

# Next-Generation Clustered Heat Maps (NG-CHM)

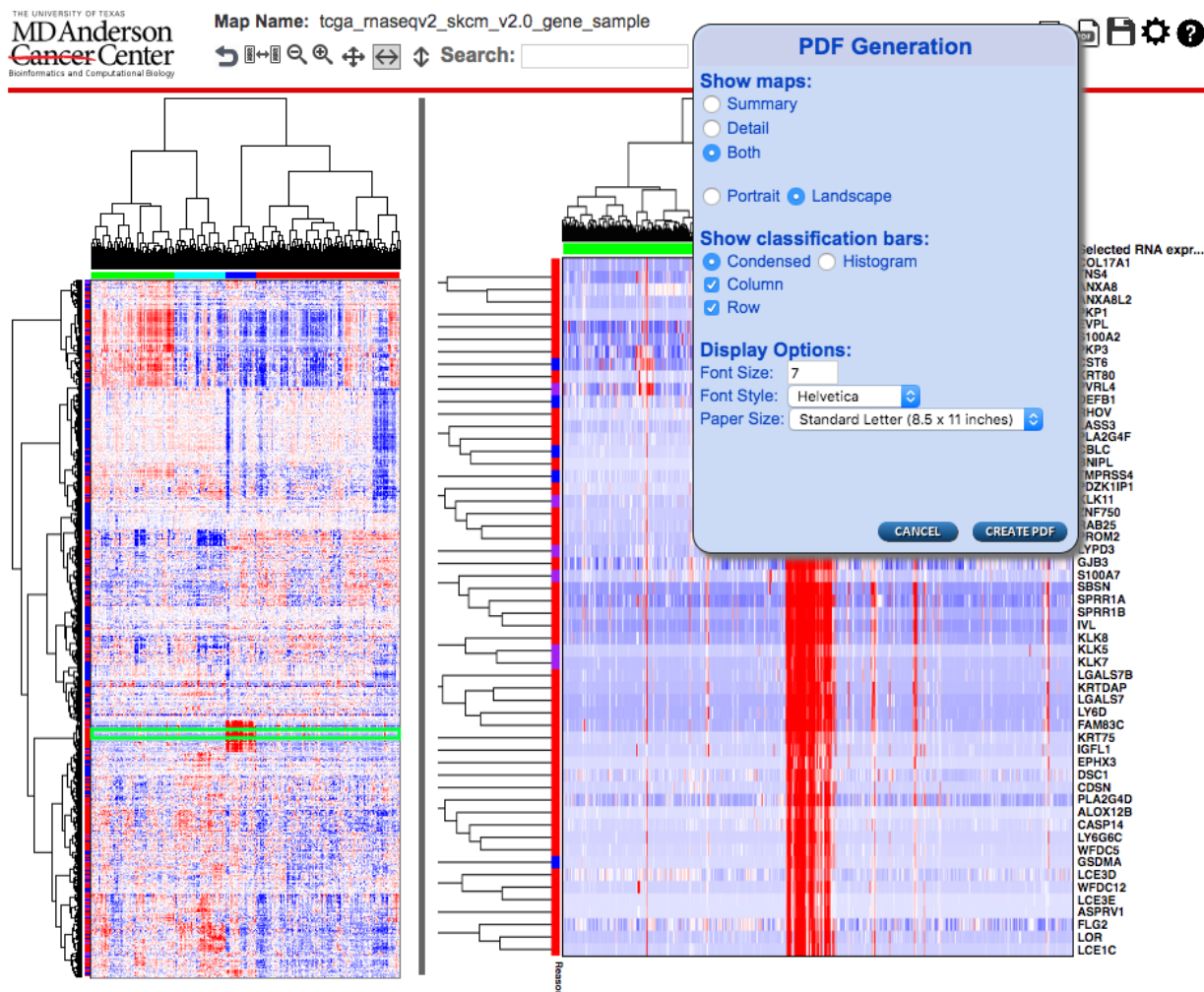
Bradley Broom and John Weinstein,

Department of Bioinformatics and Computational Biology, UT MD Anderson Cancer Center

A visualization tool for the dynamic exploration of large, omic data sets.

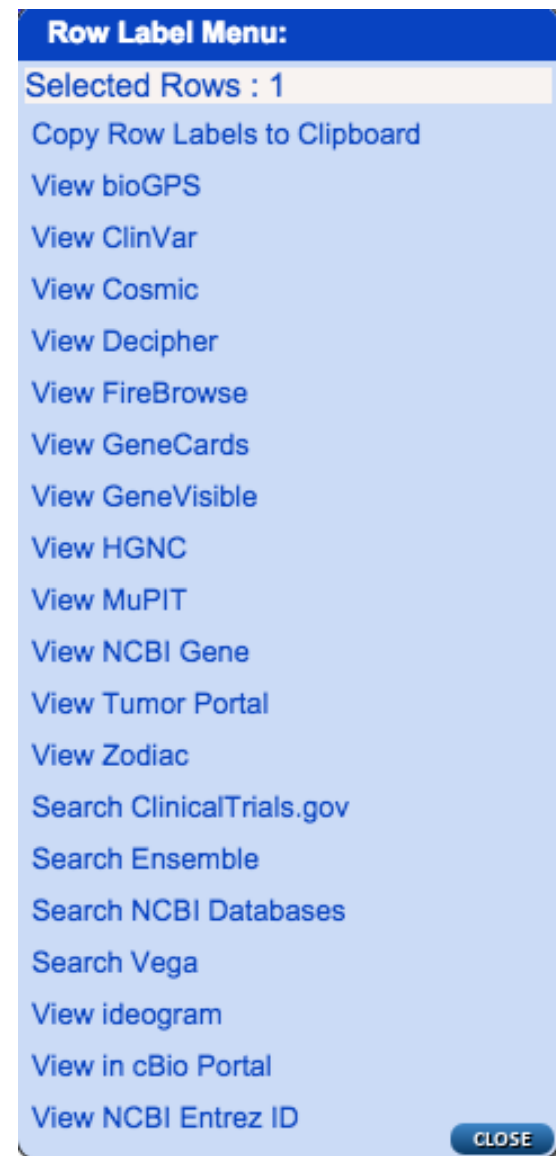
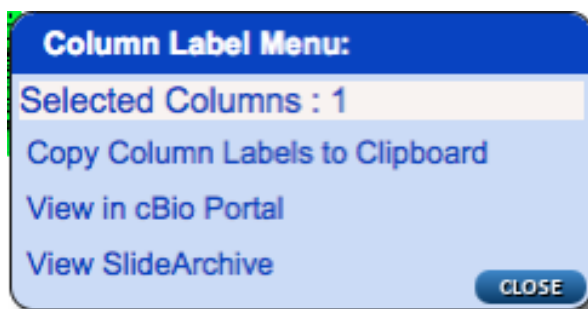
## Visualization Features

- Easy-to-use navigation
- Flexible zoomed views for inspecting small details in large data sets
- Generate publication-quality graphics
- Scalable
- Highly interactive

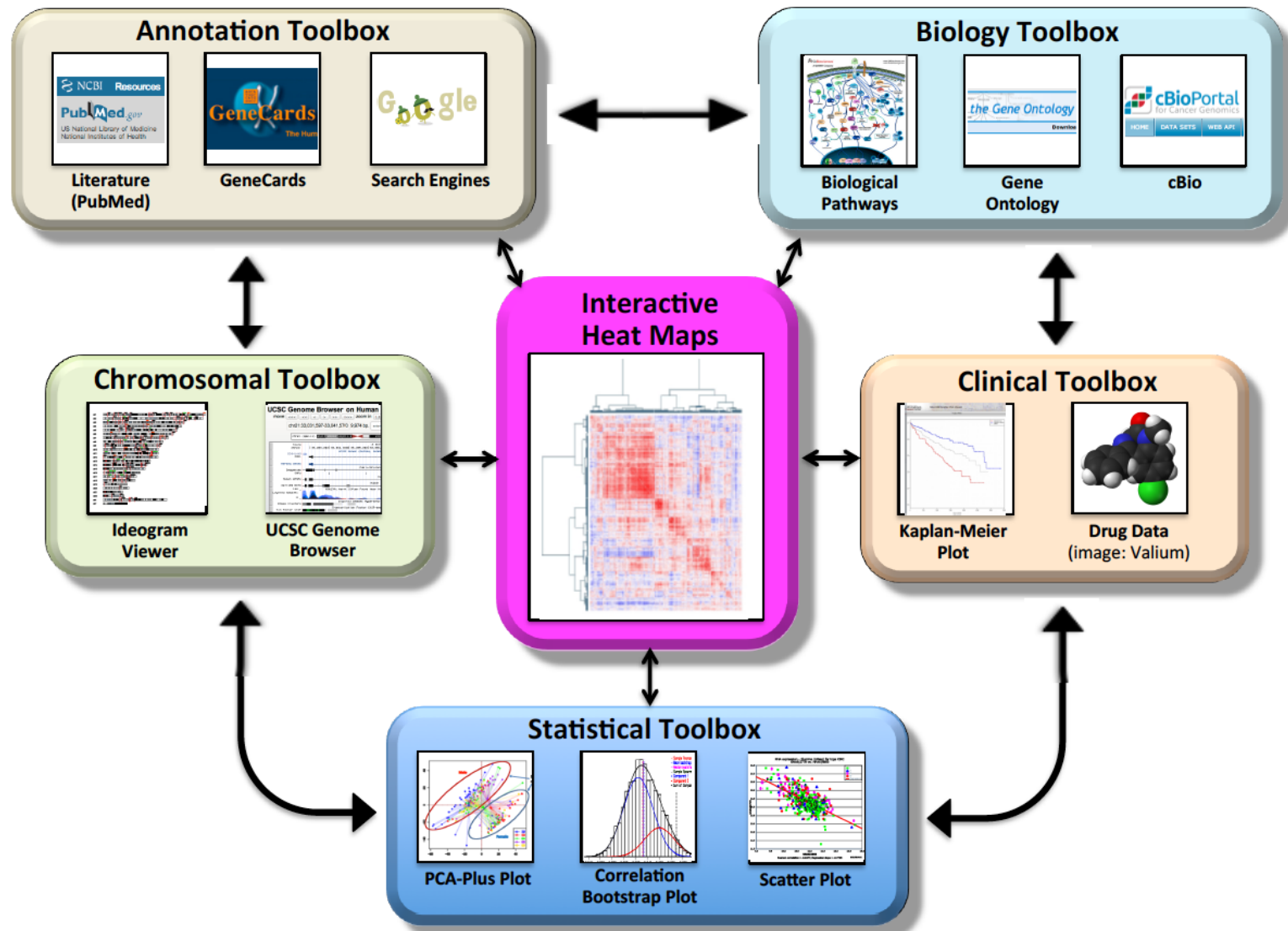


# Dynamic “Link-outs”

- Can link-out from selected **rows**, **columns**, or **matrix elements**
- To other content specific resources,
  - E.g. NCBI, Genecards, Cancer Digital Slide Archive
- Content specific
  - Associate ‘types’ with row and column labels
  - ‘Link-out’ database indexed by type
  - Only matching link-outs included in a specific NG-CHM
- Or to dynamic analyses
  - E.g. Box Plot, Points Plot
- Five major types of resource links



# An Exploratory Environment with five major types of resource links



# Creating NG-CHMs

## Four “builder” levels:

### – GUI

- New non-quantitative user
- Experienced non-quantitative user

### – R

- Quantitative user

### – Specification files

- System developer

THE UNIVERSITY OF TEXAS  
**MD Anderson**  
 Cancer Center

## NG-CHM Quick Builder

☐ Not sure where to start? Select this checkbox to use sample data [Help](#)

### Clustered Heat Map Central Image

CHM Name:

#### Matrix Data File(s)

| Matrix Label    | Matrix File Name           | Add | Clear  |
|-----------------|----------------------------|-----|--------|
| Main Data Layer | Choose File No file chosen |     | Delete |

[Show Advanced Color Options >>](#)

#### Row Ordering

Row Ordering Method: Hierarchical Clustering  
 Distance Metric: Correlation  
 Agglomeration Method: Ward

[Show Advanced Options >>](#)

#### Column Ordering

Column Ordering Method: Hierarchical Clustering  
 Distance Metric: Correlation  
 Agglomeration Method: Ward

[Show Advanced Options >>](#)

# RECENT DEVELOPMENTS

## New architecture NG-CHMs

- Previous NG-CHM system architecture
  - Was designed six years ago
  - Relied heavily on back-end server for tile rendering
  - Did not exploit recent advances in web browsers
  - Had performance and scaling issues
- New NG-CHM system architecture
  - Complete re-design to take advantage of new browser capabilities (e.g. WebGL)
  - “Tile Generation” performed entirely in browser
  - Far more responsive user experience

## New architecture NG-CHMs

- “Standalone” server-less mode possible
  - Nearly-all NG-CHM functionality is present
    - Cannot “link-out” to server-side dynamic computations
  - Allows NG-CHMs to be:
    - Saved locally (e.g. with other project data)
    - Attached as supplementary material to papers
    - Shared with colleagues via email, web sites, portals, etc.

# New architecture NG-CHMs integrated with Galaxy

- Added new architecture NG-CHMs to Galaxy:
  - BigQuery Interface (for ISB-CGC data access)
  - NG-CHM Builder tool
  - NG-CHM Visualization
- Available
  - As a ready-to-run Docker container
  - From the Galaxy Toolshed
  - From Github
- Exposes NG-CHMs to a new user community



# Galaxy NG-CHM BigQuery Interface (for ISB-CGC)

[Analyze Data](#)
[Workflow](#)
[Shared Data](#)
[Visualization](#)
[Admin](#)
[Help](#)
[User](#)

Tools

search tools

**Big Query**

[Select Genes based on Participant Samples and RPKM scores with NGCHM Samples and Genes](#)

[Query Tissue Database for RPKM values with NGCHM Samples and Genes](#)

[Define your participant cohort from NCGCHM ISB for heatmap creation](#)

**Tools**

[MDA Heatmap](#)

[Send Data](#)

[Lift-Over](#)

[Filter and Sort](#)

[Join, Subtract and Group](#)

[Convert Formats](#)

[Fetch Sequences](#)

[Fetch Alignments](#)

[Statistics](#)

[Graph/Display Data](#)

[Heatmaps Colormanipulation](#)

[Heatmaps timpalpant](#)

[Heatmap Rversion](#)

[CRAVAT](#)

[PCA-diagrams](#)

**Workflows**

[All workflows](#)

**Define your participant cohort from NCGCHM ISB for heatmap creation (Galaxy Tool Version 2.0.5)**

**Label Your New Participant Cohort**

**Study**

Click cursor in middle of this text to see choices

Select a Study by clicking curson on ACC text

**Categorical Query Info**

1: Categorical Query Info

**Categorical Query Field Choice**

Click cursor in middle of this text to see choices

Gender

Race

Ethnicity

Clinical Stage


**Pathologic Stage**

Vital Status


Country


Participant Barcode -- format TCGA- {an}{an} -nnnn where a=caps alpha, n=number {an} one or other no spaces

# Galaxy NG-CHM Builder


**Galaxy**

Analyze Data
 Workflow
 Shared Data
 Visualization
 Help
 User


 Using 0 bytes

Tools
 

search tools

**Big Query**

**Tools**

**MDA Heatmap**

[Normalize Matrix](#) Normalize values to range -1 to 1 Set non-numerics to Mean value

[Mean-Center a Matrix](#) with Samples as columns and data Values per item in Rows, non-numerics set to mean value

[MDA Heatmap Generation](#) Create HeatMap from Dendogram and Matrix

**Lift-Over**

**Filter and Sort**

**Join, Subtract and Group**

**Convert Formats**

**Fetch Sequences**

**Fetch Alignments**

**Statistics**

**Graph/Display Data**

**Heatmaps Colormanipulation**

**Heatmaps timpalpant**


**Heatmap Rversion**

**CRAVAT**

**PCA-diagrams**




**Workflows**

- All workflows


**MDA Heatmap Generation** Create HeatMap from Dendogram and Matrix (Galaxy Tool Version 5.2.4)
 

Options

**Input Matrix**




 2: Mean-Center a Matrix on data 1

**User Defined HeatMap Name**

HM\_name

**Optional HeatMap Description**

**Data Summarization Method**

average

**Row ordering method**

Hierarchical Clustering

Choices -- Hierarchical Clustering, Original Order, Random

**Row Distance Metric**

Euclidean

euclidean, binary, manhattan, maximum, canberra, minkowski, or correlation

**Row Clustering Method**

Ward

Choices: 'average' for Average Linkage, 'complete' for Complete Linkage, 'single' for Single Linkage, 'ward', 'mcquitty', 'median', or 'centroid'.

**Column ordering method**

Hierarchical Clustering

Choices -- Hierarchical Clustering, Original Order, Random

**Column Distance Metric**

Euclidean


euclidean, binary, manhattan, maximum, canberra, minkowski, or correlation


**Column Clustering Method**


Ward

Choices: 'average' for Average Linkage, 'complete' for Complete Linkage, 'single' for Single Linkage, 'ward', 'mcquitty', 'median', or 'centroid'.

**Covariate Bars**

 Insert Covariate Bars


 Execute




History
 




search datasets

**Unnamed history**

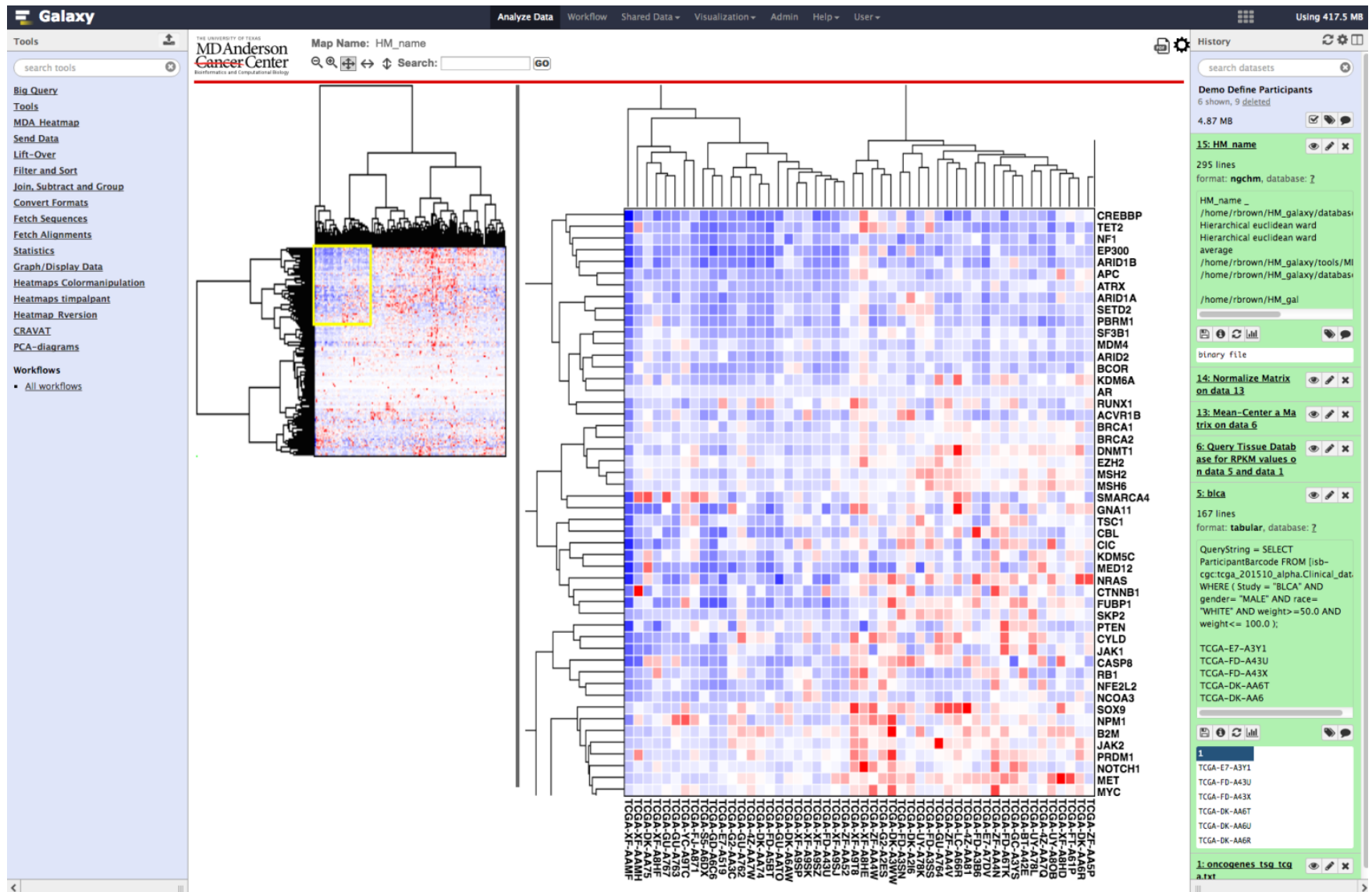
2 shown

60.55 KB 

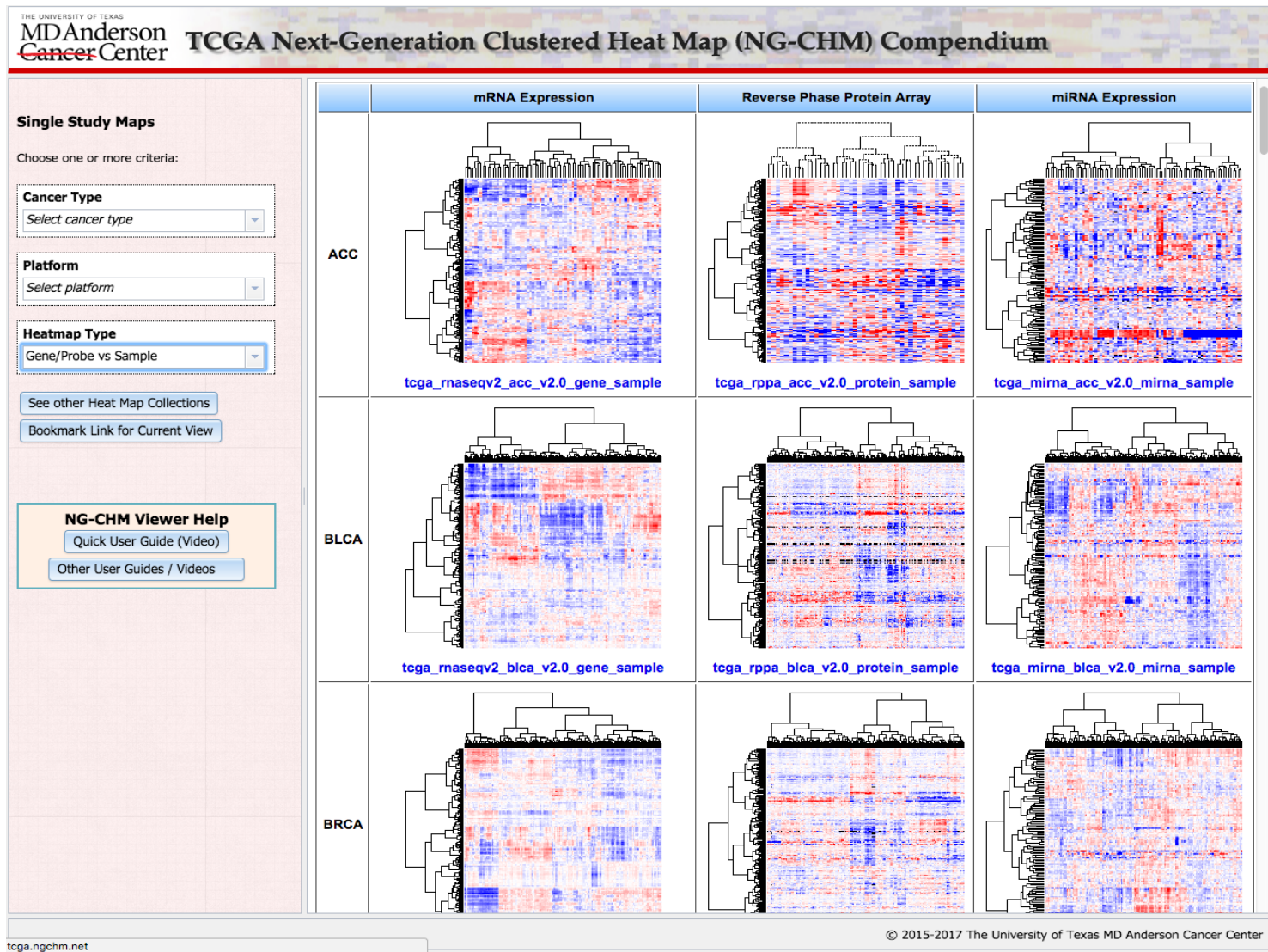
**2: Mean-Center a Matrix on data 1**   

**1: matrix.tsv**   

# Galaxy NG-CHM Visualization Tool



# New architecture TCGA NG-CHM Compendium



<http://tcga.ngchm.net/>

# Numerous Improvements to NG-CHM Viewer

- More user options for fine tuning PDF generation
- Improved rendering responsiveness for large NGCHMs
- Ability to create very large heat maps (> 20,000 rows)
  - Contributed dendrogram code to R project (in R 3.4.0)
- Added new features to NG-CHMs:
  - Can add “cuts” (extra blank rows/columns) to heat map
  - Can highlight a small number of top rows/columns in overview pane
- Numerous bug fixes and other improvements
- Created additional tutorial videos

# UPCOMING PLANS

## Upcoming plans

- Release new architecture NG-CHM manager
- Complete our new architecture NG-CHM compendium
- Further increase viewer responsiveness for very large data matrices
- Develop a tablet-optimized user interface
- Collaborate
  - Add links to/from NG-CHMs from/to other tools
  - Create more advanced tools & data exchange



# Credits

## **MD Anderson Dept of Bioinformatics and Computational Biology**

Bradley Broom

John Weinstein

Rehan Akbani

Chris Wakefield

James Melott

## **In Silico Solutions**

Michael Ryan

Robert Brown

Futa Ikeda

Mark Stucky

## **Hobsons**

David Kane