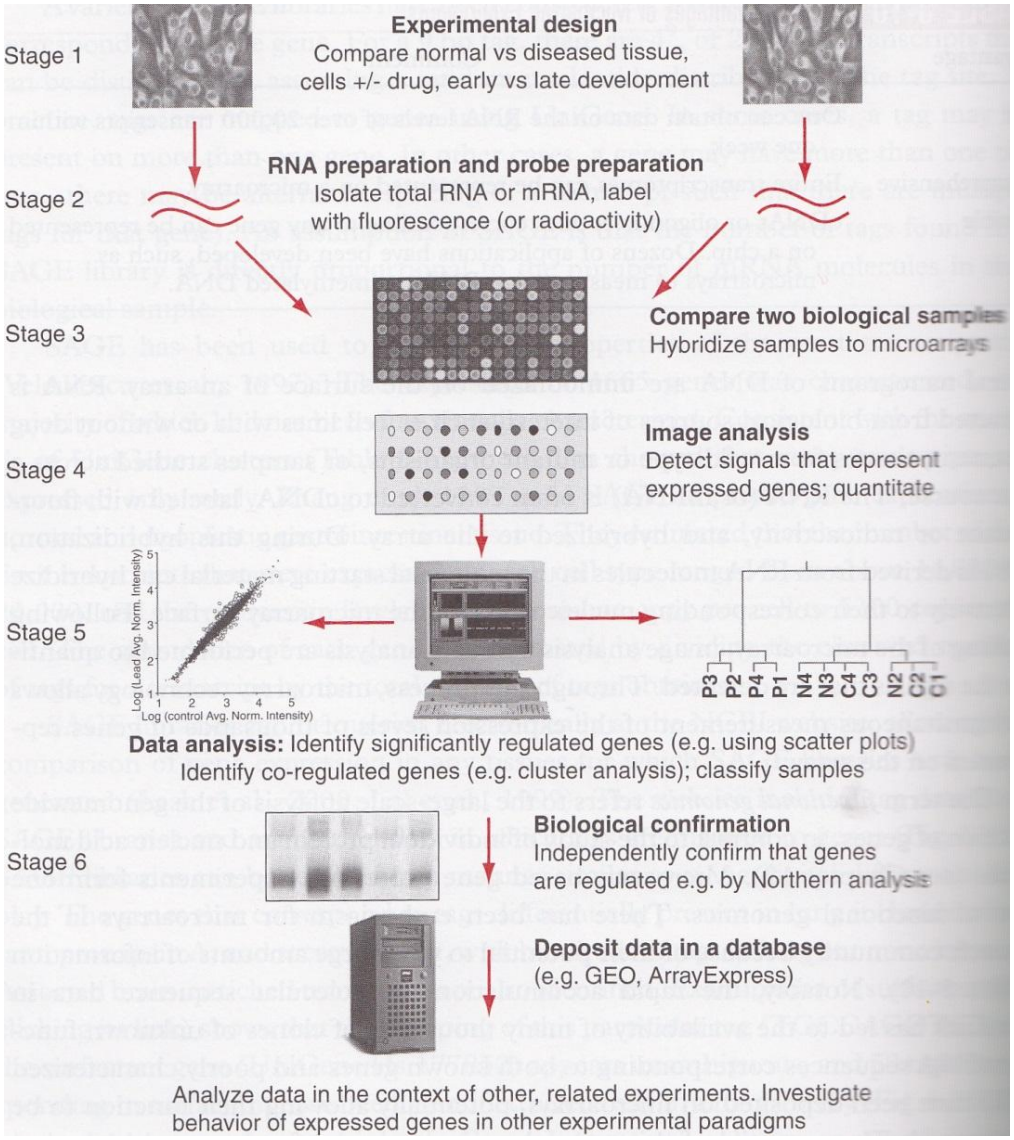




1:1 correlation plot of all peptide raw intensity values for 2 reporter ions from Libra's quantitation output

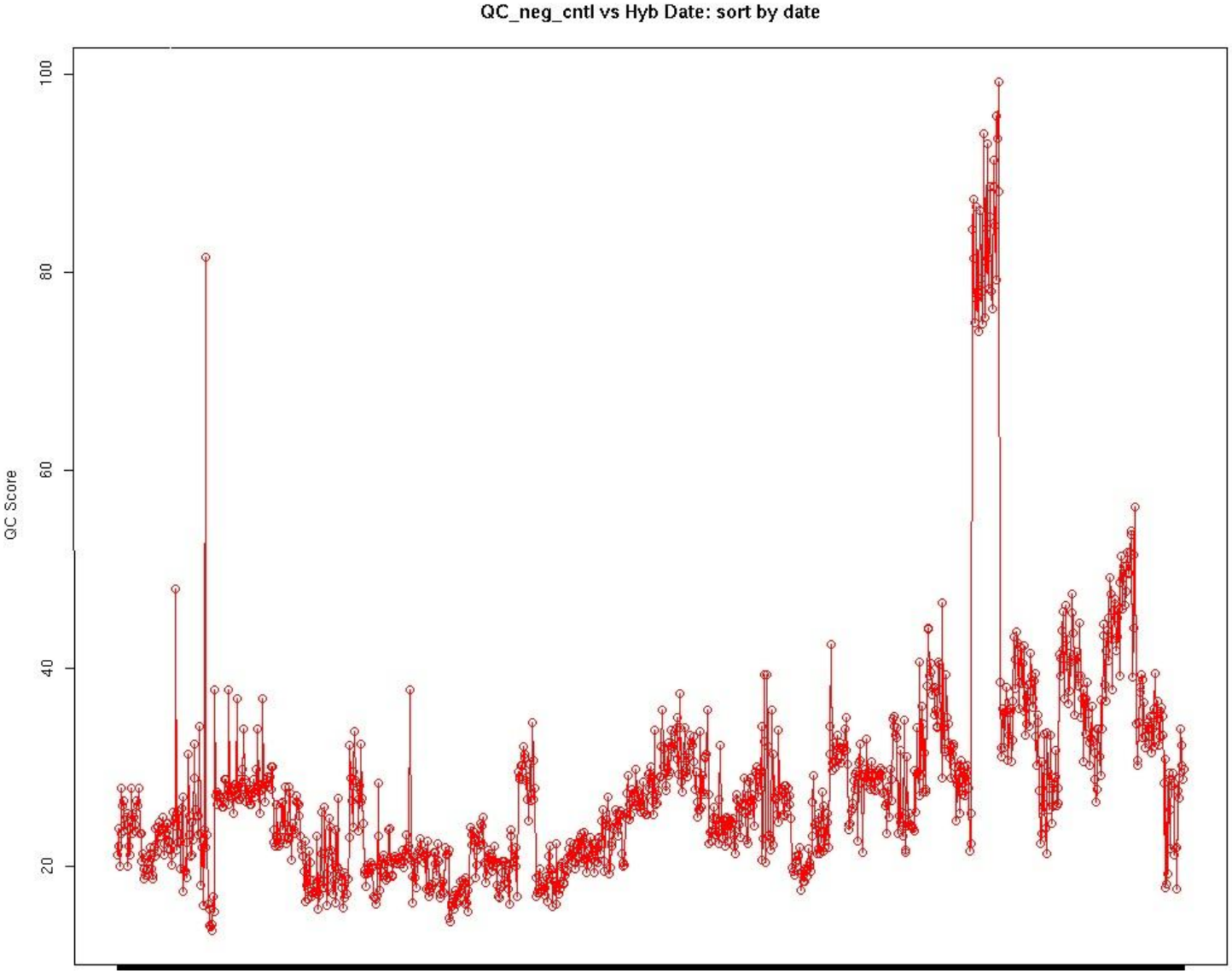
Automation and Analysis in R and Bioconductor

Standard MA Analysis Workflow

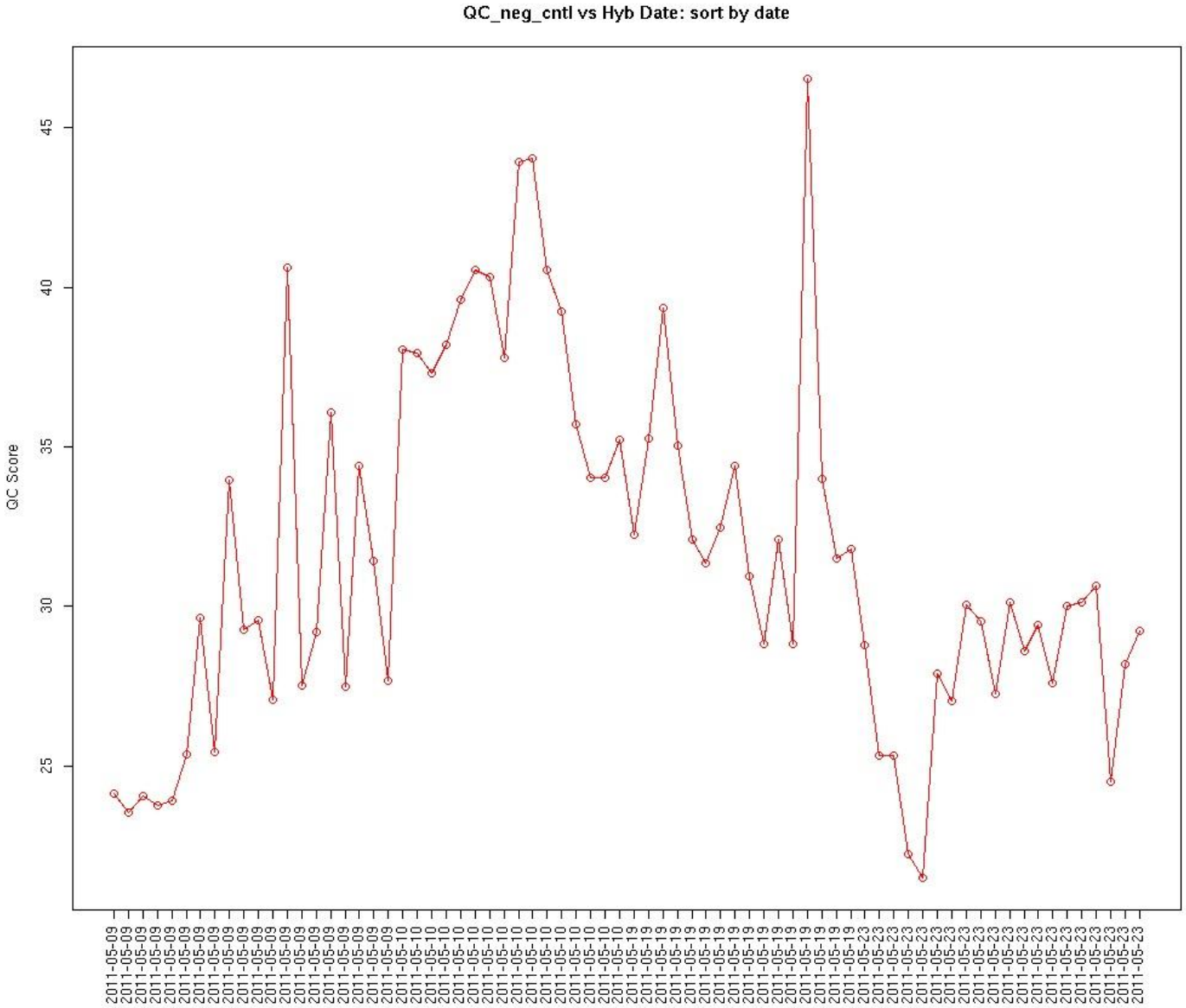


Bioinformatics and Functional Genomics , p.314 Jonathan Pevsner

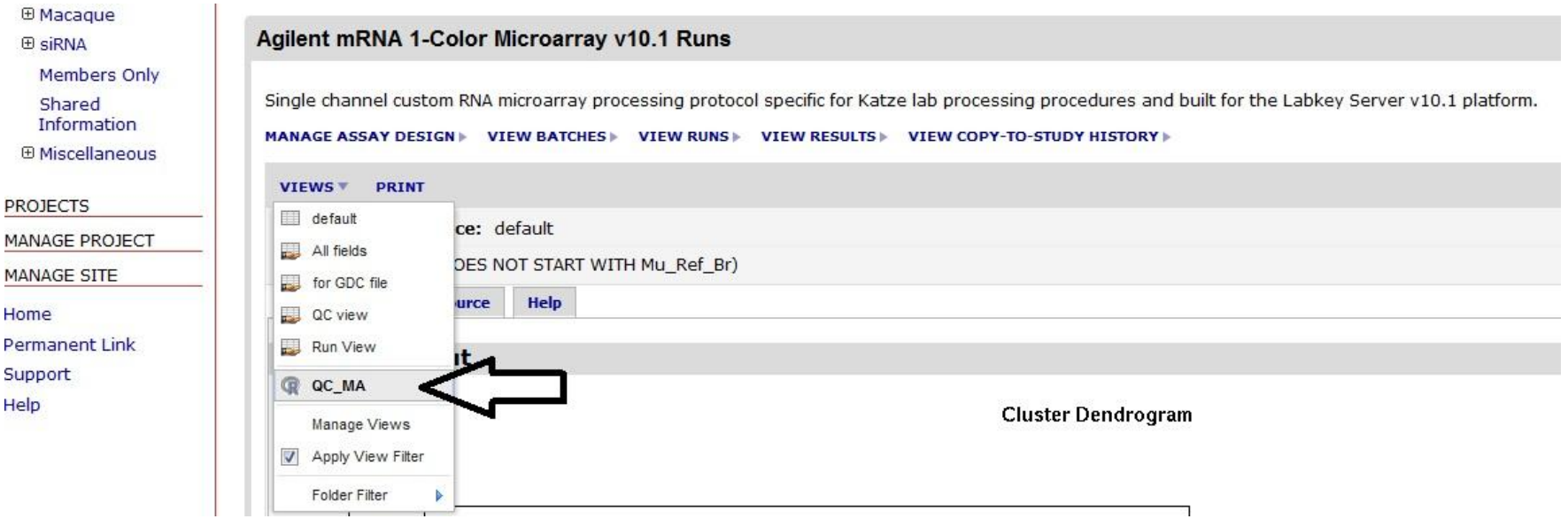
Plotting QC Scores



Plotting QC Scores

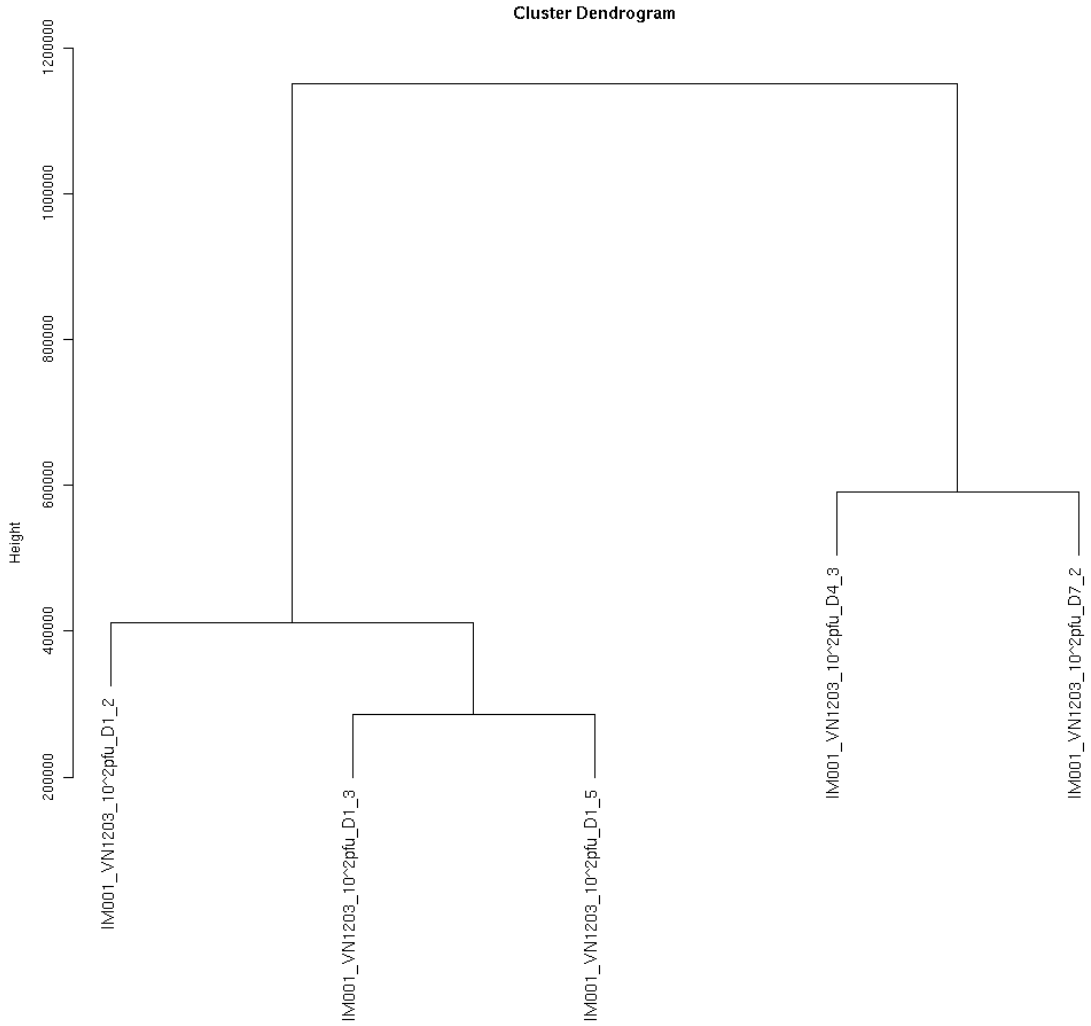


QC Analysis

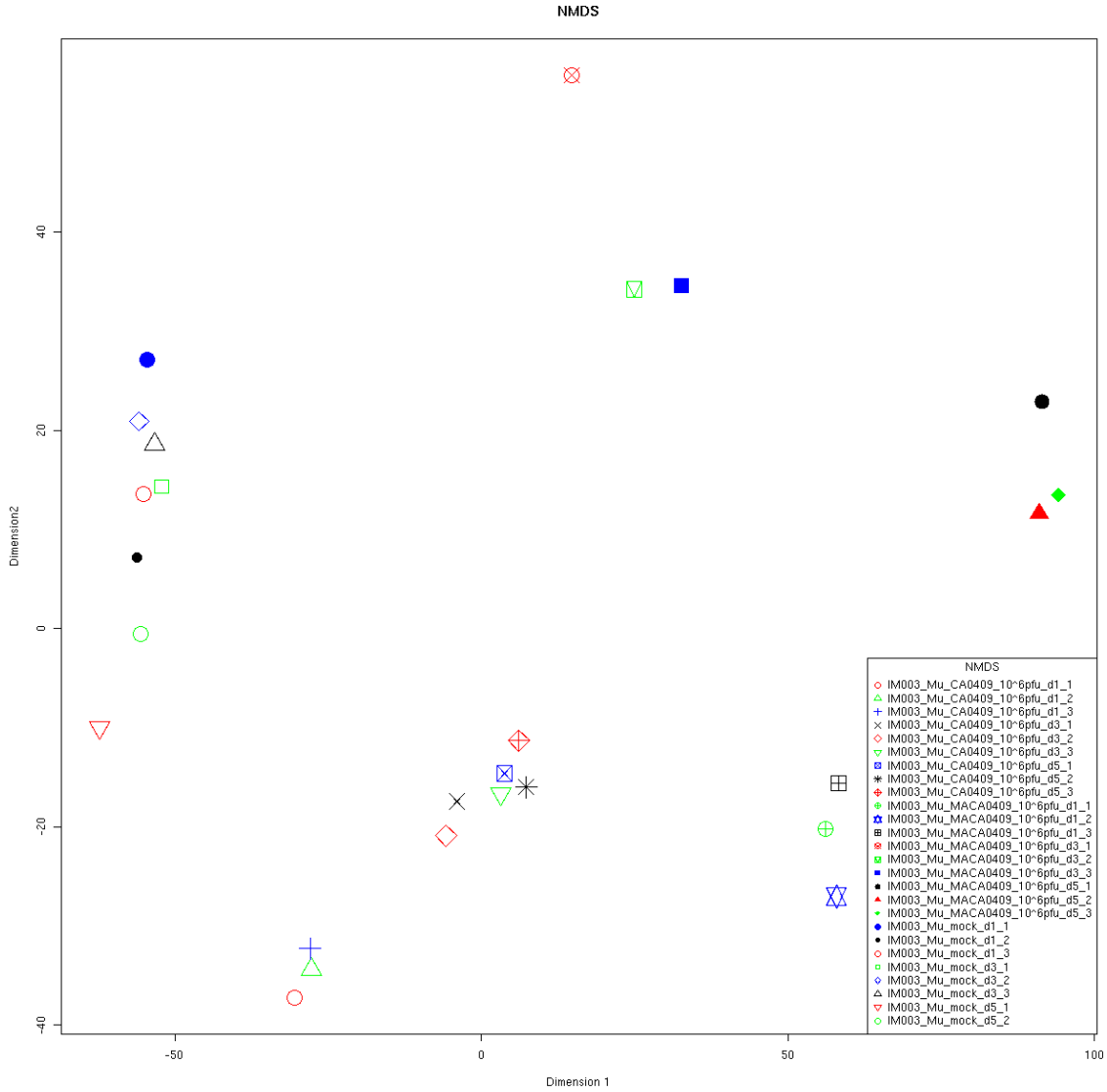

Macaque
siRNA
Members Only
Shared Information
Miscellaneous
PROJECTS
MANAGE PROJECT
MANAGE SITE
Home
Permanent Link
Support
Help
Agilent mRNA 1-Color Microarray v10.1 Runs
Single channel custom RNA microarray processing protocol specific for Katze lab processing procedures and built for the Labkey Server v10.1 platform.
MANAGE ASSAY DESIGN
VIEW BATCHES
VIEW RUNS
VIEW RESULTS
VIEW COPY-TO-STUDY HISTORY
VIEWS
PRINT
default
All fields
for GDC file
QC view
Run View
QC_MA
Manage Views
Apply View Filter
Folder Filter
ce: default
OES NOT START WITH Mu_Ref_Br)
urce
Help
Cluster Dendrogram

QC Analysis

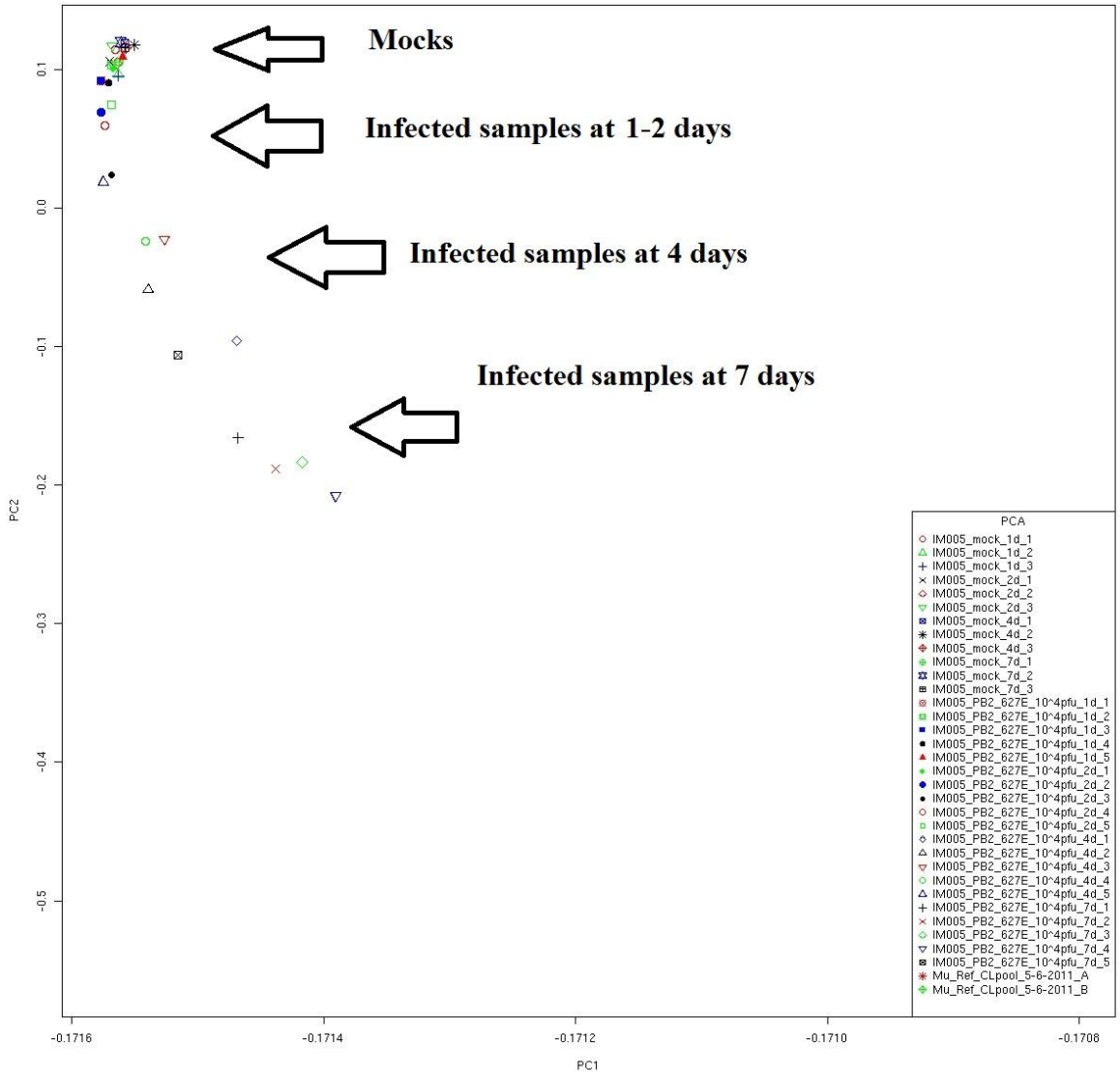
- Hierarchical Cluster Analysis



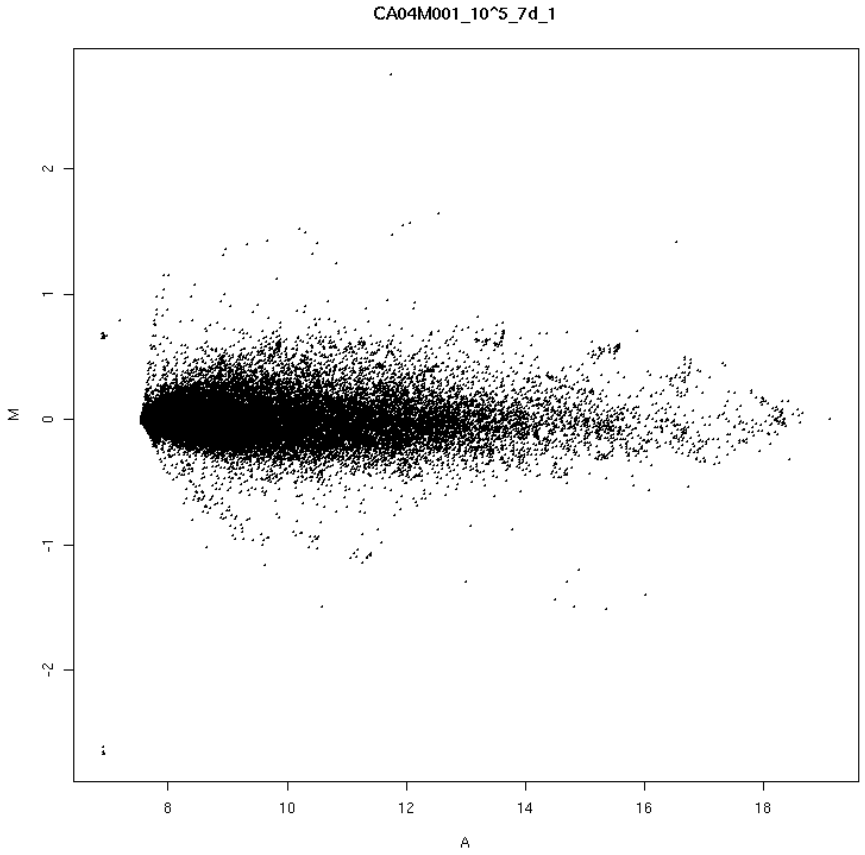
QC Analysis: NMDS



QC analysis: Principal Component Analysis



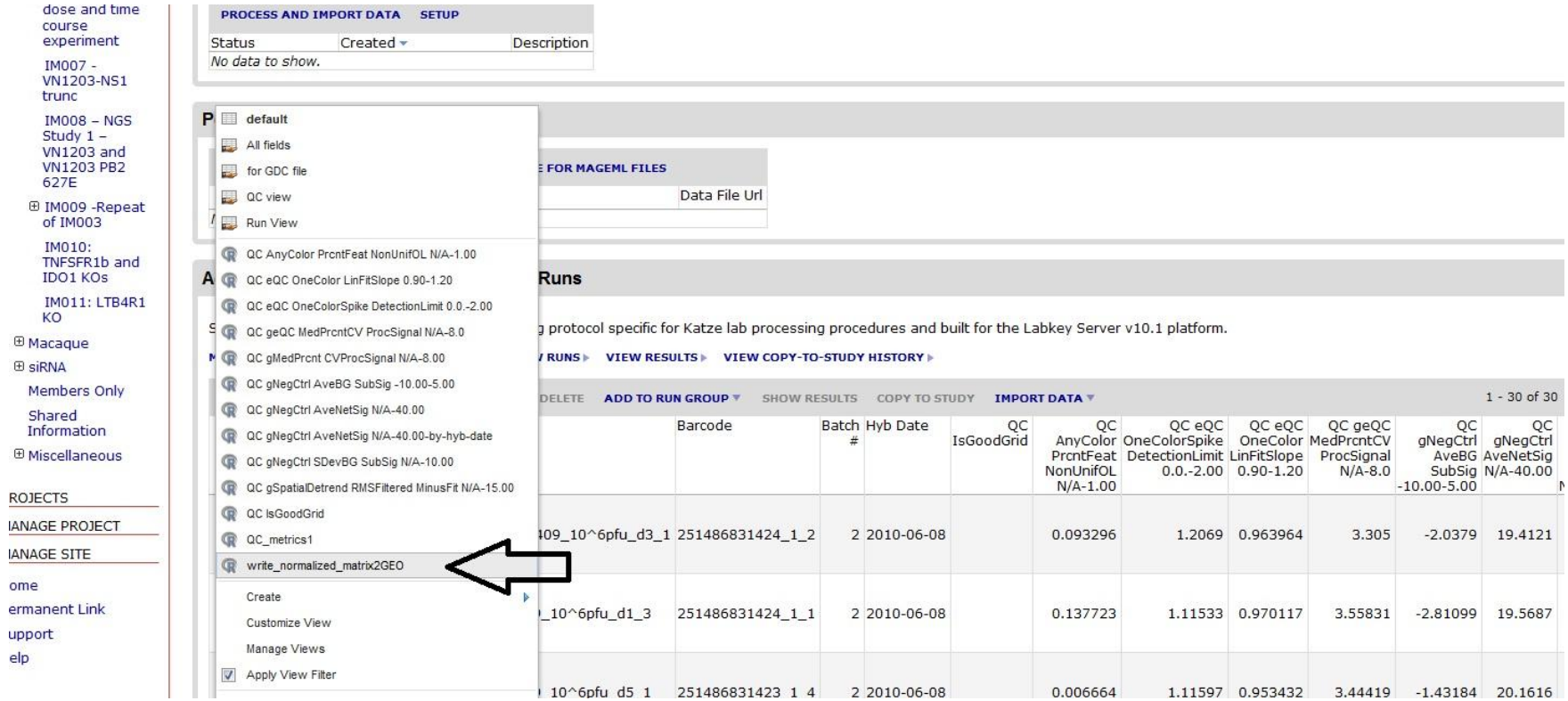
QC Analysis: MA Plots



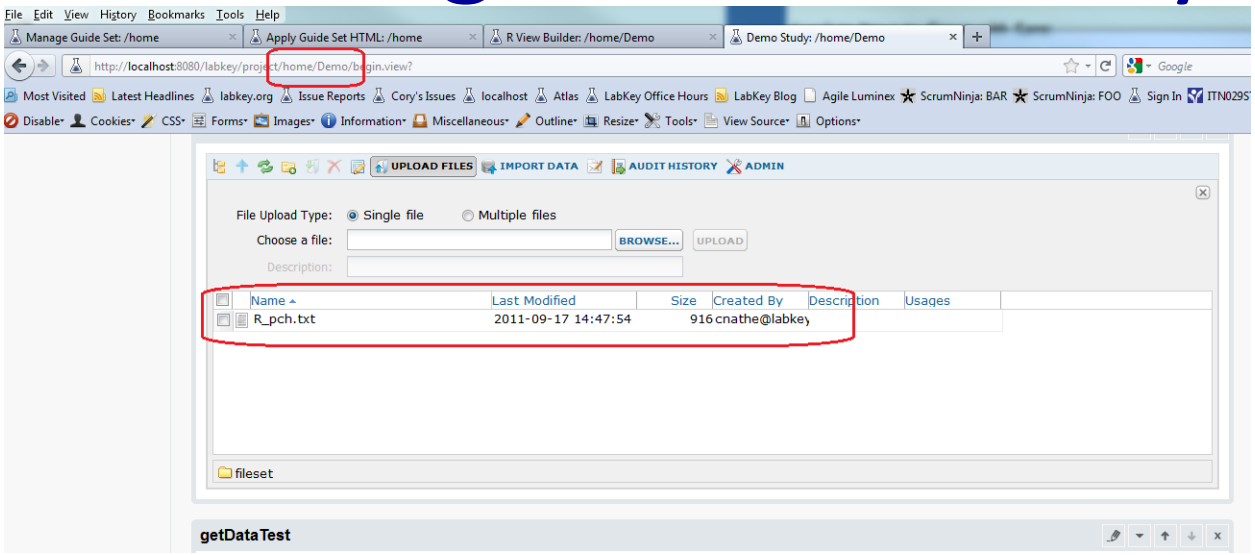
To determine whether normalization is needed, the MA-plot is generated to plot of the distribution of the red/green intensity ratio ('M') plotted by the average intensity ('A').

GEO Submission

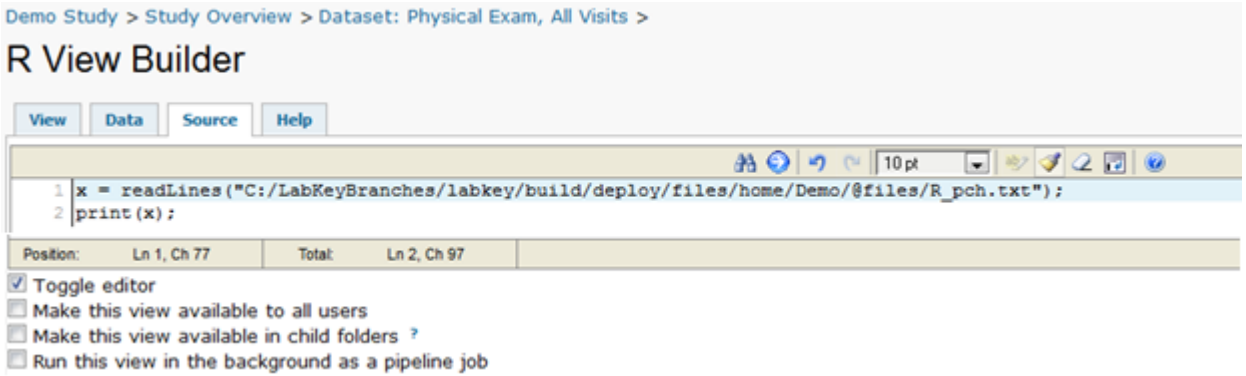
- Prepare the microarray data to be pushed into the public repository



Uploading a Configuration File Can Strengthen Our Analysis Power



- Preprocess 44x4K Agilent (target file)
- Limma (contrast matrix)



How Analysis Has Moved Forward: LabKey's Support

- Constant development on new tools
- Pushing integrated tools into a open source platform that is available to all

- Consistent support : labs of 4 to 43, either through a service contract or on the support boards online

Future Steps

- Additional GEO customization
- Adapting additional Bioconductor tools into our automated pipeline
- Further NGS integration into labkey
- NGS pipeline: (Galaxy server, QC methods)
- Visualizing libra output through R/bioconductor
- R-cytoscape
- Kineta: Integration of chemistry schema into postgres database

Acknowledgements

- Cory Nathe
- Josh Eckels
- Tim Owens
- Katze lab